OXFORD

Social Cognitive and Affective Neuroscience, 2017, 1775–1782

doi: 10.1093/scan/nsx109
Advance Access Publication Date: 14 September 2017
Original Article

Repetitive TMS of the temporo-parietal junction disrupts participant's expectations in a spontaneous Theory of Mind task

Lara Bardi,[1] Pieter Six,[2] and Marcel Brass[1]

[1]Department of Experimental Psychology, University of Ghent, 9000 Ghent, Belgium, and [2]Australian Maritime College, University of Tasmania, Australia

Correspondence should be addressed to Lara Bardi, Department of Experimental Psychology, Ghent University, Henri Dunantlaan 2, 9000 Ghent, Belgium. E-mail: lara.bardi@ugent.be

Abstract

A recent debate about Theory of Mind (ToM) concerns whether spontaneous and explicit mentalizing are based on the same mechanisms. However, only a few neuroimaging studies have investigated the neural bases of spontaneous ToM, with inconsistent results. The present study had two goals: first, to investigate whether the right Temporo-Parietal Junction (rTPJ) is crucially involved in spontaneous ToM and second, to gain insight into the role of the rTPJ in ToM. For the first time, we applied rTMS to the rTPJ while participants were engaged in a spontaneous false belief task. Participants watched videos of a scene including an agent who acquires a true or false belief about the location of an object. At the end of the movie, participants reacted to the presence of the object. Results show that, during stimulation of the control site, RTs were affected by both the participant's expectations and the belief of the agent. Stimulation of the rTPJ significantly modulated task performance, supporting the idea that spontaneous ToM, as well as explicit ToM, relies on TPJ activity. However, we did not observe a disruption of the representation of the agent's belief. Rather, the stimulation interfered with participant's predictions, supporting the idea that rTPJ is crucially involved in self-other distinction.

Key words: repetitive TMS; temporo-parietal junction; spontaneous Theory of Mind

Introduction

In social interaction, we are constantly engaged in inferring intentions, goals, desires, traits and beliefs of our interaction partners to understand and predict others' behavior. This capacity, which is commonly referred to as 'Theory of Mind' (ToM) or mentalizing (e.g. ToM; Premack and Woodruff, 1978), has been the subject of extensive investigation in children and adults through the usage of 'false belief tasks' (e.g. Sally-Anne false belief tasks, Wimmer and Perner, 1983) in which participants are explicitly asked to reason about the mental states of others. Based on this research, ToM has been characterized as a complex capacity that critically involves the ability to inhibit one's own perspective in favor of the other's perspective, which is an executive function that emerges relatively late during development. Only at around 4 years of age children are able to report that someone else may have a different perception of the reality and thus end up for example with false beliefs about an event or object (e.g. Wimmer and Perner, 1983; Baron-Cohen *et al.*, 1985; Wellman *et al.*, 2001; McKinnon and Moscovitch, 2007; Gweon *et al.*, 2012).

Another form of ToM that has received increasing attention in recent years is what is referred to as spontaneous or implicit mentalizing. The main distinction between explicit and spontaneous mentalizing tasks is that in spontaneous ToM tasks participants are never instructed to reason about the other's

Received: 7 April 2017; **Revised:** 7 September 2017; **Accepted:** 11 September 2017

beliefs. Indeed, analysis of eye movements and reaction times during simple detection tasks in a social context strongly suggests that both adults and children under 4 years may spontaneously or automatically represent information from another person's perspective. These findings show that we represent other's beliefs even when we are not required to do so and even in situations where other's mental states are completely irrelevant for our current goals (e.g. Clements and Perner, 1994; Onishi and Baillargeon, 2005; Southgate *et al.*, 2007; Surian *et al.*, 2007; Kovács *et al.*, 2010; Samson *et al.*, 2010; Schneider *et al.*, 2011, 2014a, 2014b, 2017; Senju *et al.*, 2011; van der Wel *et al.*, 2014; Nijhof *et al.*, 2016; Bardi *et al.*, 2017a; Meert *et al.*, 2017).

Whether implicit and explicit ToM rely on the same mechanisms is still a matter of debate. Some authors have questioned that spontaneous ToM tasks reflects mentalizing (Heyes, 2014; Phillips *et al.*, 2015), while others (Apperly and Butterfill, 2009; Back and Apperly, 2010) have proposed the existence of two mentalizing systems: (i) a spontaneous ToM system that is present early in life, is fast and efficient and operates spontaneously/unconsciously and (ii) an explicit form of mentalizing that develops later, is more deliberate and is more flexible. Finally, Carruthers (2016) postulates just a single mindreading system, which operates fully automatically by default but may operate in a more controlled way by invoking executive functions. Interestingly, a direct comparison between adults' performance in the same task under implicit and explicit instructions showed that participants' performance always reflects the representation of the other's perspective or belief and is not affected by task requirements (Schneider *et al.*, 2014a; Nijhof *et al.*, 2016; Bardi *et al.*, 2017a).

A number of neuroimaging studies focused on the neural correlates of explicit belief processing (e.g. Fletcher *et al.*, 1995; Gallagher *et al.*, 2000; Saxe and Kanwisher, 2003; Ruby and Decety, 2003; Saxe and Powell, 2006; Sommer *et al.*, 2007; Van Overwalle 2009; Schurz *et al.*, 2014). Tasks that require to reason about the content of other people's minds have been reported to engage a range of cortical areas, most consistently including the temporo-parietal junction (TPJ) (especially the right hemisphere; rTPJ) and the medial prefrontal cortex (MPFC) but also the superior temporal sulcus (STS) and precuneus (PC). Damage to the TPJ has been associated with impairment in ToM reasoning (Apperly *et al.*, 2004; Samson *et al.*, 2004.

