NeuroImage 196 (2019) 329–336

Contents lists available at ScienceDirect

NeuroImage

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

Causal evidence of right temporal parietal junction involvement in implicit Theory of Mind processing

Hannah L. Filmer *, Amaya Fox, Paul E. Dux

School of Psychology, The University of Queensland, St Lucia, 4072, Australia

A B S T R A C T

The ability to represent the internal thoughts, beliefs and desires of others, and recognise that these might be distinct from one's own, is crucial for adaptive social interaction. Such operations are thought to tap Theory of Mind (ToM), with its importance underscored by the link between ToM impairment and a range of neurodevelopmental disorders (e.g., Autism and Schizophrenia). Extensive investigations into the neural substrates of ToM, when individuals have to make overt/explicit judgments concerning others, have highlighted a link with a network of regions including the temporal parietal junction (TPJ), particularly in the right hemisphere. Recently, evidence has emerged that ToM can also operate implicitly and that this may be particularly impaired in Autism. However, very few studies have examined the neural basis of implicit ToM and none have employed methods allowing casual inferences to be made. Here, using brain stimulation, a Sally-Anne false-belief task, and eye-tracking we show that right TPJ is causally involved in ToM judgments that are made implicitly. These findings have implications for characterising the neural substrates of a key executive function, determining the extent to which implicit and explicit ToM draw on overlapping neural architecture and, potentially, better understanding of disorders tied to ToM impairment.

1. Introduction

To adaptively negotiate our rich and socially complex world one must possess a Theory of Mind (ToM) – the ability to represent the internal thoughts, beliefs and desires of others, and recognise that these might be distinct from one's own (Frith and Frith, 2005; Premack and Woodruff, 1978). Given the importance of social interaction for human beings and that impaired ToM is associated with a range of neurodevelopmental disorders (Baron-Cohen et al., 1985; Brüne, 2005; Moran et al., 2011), the processes associated with *mentalising* about others' internal cognitions has been a topic of intense investigation across Psychology and Neuroscience (see Schneider et al., 2015; Schneider et al., 2017).

Key to assessing ToM has been the extent to which individuals are able to pass false-belief tasks which is thought to be a uniquely human ability and an important developmental milestone (Wimmer and Perner, 1983). In the classic Sally-Anne false belief paradigm, an agent "Sally" moves a ball in one of two locations and then leaves the room. Following this, "Anne", a second agent, moves the ball to the other location. When "Sally" returns, the critical question is: where will she search for the ball? To pass subjects must realise that "Sally" has a false belief relative to themselves.

Traditionally, it was thought the children only developed the capacity to represent others' false beliefs around 3–4 years of age, however, more recently, evidence has emerged via non-verbal measures (e.g., eye-tracking), that much younger children (1–2 years) can spontaneously track the mental states of others (e.g., Senju et al., 2011; Southgate et al., 2007; Scott and Baillargeon, 2017). In addition, adults also display these spontaneous ToM effects and, importantly, when debriefed have no awareness of having engaged in mentalising (Schneider et al., 2012; Schneider et al., 2012; Schneider et al., 2014) – thus, implicit ToM (iToM) processing. Such findings support a two-pathway hypothesis for ToM (Apperly and Butterfill, 2009; Butterfill and Apperly, 2013; Clements and Perner, 1994): An early developing pathway that allows the implicit and rapid extraction of others' belief-like states, and a later developing pathway involved in the conscious and deliberate analysis of others' internal cognitions, which also draws on associated executive functions – explicit ToM (eToM) processing. Findings that people with Autism show particular impairments in iToM, but not eToM (Schneider et al., 2013; Senju et al., 2009), and vice-versa in healthy older adults (Grainger et al., 2018), further supports the two-pathway account. Having said this, recently there have been several large studies providing evidence questioning the stability/replicability of iToM findings, particularly with children (Kulke et al., 2018a,b), thus illustrating the importance of pre-registration and open science practices when conducting such studies (see below).

Here we focus on the neural substrates of iToM. Functional magnetic resonance imaging (fMRI) studies have consistently highlighted a network of regions - medial prefrontal cortex (mPFC), temporoparietal

* Corresponding author. McElwain Building, Campbell Road, St Lucia, 4072, Australia.
E-mail address: h.l.filmer@gmail.com (H.L. Filmer).

https://doi.org/10.1016/j.neuroimage.2019.04.032
Received 13 February 2019; Received in revised form 2 April 2019; Accepted 8 April 2019
Available online 11 April 2019

junction (TPJ), superior temporal sulcus (STS), precuneus and temporal poles - that preferential activate when subjects make explicit judgements regarding the mental state of a protagonist, compared with making a physical judgement (e.g., Dodell-Feder et al., 2011; Gallagher et al., 2000; Saxe and Kanwisher, 2003). In particular, the right hemisphere TPJ is thought to be a core component of the explicit mentalising network, as it is involved in a wide range of ToM tasks (Saxe, 2009; Saxe and Kanwisher, 2003; Schurz et al., 2014). Recent studies have also implicated this network in iToM (Naughtin et al., 2017; Schneider et al., 2014), however, to date, there is no casual evidence implicating the right TPJ in implicit belief processing. Here we address this question in a pre-registered study by combining an iToM Sally-Anne paradigm, eye-tracking, and transcranial direct current stimulation (tDCS; Filmer et al., 2014). Specifically, as stated in our pre-registration, if the right TPJ is causally involved in iToM processing, it is expected that active stimulation, via tDCS, will modulate the difference in fixation durations/first fixations for the no-ball location in false-belief trials compared to true-belief trials in an anticipatory looking iToM Sally-Anne paradigm. Note, no directional hypotheses are made here regarding the effect of stimulation polarity given the propensity for both anodal and cathodal stimulation influence to be determined by a wide range of task based and stimulation protocol factors (see Filmer et al., 2014).

