

Resting State Morphology Predicts the Effect of Theta Burst Stimulation in False Belief Reasoning

Charlotte E Hartwright^{a,b}, Robert M Hardwick^{b,c}, Ian A Apperly^b, Peter C Hansen^b

^a Aston Brain Centre, Aston University, Birmingham, B4 7ET

^b Behavioural Brain Sciences, School of Psychology, University of Birmingham, Birmingham, UK

^c Department of Neurology, Johns Hopkins University, Baltimore, MD, USA

Address for correspondence:

Dr Charlotte E Hartwright

School of Life and Health Sciences

Aston University

Birmingham

B4 7ET

UK

Email: c.hartwright@aston.ac.uk

Tel: +44 (0) 121 204 3000

Abstract

When required to represent a perspective that conflicts with one's own, functional magnetic resonance imaging (fMRI) suggests that the right ventrolateral prefrontal cortex (rvLPFC) supports the inhibition of that conflicting self perspective. The present task dissociated inhibition of self perspective from other executive control processes by contrasting belief reasoning – a cognitive state where the presence of conflicting perspectives was manipulated – with a conative desire state wherein no systematic conflict existed. Linear modelling was used to examine the effect of continuous theta burst stimulation (cTBS) to rvLPFC on participants' reaction times in belief and desire reasoning. It was anticipated that cTBS applied to rvLPFC would affect belief but not desire reasoning, by modulating activity in the Ventral Attention System (VAS). We further anticipated that this effect would be mediated by functional connectivity within this network, which was identified using resting state fMRI and an unbiased model-free approach. Simple reaction-time analysis failed to detect an effect of cTBS. However, by additionally modelling individual measures from within the stimulated network, the hypothesised effect of cTBS to belief (but, importantly, not desire) reasoning was demonstrated. Structural morphology within the stimulated region, rvLPFC, and right temporoparietal junction (rTPJ) were demonstrated to underlie this effect. These data provide evidence that inconsistencies found with cTBS can be mediated by the composition of the functional network that is being stimulated. We suggest that the common claim that this network constitutes the VAS explains the effect of cTBS to this network on false belief reasoning.

Of interest to Cognitive, Developmental and Clinical researchers, the ability to represent mental states such as the beliefs, desires or intentions of other people, termed having a 'Theory of Mind' (ToM), has been studied extensively with functional magnetic resonance imaging (fMRI), with particular focus on a fronto-parietal network comprising the medial prefrontal cortex (mPFC) and right temporoparietal junction, rTPJ (Carrington and Bailey, 2009; Lieberman, 2007; Mar, 2011). Two neuropsychological studies, however, demonstrated that damage affecting a large area of right lateral prefrontal cortex (LPFC) – a region not classically associated with ToM – resulted in a strong egocentric bias when reasoning about the perspectives of others: the patient suffered interference from their own perspective, making it difficult for them to infer others', if in conflict their own (Samson, et al., 2005; Samson, et al., 2015). A belief that is incongruent between assigner and assignee, termed a 'false' belief, thus appears to involve inhibition of own perspective in order to successfully adopt the other person's viewpoint, which is supported by right LPFC. This theory has to some degree been corroborated by fMRI studies, which more precisely localize this inhibitory process to right ventrolateral prefrontal cortex, rvlPFC (Hartwright, et al., 2012; Hartwright, et al., 2013; van der Meer, et al., 2011; Vogeley, et al., 2001). However, given the relatively large brain lesions in the neuropsychological case studies described above, it remains critical to obtain converging evidence that this brain region does indeed serve a causal role in resisting interference from one's own perspective. Moreover, as argued in a recent commentary (Schurz and Tholen, 2016), vlPFC is not identified in all fMRI studies wherein inhibition of own perspective would be expected to be requisite of the task. These authors urge further investigations "to gain a full understanding of the IFG's role in ToM" Schurz and Tholen (2016, p.332). One powerful approach to addressing this issue is to use an interference technique such as transcranial magnetic stimulation (TMS). TMS

permits a transient change in brain function, therefore revealing a casual link between brain and behavior, without any compensation that might occur in the case of the damaged brain.

The present study used an offline TMS protocol – continuous theta burst (cTBS) – to test whether rvlPFC serves a causal role in belief reasoning, specifically in the inhibition of self perspective. We used an established behavioural paradigm in which the valence of an agent's belief and desire state is systematically varied. In prior uses of this task, we have demonstrated consistent behavioural effects with typical adults, where negatively valenced ToM states – 'false belief' and 'avoidance desire' – are shown to attract greater response latencies and error rates than orthogonal, positively valenced 'true belief' and 'approach desire' (Apperly, et al., 2011; Hartwright, et al., 2012; Hartwright, et al., 2013). Importantly, however, in this experimental manipulation, only false belief, and not avoidance desire, features self-other incongruence (see Hartwright, et al., 2012; Hartwright, et al., 2013 for detailed discussion). When completed as an fMRI experiment, valence for both belief and desire modulates activity in classic control regions, such as anterior cingulate cortex.

Critically, the valence of belief, but not desire, additionally modulates rvlPFC (Hartwright, et al., 2012; Hartwright, et al., 2013), which further supports the view that conflict between self and other perspectives is resolved by a network in which rvlPFC forms a critical part.

Regardless, the process of inhibiting one's own knowledge in order to represent someone else's has only been causally tested using poorly-circumscribed lesions. To more precisely localise this process is important, however, as the ability to overcome self perspective distinguishes adult social cognition from young children's, and may also be a mediating factor in some clinical disorders, such as autism (Begeer, et al., 2012). Thus, using TMS with an experimental paradigm that has been shown to recruit rvlPFC provides an opportunity to seek such evidence.

TMS is a powerful tool that can address questions regarding the function of a stimulated brain region. Nevertheless, the variability in neurophysiological and behavioural responses to stimulation is remarkably high (see Nicolo, et al., 2015; Ridding and Ziemann, 2010 for reviews). Stimulation can increase plasticity metrics such as motor evoked potentials and reaction times (RT) in some participants, yet with the same stimulation parameters reduce, or have no effect, in others (e.g., Hamada, et al., 2012; McAllister, et al., 2013; Vernet, et al., 2013). Ideally, averaging the induced performance differences over all participants would still demonstrate a causal role for the stimulated region in that task, albeit with weaker statistical power. However, such contrasting effects could also result in the measured effects averaging to zero, resulting in a Type II error. Likely due to the more overt effects of stimulation to the primary motor cortex, research examining inter- and intra-subject variability in the response to TMS has been centred on studies of the motor system; thus, how individual differences might interact with the effects of TMS on cortical dynamics in the cognitive domain is less clear. Still, this caveat aside, such factors that have been shown relevant include genetic, neurobiological and environmental influences (Ridding and Ziemann, 2010).

Work to establish the bases of inter-individual differences in response to TMS has identified that one cause of cross-subject variability in the pattern of performance is associated with changes in neuronal resting network synchronicity (McAllister, et al., 2013). Indeed, electrically evoked responses following intracranial stimulation are bounded by networks that are spatiotemporally correlated, where the spatial distribution and magnitude of neural activity in regions at rest predicts the pattern and magnitude of evoked responses (Keller, et al., 2011). This suggests that resting functional connectivity may mediate individual responses to TMS, and shows convergence across different neural

indices. Neural regions which are spatiotemporally related in the absence of a task, known as resting-state networks (Damoiseaux, et al., 2006), have been shown to agree across electroencephalogram (EEG) and blood oxygen level dependant (BOLD) resting-state (R-fMRI) modalities (Musso, et al., 2010), despite the differing neurophysiological mechanisms that they measure. It would seem, then, that the differing approaches to characterising a network used within fMRI and EEG, ultimately, are both able to capture predictive measures of individual responsiveness to TMS. The utility of R-fMRI in predicting TMS outcomes has been translated into directing therapeutic application of stimulation. Fox, et al. (2014) found that, where invasive deep brain stimulation successfully alleviated clinical symptoms of a specified neurological or psychiatric disorder, the success of non-invasive TMS in producing the matched clinical outcomes could be determined on the basis of resting-state functional connectivity between the two stimulation sites. This indicates that a network perspective should be applied to stimulation. For cognitive neuroscience, wherein TMS is often used to create a 'virtual lesion' (Pascual-Leone, et al., 2000), such a perspective highlights the importance of examining the stimulation site in relation to its network constituents.