Only a few studies have investigated the neural bases of spontaneous or implicit ToM, with inconsistent results (Kovács *et al.*, 2014; Schneider *et al.*, 2014b; Hyde *et al.*, 2015; Bardi *et al.*, 2017a; Naughtin *et al.*, 2017). Schneider *et al.* (2014b) showed that during a spontaneous ToM task, only the left STS and posterior cingulate (PC), but not the rTPJ, showed the typical pattern of activity (false belief > true belief) commonly found for explicit ToM tasks. On the other hand, some studies on spontaneous ToM, reported that the rTPJ is significantly more activated during false belief processing tasks as compared to true belief conditions or no-belief conditions (Kovács *et al.*, 2014; Hyde *et al.*, 2015; Bardi *et al.*, 2017a; Naughtin *et al.*, 2017) supporting the idea that spontaneous and explicit ToM share the similar neural substrates.

Although ToM is one of the most interesting abilities of the human brain, only a surprisingly small number of studies have investigated ToM with the use of brain stimulation techniques. To the best of our knowledge, only three studies have investigated the effects of Transcranial Magnetic Stimulation (TMS) on the TPJ during tasks related to mental state attribution. These studies showed that disruption of TPJ activity could worsen our ability to explicitly attribute mental states to others in different ToM tasks, such as "faux-pas" stories and moral judgments (Costa *et al.*, 2008; Young *et al.*, 2010; Krall *et al.*, 2016).

The first question we sought to answer with this study is whether spontaneous ToM is based on the same neural mechanisms that have been described for explicit ToM. Specifically, we investigated the possibility that the rTPJ is crucially involved during spontaneous ToM processing. To this end, we applied repetitive Transcranial Magnetic Stimulation (TMS) to temporarily interfere with the activity of the rTPJ while participants were involved in a spontaneous ToM task. The advantage of using brain stimulation is that, while fMRI provides only correlational evidence of the involvement of a certain area in a task or function, brain stimulation techniques can be used to investigate a causal relationship. If the rTPJ plays a crucial role in spontaneous mentalizing, then TMS should interfere with participants' task performance in the spontaneous ToM task. This would support the idea that the two forms of ToM (spontaneous and explicit) are—at least partially—overlapping.

The second aim of the present study is to gain insight into the exact function of the rTPJ in ToM. In the last years, the TPJ has attracted the attention of an increasing number of researchers who see this area as a critical node for socio-cognitive abilities. Some authors have hypothesized that the rTPJ is a specialized region for representing others' mental states (e.g. Saxe and Kanwisher 2003; Saxe and Powell, 2006). Others maintain that rTPJ is involved in detecting a mismatch between self and other representations and switching between self and other depending on task demands. This mechanism would be common to different abilities such as ToM, perspective taking, agency attribution and control of imitation (Brass *et al.*, 2005, 2009; Santiesteban *et al.*, 2012; Cook *et al.*, 2014; Bardi *et al.*, 2017b). For example, the application of transcranial Direct Current Stimulation (tDCS) to excite neurons in the rTPJ increases participants' ability to take another person's visual perspective in a perspective-taking task (e.g. Santiesteban *et al.*, 2012, 2015) and enhances self representations when participants are required to inhibit imitation tendencies (Santiesteban *et al.*, 2012; Bardi *et al.*, 2017b). In the same vein, temporary interfering with rTPJ activity with rTMS leads to an increased interference of other representations (Sowden and Catmur, 2015). For a fine-grained conceptualization of the TPJ processing in belief attribution, neuroimaging studies do not provide enough information. In this sense, TMS provides a unique opportunity to unveil TPJ function by looking at the consequences of neural interference with the activity in this area.

We used an adapted version of a spontaneous ToM task (Deschrijver *et al.*, 2016; Nijhof *et al.*, 2016; Bardi *et al.*, 2017a), originally developed by Kovács *et al.* (2010). Here, participants are presented with a video representing an agent who obtains certain knowledge about the location of an object, this being either behind an occluder or outside of the scene. At the end of the video, the occluder is lowered and participants are requested to press a button if the object is present behind the occluder (detection). Whether the ball is behind the occluder or not is completely random, with no relation to prior events. Reaction times depend on the participant expectations: responses are faster when the participant expects the object to be present (P+ conditions) than when he/she does not (P− conditions). This difference is also referred to as self or reality-bias (Deschrijver *et al.*, 2016). More strikingly, responses are also shortened when the agent only (false belief condition) believes the object is present (P−A+), showing that participants' performance is also influenced by the other's expectations about the presence of the object. The difference in RTs when neither the participant nor the agent expects the object to be present (P−A−) and the conditions in which the agent expects the object to be present (P−A+)

reflects the pure influence of the agent's belief and is taken as an index of ToM processing (Deschrijver *et al.*, 2016; Nijhof *et al.*, 2016).

Importantly, with TMS of the rTPJ we can test whether the rTPJ is involved in representing the beliefs of others or in distinguishing between self and other. If rTPJ is specifically involved in representing the other's beliefs, we expect that interfering with activity of the rTPJ will result in a reduced influence of the other belief on the participant's performance leading to a reduction of the ToM index. If, however, rTPJ is involved in self-other distinction, we expect that TMS will affect the reality bias because the preference for self-related beliefs as indexed by the reality bias should disappear when participants cannot distinguish between self and other representations.

Materials and methods

Participants

We recruited participants who had already participated in an fMRI study and could provide high-resolution anatomical data for neuronavigation purposes. A total sample of 21 volunteers participated in the study. One participated felt unwell during the second session and was not included in the dataset. For another participant, a portion of the data failed to be saved properly, resulting in a final sample of 19 participants ($N = 19$, mean age = 24, SD = 5, 10 females, all right-handed). All participants gave informed consent prior to engaging in the task. They had no history of neurological or psychiatric disorders, had normal or corrected-to-normal vision and were prescreened for the risk factors associated with TMS (Rossi *et al.*, 2009). All participants received a €50 compensation for participating in the study. The study was granted ethical approval by the Medical Ethical Review Board of Gent University Hospital. Data are stored with the experimenter and are available for consultation upon request.

