

Attentional processes, not implicit mentalizing, mediate performance in a perspective-taking task: Evidence from stimulation of the temporoparietal junction

Idalmis Santiesteban¹, Simran Kaur², Geoffrey Bird^{3,4} and Caroline Catmur^{2,5}

¹ Department of Psychology, University of Cambridge, Downing Street, Cambridge, CB2 3EB, UK.

² School of Psychology, University of Surrey, Guildford, Surrey, GU2 7XH, UK.

³ Department of Experimental Psychology, University of Oxford, 9 South Parks Rd, Oxford, OX1 3UD, UK.

⁴ MRC Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Kings College London, DeCrespigny Park, London, SE5 8AF, UK.

⁵ Department of Psychology, Institute of Psychiatry, Psychology & Neuroscience, King's College London, SE1 1UL, UK.

Correspondence concerning this article should be addressed to Idalmis Santiesteban, is405@cam.ac.uk; or Caroline Catmur, caroline.catmur@kcl.ac.uk

Abstract

Mentalizing is a fundamental process underpinning human social interaction. Claims of the existence of 'implicit mentalizing' represent a fundamental shift in our understanding of this important skill, suggesting that preverbal infants and even animals may be capable of mentalizing. One of the most influential tasks supporting such claims in adults is the dot perspective-taking task, but demonstrations of similar performance on this task for mentalistic and non-mentalistic stimuli have led to the suggestion that this task in fact measures domain-general processes, rather than implicit mentalizing. A mentalizing explanation was supported by fMRI data claiming to show greater activation of brain areas involved in mentalizing, including right temporoparietal junction (rTPJ), when participants made self-perspective judgements in a mentalistic, but not in a non-mentalistic condition, an interpretation subsequently challenged. Here we provide the first causal test of the mentalizing claim using disruptive transcranial magnetic stimulation of rTPJ during self-perspective judgements. We found no evidence for a distinction between mentalistic and non-mentalistic stimuli: stimulation of rTPJ impaired performance on all self-perspective trials, regardless of the mentalistic/non-mentalistic nature of the stimulus. Our data support a domain-general attentional interpretation of performance on the dot perspective-taking task, a role which is subserved by the rTPJ.

Keywords: Automatic attentional orienting; attentional pop-out; dot perspective-taking task; implicit mentalizing; perspective-taking; sub-mentalizing; temporoparietal junction; repetitive transcranial magnetic stimulation.

Mentalizing, the ability to attribute mental states to oneself and others, is a fundamental process underpinning human social interaction. Although generally assumed to be an explicit process, requiring conscious thought and cognitive flexibility, there have been recent claims that mentalizing can also be *implicit* - that it is a fast and efficient process that occurs automatically, without conscious awareness (Apperly, 2011; Apperly & Butterfill, 2009, Frith & Frith, 2012). Claims of implicit mentalizing represent a fundamental shift in our understanding of this important skill, with suggestions that it is present in pre-linguistic infants (Baillargeon, Scott, & He, 2010; Onishi & Baillargeon, 2005) and in a variety of social animals (e.g. Premack & Woodruff, 1978; Call, 2012; Krupenye, Kano, Hirata, Call & Tomasello, 2016) – although, for contrasting views see De Bruin and Newen (2012), Heyes (2014a, 2014b, 2017), Penn and Povinelli (2007), Perner and Ruffman (2005), Phillips et al. (2015), and Ruffman, Taumoepeau and Perkins (2012).

Recent studies have spurred controversy by claiming that implicit mentalizing persists in adulthood. Evidence for this claim comes from visual perspective-taking studies using a paradigm known as the ‘dot perspective-taking task’ (henceforth ‘the dots task’; e.g. Samson, Apperly, Braithwaite, Andrews, & Bodley Scott, 2010; McCleery, Surtees, Graham, Richards, & Apperly, 2011; Qureshi, Apperly, & Samson, 2010).

In the dots task, participants are presented with a word cue indicating whether they will be required to adopt their own perspective (“YOU”: ‘self-perspective’ trials) or someone else’s (“SHE”/“HE”: ‘non-self-perspective’ trials), before the appearance of a number cue (0-3), followed by a picture of a room containing large circles/dots pinned on the wall. In the centre of the room, there is a human-like figure or avatar

facing either the left or right wall. The participant's task is to verify if the cued number corresponds to the number of dots that they (self-perspective trials) or the avatar (non-self-perspective trials) can see. Depending on the location of the dots, sometimes the number of dots that can be seen is the same for both participant and avatar (consistent trials), whereas sometimes the number of dots is different across the two perspectives (inconsistent trials); see Figure 1. A robust finding from all previous studies using this task is that participants' responses are slower in inconsistent compared to consistent trials. Furthermore, this effect is found even when participants make judgements on self-perspective trials and thus do not need to take into account the avatar's perspective. This 'self-consistency effect' has been interpreted as evidence of implicit mentalizing: participants automatically adopt the other person's perspective and seem unable to ignore it, even when they are only required to adopt their own perspective (Samson et al., 2010). However, the implicit mentalizing interpretation has been criticized because the task lacked a non-mentalistic control condition. When such controls are included (e.g. Cole, Atkinson, Le & Smith, 2016; Conway, Lee, Ojaghi, Catmur & Bird, 2017; Santiesteban, Catmur, Coughlan Hopkins, Bird & Heyes, 2014; Schurz et al., 2015), results suggest that domain-general attentional processes, rather than a domain-specific process such as implicit mentalizing, underlie performance on the task. However, a recent neuroimaging study claimed to have found evidence of domain specificity at the neural level using the dots task (Schurz et al., 2015). Schurz and colleagues reported greater activation of brain regions generally associated with mentalizing such as rTPJ, medial prefrontal cortex (mPFC) and ventral precuneus when participants made self-

perspective judgements in the mentalistic (avatar) but not in the non-mentalistic (arrow) condition.

We recently suggested that neuroimaging methods are ill-suited to address claims of implicit mentalizing due to the fact that, under an implicit mentalizing account, the presence of a mentalistic stimulus is sufficient to prompt the mentalizing process. Thus, it is impossible to determine whether differential activation is caused by the stimulus (the avatar), or the process of interest (mentalizing), when contrasted with a non-mentalistic stimulus such as an arrow (see Catmur, Santiesteban, Conway, Heyes & Bird, 2016). In the present study, we use both behavioural (Experiment 1) and brain stimulation (disruptive repetitive transcranial magnetic stimulation – rTMS – of rTPJ, Experiment 2) methods to provide an empirical test of the claim that rTPJ is involved in representing another’s visual perspective during self-perspective judgements for mentalistic, but not for non-mentalistic, stimuli. In both experiments all participants completed the dots task in two stimulus conditions, where the central stimulus was either mentalistic (avatar) or non-mentalistic (arrow). Should stimulation of rTPJ result in impairment of self-perspective judgements in the avatar but not in the arrow condition, this would provide support for the domain-specific claim. Conversely, if stimulation of rTPJ fails to distinguish between the avatar and arrow trials, this would favour a domain-general attentional interpretation of performance on this task.