The use of structural and functional MRI to explain variability in cognitive performance is commonplace. Numerous studies demonstrate that measures of executive function correlate with indices of gross morphology, such as voxel-wise or total grey matter (GM) and white matter (WM) volume (e.g., Li, et al., 2012; Magistro, et al., 2015; Takeuchi, et al., 2012). Similarly, the strength of connectivity within a functional network can be considered a proxy for network efficiency, where connectedness determined by R-fMRI can predict individual differences in executive function (Reineberg, et al., 2015). In the context of brain stimulation, although only explored in two papers, these converge in that individual differences in metrics taken from structural MRI can be informative to explaining variability

in a participant's response to TMS. Conde, et al. (2012) showed that individual differences in cortical thickness of the sensory motor cortex could explain approximately half of the variance in excitability changes following paired-associative stimulation, where participants with thicker cortical GM were more affected. Likewise, GM and WM volume were shown to predict motor-learning outcomes in stroke patients following stimulation to the primary somatosensory cortex (Brodie, et al., 2014). Based on these data, and the earlier described work on from resting-state data, it follows that indices taken from structural MRI, which is informed by functional MRI, might be informative in the prediction or explanation of TMS outcomes.

The purpose of the present study was to test for a causal role for rvlPFC in belief reasoning, which is the most widely studied component of ToM. Additionally, we were interested in whether it was possible to predict responsiveness to TMS in healthy participants using brain indices extracted from R-fMRI; do features of the stimulated network – the strength of connectivity across the network and gross morphology of network nodes – further explain responsiveness to TMS? Though Keller, et al. (2011) and Fox, et al. (2014) demonstrate a relationship between R-fMRI and the response to TMS in clinical groups, to our knowledge, this has not been demonstrated in non-clinical populations. This is important as differences in the engagement and morphology of resting-state networks between typical and atypical populations is well documented (see Greicius, 2008 for a review). R-fMRI and diffusion based tractography have identified a network including vlPFC and TPJ (Mars, et al., 2012), which is commonly termed the Ventral Attention System (VAS) (Vossel, et al., 2014). The VAS is suggested to be responsible for attentional reorientation during unexpected events (Corbetta, et al., 2008; Fox, et al., 2006; Vossel, et al., 2014). In the context of belief reasoning, such attentional requirements have long been proposed as

critical to judgements about an agent who holds an overtly false belief (e.g., Apperly, 2010; Frith & Frith 2003; Leslie, 1987; Schuwerk et al., 2015). Consequently, in addition to an effect of cTBS that is specific to belief and not desire reasoning, we were also interested as to whether any inter-individual variability in responsiveness to stimulation in false belief reasoning may be mediated by rTPJ, as identified by virtue of its resting state connectivity with the stimulation site, rvIPFC. Since rTPJ is both structurally and functionally connected to our rvIPFC stimulation site (Mars, et al., 2012), and rTPJ is consistently implicated in studies of belief reasoning (Schurz, et al., 2014), but also attentional switching (Scholz, et al., 2009), a secondary hypothesis was therefore that inter-individual variability in the effects of cTBS stimulation on false belief reasoning would be influenced by properties of both rvIPFC and rTPJ, and the degree of functional connectivity within the stimulated network that comprises them.

Method

Participants

Twenty-one right-handed adults (9 female; age range 19-28, mean age = 22 years) with no reported neurological or psychiatric history participated in the study. All were recruited through the University's research participation scheme, were given a safety information booklet regarding TMS and MRI prior to participating, and gave written, informed consent. Each was paid a small honorarium for their participation. The University of Birmingham STEM Ethics Committee approved the study.

Materials and Procedure

Pre-screen

A pre-screen was carried out prior to data collection in order to identify suitable participants. Suitability was determined on the basis of a TMS-MRI safety screening questionnaire (Rossi, et al., 2011) and, following in line with previous MRI work with this paradigm, their ability to perform the experimental task. Participants completed a computer-based, interactive training session that outlined the task, then completed two practice blocks. Only individuals who demonstrated no contraindications to TMS/MRI and could perform the experimental task to above chance (at $p < .05$) participated. 4 participants performed at, or below, chance so were not invited to participate beyond the pre-screen. The level is commensurate with performance levels we have seen in prior uses of this paradigm.

Neuroimaging Data Acquisition

On a separate day prior to TMS stimulation, the participants completed a R-fMRI scan, where they were instructed to lie still in the scanner with their eyes open. R-fMRI data were collected using a 3T Philips Achieva scanner with an 8-channel head coil. 192 T2*-weighted echo-planar imaging (EPI) volumes were obtained in a single ~8 minute run. These consisted of 42 axial slices obtained consecutively in a bottom up sequence, reconstructed voxel size = $3 \times 3 \times 3 \text{ mm}^3$. Whole brain coverage was achieved where TR = 2.5 s, TE = 35 ms, acquisition matrix = 80×80 , flip angle = 79° . High resolution T1-weighted structural images were acquired following collection of the functional data where 3D TFE, sagittal orientation, TR=8.4 ms, TE=3.8, 175 slices, reconstructed voxel size = $1 \times 1 \times 1 \text{ mm}^3$. Due to equipment failure, no R-fMRI data were collected for one participant (male, 22 years).

R-fMRI Data Analysis

Preprocessing and statistical analyses were performed using the FMRIB software library (FSL v.5.0.6; FMRIB Oxford, www.fmrib.ox.ac.uk/fsl). In short, initial preprocessing 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 150 s, then spatially smoothed with a 6mm full-width-half-maximum kernel, which has been shown optimal in R-fMRI (Van Dijk, et al., 2010). Prior to running independent components analysis, multi-stage registration was performed (MCFLIRT), with a resampling resolution corresponding to the functional data (3mm). Multi-session temporal concatenation was then implemented in MELODIC (v 3.14) (Beckmann and Smith, 2004), where the pre-processed R-fMRI data were whitened and projected into a 20-dimensional subspace, with the resulting group maps thresholded at $p < .001$. Using Independent Components Analysis (ICA), the data were decomposed into sets of vectors describing signal variation across temporal, subject and spatial domains. This approach was used to achieve an unbiased, model-free method of delineating region of interest (ROI) coordinates that reflected the resting-state network on which cTBS was applied.

A measure of network connectivity strength was also extracted for each participant using dual regression (Filippini, et al., 2009), implemented in FSL. First, for each subject, the group-average set of spatial maps from the ICA was regressed (as spatial regressors in a multiple regression) into the subject's 4D space-time dataset. This results in a set of subject-specific timeseries, one per group-level spatial map. Next, those timeseries were regressed (as temporal regressors, again in a multiple regression) into the same 4D dataset, resulting in a set of subject-specific spatial maps, one per group-level spatial map. The group component that reflected the network of interest was used to define an ROI, which was then applied to extract the mean strength of each individual's network. This approach

provided a measure of functional connectivity within the putative VAS for each participant, where higher values indicate a stronger degree of correlation across the areas within the network (methodology as per Sampaio-Baptista, et al., 2015; Stagg, et al., 2014).