Stimuli and task

The task consisted in an adapted version of the task originally created by Kovács *et al.* (2010) as adopted in previous studies (Nijhof *et al.*, 2016; Bardi *et al.*, 2017a). The participant observes a series of video fragments, each one involving an agent (Buzz Lightyear, a character from the movie Toy Story) in relation to an object (a ball). The videos all follow the same structure, but vary in content according to our experimental manipulation. These movies could differ in two aspects related to belief attribution. The agent's belief could be true or false (true: matching reality and participant knowledge, false: not matching reality and participant's knowledge) and belief content (positive content: the agent believes the ball is present, negative content: the agent believes the ball is absent; Figure 1).

All movies started with an agent placing a ball on a table in front of an occluder. Then the ball rolled behind the occluder. From this point onward, the movies could continue in four ways depending on the experimental conditions: (i) In the True Belief-Positive Content condition (P + A+), the ball rolled out of the scene from behind the occluder, and then rolled back behind the occluder (ball last seen by the participant at 10 s; time information is given relative to the beginning of the movie) in the agent's presence. The agent left the scene at 11 s. Thus, the agent could rightly believe the ball to be behind the occluder. (ii) In the True Belief-Negative Content condition (P−A−), the ball emerged from behind the occluder without leaving the scene, then rolled back behind the occluder, and finally left the scene (ball last seen at 10 s), all in the agent's presence. The agent left the scene at 11 s. Thus, the agent could rightly believe the ball not to be behind the occluder. (iii) In the False Belief–Positive Content condition (P−A+), the order of when the ball and the agent left the scene was reversed relative to the True Belief–Negative Content condition. Thus, the agent left the scene at 6 s. Then, the ball emerged from behind the occluder without leaving the scene, rolled back behind the occluder, and finally left the scene (ball last seen at 11 s), all in the agent's absence. Thus, the agent could wrongly believe the ball to be behind the occluder. (iv) In the False Belief-Negative Content condition (P + A−), the ball rolled out of the scene from behind the occluder in the agent's presence. Then, the agent left the scene at 9 s. In his absence, the ball rolled back behind the occluder at 11 s. Thus, the agent could wrongly believe that the ball would not to be behind the occluder. As in the original task, in order to keep participants' attention during the presentation of the movies, they were instructed to press a key with the index finger of their left hand when the agent left the scene.

At the end of each movie, the agent re-entered the scene and the occluder fell down. The four task conditions were paired with two equally probable outcomes, in which the ball was either present behind the occluder or absent (off-scene). Participants were instructed to press a key as fast as possible with the index finger of their right hand if the ball was present when the occluder fell down (object detection). The presence or absence of the ball was completely independent of the belief formation phase, because the ball was randomly present in 50% of the trials in all the conditions. There were eight different conditions (eight movies) and all movies were repeated 12 times each in random order, resulting in a total of 96 experimental trials. Responses were given through a keyboard.

Importantly, the participant is only instructed to watch the movie and to react to the presence of the ball. The beliefs of the agent are never mentioned, and are in fact completely irrelevant for the visual detection task.

In their critique on the implicit task developed by Kovács and colleagues, Phillips *et al.* (2015) raised concerns about the processes underlying the condition effects in the original Kovács study, as they conclude that these effects may be explained by differences in the timing of the attention check (i.e. the response to the agent leaving the scene). The authors suggest that the timing differences may explain the results in terms of differences in refractory period. Indeed, research has shown the influence short stimulus onset asynchrony (SOA) on response times to a second stimulus (psychological refractory period-PRP; Herman and Kantowitz, 1970). However, this effect only occurs at SOAs lasting up to several hundred milliseconds. In the movies of the current study, the shortest SOA between the two responses (i.e. the moment when the agent leaves and the moment when the ball appears) is 3.376 s (in the P + A+ and P−A− condition), which seems far beyond the reach of a PRP effect. In the present study, the two events also serve as a trigger for two TMS trains delivered on the same trial (see below). However, since the shortest time interval between the first and the second TMS train was >3 s long, we can exclude that any "summation effect" of consecutive TMS stimulations might have occurred.

TMS protocol

All TMS stimulation was administered using a biphasic Rapid2 (Magstim) magnetic stimulator, using a 70 mm figure-eight coil.

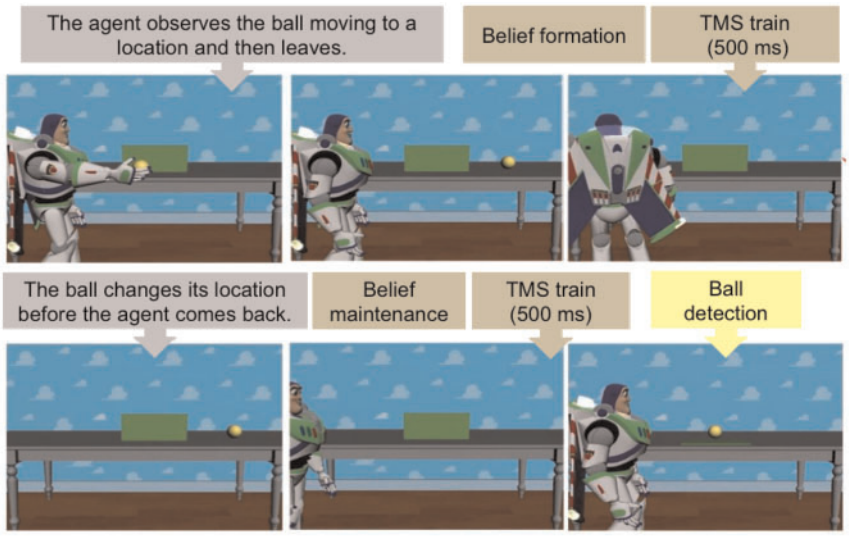

Fig. 1. Frames from a video presented during the task (P + A− condition). The agent observes the ball rolling outside the scene (A−) and then leaves the scene. In the second phase of the video, the ball comes back to the scene and rolls back to behind the occluder (P+). TMS was applied at the end of the belief formation phase (before the agent leaves the scene) and before the outcome phase (ball detection).