Although domain-general accounts of performance on the dots task have been proposed, the nature of any such domain-general processes has been relatively under-specified and, as far as we are aware, no study has provided positive evidence for their existence. Consideration of the task demands of the different conditions can help

elucidate the nature of any such processes. For example, on self-perspective trials, the participant must overcome any attentional cuing effect of the avatar and arrow, and re-orient their attention to scan the whole room for the presence of dots (both in front of and behind the central stimulus). In contrast, on non-self-perspective trials the participant does not need to reorient their attention after it has been allocated to the side of the room cued by the central stimulus, as this is the only side that must be searched for dots. This analysis would indicate that domain-general processes involved in attentional reorienting should be required on self-perspective, but not on non-self-perspective trials. Another possibility is that the saliency of the dots makes them 'pop-out' compared to the background. On self-perspective trials, participants could use attentional processes in combination with this pop-out effect to select all the dots, following which the number of dots would be automatically subitized (Sathian et al., 1999). The use of attentional selection to profit from this 'pop-out and subitization' process would be helpful on self-perspective trials, as it would result in the correct number of dots being identified; but on non-self-perspective trials, such attentional selection of all red dots would be counterproductive. Again, this analysis indicates that different domain-general attentional processes would be involved on self-perspective than on non-self-perspective trials.

Crucially, previous fMRI studies using the dots task have reported stronger activation of rTPJ for self- than for non-self-perspective judgements (Ramsey, Hansen, Apperly & Samson, 2013; Schurz et al., 2015); a finding which is consistent with the task-demand analyses above, given that the TPJ has a well-documented role in certain domain-general attentional processes including attentional reorienting and visual pop-out (Buschman & Miller, 2007; Corbetta & Shulman, 2002; Ellison, Schindler, Pattison,

& Milner, 2004; Geng & Vossel, 2013; Pollmann et al., 2003), but not in others such as endogenous orienting of attention (Thiel, Zilles & Fink, 2004). Therefore, a domain-general attentional account of performance on this task would be supported by data whereby stimulation of rTPJ fails to distinguish between mentalistic and non-mentalistic trials during self-perspective judgements, yet selectively affects self-perspective trials compared to non-self-perspective trials.

Experiment 1

The aim of this behavioural experiment was to a) replicate our previous findings (Santiesteban et al., 2014) that the consistency effect – faster responding for consistent than inconsistent trials – is also elicited by a non-mentalistic, but directional, object such as an arrow; and b) verify that optimising the number of trials for the rTMS study, by inclusion of mismatching trials (see methods below), does not eliminate the consistency effect for either self- or non-self-perspective judgements.

Method

Participants

Sixteen healthy adults (10 males; age range: 18 – 47 years, $M = 24.6$, $SD = 7.6$) volunteered to take part in this study. Fifteen were right-handed. Since performance of the left-handed participant did not differ from the group mean, their data were included in the reported analysis.

Stimuli and Procedure

Figure 1 shows examples of the stimuli presented to participants. The image files were those used by Samson et al. (2010) and Santiesteban et al. (2014). The central

stimulus was either an avatar or an arrow. There was a male and a female avatar (presented to male and female participants, respectively), and two arrows with colour palettes and colour distributions matched to those of the male and female avatars. The arrows also matched the avatars in height (5.84° of visual angle) and area.

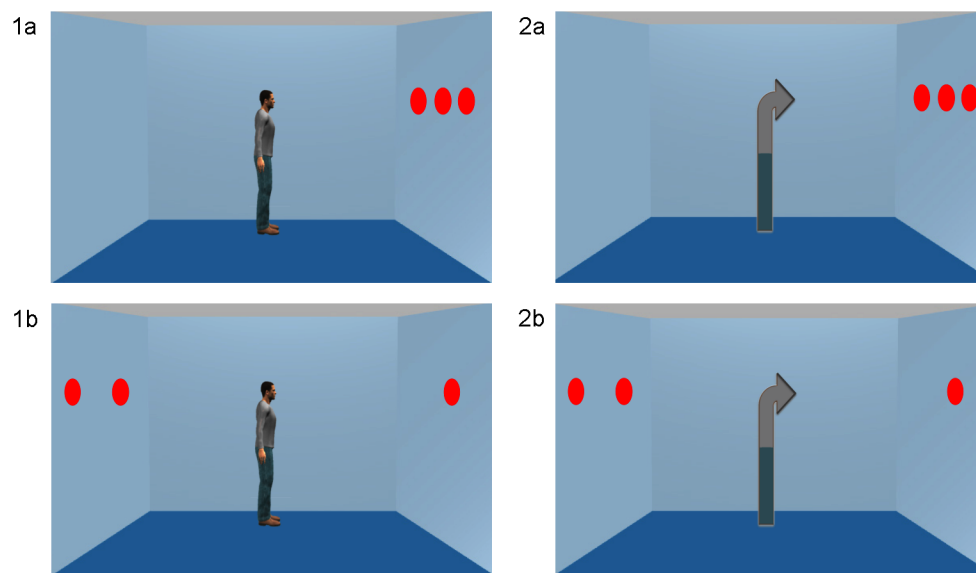


Figure 1. Examples of the stimuli. The avatar and arrow trials were either consistent (1a, 2a) or inconsistent (1b, 2b) with the participant's perspective.

Details of the task procedure are described in Samson et al. (2010, Experiment 1) and Santiesteban et al. (2014). As described in the Introduction, participants were required to verify if a previously seen number cue corresponded to the number of dots displayed in the stimulus picture either from their own visual perspective (self-perspective trials), the avatar's perspective (non-self-perspective avatar trials), or to which the arrow was pointing (non-self-perspective arrow trials). Participants made their responses by pressing 1 for 'yes' if the number cue matched the announced perspective/ arrow pointing and 2 for 'no' if these did not match. Trial types were defined not only by the perspective participants were asked to verify (self, non-self

avatar, non-self arrow) but also by whether the avatar's perspective / arrow pointing was consistent or inconsistent with the participant's perspective (see Figure 1).