ROI Generation and Structural Data Analyses

ROI masks were created using FSL command line tools. Each ROI contained voxels within a sphere with a 6mm radius, centred on the coordinates of interest. The first ROI reflected the TMS stimulation site – rvlPFC – MNI152 coordinates 50, 20, -6. This was based on group analyses from a different participant group who are described in Hartwright, et al. (2012). The coordinates for the rTPJ ROI were extracted based on group ICA of the R-fMRI data. First, the single R-fMRI component that yielded the largest Z-value in the stimulation coordinate set for rvlPFC was identified. The cluster peak for rTPJ was then extracted from within this component, to form the centre coordinates of the mask. By virtue of their featuring in a single component, these ROIs captured voxels that showed strong resting functional connectivity. Lastly, a control ROI was made. Also based on the group ICA, this was a cortical region that showed little functional connectivity with the TMS stimulation site, indexed by minimal correlation with rvlPFC. Having identified those voxels that showed minimal resting connectivity with rvlPFC, we selected from these the single voxel that had the closest proximity to rvlPFC to form the centroid of a control site ROI. Each mask was transformed into each participant's anatomical space using a linear transformation, FLIRT (v.6.0; Jenkinson, et al., 2002; Jenkinson and Smith, 2001) and then masked with the participant's brain extracted anatomical image to ensure that it comprised only voxels within the brain. Thus, a set of ROI masks, derived from group statistics, but spatially

transformed to reflect individual morphology, was produced for each participant.

Coordinates for the resultant ROIs are outlined in Table 1 in the Results section.

Non-brain tissue was extracted from each T1 structural image using BET (v.2.1; Smith, 2002). Automated tissue segmentation was performed on each brain extracted T1 structural image using FAST (v.5.0; Zhang, et al., 2001). This resulted in a value equating to the proportion of cerebrospinal fluid, GM and WM within each voxel, across the whole brain, for each participant. Using FSL command line tools, mean values of GM and WM were extracted from within each of the subject-specific ROIs. To control for any effects of brain size, total brain volume – defined as the combined volume of GM and WM – was also extracted for each participant from the brain extracted T1 structural image to use as a covariate of no interest.

TMS Procedure

While the effects of TMS on the primary motor cortex or visual areas are readily apparent (generating muscle twitches or phosphenes, respectively), other regions are typically 'silent' to TMS (de Graaf and Sack, 2011; Hardwick, et al., 2014; Jahanshahi and Rothwell, 2000). As the excitability of the primary motor cortex is easily quantified, many studies stimulating 'silent' regions have normalized the intensity of TMS they deliver based on motor excitability; cTBS is typically delivered at 80% of the active motor threshold. However, as the correlation between the excitability threshold of different brain regions is weak (Boroojerdi, et al., 2002; Gerwig, et al., 2003; Stewart, et al., 2001), it is questionable whether this normalization process is effective. Furthermore, our pilot work revealed that stimulating at the typical intensity of 80% of the active motor threshold led to painful facial contractions when targeting the vIPFC. The cTBS protocol used in the present study was

thus modified to 30-35% of stimulator output. This intensity was chosen as it minimized facial contractions during stimulation, and has previously been shown to be effective in modulating vIPFC function in the domain of attention (Verbruggen, et al., 2010).

The participants completed two TMS sessions, which were separated by one week. Each participant was assigned at random to receive TMS to either the site of interest – rvlPFC – or stimulation to a control site at the vertex, Cz, first. All participants completed both target and control site sessions. Participants were taken to a nearby computer to complete the behavioural task immediately following stimulation. The MNI152 coordinates for rvlPFC were transformed into individual coordinate sets in the current participant group using a series of transformation matrices generated with FLIRT (v.6.0; Jenkinson, et al., 2002; Jenkinson and Smith, 2001). The target site was then marked on each participant's anatomical image, visualised through Brainsight 2, a system for frameless stereotaxy (v2.2; Rogue Research Inc., Canada). The control site, Cz, was identified using skull landmarks and labelled with skin markers. During stimulation, participants were seated with their chin lowered onto a padded rest. A foam block was placed at the left side of the head to minimise movement. A cTBS paradigm comprising a 40 second train of uninterrupted TBS (600 pulses) – as outlined in Huang, et al. (2005)– was administered using a Magstim Rapid2 system (The Magstim Company, Whitland, UK) with a 70 mm figure-of-eight coil.

Theory of Mind Experiment Design

The experimental task was based on Hartwright, et al. (2012). Participants watched a sequence of events on a computer and identified whether a protagonist would feel happy or sad about the potential outcome of a virtual game show. The protagonist could win a finite

number of prizes in the game show, and these were hidden in various boxes. Prizes varied in desirability, so sometimes the protagonist would prefer not to win a particular prize in the hope of getting something more desirable. The task comprised an orthogonal design where the protagonist's Belief (true (B+) or false (B-)) and Desire state (approach (D+) or avoid (D-)) was systematically manipulated. This resulted in four equally occurring conditions B+D+, B+D-, B-D+, B-D-, each of which was repeated 16 times across the experiment. The protagonist's belief state was created by overtly stating whether they "correctly" (B+) or "mistakenly" (B-) believed that X was the case. Similarly, their desire state was created by stating whether the protagonist hoped to "find" (D+) or "avoid" (D-) the X, (see Fig. 1 for examples). The participants were required to make a left/right button response indicating whether they thought that the protagonist would be happy (left) or sad (right) about what the protagonist *believed* to be inside each prize box that was selected for him. Note that the decision was made prior to the protagonist knowing the true contents of the prize box. Presentation software (v. 14.1; Neurobehavioral Systems, CA) was used to randomise the presentation of trials, present the stimuli and record the behavioural response data simultaneously. RT and error data were collected. Behavioural analyses were performed in SPSS v. 22.0.0.1 64 bit (IBM, NY).

[FIGURE 1 ABOUT HERE]

Generation of Values for Linear Modelling of Structural Morphology and Responsiveness to TMS

On inspection of the participant's RT data it was apparent that the distribution function of RTs demonstrated significant positive skew, as is common with such data. Since

linear modelling operates under the assumption that the sample data are normally distributed, analysing non-normal data without taking this into account may affect the modelled results. RT data was therefore log transformed (see Hamada, et al., 2012 who also adopt such an approach with cTBS data). Summary measures of Belief (B) and Desire (D) by TMS Site (vIPFC/Cz) were then further computed as follows:

$$B_{vIPFC} = (RT(B-D+_{vIPFC}) + RT(B-D-_{vIPFC})) - (RT(B+D+_{vIPFC}) + RT(B+D-_{vIPFC}))$$

$$D_{vIPFC} = (RT(B+D-_{vIPFC}) + RT(B-D-_{vIPFC})) - (RT(B+D+_{vIPFC}) + RT(B-D+_{vIPFC}))$$

$$B_{Cz} = (RT(B-D+_{Cz}) + RT(B-D-_{Cz})) - (RT(B+D+_{Cz}) + RT(B+D-_{Cz}))$$

$$D_{Cz} = (RT(B+D-_{Cz}) + RT(B-D-_{Cz})) - (RT(B+D+_{Cz}) + RT(B-D+_{Cz}))$$

The effect of TMS was then calculated as a percentage ratio for B and D as follows:

$$\Delta B = (B_{vIPFC} / B_{Cz}) * 100$$

$$\Delta D = (D_{vIPFC} / D_{Cz}) * 100$$

Since the values of GM and WM are calculated as proportions (varying between 0-1) these measures were also transformed using a standard logit function,

$$\log(p/1-p)$$

where p is the proportion of each type of tissue in order to make them suitable for inclusion in the linear model.

Results

Effect of TMS on ToM Behavioural Data

Comparison with pilot data suggested that one participant's untransformed RT data and error pattern following TMS to the control site, Cz, was anomalous in all conditions; thus, this participant was excluded from further analyses presented here (male, 20 years). Their exclusion did not, however, alter any of the trends reported. All other participants' data were comparable with previous uses of this task. Overall, error rates were low, as highlighted in Fig. 2A. For ease of comparison with prior work using this experimental paradigm (Apperly, et al., 2011; Hartwright, et al., 2012; Hartwright, et al., 2013), the error and RT data were initially analysed each using a 3 factor ANOVA. The frequency of errors made per participant were input into a repeated measures ANOVA, with Belief (B+/B-), Desire (D+/D-) and Site (vlPFC/Cz) as within-subject factors. There was no effect of Belief or Site (both $F < 2.72$; $p > .12$). A significant main effect of Desire was identified, where error rate in D- > D+ ($F(1,19) = 9.19$, $p < .01$, $\eta^2 = .33$). No significant two or three way interactions between any of the factors were identified (all $p > .53$).