2. Materials and methods

All aspects of the study design, methods, data exclusions, and analysis plan were pre-registered at the Open Science Framework (https://osf.io/rxpb2/).

2.1. Subjects

One hundred and twenty-one subjects (mean age = 22.28, range = 17–42 years, 71 females) were recruited through the paid participation pool of The University of Queensland's School of Psychology. Each subject was paid $40 for taking part in the study. Table 1 shows the subject demographics for each stimulation group. All subjects had normal or corrected to normal vision and were right-handed.

2.2. Session outline

Subjects completed the Autism-Spectrum Quotient questionnaire (AQ; Baron-Cohen et al., 2001) at the end of the session to screen for autistic traits. This was completed at the end of the session to ensure that taking the questionnaire did not hint to subjects that our task examined ToM. Three subjects scored above the clinical cut-off of 32/50 and were excluded from analysis (mean AQ score of remaining subjects = 19.13). Importantly the groups did not differ in terms of AQ scores ($BF_{01} = 5.964$; $F < 1$, $p = .598$). Twenty-four subjects were also excluded from analysis as their responses in the funnelled debriefing procedure (see below) indicated they might have engaged in explicit ToM processing. A further 16 subjects were excluded for not looking at the relevant part of the screen during critical points in the task and one subject was excluded due to eye tracker calibration issues. Thus, the final sample consisted of 77 subjects (mean age = 22.29, range = 17–42 years, 41 females). All subjects gave informed written consent, passed an in-house and ethically approved tDCS safety-screening questionnaire, and were paid for their time. The University of Queensland Human Research Ethics Committee approved the study.

Table 1
Number of subjects (n), mean age, and gender for the different stimulation groups.

Stimulation Type	*n*	Age	Gender
Anodal	27	22.19 (4.57)	16F
Cathodal	27	22.22 (3.08)	13F
Sham	23	22.48 (3.48)	12F

Note: Standard deviations are presented in parentheses. (F = female).
Task.

Subjects completed an anticipatory looking iToM task involving Sally-Anne-like movie clips, adapted from the second experiment in Schneider et al. (2012a,b). The task consisted of 40 filler trials and 20 experimental trials, of which 10 were false-belief and 10 were true-belief trials, presented randomly over 50 min.

A black fixation cross was presented in the centre of the screen for 500 ms at the beginning of each trial. All movies began with an actor seated behind a desk with two boxes on it. The actor also wore a visor in all movies to conceal her eyes to eliminate gaze-cuing effects (Frischen et al., 2007; Schneider et al., 2012a,b). In the false-belief trials (Fig. 1; http://youtu.be/HMaLIBRwN-Q) the actor watched the puppet place the ball in one box and then transfer it into the other box. Next, the actor exited the room and the puppet switched the ball back into the original box. Therefore, when the actor returned she had an incorrect belief about the location of the ball. The true-belief trials (Fig. 1, http://youtu.be/yf2vVSaaF9Q) were virtually identical to the false belief trials; however, the actor left after the ball was placed into the first box. Thus, the actor was not present when the ball was transferred to the other box and returned to the original box. Thus, in these trials, when the actor returned her belief about the ball location was correct. In all experimental trials the actor sat down when she returned to the room and the final frame froze for approximately 6s. The experimental trials had duration of 66s–73s. Two versions of each type of experimental trial were used to counterbalance the initial and final location of the ball: true belief right, false belief right, true belief left, and false belief left. Two types of filler trials were used (Fig. 2). In the 20 'moving ball' trials, a koala-shaped hand puppet hid a red ball in one of the boxes (duration: 29s). In the 20 'static ball' trials, the red ball is sitting on the lid of one of the boxes (duration: 3s). All filler trials conclude with the actor picking up the ball to illustrate that they were engaged with this object.

Subjects were instructed to watch the videos; however, their main task was to differentiate between high (2000 Hz) and low (500 Hz) frequency tones as quickly and accurately as possible (tone duration: 100 ms). These were included to distract subjects from the belief manipulation to increase the likelihood that only iToM operations were engaged. Subjects responded as quickly and as accurately as possible by pressing the '1' key for high tones and the '2' key for low tones, using their right hand on the number pad of a standard keyboard. The tones were presented throughout all experimental trials and all 29s filler trials. No tones were presented during the 3s filler trials. A maximum of 10 tones sounded per trial and they played at least 3s apart. Each tone sequence began after the first 1s of the trial had elapsed and concluded with at least 6s remaining of the trial. Within these specifications, tone sequence was pseudo-randomly allocated within each trial.