In the experimental condition, rTMS was delivered to the right temporo-parietal junction (RTPJ). For each participant, RTPJ stimulation location was defined using individual high-resolution T1-weighted structural MRI scans. We used the MNI reference space coordinates provided by a recent meta-analysis (Kovács *et al.*, 2014) which found that the RTPJ cluster is centered around the following MNI coordinate: 56; –47; 33. Because of an error during the testing procedure, five participants were stimulated slightly posteriorly (average MNI coordinates: 50, −58, 37). The average coordinates of our study (54, −54, 34) well overlap with those used the study of Young *et al.* (2010; exp. 1: 60, −54, 34, exp. 2: 52, −52, 28). We co-registered the TPJ location with scalp coordinates using a Brainsight 2.0 frameless stereotaxy system (Rogue Research, Montreal, Canada) to guide coil placement. In the control condition, rTMS was delivered to the Cz according to the 10–20 EEG system for electrode placement. This control region was defined as the crossing of the midline between the inion and the nasion, and the midline between the left and the right preauricular points.

Our rTMS-protocol involved on-line stimulation: trains of stimulation were administered during the presentation of the videos, with two specific events serving as triggers for a stimulation sequence (see below). In each condition, rTMS was delivered at 10 Hz for 500 ms (five pulses) at a default intensity of 75% of the maximum output of the device. However, when the participant indicated clearly that they experienced the stimulation to be unpleasant, or when the participant showed facial muscle contractions, we lowered the intensity until the participant reported no unconfortable sensations induced by TMS. These results in an average intensity of 72.2% of the maximum output of the stimulator [mean intensity of 72.8% in the control condition (Cz) and 71.5% in the RTPJ condition].

Procedure

All subjects participated in two different sessions, with a different region being targeted each session. Each session took place on a different day, and lasted ∼1 h. The order of the two sessions (RTPJ and Cz) was counterbalanced between participants.

Participants were presented with 96 video stimuli per session, divided in six blocks of 16 trials each. Each block contained all eight conditions (observer belief × agent belief state × outcome) in random order. Before engaging in the task itself, there was a training block featuring four practice trials without stimulation. Each stimulus video lasted 13.65 s. Between blocks a short break was included. In these breaks, the coil position was readjusted or the coil was replaced (to avoid overheating) if necessary. TMS was delivered at two specific points during the videos: (i) The point when the agent is about to leave the scene. This moment is critical for a successful formation of belief representation (when the agent leaves the scene, his belief is formed: he thinks that the ball is behind the occluder or not). (ii) The moment in which the agent re-enters the scene. Again, this point, immediately preceding the lowering of the occluder, is a critical moment where the representation of the agent's belief is thought to influence participant's behavior (RTs in the ball detection). In both cases, an rTMS train of 500 ms ensued this trigger.

At the end of the last session of the experiment, an informal debriefing checked if the participant did not process the belief state of the agent consciously. Specifically, the experimenter asked the participant whether she/he could guess what the experiment was about. All participants indicated to be unaware of the goals of the study and of the beliefs of the agent.

Results

First, we did outlier removal using a cut-off at 1000 ms on reaction times. Overall there was a very high accuracy rate of 96.4%. Reaction times were analyzed with a repeated-measure analysis of variance (ANOVA) with participant's belief (ball present P+, ball absent P−), agent's belief (ball present A+, ball absent A−) and TMS target as within-subject factors. Furthermore, we computed the ToM index by comparing the P−A− with the P−A+ condition and the reality bias by comparing P+ *vs* P− conditions (Deschrijver *et al.*, 2016; Nijhof *et al.*, 2016). Alpha level was set to 0.05. We first performed the analysis on the 14 participants who have been stimulated on the main TPJ target and then added the additional five participants who were stimulated slightly

posterior. Since we found the same effect of stimulation in both cases, we decided to pull the participants together ($N = 19$). However, for sake of clarity, we also report below the values of the statistics performed with the smaller sample for any significant interaction with TMS.

We found a main effect of participant's belief [$F(1,18) = 8.36$, $P < 0.05$, $\eta 2 = 0.32$]. Responses were faster when participants expected the ball to be present behind the occluder (P + A− and P + A+ conditions, $M = 339$, $SD = 73$) than when participants did not expect the ball to be there (P−A− and P−A+ conditions, $M = 347$, $SD = 63$). We also found a significant interaction effect between participant and agent belief [$F(1, 18) = 6.07$, $P < 0.05$, $\eta 2 = 0.25$], showing that the agent belief also influenced participants' performance (report mean RTs and t test comparisons). Follow-up pairwise t-tests with Holm-correction revealed that RTs were slower when neither the participant and the agent believed the ball would be behind the occluder as compared to when only the agent, only the participant, or both believed the ball was behind the occluder. In effect, we found a significant difference between the P−A− condition ($M = 355$, $SD = 65$) and the P−A+ ($M = 338$, $SD = 57$, $P = 0.009$) and the P + A− ($M = 334$, $SD = 67$, $P = 0.036$) conditions as well as a marginally significant difference between the P−A− and the P + A+ condition ($M = 343$, $SD = 69.99$, $P = 0.077$). We found no differences between the P−A+, P + A− and the P + A+ conditions (all P's > 0.05). This pattern of data is in line with previous studies using the same task (e.g. Kovács *et al.*, 2010; Bardi *et al.*, 2017a). More importantly, there was a significant interaction between participant belief and TMS target [$F(1,18) = 6.05$, $P < 0.05$, $\eta 2 = 0.25$] (for $N = 14$: $F(1,13) = 5.13$, $P < 0.05$, $\eta 2 = 0.28$). No other interactions with stimulation were significant. These result that TMS significantly interfered with participants' performance, supports the hypothesis that rTPJ is critically involved in spontaneous ToM processing. Below we will explore further the data by contrasting the two alternative hypotheses proposed in the introduction.