A 2x2x2 repeated measures ANOVA was conducted on the RT data, with Belief (B+/B-), Desire (D+/D-) and Site (vlPFC/Cz) as within-subject factors. As with previous work using this belief-desire reasoning framework (Apperly, et al., 2011; Hartwright, et al., 2012; Hartwright, et al., 2013), RT showed significant main effects of Belief, where B- > B+ ($F(1,19) = 58.05$, $p < .001$, $\eta^2 = .75$) and Desire, where D- > D+ ($F(1,19) = 188.58$, $p < .001$, $\eta^2 = .91$) (see Fig. 2B). There was, however, no effect of Site ($F(1,19) = .003$, $p = .96$). No statistically significant interactions were identified, including Site by Belief ($F(1,19) = .24$, $p = .63$).

[FIGURE 2 ABOUT HERE]

R-fMRI Data

Model free analysis yielded components from the R-fMRI data that were largely consistent with the prior literature, including Default Mode and attentional networks (e.g., Van Dijk, et al., 2010). With a Z-score of 7.92, the TMS stimulation site was identified in a bilateral fronto-parietal network comprising vIPFC, TPJ and dorsal mPFC in a component that explained 2.45% of the total variance (Fig. 3) (see Mars, et al., 2012 for a similar network configuration). Table 1 lists the local maxima from this network, including which of these were used to define the structural ROIs (rTPJ and the control ROI).

[TABLE 1 ABOUT HERE]

[FIGURE 3 ABOUT HERE]

There was some variability in network connectivity strength data; however, these data were consistent with the range of values returned previously published work (e.g., see Sampaio-Baptista, et al., 2015; Stagg, et al., 2014) and were normally distributed (mean = 3.71, SD = 2.86, Shapiro-Wilk=.923, $p = .13$). We performed a linear regression analysis using network strength as a predictor against the summary measure, BvIPFC. 95% confidence intervals (CIs) were estimated using the bias-corrected and accelerated (BCa) percentile bootstrap method (10,000 samples). This demonstrated that network strength was a significant predictor of false belief reasoning following stimulation to vIPFC ($F(1, 17) = 6.69$, $p < .05$, $R^2 = .28$, CI $-.116, -.005$). Consistent with our hypotheses, network strength was not a significant predictor for the other three summary measures, DvIPFC BCz DCz, (all F

$< .78, p > .39$; see Supporting Information Fig. S1). However, there was no statistically significant difference between regression slopes for BvIPFC and DvIPFC, suggesting that the strength of association with network connectivity strength did not differ between experimental conditions following cTBS ($Z = -.74, p = .46$)¹.

[FIGURE 4 ABOUT HERE]

Structural Morphology and Responsiveness to TMS

A Pearson correlation matrix comprising the exploratory variables – ΔB , connectivity strength, brain index (GM/WM) for ROI (rvIPFC/rTPJ) – was produced with 95% CI estimates using the BCa percentile bootstrap method (10,000 samples). Following adjustment of p -values for multiple comparisons using the Holm-Bonferroni procedure (Holm, 1979), this indicated that ΔB was positively correlated with WM in rvIPFC ($r(19) = .57, p' = .036, CI .03, .83$) and GM in rTPJ ($r(19) = .62, p' = .015, CI .02, .86$) and negatively correlated with network connectivity strength ($r(19) = -.57, p' = .036, CI -.81, .18$). The remaining variables correlated with ΔB were ns (all $p' > .54$; see Supporting Information Fig. S2a). When the procedure was repeated for ΔD , all were ns (all $p' > .27$; see Supporting Information Fig. S2b). A stepwise, backward elimination multiple regression was conducted to predict the effect of TMS on belief reasoning RT (ΔB) using those values that were correlated with ΔB : the measure of connectivity strength and the neural indices, WM in rvIPFC and GM in rTPJ. The model indicated that an increase in ΔB was linearly associated with increased WM in rvIPFC and GM in rTPJ ($F(2, 19) = 14.90, p < .001, R^2 = .64$), whereas connectivity strength

¹ Where $Z = (\text{beta1} - \text{beta2}) / \sqrt{\text{SE}[\text{beta1}]^2 + \text{SE}[\text{beta2}]^2}$

was excluded from the model as it held no predictive value. rTPJ GM contributed slightly more to the model than WM in vlPFC. All possible interaction effects encompassing any two- or three-way combination of the three initial predictors were ns. Table 2 shows regression coefficients for the resulting two predictor model (see Supporting Information Tables S1 and S2 for bootstrapped coefficients to compare predictors of ΔB applied to ΔD). To confirm that these effects were not simply driven by overall brain size, the above model was repeated including a measure of total brain volume as an additional predictor. Backwards regression analyses confirmed that this measure was not a significant predictor of ΔB ($SE B = -.02$; $t = -.09$; $p = .93$) and that WM in vlPFC and GM in rTPJ remained significant predictors when overall brain volume was included in the model.

In prior work, we have shown that variation in belief valence and desire valence both modulate TPJ (Hartwright, et al., 2012; Hartwright, et al., 2013). The lack of predictive value for TPJ GM in ΔD suggests that our model was not simply capturing underlying gross morphology mediating, for example, processing speed. However, to rule out this possibility, we confirmed that there was no linear relationship between rTPJ GM and B_{vlPFC} or B_{Cz} (both $p > .22$). This also held for vlPFC WM (both $p > .13$). This is important as, unlike B_{vlPFC} or B_{Cz} , the measure ΔB captured the *cost* of cTBS to vlPFC on RT, thus making it unlikely that the model was simply capturing base RT being driven by overall brain structure. As a further check, we assessed whether there was any linear relationship between ΔB and the control ROI, central opercular cortex – a region that demonstrated no functional connectivity, but close proximity – with the TMS site. None of the neural indices demonstrated a linear relationship with ΔB (all $p > .37$). Together, these analyses validate that TPJ GM and vlPFC WM were relevant to the measure of TMS cost to RT in belief reasoning, ΔB . The neural indices were not simply predictive of individual differences in base RT, but instead were able

to demonstrate how the behavioural consequence of TMS is mediated through an effect on a functional network.

[TABLE 2 ABOUT HERE]

Discussion

The present study provides the first evidence from TMS that vIPFC has a causal role in inhibition of self perspective during a “theory of mind” task. By applying cTBS, a TMS protocol thought to depress cortical excitability (Huang, et al., 2005; Ridding and Ziemann, 2010), we anticipated that stimulation to rvIPFC would affect those ToM states which feature incongruence between self and other, such as false belief, but that it would not impact on other behaviourally effortful ToM states, like avoidance desire. The induction of plasticity following TMS is mediated by numerous neurobiological, genetic and environmental factors (Nicolo, et al., 2015; Ridding and Ziemann, 2010). Among these factors, differing patterns of cTBS induced changes have been shown to relate to variation in at-rest oscillatory networks (McAllister, et al., 2013). Convergence of resting-state networks across EEG and fMRI modalities (Musso, et al., 2010), combined with the value of R-fMRI and structural MRI in predicting responsiveness to TMS (Brodie, et al., 2014; Conde, et al., 2012; Fox, et al., 2014), led the present study to evaluate whether individual differences in the strength of connectivity within the stimulated network, and structural brain indices extracted from relevant R-fMRI regions, could predict individual responsiveness to cTBS. Model free analysis was used to identify the stimulated network, and the degree of connectivity within this network was identified for each participant. The proportion of GM and WM was extracted from the stimulation site, rvIPFC, and from the area of rTPJ which

demonstrated resting functional connectivity with the stimulation site. Linear modelling was used to examine the predictive value of these brain indices against the cost of cTBS to participants' reaction times in belief and desire reasoning.