Eye movement data was collected from three regions of interest: the empty box, the box holding the ball, and the actor's face. Fixations on other areas of the screen were not recorded. Anticipatory looking behaviour was measured via subjects' first fixations and the duration of fixation on the three regions of interest during the final 6s of each experimental trial. Both these dependent variables were included as each has been employed when studying iToM. Fixation duration was the length of fixation on each of the three regions of interest relative to total length of fixation on the three regions during the 6s period. Thus, it can be determined whether subjects first fixated and fixated longer on the empty box when the actor believed the ball was there (false-belief trial), compared to when the actor correctly believed the ball was not there (true-belief trial).

2.3. Stimulation protocol

Before completing the iToM task, subjects were pseudo-randomly allocated to receive anodal, cathodal, or sham tDCS (see Table 1). Allocation to stimulation groups occurred sequentially in the order of anodal,

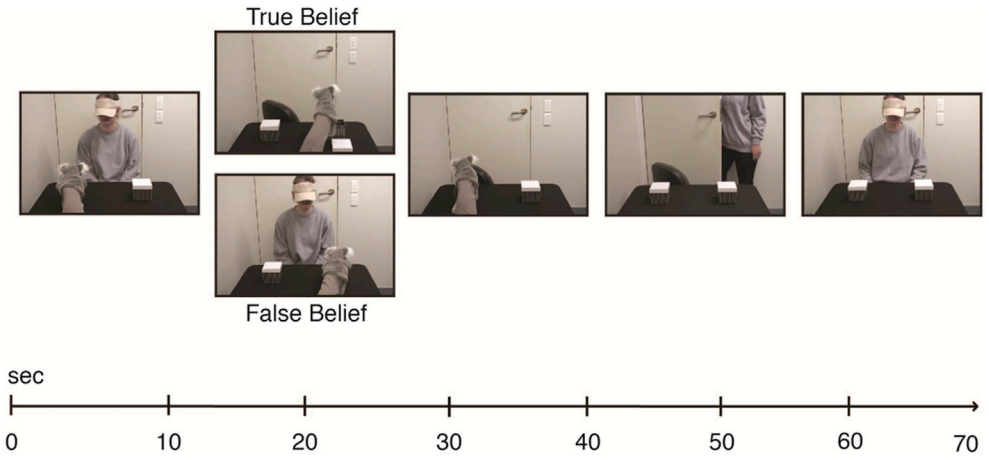

Fig. 1. Still images from the false and true belief trial type movies. Subjects passively watched the movies while completing a tone task and having their eye-movements recorded. Note, the time point where the actor leaves the room determines the nature of their belief.

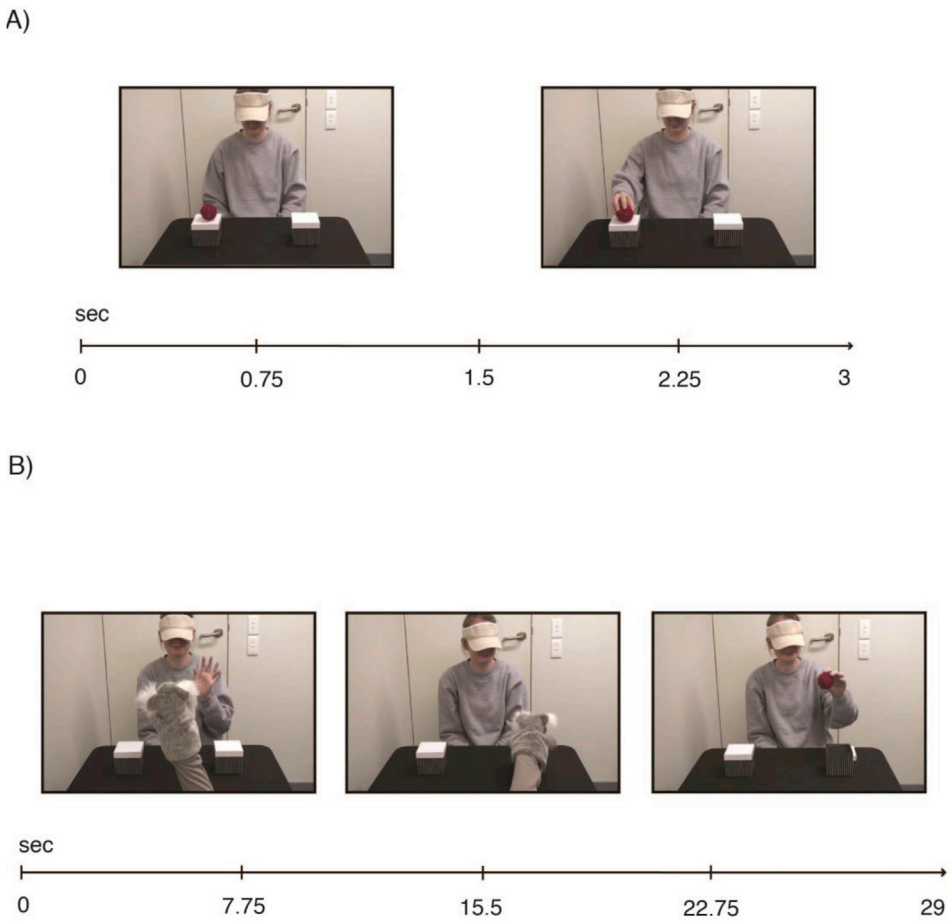

Fig. 2. Still images from two types (A, B) of filler movies. These were employed to make it clear to subjects that the actor was engaged with the ball.

cathodal, and sham, such that the same stimulation type was administered to every third subject. This was a single-blinded procedure, with subjects unaware of the three different types of stimulation and which type they were administered.