The first hypothesis that the rTPJ is involved in representing the mental states of the others predicts that temporarily interfering with rTPJ would reduce the influence of the belief on participants' RTs, that is, the ToM index (Deschrijver *et al.*, 2015; Nijhof *et al.*, 2016). We therefore performed an ANOVA with TMS target (rTPJ, Cz) and condition (P−A−, P−A+) as within-subjects factors. Results confirmed the presence of a significant ToM index as attested by the main effect of condition [$F(1,18) = 13.84$, $P < 0.05$, $\eta 2 = 0.25$]. However rTPJ stimulation did not affect the ToM index, that is no effect of TMS target, or interaction between TMS target and condition was found. Follow-up paired t-tests with a Holm-correction found that the mean reaction times in the P–A– condition ($M = 352$, $SD = 69$) were significantly higher as compared to the P–A+ condition ($M = 333$, $SD = 54$) when the target was rTPJ (ToM index = 19 ms; $P = 0.021$). Similarly, we found that the mean reaction time in the P−A−- condition ($M = 358$, $SD = 53.51$) was marginally significantly higher when compared to the P−-A+ condition ($M = 343$, $SD = 52$) when the Cz was targeted ($P = 0.068$; ToM index = 15 ms). Our results show that rTPJ stimulation did not affect the representation of the agent's belief (Figure 2).

The second hypothesis is that rTPJ is dealing with self-other distinction, i.e. the ability to keep self and other representations apart and switch between them based on task demands. This hypothesis predicts that disruption of rTPJ activity would affect correct participant's predictions based on his/her acquired knowledge. This hypothesis is supported by the significant interaction between TMS target and participant's belief reported above. A series of follow-up pairwise *t*-tests on the interaction using Holm-correction found that mean reaction times for Cz stimulation were significantly slower when the participant did not expect the ball to be behind the occluder (P− condition $M = 351$, $SD = 55$) as compared to when he/she expected the ball to be present (P+ condition $M = 334$, $SD = 56$, $P = 0.006$). The reality bias effect was 17 ms. Crucially, stimulation of the rTPJ led to a suppression of the reality bias: no difference ($P = 0.985$) was found between P− and P+ when the rTPJ was targeted (P− $M = 343$, $SD = 64$; P+ $M = 343$, $SD = 80$) (Figure 3). These results support the hypothesis that disruption of rTPJ activity would disrupt self-other distinction and therefore affects correct participant's predictions based on his/her acquired knowledge.

Discussion

To the best of our knowledge, this is the first study where brains stimulation is applied during a spontaneous ToM task. Our results clearly show that TMS of the rTPJ affects participant's performance in the task. In line with previous studies on explicit ToM (Costa *et al.*, 2008; Young *et al.*, 2010; Krall *et al.*, 2016), this outcome supports the idea that the rTPJ is causally involved in spontaneous ToM, suggesting that neural mechanisms for explicit and spontaneous ToM are, at least partially, overlapping (Hyde *et al.*, 2015; Bardi *et al.*, 2017a; Naughtin *et al.*, 2017).

Moreover, our results show that stimulating TPJ activity during a spontaneous ToM task does not interfere with the participant's ability to represent the other's belief. However, it seems that TMS of the rTPJ strongly affected the predictions the participant held on future events based on his/her acquired knowledge. Our data fit well with a recent conceptualization of the TPJ as a key region for self-other distinction and regulating self and other representations (Brass *et al.*, 2009; Santiesteban *et al.*, 2012). In a recent study on the control of imitation, we applied tDCS to modulate activity in the TPJ area and then we analyzed motor representations related to the self and the other. Results show that increasing activity in the TPJ (anodal tDCS) boosts the self-related representations. Note that in the adopted task the self-representation was task-relevant as participants were instructed to respond to a cue by performing a movement, while ignoring a movement presented on the screen (the other-related representation). This supports the idea that TPJ is involved in detecting a mismatch between self and other depending on task context. When TPJ activity is perturbed or enhanced, this result in a modulation of relevant representations, that is the self or the other, depending on task requirements (Santiesteban *et al.*, 2012; Cook *et al.*, 2014; Bardi *et al.*, 2017b). In this sense, our results are not in contrast with previous brain stimulation studies of explicit ToM showing that performance was worsened when rTPJ was stimulated with TMS (Costa *et al.*, 2008; Young *et al.*, 2010; Krall *et al.*, 2016). The main difference here is that in explicit ToM tasks, participants are instructed to report the mental states of an other person so that the other representations are task-relevant.