In previous studies using the dots task, the data from those trials where the participant's response should be 'no' (*mismatching* trials) were not included in any reported analyses. This is because of a disparity in the experimental design. In consistent 'no' trials the number cue displayed was irrelevant to both perspectives. For example, if both the avatar and participant could see (or the arrow was pointing towards) 2 dots, the number cue was either 1 or 3. The inconsistent 'no' trials, however, displayed a number cue representing the inverse perspective. For example, if the participant could see 2 dots and the avatar could see (or the arrow was pointing towards) only 1, the number cue in the 'no' trial would always represent the inverse perspective, being either 1 for self-perspective or 2 for non-self-perspective judgements.

In order to optimize the experimental design for use in the rTMS study (Experiment 2), it was crucial to be able to include all trial types in analysis. This was in order to keep the number of TMS pulses within acceptable tolerance and safety limits: discarding data from half of the experimental trials would have entailed delivering twice as many TMS pulses. Therefore, we modified the inconsistent 'no' trials so that the number cue was irrelevant to both perspectives, as it was in the consistent 'no' trials. Hence, when the participant could see 2 dots but the avatar could see (or the arrow was pointing towards) only 1, the number cue was 3. This modification allowed us to collapse across matching (yes) and mismatching (no) trials. Also for design optimization for rTMS, the filler trials (where no dots were displayed) included in the study by Samson et al. (2010) were excluded from this experiment, and the number

cue was never 0. Experiment 1 therefore tested whether the consistency effect was still present for self-perspective and non-self-perspective judgements when these minor alterations were made to the procedure.

There were 4 consecutive blocks of trials for each stimulus condition (avatar and arrow) and each block consisted of 48 trials. The order of stimulus condition was counterbalanced across participants. The experimental trials for each stimulus condition were preceded by 26 practice trials. Accuracy feedback was given during practice trials only. In half of the experimental trials the avatar/arrow pointed to the left and in half it pointed to the right. Half of the trials required a 'yes' response and half required a 'no' response. Response time was measured from the onset of the stimulus picture.

Results and Discussion

Due to the small percentage of errors (3.8% in total) we did not submit these data to any statistical analyses. The response time (RT) data were analysed with a $2 \times 2 \times 2$ repeated measures ANOVA with the factors Stimulus (avatar, arrow), Perspective (self, non-self) and Consistency (consistent, inconsistent). Trials for which RTs were more than 2 standard deviations from the mean (0.6%) and incorrect responses (3.8%) were excluded from the analysis.

Figure 2 illustrates the mean RT for each of the conditions and trial types. The analysis revealed that after collapsing the matching ('yes' response) and mismatching ('no' response) trials, the main effect of Consistency was significant, $F_{(1,15)} = 54.93$; $p < .001$; $\eta^2_p = .79$. RTs were longer in inconsistent ($M = 618$ ms, $S.E.M. = 29$) than in consistent ($M = 581$ ms, $S.E.M. = 29$) trials. The main effect of Perspective was also

significant, $F_{(1,15)} = 12.90$; $p = .003$; $\eta^2_p = .46$. Participants responded faster to self ($M = 585\text{ms}$, $S.E.M. = 30$) than to non-self trials ($M = 614\text{ms}$, $S.E.M. = 29$). Consistent with our previous study, neither the main effect of Stimulus ($p = .187$) nor any of its interactions were significant (all $ps > .250$).

This pattern of results was replicated when we performed a mixed analysis with Stimulus as a between-subjects factor, taking into account only the first stimulus condition. In this analysis we found a main effect of Consistency ($F_{(1,14)} = 19.93$; $p = .001$; $\eta^2_p = .59$), a main effect of Perspective ($F_{(1,14)} = 10.82$; $p = .005$; $\eta^2_p = .44$), but no main effect of Stimulus ($p = .836$). The only significant interaction was that between Perspective \times Consistency; ($F_{(1,14)} = 5.41$; $p = .036$; $\eta^2_p = .28$). Post-hoc analysis showed that while self-perspective judgements ($M = 588\text{ms}$, $S.E.M. = 29$) were faster than non-self-perspective judgements ($M = 637\text{ms}$, $S.E.M. = 30$) in the consistent trials ($p < .001$), this comparison was not significant in the inconsistent trials (self: $M = 631\text{ms}$, $S.E.M. = 32$; non-self: $M = 658\text{ms}$, $S.E.M. = 31$; $p = .12$). This mixed analysis confirms our previous results (Santesteban et al., 2014) that the consistency effect seen in the arrow condition is not due to participants' exposure to the avatar condition.

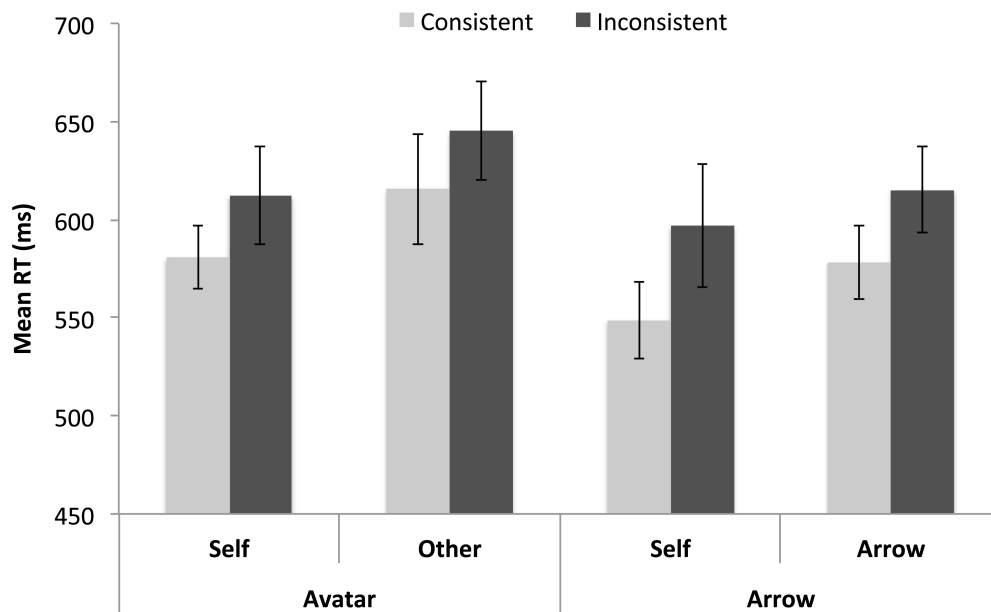


Figure 2. Mean RT for each of the trial types. The light bars represent consistent and the dark bars represent inconsistent trials. The error bars illustrate within-subject *S.E.M.*

The results from Experiment 1 confirmed that inclusion of the mismatching ('no' response) trials in the analysis did not eliminate the consistency or the perspective effects. This pattern of results gave us the confidence to include this trial type in Experiment 2, allowing us to optimize the design for rTMS and include all experimental trials in the analysis. Crucially, the results from Experiment 1 also support our previous findings (Santesteban et al., 2014) that an arrow is just as effective as a human-like figure to elicit the consistency effect.