The tDCS was performed offline with both anodal and cathodal stimulation at 1 mA intensity. This type of protocol has previously been shown to induce after-effects in the motor cortex for at least an hour (Nitsche et al., 2003; Nitsche and Paulus, 2001). Therefore, the after-effects induced by the 20-min stimulation at 1 mA in the current study should have lasted for the duration of the task (approximately 50 min).

A Neuroconn stimulator was used to administer the 1 mA current to the brain via two 5×5 cm rubber electrodes. The electrodes were encased in synthetic saline-soaked sponges and secured onto the scalp using rubber bands. The current density was 0.04 mA/cm^2, which falls within the present safety guidelines (Bikson et al., 2009; Bikson et al., 2010). To target the right TPJ, the target electrode was placed in a vertical orientation over CP6 (Santiesteban et al., 2015), located using the international 10–20 EEG system (Jasper, 1958). The reference electrode was placed in a horizontal orientation over the contralateral mastoid, as this is a non-brain region and would not influence iToM processes (Fig. 3;

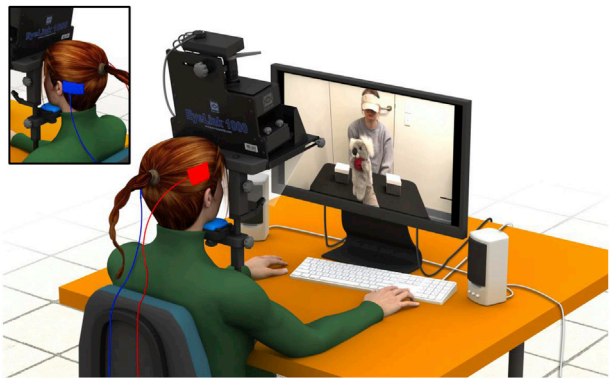

Fig. 3. Graphical depiction of the key elements of the experimental set up. Electrode configuration (Anode), eye-tracking rig and computer/response set up.

Filmer et al., 2014).

During the stimulation, the current was ramped up over 30 s and held constant at 1 mA during the active stimulation period, before ramping down over 30 s. The active stimulation period lasted for 19 min in the anodal and cathodal stimulation conditions, while in the sham stimulation condition the active stimulation period lasted for 15 s. This sham stimulation protocol produces the initial itching and/or tingling sensations normally experienced during tDCS, which largely prevents subjects from differentiating between the active and sham stimulation conditions, without influencing cortical excitability (Gandiga et al., 2006; Woods et al., 2016). Therefore, the sham stimulation was used as a baseline control measure. Subjects in the sham condition were pseudo-randomly allocated to either anodal ($n = 10$) or cathodal ($n = 13$) sham stimulation, with every second sham subject assigned to the cathodal sham group.

2.4. Questionnaire measures

AQ. The AQ measures the degree of autistic-like traits found in members of the general population (Baron-Cohen et al., 2001). The questionnaire contained 50 items (e.g., "I find social situations easy") and responses were made on a 4-point scale (1 = *definitely agree* and 4 = *definitely disagree*). Any subjects who scored above the clinical threshold of 32 were excluded from analysis.

Funnelled Debriefing Procedure. The Funnelled Debriefing Questionnaire, adapted from Bargh and Chartrand (2000), assesses whether subjects consciously monitored the actor's belief state throughout the task and/or had knowledge of ToM processing (Schneider et al., 2012a, b). The questionnaire begins with broad questions (e.g., "What do you think the purpose of the experiment was?"), which become increasingly more specific as the questionnaire progresses (e.g., "What do you think was the story in the videos?"). If any subject responses indicated they had engaged in mentalising or had explicit awareness of the belief manipulations during the task, they were excluded from analysis (e.g. "The koala was trying to trick the person").

2.5. Procedure

Fig. 3 provides a graphical depiction of all aspects of the experimental set up. All sessions were conducted on a 17" Samsung LCD monitor running the Windows operating system. The movie stimuli were presented using MATLAB R2009b software and the Psychophysics Toolbox 3.0.9 extension (Brainard, 1997; Pelli, 1997). Subjects' right pupil movements were recorded using an Eyelink 1000, with a sampling rate of 500 Hz (SR Research, Mississauga, Ontario, Canada). Subjects were seated 58 cm from the monitor, which was controlled by a chin rest. The audio was played through external speakers connected to the computer.

At the beginning of the session, subjects read two information sheets: a Plain language statement and general information sheet. These documents provided information about the task and tDCS (without mentioning anything related to ToM). Subjects then filled out the Safety Screening Questionnaire to establish their eligibility to receive stimulation. They also completed the 'before tDCS' columns of an Adverse Effects Questionnaire in order to compare any symptoms present before receiving the stimulation to those present after receiving tDCS. Finally, subjects signed the Statement of Informed Consent.