Behavioural performance in the present study was largely consistent with our prior work conducted without brain stimulation (Apperly, et al., 2011; Hartwright, et al., 2012; Hartwright, et al., 2013). Although few errors were made, longer response latencies suggested that participants found it more difficult to judge the feelings of an agent who was acting under a negatively valenced ToM state: a misinformed state, such as a false belief; or an aversive state, such as an avoidance desire. Previous behavioural work suggests that, whilst negatively valenced belief and desire states pose some shared cognitive demands, when the participant's own perspective is in conflict with the agent's, a distinctive cost is realised (Apperly, et al., 2011; German and Hehman, 2006). In the case of false belief reasoning, for example, one's own knowledge of the real state of affairs can interfere with making decisions about what another person might believe or do on the basis of that belief (Samson, et al., 2005). This egocentric bias, has been reported extensively in behavioural work with children and adults (see Birch and Bloom, 2007 for a review). Prior to the current experiment, a single-case study demonstrated that egocentrism in false belief reasoning follows damage to right lateral PFC (Samson, et al., 2005). fMRI has more finely localized this process to rvlPFC (van der Meer, et al., 2011) and this can be dissociated from negatively valenced desire reasoning where no self-other conflict exists (Hartwright, et al., 2012; Hartwright, et al., 2013). Nonetheless, despite this strong evidential background and replication of behavioural findings, the present study found no simple increase in egocentrism following cTBS. This might have led to the premature conclusion that, contrary

to the aforementioned literature, rvIPFC is not functionally relevant when processing self-other incongruent mental states.

Individual Responsiveness to cTBS is Informative to Understanding Brain Networks that Support ToM

R-fMRI suggested that the TMS site – rvIPFC – was part of a wider fronto-parietal network additionally encompassing rTPJ and dorsal mPFC. This replicates other studies that show resting functional connectivity across these regions (Fox, et al., 2006; Mars, et al., 2012). The current study demonstrated that the degree of individual connectedness within this network was linearly related with the ability to reason about an agent holding a false belief following stimulation to vIPFC. This relationship was not identified in the desire conditions. Furthermore, the use of structural brain indices extracted from this network supported a linear model of belief reasoning, where WM in rvIPFC and GM in rTPJ were significant predictors of the effect of TMS on RT. Crucially, this predictive value held for belief, but not desire, reasoning. It is important to reiterate here that the design of the experimental paradigm meant that only belief valence featured incongruence between self-other perspectives; desire valence in the format used here, and self-other congruence of desire, which was not a systematic variable in the present experiment (see Samson, et al., 2015 for such a manipulation within desire reasoning), are logically orthogonal factors (Hartwright, et al., 2012). As well as no association with desire reasoning, no predictive effect was found for base RT, or using indices extracted from the central opercular cortex – a spatially proximal, but functionally uncorrelated, region. Thus, among these variables that might plausibly have predicted the cost of TMS to belief RT, only morphological data from

the stimulation site, and a second region within the same functional network were significant predictors.

These data support our initial hypothesis that rvlPFC is functionally relevant when required to inhibit one's own perspective, and in addition the present result is informative because it casts light on the role of both vlPFC and TPJ in achieving this function. The interpretation of activation in rTPJ during ToM tasks has been a source of contention, where fMRI has implicated this region in both ToM and generic attentional selection (Mitchell, 2007; Rothmayr, et al., 2011). Mars, et al. (2012) were able to provide a potential resolution to this, by demonstrating subdivisions within rTPJ on the basis of functional and anatomical connectivity. Their data suggested that TPJ features an anterior/posterior divide, with the anterior region showing strongest connectivity with rvlPFC – typically a network associated with attention (Fox, et al., 2006) – and the posterior component showing stronger connectivity with mPFC – a network more commonly associated with ToM (Amodio and Frith, 2006). In the present experiment, there was a relationship between structural indices for rTPJ and the effect of TMS to rvlPFC on RT. Given evidence that structural characteristics, such as tissue anisotropy, have a considerable effect on the distribution of a TMS-induced field (De Lucia, et al., 2007), we believe that direct TMS stimulation to vlPFC not only affects the function of vlPFC, but also activity in the TPJ, with the size and nature of this effect depending upon structural characteristics of both regions. ICA of the R-fMRI data suggested that the stimulation site formed part of network additionally comprising TPJ, consistent with a fronto-parietal network termed the VAS (Corbetta, et al., 2008; Fox, et al., 2006; Vossel, et al., 2014). Although the source of debate, the VAS has been suggested to reorient attention away from one's own perspective towards the other's perspective, with reorientation being more demanding when these perspectives are conflicting (as in false belief trials) (Corbetta,

et al., 2008). This viewpoint is in line with work that has demonstrated variation in cortical excitability as a function of attentional state (Conte, et al., 2007): the current paradigm was designed such that only in the case of false belief reasoning featured a conflicting perspective, giving rise to broader attentional demands not present in the other experimental conditions. Hence, following cTBS, the participants' reaction times in false belief reasoning become more tightly coupled with the stimulated – and functionally relevant – network.

Returning to the controversy about the role of rTPJ in ToM (e.g., Mitchell, 2007; Rothmayr, et al., 2011; Young, et al., 2010), Geng and Vossel (2013) propose a unifying framework which characterises TPJ more broadly, taking into account the known subdivisions described previously. They suggest that TPJ constitutes an attentional mechanism for 'contextual updating' – that is, to update one's internal environment in order to construct appropriate responses or expectations. When considered alongside relevant anatomy, connections within the superior longitudinal fasciculus between TPJ and vIPFC are consistent with an attentional process that engages TPJ for environmental preparation and, where required, employing vIPFC to regulate or resolve conflict respectively. For the present data, we suggest that the GM density of TPJ is relevant because of the role of TPJ in computing responses according to the target character's mental states, which in turn determine the degree to which there is competition that must be resolved at a later point in processing. Indeed, the spatio-temporal dynamics of ToM support a process where activation in TPJ precedes that in vIPFC (McCleery, et al., 2011). In the present study, cTBS added variance to the normal functioning of rvIPFC, which was amplified by individual differences in the underlying network. What is important to reiterate, however, is that these data suggest that TPJ is relevant to contextual updating,

but far from being an unintended confound in ToM tasks, attentional control is integral to the ability to think about perspectives other than our own.

Conclusion and Applications

The current study provides critical new evidence on the causal role for rvIPFC in ToM. While previous fMRI studies provided evidence that activity in rvIPFC was correlated with demands for self-perspective inhibition, the only evidence for a causal role came from neuropsychological studies of patients with lesions affecting many other brain areas besides rvIPFC. Furthermore, of broader relevance, the data presented here highlight how individual differences in functional connectivity and structural morphology may affect behavioural outcomes following cTBS. Our work suggests that the effect of cTBS was anti-correlated with R-fMRI network connectivity strength, where weak network strength was associated with an increased effect of CTBS during false belief reasoning. Nonetheless, there was no difference in the degree of association between network strength and belief versus desire reasoning. Moreover, the full model indicated that network strength held no predictive value over and above the morphology data. One possible explanation for this is that functional connectivity reflects the underlying neural morphology; thus, the present result may mirror this coupling. Regardless, though not something anticipated in the current study, our approach highlights the need for further work to verify such network effects of stimulation.

What cannot be determined from the present study are the physiological mechanisms underlying the different response patterns to TMS. What has been measured here is mediation of an activated network, on the basis of its underlying morphology and state. These interact with the demands of a task to produce the outcome across a

population of neurons (Miniussi, et al., 2010). Likewise, it is not possible to tell with these measures how GM is mediating responsiveness, for example, in terms of atrophy or gyrification. It remains for future work to determine whether other indices that may also influence the response to TMS, such as attentional state, genetic factors and physical fitness (Nicolo, et al., 2015; Ridding and Ziemann, 2010) might affect stimulation outcomes. Nonetheless, although network state, for example, may further mediate both inter- and intra-individual responses to TMS in overt outcomes such as RT (Vernet, et al., 2013), the fact that both relatively static, gross indices of network morphology extracted from broader measures of functional connectivity explain responsiveness to TMS makes the present finding methodologically important. For studies already employing neuro-navigated TMS, R-fMRI presents a short, straightforward acquisition, permitting a relatively inexpensive, but informative, inclusion within a standard MRI protocol. We demonstrate here that the inclusion of such data may be particularly fruitful for understanding the behavioural changes following cTBS. The nuanced nature of the present model means that it may not be valid in other domains; this is something worthy of evaluation. However, what is being put forward is the utility of including relevant, data-driven neural indices when examining the effect of cTBS. In the present circumstance, such an approach was critical to understanding network phenomena in a complex neurocognitive domain.