Experiment 2

The main objective of Experiment 2 was to determine whether the role of the rTPJ in the dots task (Ramsey et al., 2013; Schurz et al., 2015) is to support mentalizing during self-perspective trials with mentalistic stimuli, or to support domain-general

attentional processes on self-perspective trials. Accordingly, participants completed both avatar and arrow conditions of the dots task while undergoing rTMS stimulation (see Methods) to either the rTPJ, or a control mid-occipital site. The two hypotheses concerning rTPJ function during the dots task make opposing predictions. If rTPJ supports mentalizing during self-perspective trials then one would expect a selective effect of rTPJ stimulation (when compared to stimulation of the mid-occipital control site) only for trials with mentalistic stimuli (avatar trials). Conversely, if rTPJ supports attentional processes such as visual pop-out or reorienting that are required on self-perspective but not non-self-perspective trials, then one would expect a selective effect of rTPJ stimulation on self-perspective trials (both arrow and avatar), but not on non-self-perspective trials.

Method

Participants

Nineteen healthy adults (12 females) were recruited to take part in this study for a small monetary reward. Age ranged between 19 and 42 years ($M = 24.9$, $SD = 6.0$). We screened all participants to ensure that there were no contraindications to TMS. Prior to the experimental session, structural T1-weighted MRI scans were obtained to aid localization of the targeted regions. All participants provided written informed consent prior to the study. The experimental procedures were approved by the local ethics committee and were carried out in accordance with the principles of the revised Helsinki Declaration (World Medical Associations General Assembly 2008).

Stimuli and Procedure

The stimuli and procedure replicated those of Experiment 1. A within-subjects design was employed, with each participant undergoing stimulation of both the rTPJ and a control site in the mid occipital cortex (MOC). However, for safety reasons we had to reduce the number of trials from the total presented in Experiment 1. The task consisted of 48 trials per stimulus type (avatar/arrow) for each of the stimulation sites (rTPJ/MOC), therefore, each participant completed 192 experimental trials in total. Stimulus type was blocked within each stimulation site. Both the order of stimulation site (rTPJ or MOC) and of stimulus type (avatar or arrow) were counterbalanced across participants.

TMS Protocol

Prior to the experiment, the structural MRI scans were manually registered to the standard MNI-152 template in the Brainsight2 neuronavigation system (Rogue Research, Montreal, Canada) and stimulation targets set using predefined MNI coordinates (rTPJ = 54, -47, 26; MOC = 0, -95, 26; Figure 3). Right TPJ coordinates were taken from Sowden and Catmur (2013), who demonstrated a disruptive effect of rTMS to rTPJ on social cognitive function. Appropriate trajectories of stimulation were set for each individual, and landmarks were set on the surface reconstruction of the participant's head.

Before the experiment began, each participant's resting motor threshold (rMT) was identified, defined as the lowest intensity of stimulation required to elicit motor evoked potentials (MEPs) of at least 50 μ V in the first dorsal interosseous muscle in the right hand, on 3 out of 5 trials. MEPs were recorded using surface skin electrodes and Brain Vision software (Brain Products, Gilching, Germany).

The participant's head was then registered in the neuronavigation system using an infrared camera and participant tracker. Repetitive TMS (6 pulses at 10 Hz per trial) was delivered using a figure-of-eight coil and a Magstim Rapid2 stimulator (The Magstim Company, Whitland, UK) at 110% of each participant's rMT. Participants received the stimulation 100ms after stimulus scene onset, ensuring that the disruptive effects of rTMS were present throughout the response preparation period identified in Experiment 1. The location of the coil with respect to the target site was monitored online, allowing precise coil location to be maintained throughout the experiment. The TMS coil was replaced and re-calibrated between stimulation sites, or if the stimulator indicated overheating of the coil. The experimental trials for each stimulation site and stimulus condition were preceded by 13 practice trials with rTMS, in order to familiarise participants with the sensation of rTMS to each site during the task. Accuracy feedback was given during practice trials only.

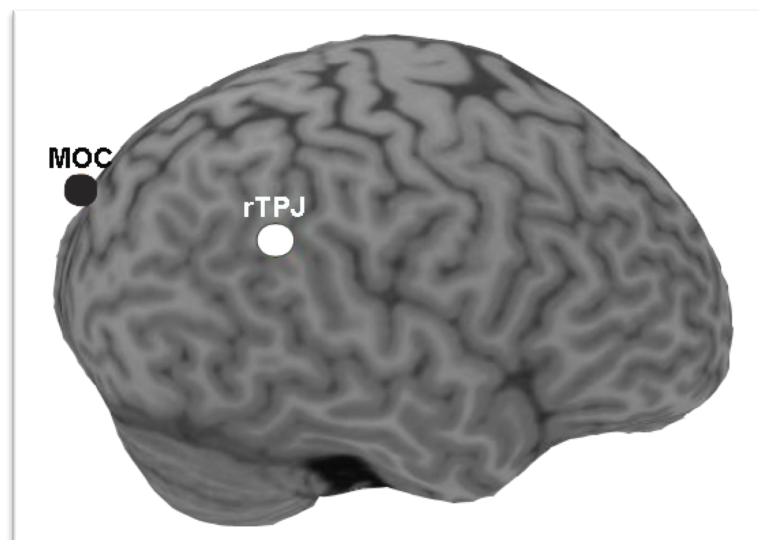


Figure 3. Graphical illustration of the rTMS targeted brain areas. MNI coordinates: rTPJ 54, -47, 26; MOC = 0, -95, 26.

Results and Discussion

As in Experiment 1, participants made very few errors (1.5%), and therefore the error data were not submitted to further statistical analysis. Trials for which RTs were more than 2 standard deviations from the mean (0.2%) and incorrect responses (1.5%) were excluded from the analysis.