Before undergoing tDCS, subjects completed a practice block of the iToM task, consisting of six trials. Once subjects had comfortably positioned their chin on the chin rest, the eye tracker was calibrated to ensure their pupil movements could be successfully recorded. The practice block began with an instruction screen that detailed the responses required for each tone (i.e., press '1' for high tones and press '2' for low tones). These directions were also verbally relayed to the subject, along with additional instructions to look at the fixation cross when it appeared and to minimise blinking when observing the movie clips. The subsequent screen presented subjects with examples of the high and low tones, after they initiated by pressing the spacebar. At the end of the practice block subjects received feedback about their tone detection accuracy.

After head measurements had been performed to locate the stimulation site and the electrodes were in place for the tDCS, subjects were instructed to remain still, keep their eyes open, and limit speaking during the stimulation. Subjects were also informed that they may experience itching or tingling sensations at the electrode sites. Once the 20-min stimulation had elapsed the electrodes were immediately removed and two additional eye tracker calibrations were performed. The iToM task typically commenced between two to 5 min after the stimulation ended. The task involved ten blocks of six trials. Subjects were presented with a rest screen displaying their tone detection accuracy at the end of each block and they pressed any key when they felt ready to begin the next block.

Once they had finished the iToM task subjects completed the AQ. The experimenter then verbally asked the subjects the funnelled debriefing questions and recorded each response. Finally, each completed the 'after tDCS' columns of the Adverse Effects Questionnaire and a debriefing questionnaire. Each session ran for approximately 2 h.

2.6. Analyses

As outlined in the OSF pre-registration, all analyses were conducted using Bayesian statistics. In addition, in order to allow comparison with previous studies, each analysis was also performed using a frequentist approach (i.e., null hypothesis significance testing), with a two-tailed alpha level of 0.05 employed. Interpretations of the data are based on the Bayesian analyses, as Bayesian inference is more conservative than frequentist inference and quantifies the evidence in favour of both the null and the alternative hypotheses (Van de Schoot et al., 2014). Importantly, Bayesian analysis also reduces the likelihood of false positives emerging from multiple comparisons, as it quantifies the evidence for each hypothesis instead of providing a probability score (Gelman et al., 2012). Here, the inverse Bayes Factor (BF_{10}) was calculated. Thus, evidence for the alternative hypothesis is considered strong for BF_{10} values greater than 10, meaningful/moderate for values between 3 and 10, and anecdotal for values between 1 and 3. Conversely, BF_{10} values less than 1 demonstrate evidence for the null hypothesis. Values of approximately one indicate no meaningful evidence for either hypothesis (Jeffreys, 1961; Wasserman, 2000).

As outlined in the analysis plan of the pre-registration (https://osf.io/rxpb2/), while subjects' eye movement data was collected from three regions (the empty box or no ball location, the box holding the ball or ball location, and the actor's face), the key region of interest for anticipatory looking behaviour – indicative of implicit false belief reasoning – is the no ball location. Therefore, the difference in the percentage of first fixations and fixation duration at the no ball location between the false-belief and true-belief trials (i.e., false-belief minus true-belief) were calculated for

each subject. All analyses were conducted using these difference scores, where positive values demonstrate greater first fixations or fixation duration towards the no ball location in false-belief compared to true-belief trials. Furthermore, to determine whether any stimulation effects remained uniform throughout the entire task each dependent variable analysis was tested at three different time windows: the whole task, the first five blocks only, and the last five blocks only.

2.7. Data Availability

The task scripts, data extraction scripts, behavioural data, and MRI data are publicly available (Filmer et al., 2019).

3. Results

3.1. First fixations

iToM analysis. This initial analysis checked whether, in the absence of stimulation effects (i.e., for the sham subjects), evidence of implicit false belief reasoning was present. Bayesian one-sample t-tests were employed to evaluate, for a Jeffrey-Zellner-Siow (JZS; i.e., non-informative; Zellner and Siow, 1980) prior distribution, the proportion of evidence supporting the null hypothesis (i.e., first fixation difference scores are not significantly different from 0) compared to that supporting the alternative hypothesis (i.e., first fixation difference scores are significantly different from 0).

Fig. 4 shows the mean percentage of first fixations on the no ball location for false- and true-belief trials in the sham condition. There was strong evidence for the alternative hypothesis across the whole task, $BF_{10} = 10.67$, $t(22) = 3.21$, $p = .004$, and moderate evidence for the alternative hypothesis in the last five blocks, $BF_{10} = 3.00$, $t(22) = 2.56$, $p = .018$. Evidence for the null hypothesis was demonstrated in the first five blocks, $BF_{01} = 0.36$, $t(22) = 1.06$, $p = .299$. These findings suggest that subjects did indeed engaged in implicit false belief processing, however this increased throughout the experiment.