Acknowledgements

This research was supported by the Economic and Social Research Council (grant number ES/G01258X/1). The authors would like to thank Francesco Sella for his helpful comments on this manuscript.

Conflict of Interest

The authors have no conflict of interest to declare.

References

- 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., 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 Development*, 82:1691-1703.
- Beckmann, C.F., Smith, S.M. (2004) Probabilistic independent component analysis for functional magnetic resonance imaging. *IEEE transactions on medical imaging*, 23:137-52.
- Begeer, S., Bernstein, D.M., van Wijhe, J., Scheeren, A.M., Koot, H.M. (2012) A continuous false belief task reveals egocentric biases in children and adolescents with Autism Spectrum Disorders. *Autism*, 16:357-366.
- Birch, S.A., Bloom, P. (2007) The curse of knowledge in reasoning about false beliefs. *Psychol Sci*, 18:382-6.
- Boroojerdi, B., Meister, I.G., Foltys, H., Sparing, R., Cohen, L.G., Topper, R. (2002) Visual and motor cortex excitability: a transcranial magnetic stimulation study. *Clinical neurophysiology : official journal of the International Federation of Clinical Neurophysiology*, 113:1501-4.
- Brodie, S.M., Borich, M.R., Boyd, L.A. (2014) Impact of 5-Hz rTMS over the primary sensory cortex is related to white matter volume in individuals with chronic stroke. *Eur J Neurosci*, 40:3405-12.
- 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.
- Conde, V., Vollmann, H., Sehm, B., Taubert, M., Villringer, A., Ragert, P. (2012) Cortical thickness in primary sensorimotor cortex influences the effectiveness of paired associative stimulation. *Neuroimage*, 60:864-70.
- Conte, A., Gilio, F., Iezzi, E., Frasca, V., Inghilleri, M., Berardelli, A. (2007) Attention influences the excitability of cortical motor areas in healthy humans. *Exp Brain Res*, 182:109-17.
- Corbetta, M., Patel, G., Shulman, G.L. (2008) The Reorienting System of the Human Brain: From Environment to Theory of Mind. *Neuron*, 58:306-324.
- Damoiseaux, J.S., Rombouts, S., Barkhof, F., Scheltens, P., Stam, C.J., Smith, S.M., Beckmann, C.F. (2006) Consistent resting-state networks across healthy subjects. *Proc. Natl. Acad. Sci. U. S. A.*, 103:13848-13853.
- de Graaf, T.A., Sack, A.T. (2011) Null results in TMS: from absence of evidence to evidence of absence. *Neuroscience and biobehavioral reviews*, 35:871-7.
- De Lucia, M., Parker, G.J.M., Embleton, K., Newton, J.M., Walsh, V. (2007) Diffusion tensor MRI-based estimation of the influence of brain tissue anisotropy on the effects of transcranial magnetic stimulation. *Neuroimage*, 36:1159-1170.
- Filippini, N., MacIntosh, B.J., Hough, M.G., Goodwin, G.M., Frisoni, G.B., Smith, S.M., Matthews, P.M., Beckmann, C.F., Mackay, C.E. (2009) Distinct patterns of brain activity in young carriers of the APOE-ε4 allele. *Proc. Natl. Acad. Sci. U. S. A.*, 106:7209-7214.
- Fox, M.D., Buckner, R.L., Liu, H., Chakravarty, M.M., Lozano, A.M., Pascual-Leone, A. (2014) Resting-state networks link invasive and noninvasive brain stimulation across diverse psychiatric and neurological diseases. *Proceedings of the National Academy of Sciences*.

- Fox, M.D., Corbetta, M., Snyder, A.Z., Vincent, J.L., Raichle, M.E. (2006) Spontaneous neuronal activity distinguishes human dorsal and ventral attention systems. *Proceedings of the National Academy of Sciences*, 103:10046-10051.
- Geng, J.J., Vossel, S. (2013) Re-evaluating the role of TPJ in attentional control: Contextual updating? *Neuroscience and biobehavioral reviews*, 37:2608-2620.
- German, T., Hehman, J. (2006) Representational and executive selection resources in 'theory of mind': Evidence from compromised belief-desire reasoning in old age. *Cognition*, 101:129-152.
- Gerwig, M., Kastrup, O., Meyer, B.U., Niehaus, L. (2003) Evaluation of cortical excitability by motor and phosphene thresholds in transcranial magnetic stimulation. *Journal of the neurological sciences*, 215:75-8.
- Greicius, M. (2008) Resting-state functional connectivity in neuropsychiatric disorders. *Current opinion in neurology*, 21:424-430.
- Hamada, M., Murase, N., Hasan, A., Balaratnam, M., Rothwell, J.C. (2012) The Role of Interneuron Networks in Driving Human Motor Cortical Plasticity. *Cerebral Cortex*, 23:1593-1605.
- Hardwick, R.M., Lesage, E., Miall, R.C. (2014) Cerebellar transcranial magnetic stimulation: the role of coil geometry and tissue depth. *Brain stimulation*, 7:643-9.
- Hartwright, C.E., Apperly, I.A., Hansen, P.C. (2012) Multiple roles for executive control in belief-desire reasoning: Distinct neural networks are recruited for self perspective inhibition and complexity of reasoning. *Neuroimage*, 61:921-930.
- Hartwright, C.E., Apperly, I.A., Hansen, P.C. (2013) Representation, Control, or Reasoning? Distinct Functions for Theory of Mind within the Medial Prefrontal Cortex. *J. Cogn. Neurosci.*
- Holm, S. (1979) A Simple Sequentially Rejective Multiple Test Procedure. *Scandinavian Journal of Statistics*, 6:65-70.
- Huang, Y.-Z., Edwards, M.J., Rounis, E., Bhatia, K.P., Rothwell, J.C. (2005) Theta Burst Stimulation of the Human Motor Cortex. *Neuron*, 45:201-206.
- Jahanshahi, M., Rothwell, J. (2000) Transcranial magnetic stimulation studies of cognition: an emerging field. *Experimental Brain Research*, 131:1-9.
- Jenkinson, M., Bannister, P., Brady, M., Smith, S. (2002) Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage*, 17:825-41.
- Jenkinson, M., Smith, S. (2001) A global optimisation method for robust affine registration of brain images. *Med Image Anal*, 5:143-56.
- Keller, C.J., Bickel, S., Entz, L., Ulbert, I., Milham, M.P., Kelly, C., Mehta, A.D. (2011) Intrinsic functional architecture predicts electrically evoked responses in the human brain. *Proc Natl Acad Sci U S A*, 108:10308-13.
- Li, R., Qin, W., Zhang, Y., Jiang, T., Yu, C. (2012) The neuronal correlates of digits backward are revealed by voxel-based morphometry and resting-state functional connectivity analyses. *PLoS One*, 7:e31877.
- Lieberman, M.D. (2007) Social Cognitive Neuroscience: A Review of Core Processes. *Annual Review of Psychology*, 58:259-289.
- Magistro, D., Takeuchi, H., Nejad, K.K., Taki, Y., Sekiguchi, A., Nouchi, R., Kotozaki, Y., Nakagawa, S., Miyauchi, C.M., Iizuka, K., Yokoyama, R., Shinada, T., Yamamoto, Y., Hanawa, S., Araki, T., Hashizume, H., Sassa, Y., Kawashima, R. (2015) The Relationship between Processing Speed and Regional White Matter Volume in Healthy Young People. *PLoS One*, 10:e0136386.
- Mar, R.A. (2011) The neural bases of social cognition and story comprehension. *Annual Review of Psychology*, 62:103-34.