Our results are not in line with the study of Santiesteban *et al.* (2012, 2015) concerning the effect of stimulation on ToM task. In their study, participants watched a 15-min video and were required to make inferences about the mental states of the characters during a social interaction. The task was administered immediately after tDCS stimulation of the TPJ area. They found no effect of stimulation on task performance. This could be explained by the fact that typical explicit ToM tasks are often subjected to ceiling effects in normal adults, reducing the possibility for brain stimulation interventions to successfully modulate performance in these tasks. This might also explain why a

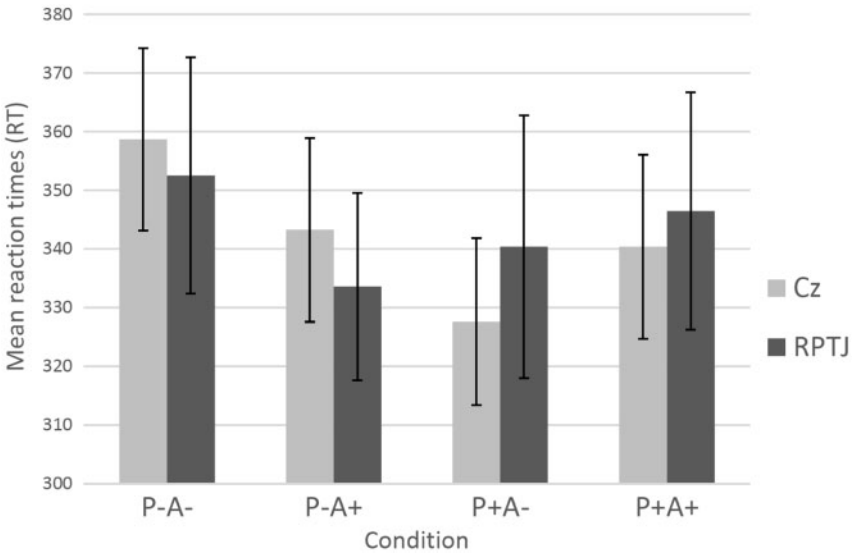

Fig. 2. Effect of TMS (rTPJ *vs* Cz) in all task conditions. The ToM-index is given by the comparison of the conditions on the left-hand side of this graph (P−A− and P−A+). Error-bars represent the SEM.

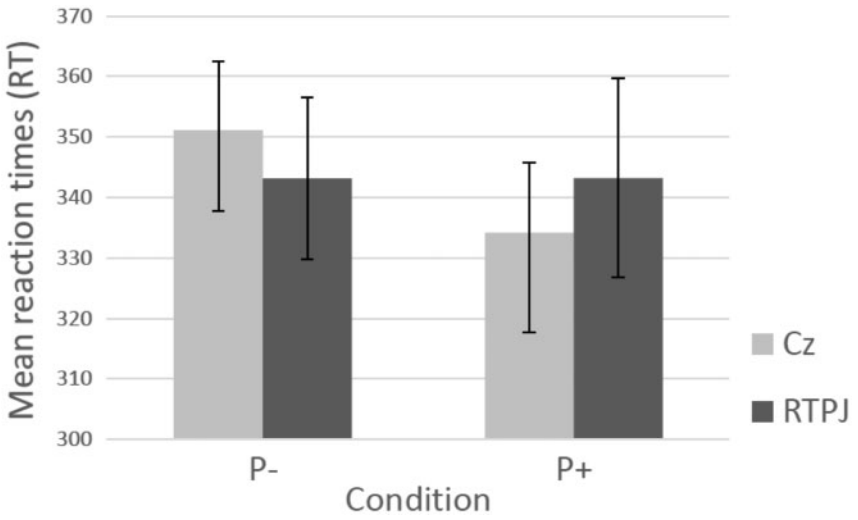

Fig. 3. Interaction between TMS target and participant's belief. In the control condition (Cz stimulation), there was a significant difference between the P− and the P+ conditions. This means that participants' RTs were faster when they expected the ball to be present when the occluder was lowered. When he right TPJ was stimulated with TMS, participants' performance was no longer affected by their knowledge/prediction about the presence of the ball.

surprisingly low number of papers have been published so far using brain stimulation in ToM tasks (Costa *et al.*, 2008; Young *et al.*, 2010; Krall *et al.*, 2016). In this sense, spontaneous ToM tasks, measuring RTs and eye movements could provide a more sensitive measure, allowing a more intensive investigation of the neural mechanisms supporting beliefs processing and social attribution. Understanding the mapping of brain involvement to cognitive models of social cognition is likely to necessitate the use of multiple techniques, like neuroimaging and brain stimulation techniques, and a combination of both. Indeed, such combined approaches could provide a better insight into the specific role of the TPJ in the social domain.

The TPJ has been related to key computations in the social domain, such as ToM, self-other distinction in the control of imitation, agency processing and perspective taking (e.g. Ruby and Decety 2001; Blanke *et al.* 2002; Farrer and Frith 2002; Farrer *et al.* 2003; Saxe and Wexler 2005; Brass *et al.*, 2009; Legrand and Ruby 2009) but also in other non-social processes, such as spatial attention (Corbetta *et al.*, 2000; Mitchell, 2008). Although it is out of the scope of this work to enter the discussion about domain-specific or domain-general neural computation of the TPJ, recent models try to reconcile observations from the social and other field of cognitive neuroscience suggesting that TPJ is involved in reorienting of attention towards unexpected relevant events (Corbetta *et al.*, 2008) or "contextual updating, updating of internal models based on incoming incongruent information" (Geng and Vossel, 2013; Mengotti *et al.*, 2017). Accordingly, TPJ would help updating representations based on changes (we did not necessary attend to) occurring in the environment. In line with this idea, in a previous fMRI study (Bardi *et al.*, 2017a) we have found preferential activation of rTPJ during the tracking period of other's beliefs (belief formation phase). Moreover, following the same idea, we should not expect differences between spontaneous and explicit ToM. This outcome is

in line with the self-other distinction model were TPJ detects an incongruence between internally generated representation and externally triggered representations (e.g. Brass *et al*., 2009).

In the present study, TMS stimulation was applied in two time windows during the presentation of the videos, i.e. immediately after the agent's belief is formed and immediately before the response. This was done to interfere with the belief processing and prevent that the participant could a posteriori reconstruct (perhaps unconsciously) the agent's belief before the response. To the best of our knowledge, a possible dissociation between the cognitive mechanisms involved in forming a belief representation and the maintenance of belief representation has not been deeply addressed in the literature. Future studies should confirm and extend the results of the present study, by using different tasks and investigating possible differences/similarities between belief formation and maintenance processes.