Influence of tDCS on iToM. Bayesian one-way between-groups analyses of variance (ANOVAs, JZS prior), assessed the ratio of evidence supporting the null hypothesis (i.e., active stimulation does not modulate the difference in first fixations at the no ball location in false-belief compared to true-belief trials) compared to that supporting the alternative hypothesis (i.e., active stimulation modulates the difference in first fixations at the no ball location in false-belief compared to true-belief trials). Of note, in the OSF preregistration, we planned to run these analyses along with pairwise comparisons to contrast the difference between the three stimulation conditions. Therefore, Bayesian independent-groups t-tests using a JZS prior were conducted to assess the proportion of evidence supporting the null hypothesis (i.e., first fixation difference scores do not significantly differ between the conditions) relative to that supporting the alternative hypothesis (i.e., first fixation difference scores significantly differ between the conditions).

First fixations across all trials. The ANOVA revealed anecdotal evidence for the alternative hypothesis: that active stimulation modulates the difference in first fixations at the no ball location between false- and true-belief trials, $BF_{10} = 1.32$, $F(2, 74) = 3.23$, $p = .045$ (Fig. 5). The independent-groups *t*-test demonstrated moderate evidence for the alternative hypothesis, a significant difference in first fixation difference scores, for the anodal relative to the sham stimulation conditions across the whole task, $BF_{10} = 3.38$, $t(48) = -2.50$, $p = .016$. Specifically, the difference between false- and true-belief conditions at the no ball location was greater for the sham group than the anodal group. Thus, anodal tDCS disrupted iToM processing. The other t-tests revealed evidence for the null hypothesis when contrasting the anodal and cathodal conditions, $BF_{10} = 0.71$, $t(52) = -1.53$, $p = .13$ and the cathodal and sham conditions, $BF_{10} = 0.45$, $t(48) = -1.07$, $p = .291$. At the request of a reviewer, as exploratory analyses, we also assessed whether including the AQ scores as a random factor would account for meaningful variance in the data (see Donaldson et al., 2018). When comparing the key anodal and sham stimulation conditions, adding in AQ scores as a random factor increased the BF10 from 3.38 to 7.69. Thus it appears that accounting for variance with AQ scores leads to a meaningful increase in evidence for the effect of anodal stimulation on the first fixation values.

First fixations across the first half of trials. The ANOVA provided evidence for the null hypothesis, demonstrating, that across the first half of trials, active stimulation did not modulate the difference in first fixations at the no ball location between false-belief and true-belief trials, $BF_{10} = 0.12$, $F(2, 74) = 0.10$, $p = .910$ (see Fig. 6). Similarly, all follow-up independent-groups t-tests revealed evidence for the null hypothesis, no significant difference in first fixation difference scores, across all stimulation conditions, $BF_{10} < 0.31$, $ts < 0.45$, $ps > .652$.

First fixations across the last half of trials. The ANOVA revealed anecdotal evidence for the alternative hypothesis, that active stimulation modulates the difference in first fixations at the no ball location for false-belief compared to true-belief trials, in the last five blocks of the task, $BF_{10} = 1.95$, $F(2, 74) = 3.75$, $p = .028$ (see Fig. 6). A follow-up independent-groups *t*-test demonstrated moderate evidence for the

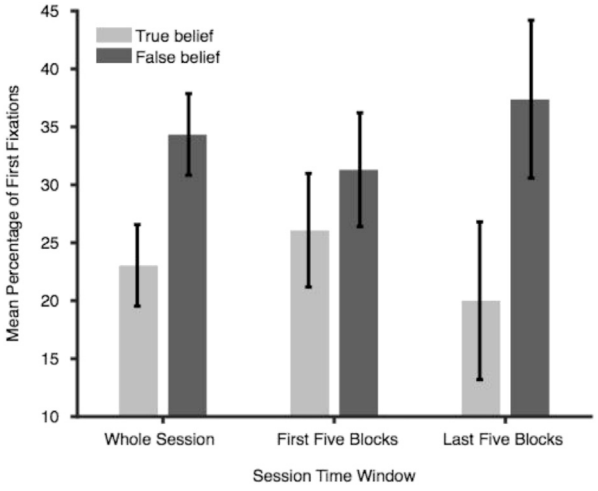

Fig. 4. Mean percentage of first fixations towards the no ball location in the sham condition as a function of time window (whole task, first five blocks, or last five blocks) and belief condition (true- or false-belief). Error bars denote one standard error of the difference.

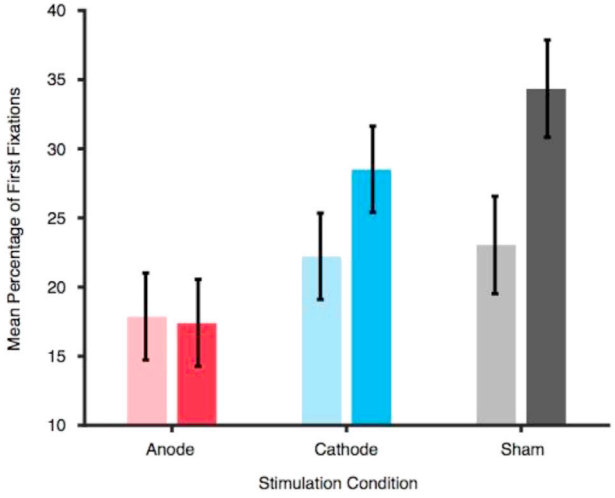

Fig. 5. Mean percentage of first fixations towards the no ball across all trials as a function of belief and stimulation condition. Pale bars represent true belief trials, darker bars false belief. Error bars denote one standard error of the difference.