In order to address our experimental question of whether rTMS of rTPJ would impair self-perspective judgements in the avatar but not in the arrow condition, we first analysed the RT data from the self-perspective trials using a $2 \times 2 \times 2$ repeated measures ANOVA with Stimulation Site (rTPJ, MOC), Stimulus Type (avatar, arrow), and Consistency (consistent, inconsistent) as the within-subjects factors. The RT data are shown in Figure 4. Our results replicated the key finding from studies using the dots task with faster responses for consistent ($M = 598\text{ms}$, $S.E.M. = 29$) than for inconsistent trials ($M = 655\text{ms}$, $S.E.M. = 35$); $F_{(1,18)} = 21.41$; $p < .001$; $\eta^2_p = .54$. There was also a main effect of stimulation site, $F_{(1,18)} = 4.60$; $p = .046$; $\eta^2_p = .20$: responding was slower for self-perspective trials following stimulation of rTPJ ($M = 651\text{ms}$, $S.E.M. = 38$) compared to MOC ($M = 602\text{ms}$, $S.E.M. = 28$). Crucially, we did not find either a 3-way interaction between stimulation site, stimulus type and consistency, $F_{(1,18)} = .035$; $p = .853$; $\eta^2_p = .002$, or a 2-way interaction between stimulation site and stimulus type, $F_{(1,18)} = .871$; $p = .363$; $\eta^2_p = .046$, demonstrating that stimulation of rTPJ did not selectively impair self-perspective judgements in the avatar condition. In order to establish the strength of evidence for the null hypothesis of no interaction between stimulation site and stimulus type, Bayes Factors were calculated using JASP (<https://jasp-stats.org/>; JASP Team, 2016). JASP default priors were used as model for H1. A Bayes Factor of 0.015 was associated with the inclusion of the 3-way interaction

into a model containing the main effects and all constituent 2-way interactions. For the 2-way interaction (which is present in multiple possible models), Bayesian model averaging revealed a Bayes Factor of 0.160 when comparing all models containing the Stimulation Site \times Stimulus Type interaction to all other candidate models. Thus for both the 3-way and 2-way interactions, the data were over 6 times as likely under the null hypothesis as under the alternative hypotheses. No other main effects or interactions were significant, all $ps > .36$. These results therefore failed to support the claim that rTPJ is selectively involved in processing the spontaneous representation of another's visual perspective during self-perspective judgements.

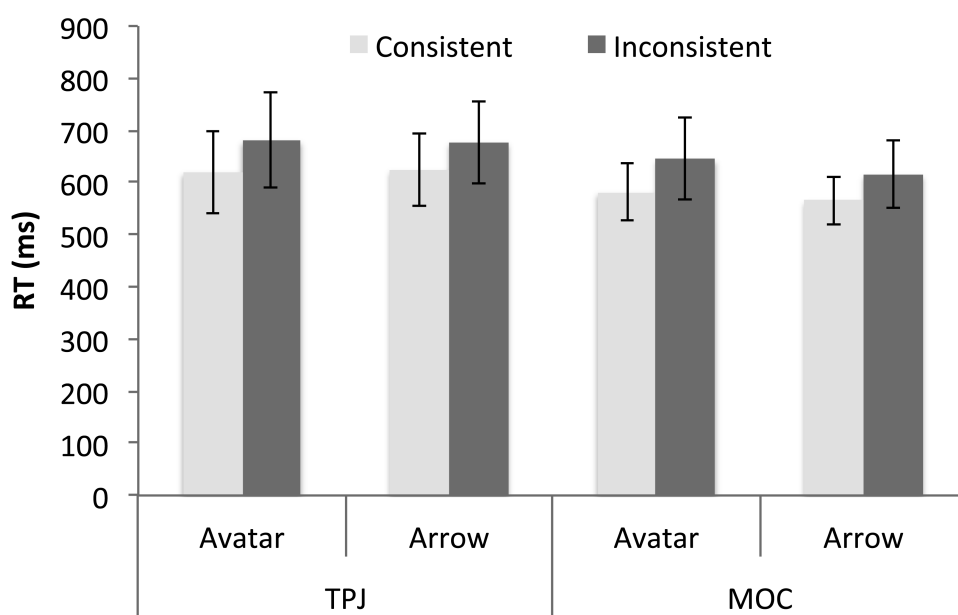


Figure 4. Mean RTs for each stimulation site during self-perspective judgements. The light bar represents consistent trials and the dark bars illustrate inconsistent trials. The error bars illustrate within-subject *S.E.M.*

The above analysis did, however, show a main effect of stimulation site when only the self-perspective judgement trials were included. Responses were slower for the rTPJ compared to the MOC stimulation site. This is consistent with a domain-

general attentional role for rTPJ on self-perspective trials. In order to investigate if this effect was selective to self-perspective compared to non-self-perspective trials, consistent with the task-demand analyses above, in our next analysis we included the non-self-perspective trials. The $2 \times 2 \times 2 \times 2$ ANOVA (factors as in the above analysis, with the addition of Perspective: self, non-self) revealed a significant main effect of Perspective, $F_{(1,18)} = 29.33$; $p < .001$; $\eta^2_p = .62$. Overall, responses were faster for self ($M = 626$ ms, $S.E.M. = 31$) than for non-self trials ($M = 673$ ms, $S.E.M. = 36$). Again, the main effect of Consistency remained significant, $F_{(1,18)} = 40.47$; $p < .001$; $\eta^2_p = .69$. Neither the main effects of Stimulus Type ($p = .52$) nor Stimulation Site ($p = .23$) were significant. However, there was a significant interaction between the Stimulation Site and Perspective factors, $F_{(1,18)} = 8.80$; $p = .008$; $\eta^2_p = .33$, supported by a Bayes factor of 3.14 in favour of inclusion of the Stimulation Site \times Perspective interaction (when averaging over all models containing the interaction compared to all other models). Post-hoc analysis revealed that under stimulation of the rTPJ, RTs for self-perspective trials were slower than under MOC stimulation, $F_{(1,18)} = 4.60$; $p = .046$; $\eta^2_p = .20$, see Figure 5 (although it should be noted that a Bayesian analysis revealed only anecdotal evidence for this follow-up test, with a Bayes Factor of 1.51 in favour of the alternative hypothesis of an effect of stimulation on these trials). The equivalent comparison for non-self-perspective trials was not significant ($p = .95$, Bayes Factor of 0.238 associated with the alternative hypothesis of an effect of stimulation on these trials). No other main effects or interactions reached significance. The results from this analysis are therefore consistent with the hypothesis that the rTPJ's involvement in the dots task is in domain-general attentional processing on self-perspective trials, irrespective of whether the central stimulus is an avatar or an arrow.