Acknowledgements

This study was founded by a Pegasus Marie Skłodowska-Curie Fellowship (Research Foundation—Flanders-FWO) to Lara Bardi and grant "331323-Mirroring and ToM", FP7 Marie Skłodowska-Curie fellowship to Lara Bardi.

References

Apperly, I. a, Samson, D., Chiavarino, C., Humphreys, G. W. (2004). Frontal and temporo-parietal lobe contributions to theory of mind: neuropsychological evidence from a false-belief task with reduced language and executive demands. *Journal of Cognitive Neuroscience*, **16**(10), 1773–84.

Apperly, I.A., Butterfill, S.A. (2009). Do humans have two systems to track beliefs and belief-like states? *Psychological Review*, **116**(4), 953–70.

Back, E., Apperly, I.A. (2010). Two sources of evidence on the non-automaticity of true and false belief ascription. *Cognition*, **115**(1), 54–70.

Bardi, L., Desmet, C., Nijhof, A., Wiersema, J., Brass, M. (2017a). Brain activation for spontaneous and explicit false belief tasks overlaps: new fMRI evidence on belief processing and violation of expectation. *Social Cognitive and Affective Neuroscience*, **12** (3), 391–400.

Bardi, L., Gheza, D., Brass, M. (2017b). TPJ-M1 interaction in the control of shared representations: new insights from tDCS and TMS combined. *NeuroImage*, **146**, 734–40.

Baron-Cohen, S., Leslie, A.M., Frith, U. (1985). Does the autistic child have a "theory of mind"? *Cognition*, **21**(1), 37–46.

Blanke, O., Ortigue, S., Landis, T., Seeck, M. (2002). Stimulating illusory own-body perceptions. *Nature*, **419**, 269–70.

Brass, M., Derrfuss, J., von Cramon, D.Y. (2005). The inhibition of imitative and overlearned responses: A functional double dissociation. *Neuropsychologia*, **43**(1), 89–98.

Brass, M., Ruby, P., Spengler, S. (2009). Inhibition of imitative behaviour and social cognition. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, **364**(1528), 2359–67.

Carruthers, P. (2016). Two systems for mindreading? Review in. *Philosophy and Psychology*, **7**, 141–62.

Clements, W.A., Perner, J. (1994). Implicit understanding of belief. *Cognitive Development*, **9**, 377–95.

Cook, J.L. (2014). Task-relevance dependent gradients in medial prefrontal and temporoparietal cortices suggest solutions to paradoxes concerning self/other control *Neuroscience and Biobehavioral Reviews*, **42**, 298–302.

Corbetta, M, Kincade, J.M., Ollinger, J.M., McAvoy, M.P., Shulman, G.L. (2000). Volun- tary orienting is dissociated from target detection in human posterior parietal cortex. *Nat. Neurosci.*, **3**, 292–97.

Corbetta, M., Patel, G.H., Shulman, G.L. (2008). The reorienting system of the human brain: from environment to theory of mind. *Neuron*, **58**, 306–24.

Costa, A., Torriero, S., Oliveri, M., Caltagirone, C. (2008). Prefrontal and temporo-parietal involvement in taking others' perspective: TMS evidence. *Behavioural Neurology*, **19**(1-2), 71–4.

Deschrijver, E., Bardi, L., Wiersema, J.R., Brass, M. (2016). Spontaneous theory of mind in adults with autism spectrum disorder: autistic traits predict lesser belief attribution to others. *Cognitive Neuroscience*, **7**, 192–202.

Farrer, C., Franck, N., Georgieff, N., Frith, C.D., Decety, J., Jeannerod, M. (2003). Modulating the experience of agency: a positron emission tomography study. *NeuroImage*, **18**, 324–33.

Farrer, C., Frith, C.D. (2002). Experiencing oneself vs another person as being the cause of an action: the neural correlates of the experience of agency. *NeuroImage*, **15**, 596–603.

Fletcher, P.C., Happé, F., Frith, U., *et al*. (1995). Other minds in the brain: a functional imaging study of "theory of mind" in story comprehension. *Cognition*, **57**, 109–28.

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

Geng, J.J., Vossel, S. (2013). Re-evaluating the role of TPJ in attentional control: contextual updating? *Neuroscience and Biobehavioral Reviews*, **37(10 Pt 2)**, 2608–20.

Gweon, H., Dodell-Feder, D., Bedny, M., Saxe, R. (2012). Theory of mind performance in children correlates with functional specialization of a brain region for thinking about thoughts. *Child Development*, **83**, 1853–68.

Heyes, C. (2014). Submentalizing: I'm not really reading your mind. *Psychological Science*, **9**, 121–43.

Hyde, D.C., Aparicio Betancourt, M., Simon, C.E. (2015). Human temporal-parietal junction spontaneously tracks others' beliefs: a functional near-infrared spectroscopy study. *Human Brain Mapping*, **36**, 4831–46.

Herman, L.M., Kantowitz, B.H. (1970). The psychological refractory period effect: only half the double-stimulation story? *Psychological Bulletin*, **73**(1), 74–88.

Legrand, D., Ruby, P. (2009). What is self-specific? Theoretical investigation and critical review of neuroimaging results. *Psychological Review*, **116**, 252–82.

Kovács, Á.M., Téglás, E., Endress, A.D. (2010). The social sense: susceptibility to others' beliefs in human infants and adults. *Science (New York, N.Y.)*, **330**(6012), 1830–4.

Kovács, Á.M., Kühn, S., Gergely, G., Csibra, G., Brass, M. (2014). Are all beliefs equal? spontaneous belief attributions recruiting core brain regions of theory of mind. *PLoS One*, **9**(9), e106558.

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.