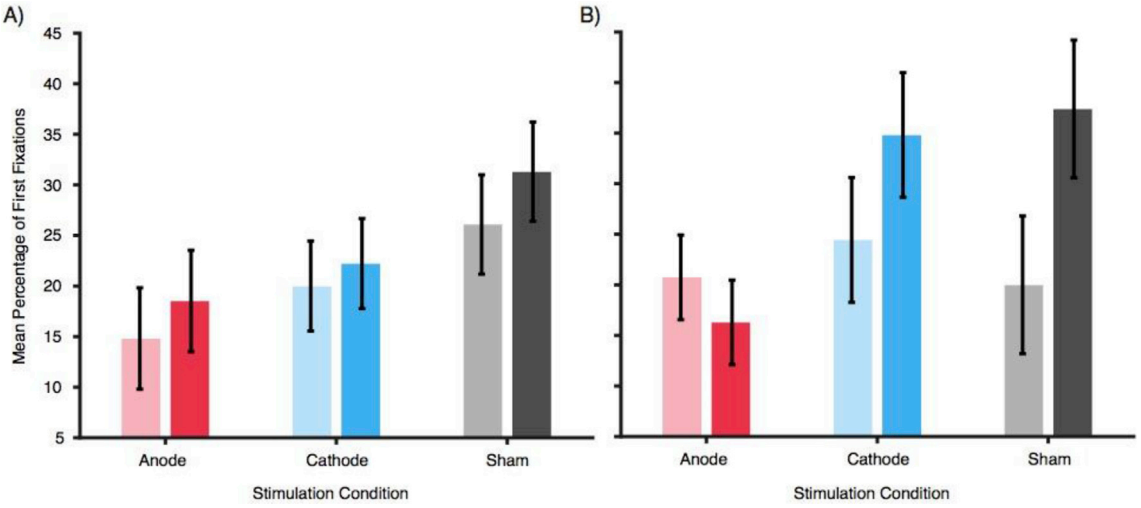

Fig. 6. Mean percentage of first fixations towards the no ball across the first (A) and second half of the trials (B) as a function of belief and stimulation condition. Pale bars represent true belief trials, darker bars false belief. Error bars denote one standard error of the difference.

alternative hypothesis, that a meaningful difference in first fixation difference scores, was present when contrasting the anodal and sham stimulation conditions, $BF_{10} = 6.45$, $t(48) = -2.82$, $p = .007$. Again, subjects in the sham condition first fixated on the no ball location more in the false-belief compared to true-belief trials in the sham condition relative to the anodal condition. A second *t*-test revealed anecdotal evidence for the alternative hypothesis when comparing anodal and cathodal conditions, $BF_{10} = 1.37$, $t(52) = -1.99$, $p = .052$, with subjects in the latter condition first fixating on the no ball location in false-belief relative to true-belief trials more than subjects in the anodal condition. The final *t*-test showed evidence for the null hypothesis when contrasting the cathodal and sham conditions, $BF_{10} = 0.36$, $t(48) = -0.77$, $p = .448$.

Exploratory Analysis. At the suggestion of the reviewer, we also conducted analyses to assess if our tDCS protocol influenced iToM specifically, or whether it impacted another general cognitive process (e.g., attention). First, we examined the extent to which TPJ anodal stimulation influenced false-belief performance as opposed to true-belief performance. When comparing first fixations at the no ball location there was evidence of a meaningful difference between false-belief anodal and false-belief sham conditions ($BF_{10} = 8.03$, $p = .006$), whereas there was anecdotal for the null when comparing the true-belief anodal and true-belief sham conditions ($BF_{01} = 2.28$, $p = .3$). Thus, our stimulation protocol specifically influenced false-belief processing, the condition that puts the greatest load on iToM. Second, when contrasting performance via an ANOVA on the tone detection task across the stimulation conditions (Anode: 97.73%; Cathode: 97.3%; Sham: 97.22%) there was evidence for the null hypothesis that the conditions do not differ ($BF_{01} = 6.73$, $F < 1$). Collectively it appears that our tDCS approach specifically influenced iToM processes rather than another general cognitive operation.

3.2. Fixation duration

iToM analysis; Fixation duration across all trials; Fixation duration across the first half of trials; Fixation duration across the last half of trials. We conducted identical analyses for fixation duration as were conducted above for first fixation**.** The ANOVAs revealed evidence for the null hypothesis, with active stimulation not modulating the difference in fixation duration at the no ball location in false-belief relative to true-belief trials, across all three time windows, $BF_{10} < 0.18$, $Fs < 0.61$, $ps > .549$. Similarly, all follow-up independent-groups t-tests, for all three time windows, demonstrated evidence for the null hypothesis, with no meaningful differences in fixation duration difference scores, across the stimulation conditions, $BF_{10} < 0.48$, $ts < 1.14$, $ps > .260$.