- Mars, R.B., Sallet, J., Schuffelgen, 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* (New York, N.Y. : 1991), 22:1894-903.
- McAllister, C.J., Ronnqvist, K.C., Stanford, I.M., Woodhall, G.L., Furlong, P.L., Hall, S.D. (2013) Oscillatory Beta Activity Mediates Neuroplastic Effects of Motor Cortex Stimulation in Humans. *J. Neurosci.*, 33:7919-7927.
- McCleery, J.P., Surtees, A.D., Graham, K.A., Richards, J.E., Apperly, I.A. (2011) The neural and cognitive time course of theory of mind. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 31:12849-54.
- Miniussi, C., Ruzzoli, M., Walsh, V. (2010) The mechanism of transcranial magnetic stimulation in cognition. *Cortex*, 46:128-30.
- Mitchell, J.P. (2007) Activity in Right Temporo-Parietal Junction is Not Selective for Theory-of-Mind. *Cerebral Cortex*, 18:262-271.
- Musso, F., Brinkmeyer, J., Mobascher, A., Warbrick, T., Winterer, G. (2010) Spontaneous brain activity and EEG microstates. A novel EEG/fMRI analysis approach to explore resting-state networks. *Neuroimage*, 52:1149-61.
- Nicolo, P., Ptak, R., Guggisberg, A.G. (2015) Variability of behavioural responses to transcranial magnetic stimulation: Origins and predictors. *Neuropsychologia*, 74:137-144.
- Pascual-Leone, A., Walsh, V., Rothwell, J. (2000) Transcranial magnetic stimulation in cognitive neuroscience--virtual lesion, chronometry, and functional connectivity. *Current Opinion in Neurobiology*, 10:232-7.
- Reineberg, A.E., Andrews-Hanna, J.R., Depue, B.E., Friedman, N.P., Banich, M.T. (2015) Resting-state networks predict individual differences in common and specific aspects of executive function. *Neuroimage*, 104:69-78.
- Ridding, M.C., Ziemann, U. (2010) Determinants of the induction of cortical plasticity by non-invasive brain stimulation in healthy subjects. *The Journal of Physiology*, 588:2291-2304.
- Rossi, S., Hallett, M., Rossini, P.M., Pascual-Leone, A. (2011) Screening questionnaire before TMS: An update. *Clinical Neurophysiology*, 122:1686-1686.
- Rothmayr, C., Sodian, B., Hajak, G., Döhnell, K., Meinhardt, J., Sommer, M. (2011) Common and distinct neural networks for false-belief reasoning and inhibitory control. *Neuroimage*, 56:1705-1713.
- Sampaio-Baptista, C., Filippini, N., Stagg, C.J., Near, J., Scholz, J., Johansen-Berg, H. (2015) Changes in functional connectivity and GABA levels with long-term motor learning. *Neuroimage*, 106:15-20.
- Samson, D., Apperly, I.A., Kathirgamanathan, U., Humphreys, G.W. (2005) Seeing it my way: a case of a selective deficit in inhibiting self-perspective. *Brain*, 128:1102-11.
- Samson, D., Houthuys, S., Humphreys, G.W. (2015) Self-perspective inhibition deficits cannot be explained by general executive control difficulties. *Cortex*.
- 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:e4869.
- Schurz, M., Radua, J., Aichhorn, M., Richlan, F., Perner, J. (2014) Fractionating theory of mind: A meta-analysis of functional brain imaging studies. *Neuroscience and biobehavioral reviews*.
- Schurz, M., Tholen, M.G. (2016) What brain imaging did (not) tell us about the Inferior Frontal Gyrus in theory of mind – A commentary on Samson et al., (2015). *Cortex*, 74:329-333.

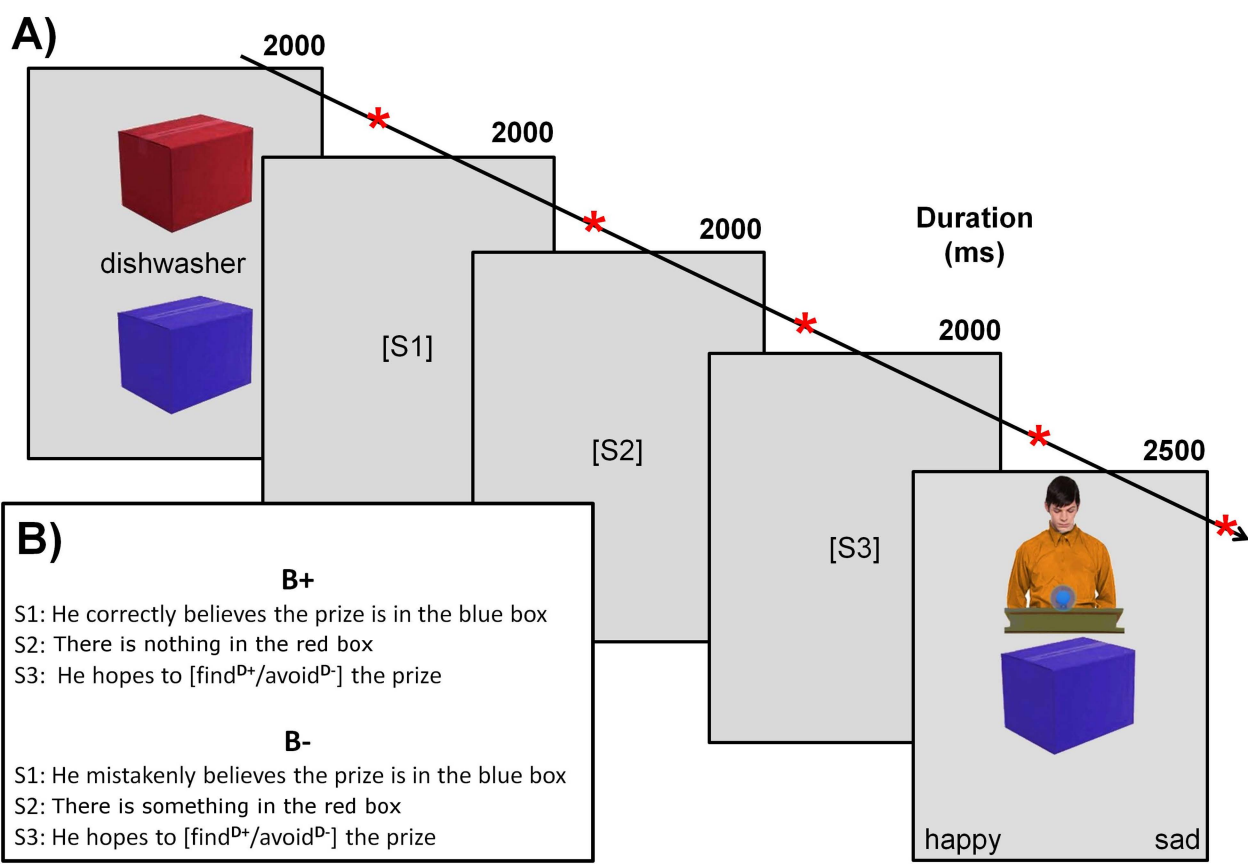
- Smith, S.M. (2002) Fast robust automated brain extraction. *Human Brain Mapping*, 17:143-155.
- Stagg, C.J., Bachtiar, V., Amadi, U., Gudberg, C.A., Ilie, A.S., Sampaio-Baptista, C., O'Shea, J., Woolrich, M., Smith, S.M., Filippini, N., Near, J., Johansen-Berg, H. (2014) Local GABA concentration is related to network-level resting functional connectivity. *eLife*, 3:e01465.
- Stewart, L.M., Walsh, V., Rothwell, J.C. (2001) Motor and phosphene thresholds: a transcranial magnetic stimulation correlation study. *Neuropsychologia*, 39:415-9.
- Takeuchi, H., Taki, Y., Sassa, Y., Hashizume, H., Sekiguchi, A., Nagase, T., Nouchi, R., Fukushima, A., Kawashima, R. (2012) Regional gray and white matter volume associated with Stroop interference: evidence from voxel-based morphometry. *Neuroimage*, 59:2899-907.
- 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*, 56:2364-2374.
- Van Dijk, K.R., Hedden, T., Venkataraman, A., Evans, K.C., Lazar, S.W., Buckner, R.L. (2010) Intrinsic functional connectivity as a tool for human connectomics: theory, properties, and optimization. *Journal of neurophysiology*, 103:297-321.
- Verbruggen, F., Aron, A.R., Stevens, M.A., Chambers, C.D. (2010) Theta burst stimulation dissociates attention and action updating in human inferior frontal cortex. *Proceedings of the National Academy of Sciences*, 107:13966-13971.
- Vernet, M., Bashir, S., Yoo, W.K., Oberman, L., Mizrahi, I., Ifert-Miller, F., Beck, C.J., Pascual-Leone, A. (2013) Reproducibility of the effects of theta burst stimulation on motor cortical plasticity in healthy participants. *Clinical Neurophysiology*.
- Vogeley, K., Bussfeld, P., Newen, A., Herrmann, S., Happé, F., Falkai, P., Maier, W., Shah, N.J., Fink, G.R., Zilles, K. (2001) Mind Reading: Neural Mechanisms of Theory of Mind and Self-Perspective. *Neuroimage*, 14:170-181.
- Vossel, S., Geng, J.J., Fink, G.R. (2014) Dorsal and Ventral Attention Systems: Distinct Neural Circuits but Collaborative Roles. *The Neuroscientist*, 20:150-159.
- 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.
- Zhang, Y., Brady, M., Smith, S. (2001) Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE transactions on medical imaging*, 20:45-57.