McKinnon, M.C., Moscovitch, M. (2007). Domain-general contributions to social reasoning: theory of mind and deontic reasoning re-explored. *Cognition*, **102**, 179–218.

Meert, G., Wang, J., Samson, D. (2017). Efficient belief tracking in adults: The role of task instruction, low-level associative processes and dispositional social functioning. *Cognition*, **168**, 91–8.

Mengotti, P., Dombert, P.L., Fink, G.R., Vossel, S. (2017). Disruption of the right temporoparietal junction impairs probabilistic belief updating. *The Journal of Neuroscience*, **37**(22), 3683–16.

Mitchell, J.P. (2008). Activity in right temporo-parietal junction is not selective for theory-of-mind. *Cerebral Cortex*, **18**, 262–71.

Naughtin, C.K., Horne, K., Schneider, D., Venini, D., York, A., Dux, P.E. (2017). Do implicit and explicit belief processing share neural substrates? *Human Brain Mapping*, **38**, 4760–72.

Nijhof, A.D., Brass, M., Bardi, L., Wiersema, J.R. (2016). Measuring mentalizing ability: a within-subject comparison between an explicit and spontaneous version of a ball detection task. *PLoS One*, **10**, e0164373.

Onishi, K.H., Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science (New York, N.Y.)*, **308**(5719), 255–8.

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, teglas, and endress (2010). *Psychological Science*, **26**(9), 1353–67.

Premack, D., Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, **1**(4), 515–26.

Rossi, S., Hallett, M., Rossini, P.M., Pascual-Leone, A. (2009). Safety, ethical considerations, and application guidelines for the use of transcranial magnetic stimulation in clinical practice and research. *Clinical Neurophysiology*, **120**(12), 323–30.

Ruby, P., Decety, J. (2001). Effect of subjective perspective taking during simulation of action: a PET investigation of agency. *Nature Neuroscience*, **4**, 546–50.

Ruby, P., Decety, J. (2003). What you believe versus what you think they believe: A neuroimaging study of conceptual perspective-taking. *European Journal of Neuroscience*, **17**(11), 2475–80.

Samson, D., Apperly, I.A., Chiavarino, C., Humphreys, G.W. (2004). Left temporoparietal junction is necessary for representing someone else's belief. *Nature Neuroscience*, **7**(5), 499–500.

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), 1255–66.

Santiesteban, I., Banissy, M.J., Catmur, C., Bird, G. (2012). Enhancing social ability by stimulating right temporoparietal junction. *Current Biology: CB*, **22**(23), 2274–7.

Santiesteban, I., Banissy, M.J., Catmur, C., Bird, G. (2015). Functional lateralization of temporoparietal junction - imitation inhibition, visual perspective-taking and theory of mind. *European Journal of Neuroscience*, **42**(8), 2527–33.

Saxe, R., Kanwisher, N. (2003). People thinking about thinking people. The role of the temporo-parietal junction in "theory of mind". *Neuroimage*, **19**, 1835–42.

Saxe, R., Powell, L.J. (2006). It's the thought that counts: specific brain regions for one component of theory of mind. *Psychological Science*, **17**, 692–9.

Saxe, R., Wexler, A. (2005). Making sense of another mind: the role of the right temporo-parietal junction. *Neuropsychologia*, **43**, 1391–9.

Schneider, D., Bayliss, A.P., Becker, S.I., Dux, P.E. (2011). Eye movements reveal sustained spontaneous processing of others' mental states. *Journal of Experimental Psychology: General*, **141**(3), 433–8.

Schneider, D., Nott, Z.E., Dux, P.E. (2014a). Task instructions and spontaneous theory of mind. *Cognition*, **133**(1), 43–7.

Schneider, D., Slaughter, V.P., Becker, S.I., Dux, P.E. (2014b). Spontaneous false-belief processing in the human brain. *NeuroImage*, **101**, 268–75.

Schneider, D., Slaughter, V.P., Dux, P.E. (2017). Current evidence for automatic Theory of Mind processing in adults. *Cognition*, **162**, 27–31.

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*, **42**, 9–34.

Senju, A., Southgate, V., Snape, C., Leonard, M., Csibra, G. (2011). Do 18-month-olds really attribute mental states to others? A critical test. *Psychological Science: A Journal of the American Psychological Society/APS*, **22**(7), 878–80.

Sommer, M., Döhnel, K., Sodian, B., Meinhardt, J., Thoermer, C., Hajak, G. (2007). Neural correlates of true and false belief reasoning. *NeuroImage*, **35**(3), 1378–84.

Southgate, V., Senju, A., Csibra, G. (2007). Action anticipation through attribution of false belief by 2-year-olds. *Psychological Science*, **18**(7), 587–92.

Sowden, S., Catmur, C. (2015). The role of the right temporoparietal junction in the control of imitation. *Cerebral Cortex (New York, N.Y.: 1991)*, **25**(4), 1107–13.

Surian, L., Caldi, S., Sperber, D. (2007). Attribution of beliefs by 13-month-old infants. *Psychological Science*, **18**(7), 580–6.

Van Overwalle, F. (2009). Social cognition and the brain: A meta-analysis. *Human Brain Mapping*, **30**(3), 829–58.

van der Wel, R.P.R.D., Sebanz, N., Knoblich, G. (2014). Do people automatically track others' beliefs? Evidence from a continuous measure. *Cognition*, **130**(1), 128–33.

Wellman, H.M., Cross, D., Watson, J. (2001). Meta-analysis of theory-of-mind development: the truth about false belief. *Child Development*, **72**(3), 655–84.

Wimmer, H., Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, **13**, 103–28.

Young, L., Camprodon, J.a., Hauser, M., Pascual-Leone, a., Saxe, R. (2010). Disruption of the right temporoparietal junction with transcranial magnetic stimulation reduces the role of beliefs in moral judgments. *Proceedings of the National Academy of Sciences*, **107**(15), 6753–8.