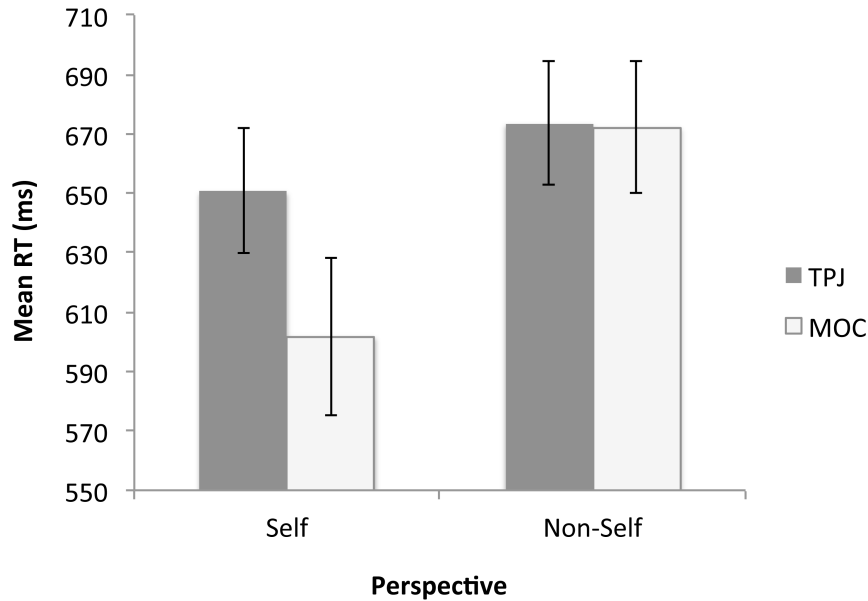


Figure 5. Stimulation Site \times Perspective interaction. Mean RT during rTMS of rTPJ (darker bars) and MOC (lighter bars) for self-perspective and non-self-perspective judgements. Compared to MOC, stimulation of rTPJ significantly increased RTs for self-perspective judgements. The error bars illustrate within-subject S.E.M.

General Discussion

The results from Experiments 1 and 2 replicate previous findings (Cole et al., 2016; Conway et al., 2017; MacDorman, Srinivas & Patel, 2013; Santiesteban et al., 2014, Schurz et al., 2015) that a non-mentalistic stimulus such as an arrow is able to elicit a consistency effect of similar magnitude to that of a human-like figure in the dots task. Of course, it is possible that equivalent consistency effects in the mentalistic and non-mentalistic conditions arise through different mechanisms: implicit mentalizing in the avatar condition, and domain-general attentional processing in the arrow condition. This question was investigated here using neurostimulation methods to test the competing predictions of two hypotheses: that the role of the rTPJ in the dots task is to support implicit mentalizing in the mentalistic condition; or, that recruitment of the rTPJ during performance of the dots task relates to domain-general attentional

processes, occurring on self-perspective trials irrespective of whether the central stimulus is mentalistic or not. Results supported the second hypothesis: stimulation of rTPJ selectively impacted self-perspective versus non-self-perspective trials, but did not distinguish between mentalistic and non-mentalistic trials.

An attentional explanation of rTPJ involvement on self-perspective trials is consistent with a large body of literature demonstrating the role of the TPJ in several aspects of attention. The role of the TPJ in attentional reorienting is well-established; for example, TPJ activity is observed on invalid trials of the Posner (1980) attentional cuing task (which require attentional reorienting) but not on valid trials (Thiel, Zilles & Fink, 2004). One suggestion put forward in the Introduction was that, on self- but not on non-self-perspective trials, participants must reorient their attention from the side of the room cued by the arrow or avatar in order to check for more dots on the other side of the room. However, other rTPJ-mediated attentional processes are also possible explanations of the effects of stimulation on self-perspective trials. Previous studies have consistently found TPJ recruitment during visual pop-out tasks, where a target 'pops out' because of its saliency and novelty when surrounded by distracting stimuli (Buschman & Miller, 2007; Ellison et al., 2004; Pollmann et al., 2003). In the dots task, it is possible that the saliency of the targets (large red dots against a light blue background) makes them 'pop-out' and they are quickly subitized (Sathian et al., 1999). On self-perspective trials this TPJ-mediated attentional selection of all the dots would be helpful; but on non-self-perspective trials, such attentional selection of all red dot targets would be counterproductive. It is possible, therefore, that stimulation of the rTPJ interfered with efficient target selection on self-perspective trials, resulting

in slower performance due to a reduced 'pop-out' effect, but did not affect non-self-perspective trials on which 'pop out' processes do not govern performance.

It should be noted therefore that the lack of stimulation effects on non-self-perspective trials does not imply that domain-general attentional processes are not required in this type of trial. Non-self-perspective trials are indeed likely to rely on domain-general attentional processes, but these processes may not involve recruitment of rTPJ. In order to establish which processes are involved in these trials it is again informative to consider the demands of the task on these trials. Recall that, before making their responses, participants are presented with a perspective cue. For non-self-perspective trials, the cue is 'She', 'He', or 'Arrow'. So, before they see the picture of the room with the dots, on non-self-perspective trials (unlike on self-perspective trials) participants know they have to pay attention to the *direction* of the central stimulus and verify the number of dots to which the avatar is facing or the arrow is pointing. The presence of the perspective cue before non-self-perspective judgements renders this trial type similar to a 'valid' trial in the Posner task (Posner, 1980). For valid trials of the Posner task, the location of a prime cue and the target stimulus is the same. This type of trial requires endogenous orienting of attention to the cued location. Previous neuroimaging research has found that this type of attentional orienting during valid trials of the Posner task engages the anterior cingulate cortex (Thiel et al., 2004), whereas attentional re-orienting during *invalid* trials (where the location of the prime cue differs from the target's location) recruits rTPJ. This could explain why performance on non-self-perspective trials in the dots task remains unaffected following stimulation of the rTPJ.