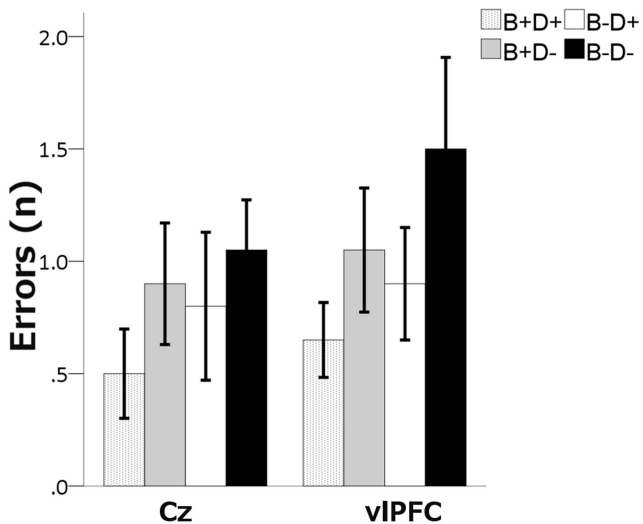
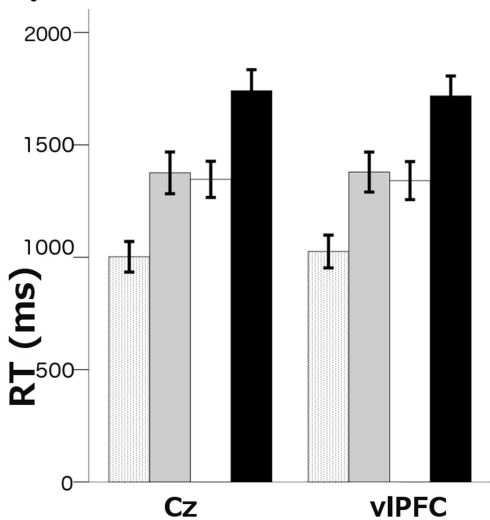
Figure Captions

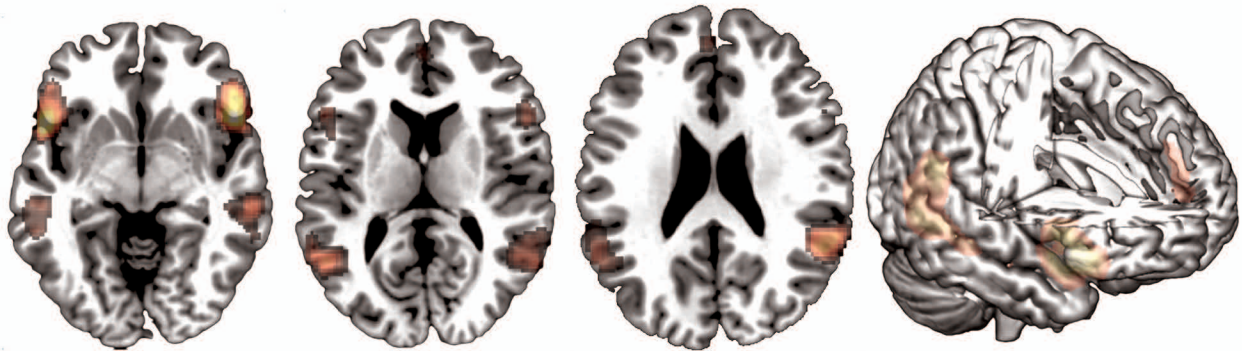
Fig. 1. (Panel A) Schematic example of a single trial. The vertical presentation of the box colour (red/blue) was randomised. The red star * indicates a blank screen shown for 500ms to reduce eyestrain. The colour (red/blue) of the final box in the sequence was randomised. (Panel B) Example statements for true (B+) and false (B-) belief scenarios. The temporal order of these statements was randomised. Where text is written within [], this denotes that the statement would contain only one of those options, dependent on whether the trial was an approach (D+) or avoidance desire (D-) condition; e.g., He hopes to avoid the prize.

Fig. 2. Error bars reflect ± 1 SE of the mean. (Panel A) Mean frequency of errors made within condition and site; total trials per condition = 16 (Panel B) Mean RT for each condition and site in milliseconds. Data shown is untransformed for ease of interpretation.

Fig. 3. Group R-fMRI data overlaid onto a template brain, where $Z > 4.0$. Slices from left to right show $Z=-6$, $Z=12$, $Z=24$, which reflect centre coordinates for ROI in rvlPFC, left TPJ and rTPJ respectively. Lateral view shows right hemisphere.



A)**B)**



Supporting Information

Figure caption S1 (see S1.tif)

Fig. S1. Correlation matrix for R-fMRI network strength plotted against transformed RT measures for Belief and Desire, by stimulation site (Cz/vIPFC).

Figure caption S2 (see S2a.tif and S2b.tif)

Fig. S2. Correlation matrices for showing cost of stimulation to Belief (ΔB ; panel A) and Desire (ΔD ; panel B) plotted against network strength and structural and indices.

Table S1

Forced Entry Regression with Bootstrap for Coefficients: Model ΔB

	<i>B</i>	<i>SE B</i>	β	<i>p</i> ¹	CI ²
Constant (ΔB)	246.17	60.51		0.002	147.37, 322.11
WM rvIPFC	58.78	20.10	.50	0.011	22.19, 85.11
GM rTPJ	203.99	76.73	.56	0.049	51.28, 293.85

Note. $adjR^2 = 0.594$. ¹ Values reflect 2-tailed. ² CI reflect 95%; BCa percentile bootstrapping

10,000 samples

Table S2Forced Entry Regression with Bootstrap for Coefficients: Model ΔD

	<i>B</i>	<i>SE B</i>	β	<i>p</i> ¹	CI ²
Constant (ΔD)	124.00	86.62		0.180	-54.80, 281.86
WM rvIPFC	4.23	25.47	.04	0.876	-52.39, 53.45
GM rTPJ	44.61	79.15	.14	0.581	-94.92, 302.63

Note. $adjR^2 = -0.094$. ¹ Values reflect 2-tailed. ² CI reflect 95%; BCa percentile bootstrapping

10,000 samples