Finally, there is another possible explanation for the lack of a selective effect of rTPJ stimulation on avatar and arrow trials: that participants were anthropomorphising the arrow stimulus and treating it as if it had mental states. We have previously argued against such an explanation (Santiesteban et al., 2014); furthermore, such an effect, if present, may be more likely for those participants who saw the avatar stimulus before the arrow stimulus, and yet there were no signs of stimulus order effects in either this study or in earlier studies with avatar and arrow stimuli (Conway et al., 2017; Santiesteban et al., 2014). Perhaps more convincingly, this possibility was directly investigated by Conway et al. (2017) who used a variant of the dots task which is able to detect the attribution of mental states to either the avatar or arrow stimulus should it occur. Specifically, participants completed the standard arrow or avatar conditions of the task but either an opaque or a transparent telescope was used to render dots in front of the avatar invisible or not; assuming the anthropomorphising explanation is true, the same would be true for the arrow. With such a design, implicit mentalizing would be revealed by the presence of the standard consistency effect in the visible condition, but an absence of the consistency effect in the invisible condition. In fact, a consistency effect was observed in all conditions, a pattern of data which does not support the implicit mentalizing account (and which is therefore also inconsistent with the anthropomorphising account of the consistency effect in the arrow condition), but which is instead consistent with a domain-general attentional account of performance on the dots task.

It is also important to clarify that our interpretation that performance on the dots task is likely to be mediated by domain-general attentional processes subserved by the rTPJ does not undermine or negate the well-established role of this brain region

in the social domain, and particularly in mentalizing processes. There is converging empirical evidence that the rTPJ is a functionally heterogeneous brain region (Scholz et al., 2009; Mars et al., 2012; Bzdok et al., 2013; Igelström et al., 2015; Krall et al., 2015, 2016; Lee & McCarthy, 2016). For example, a recent meta-analysis by Krall et al. (2015) of neuroimaging data from attention reorienting and false belief studies showed recruitment of the anterior subregion of the rTPJ in both types of task, whereas higher activation was found in the posterior rTPJ for false belief compared to attention reorienting tasks. These findings were supported by meta-analytic connectivity mapping and resting-state functional connectivity analyses, which converged on the separation of the rTPJ into anterior ($x = 54, y = -44, z = 18$) and posterior ($x = 54, y = -52, z = 26$) subdivisions. The stimulated portion of the rTPJ in the current study lies on the border of these anterior and posterior subdivisions. According to the findings of Krall et al., therefore, this area supports both attentional and social processing. This makes it an ideal target for the present study because disruptive stimulation of this area had the capacity for interference with both social and attentional processes, allowing us to distinguish between the two hypotheses concerning rTPJ function during the dots task, on the basis of the pattern of effects of disruptive stimulation that we found. If rTPJ function during the dots task supports mentalizing then one would expect a selective effect of rTPJ stimulation only for trials with mentalistic stimuli (avatar trials). Conversely, if rTPJ function during the dots task supports attentional processes such as visual pop-out or attention reorienting that are required on self-perspective but not non-self-perspective trials, then one would expect a selective effect of rTPJ stimulation on self-perspective trials (both arrow and avatar), but not on non-self-perspective trials. The finding that only self-perspective

(but not non-self-perspective) trials for both mentalistic and non-mentalistic conditions were affected by rTMS of this subregion of the rTPJ strongly supports the view that mentalizing is not required in the dots task.

In summary, the findings reported here provide further evidence of the robustness of the self-consistency effect in the dots task. However, rather than supporting the view that this effect is driven by participants automatically adopting the perspective of the other person in the room, as claimed under the implicit mentalizing account, two key findings in our study indicate that domain-general attentional processes mediate performance on this task. The first is the replication of previous findings that a non-mentalistic stimulus - an arrow - is as effective as a mentalistic stimulus - an avatar - to elicit the self-consistency effect (Cole et al., 2016; Conway et al., 2017; MacDorman et al., 2013; Santiesteban et al., 2014). Crucially, here we demonstrate using a causal brain stimulation technique that the right TPJ does not distinguish between the mentalistic and non-mentalistic nature of the stimulus producing the consistency effect. The second finding, that disruption of the right TPJ impairs performance only during self-perspective trials (for both mentalistic and non-mentalistic stimuli), suggests that rather than perspective taking or self-other processing *per se*, the dots task taps into domain-general attentional effects. Hence, our results lend support to the view that often what is perceived as mentalizing in everyday social interactions is instead mediated by domain-general processes, or sub-mentalizing (Heyes, 2014b; Santiesteban et al., 2014).

Acknowledgements

This work was supported by a Royal Society Research Grant awarded to CC. IS contributed to this project during a Fellowship awarded by the ESRC [ES/N00325X/1].

GB contributed while supported by the Baily Thomas Charitable Trust.

References

- Apperly, I. A. (2011). *Mindreaders: The Cognitive Basis of "theory of Mind"*. Hove and New York: Psychology Press.
- Apperly, I. A., & Butterfill, S. A. (2009). Do humans have two systems to track beliefs and belief-like states? *Psychological Review*, 116(4), 953.
- Baillargeon, R., Scott, R. M., & He, Z. (2010). False-belief understanding in infants. *Trends in Cognitive Sciences*, 14(3), 110-118.
- Buschman, T. J., & Miller, E. K. (2007). Top-down versus bottom-up control of attention in the prefrontal and posterior parietal cortices. *Science*, 315(5820), 1860-1862.
- Bzdok, D., Langner, R., Schilbach, L., Jakobs, O., Roski, C., Caspers, S., ... & Eickhoff, S. B. (2013). Characterization of the temporo-parietal junction by combining data-driven parcellation, complementary connectivity analyses, and functional decoding. *Neuroimage*, 81, 381-392.
- Call, J. (2012). Theory of Mind in Animals. In *Encyclopedia of the Sciences of Learning* (pp. 3316-3319). Springer US.
- Catmur, C., Santiesteban, I., Conway, J. R., Heyes, C., & Bird, G. (2016). Avatars and arrows in the brain. *NeuroImage*, 132, 8-10.
- Cole, G. G., Atkinson, M., Le, A. T., & Smith, D. T. (2016). Do humans spontaneously take the perspective of others? *Acta psychologica*, 164, 165-168.
- Conway, J.R., Lee, D., Ojaghi, M., Catmur, C., & Bird, G. (2017). Submentalizing or mentalizing in a Level 1 perspective-taking task: A cloak and goggles test. *Journal of Experimental Psychology: Human Perception and Performance*. 43(3):454-465.

- Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature reviews neuroscience*, 3(3), 201-215.
- De Bruin, L. C., & Newen, A. (2012). An association account of false belief understanding. *Cognition*, 123(2), 240-259.
- Ellison, A., Schindler, I., Pattison, L. L., & Milner, A. D. (2004). An exploration of the role of the superior temporal gyrus in visual search and spatial perception using TMS. *Brain*, 127(10), 2307-2315.
- Frith, C. D., & Frith, U. (2012). Mechanisms of social cognition. *Annual Review of Psychology*, 63, 287-313. doi:10.1146/annurev-psych-120710-100449.
- Geng, J. J., & Vossel, S. (2013). Re-evaluating the role of TPJ in attentional control: contextual updating? *Neuroscience & Biobehavioral Reviews*, 37(10), 2608-2620.
- Heyes, C. (2014a). False belief in infancy: a fresh look. *Developmental Science*, 17(5), 647-659.
- Heyes, C. (2014b). Submentalizing I am not really reading your mind. *Perspectives on Psychological Science*, 9(2), 131-143.
- Heyes, C. (2017). Apes submentalise. *Trends in Cognitive Sciences* 21(1):1-2.
- Igelström, K. M., Webb, T. W., & Graziano, M. S. (2015). Neural processes in the human temporoparietal cortex separated by localized independent component analysis. *Journal of Neuroscience*, 35(25), 9432-9445.
- JASP Team (2016). JASP (Version 0.8.0.0) [Computer software].
- Krall, S. C., Rottschy, C., Oberwelland, E., Bzdok, D., Fox, P. T., Eickhoff, S. B., ... & Konrad, K. (2015). The role of the right temporoparietal junction in attention

- and social interaction as revealed by ALE meta-analysis. *Brain Structure and Function*, 220(2), 587-604.
- Krall, S. C., Volz, L. J., Oberwelland, E., Grefkes, C., Fink, G. R., & Konrad, K. (2016). The right temporoparietal junction in attention and social interaction: A transcranial magnetic stimulation study. *Human brain mapping*, 37(2), 796-807.
- Krupenye, C., Kano, F., Hirata, S., Call, J., & Tomasello, M. (2016). Great apes anticipate that other individuals will act according to false beliefs. *Science*, 354(6308), 110-114.
- Lee, S. M., & McCarthy, G. (2016). Functional heterogeneity and convergence in the right temporoparietal junction. *Cerebral Cortex*, 26(3), 1108-1116.
- Mars, R. B., Sallet, J., Schüffelen, U., Jbabdi, S., Toni, I., & Rushworth, M. F. (2012). Connectivity-based subdivisions of the human right “temporoparietal junction area”: evidence for different areas participating in different cortical networks. *Cerebral cortex*, 22(8), 1894-1903.
- McCleery, J. P., Surtees, A. D. R., Graham, K. A., Richards, J. E., & Apperly, I. A. (2011). The neural and cognitive time course of theory of mind. *The Journal of Neuroscience*, 31(36), 12849-12854.
- MacDorman, K. F., Srinivas, P., & Patel, H. (2013). The uncanny valley does not interfere with level 1 visual perspective taking. *Computers in human behavior*, 29(4), 1671-1685.
- Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science*, 308(5719), 255-258.
- Penn, D. C., & Povinelli, D. J. (2007). On the lack of evidence that non-human animals possess anything remotely resembling a ‘theory of mind’. *Philosophical*

- Transactions of the Royal Society of London B: Biological Sciences*, 362(1480), 731-744.
- Perner, J., & Ruffman, T. (2005). Infants' insight into the mind: How deep? *Science*, 308(5719), 214-216.
- Phillips, J., Ong, D. C., Surtees, A. D., Xin, Y., Williams, S., Saxe, R., & Frank, M. C. (2015). A Second Look at Automatic Theory of Mind: Reconsidering Kovács, Téglás, and Endress (2010). *Psychological Science*, 26(9), 1353-1367.
- Pollmann, S., Weidner, R., Humphreys, G. W., Olivers, C. N., Müller, K., Lohmann, G., ... & Watson, D. G. (2003). Separating distractor rejection and target detection in posterior parietal cortex—an event-related fMRI study of visual marking. *Neuroimage*, 18(2), 310-323.
- Posner, M. I. (1980). Orienting of attention. *Quarterly journal of experimental psychology*, 32(1), 3-25.
- Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, 1(04), 515-526.
- Qureshi, A. W., Apperly, I. A., & Samson, D. (2010). Executive function is necessary for perspective selection, not Level-1 visual perspective calculation: Evidence from a dual-task study of adults. *Cognition*, 117(2), 230-236.
- Ramsey, R., Hansen, P., Apperly, I., & Samson, D. (2013). Seeing it my way or your way: frontoparietal brain areas sustain viewpoint-independent perspective selection processes. *Journal of Cognitive Neuroscience*, 25(5), 670-684.
- Ruffman, T., Taumoepeau, M., & Perkins, C. (2012). Statistical learning as a basis for social understanding in children. *British Journal of Developmental Psychology*, 30(1), 87-104.

- Samson, D., Apperly, I. A., Braithwaite, J. J., Andrews, B. J., & Bodley Scott, S. E. (2010). Seeing it their way: evidence for rapid and involuntary computation of what other people see. *Journal of Experimental Psychology: Human Perception and Performance*, 36(5), 1255-1266.
- Santiesteban, I., Catmur, C., Hopkins, S. C., Bird, G., & Heyes, C. (2014). Avatars and arrows: Implicit mentalizing or domain-general processing? *Journal of Experimental Psychology: Human Perception and Performance*, 40(3), 929.
- Sathian, K., Simon, T. J., Peterson, S., Patel, G. A., Hoffman, J. M., & Grafton, S. T. (1999). Neural evidence linking visual object enumeration and attention. *Journal of Cognitive Neuroscience*, 11(1), 36-51.
- Scholz, J., Triantafyllous, C., Whitfield-Gabrieli, S., Brown, E. N., & Saxe, R. (2009). Distinct regions of the right temporo-parietal junction are selective for theory of mind and exogenous attention. *PLoS ONE*, 4(3), 1-7.
- Schurz, M., Kronbichler, M., Weissengruber, S., Surtees, A., Samson, D., & Perner, J. (2015). Clarifying the role of theory of mind areas during visual perspective taking: Issues of spontaneity and domain-specificity. *NeuroImage*, 117, 386-396.
- Sowden, S., & Catmur, C. (2013). The role of the right temporoparietal junction in the control of imitation. *Cerebral Cortex*, 25(4), 1107-1113.
doi:10.1093/cercor/bht306.
- Thiel, C. M., Zilles, K., & Fink, G. R. (2004). Cerebral correlates of alerting, orienting and reorienting of visuospatial attention: an event-related fMRI study. *Neuroimage*, 21(1), 318-328.