4. Discussion

We examined if right hemisphere TPJ, a key cortical node associated with both explicit and iToM processing (e.g., Naughtin, et al., 2017), is causally implicated in implicit false-belief processing. In order to do this, we combined tDCS, eye tracking and an established anticipatory looking false-belief task and, specifically, focussed on the extent to which individuals looked at a location that *did not* contain an object under conditions when an actor did not (true-belief) or did believe (false-belief) an object was present there. Importantly, given recent questions about the reliability of iToM measures (Kulke et al., 2018; Kulke et al., 2018), ours was a preregistered study that employed a relatively large sample and a conservative statistical approach. On the whole, anodal tDCS applied to TPJ disrupted implicit false belief processing relative to sham stimulation, however this was only observed when measuring first fixations and only manifest itself in the second half of the trials subjects completed.

Within the broader ToM literature, it has been suggested that previous anticipatory looking findings, taken as evidence for iToM, can rather be explained by other, low-level variables. For example, one such argument posits that attention is sensitive to stimulus features, such as colour and movement, within the test trials and these influence subjects' looking behaviour (Heyes, 2014). Here, to address this, we tightly controlled the visual elements in the true- and false-belief trials such that stimulus features, the trajectory of the ball and number of puppet movements were matched across trials (Grainger et al., 2018). Therefore, any variation in looking behaviour between true- and false-belief conditions as observed in the sham group were most likely due to the manipulation of the protagonist's belief state as opposed to extraneous stimulus variables. Similarly, gaze cues from the actor/protagonist were unlikely to lead to our pattern of results as the subjects never saw the actor's eyes which were occluded via the use of a visor. Moreover, the gaze cueing literature suggests that these effects only last for an extended period when using famous faces (Frischen and Tipper, 2006). Thus, collectively, the most parsimonious account is that our present results do reflect genuine, implicit social processing, rather than early processes that do not tap mentalising.

Furthermore, two components of our design ensured subjects' eye movements were reflective of implicit, rather than explicit, ToM processing. The attentionally demanding tone discrimination task overlapped temporally with the critical points in the movies where the protagonist's belief is formed. Thus, subjects were distracted from our belief state manipulation (Schneider et al., 2012). Secondly, the funnelled debriefing procedure identified (and was used to exclude) subjects who consciously processed of the protagonist's belief state.

Together, these elements make it highly likely our paradigm tapped implicit false belief processing. It should also be noted that previous work has shown that even when subjects are given instructions which are incongruent with tracking beliefs in this task (i.e., track the ball) they nevertheless display eye-movement behaviour that is consistent with implicit false-belief processing suggesting the behaviour is efficient and uncontrollable, key characteristics of implicit operations (Schneider et al., 2014a,b).

Along with the possible alternative explanations for subjects' eye-movement behaviour in the anticipatory looking paradigm, there are potential alternate explanations for our stimulation effects that must also be addressed. Typically, tDCS protocols employ relatively large target electrode and the reference electrode is placed over another brain region, thus, the technique has a reduced spatial resolution relative to other stimulation techniques such as transcranial magnetic stimulation (Nitsche et al., 2003). Consequently, could the present results not reflect TPJ stimulation but the influence of another area? Here, the reference electrode was positioned over the mastoid and only the target electrode was placed on the scalp. Therefore, the reference electrode would not have modulated neural excitability (Nitsche and Paulus, 2001). Additionally, as the reference electrode was contralateral relative to the target electrode, it is more likely the current reached the right TPJ rather than being shunted across the scalp (Bikson et al., 2010). In short, other areas may have been stimulated but, given the high inter-connected nature of the brain, it is highly likely that TPJ was influenced by our stimulation protocol.

It can also be questioned whether the use of sham stimulation as a control measure is adequate. Though sham tDCS stimulation is thought to effectively blind subjects (Gandiga et al., 2006), studies have also reported conflicting evidence (e.g., O'Connell et al., 2012). In our experiment subjects correctly guessed their stimulation condition above chance level. However, while ineffective blinding can result in expectancy effects (Horvath et al., 2014), these cannot explain the present findings. First, included subjects were unaware of the study's key aim until debriefing. Second, cathodal stimulation, had no influence on our measure of implicit false belief processing. Furthermore, in the last half of the task there was a meaningful difference between the anodal and cathodal stimulation conditions. Specifically, anodal stimulation disrupted implicit false belief processing relative to cathodal stimulation, while cathodal and sham stimulation effects did not differ. Thus, under the conditions we employed and, in particular, given the present results, the cathodal stimulation acts an active control condition in this context, further indicating the observed effect is a result of anodal stimulation and not simply general excitation following tDCS. One might also question, only anodal stimulation having an influence and that it was disruptive, but the directionality of the effects of tDCS are determined by a range of factors; including type of tasks employed (e.g., perceptual, motor or executive), brain region targeted, electrode montage, stimulation intensity and whether the protocol is online or offline. Indeed, there are many examples of both anodal and cathodal enhancing performance, inhibiting performance, and only one having an influence on a particular paradigm in the literature (Filmer et al., 2014). It should also be noted, that all elements of our design and the analysis approach were pre-registered. This protects against any posthoc bias occurring at the analysis, reporting, or interpretation stage.

However, despite finding evidence of iToM here and, indeed, disrupting this operation via stimulating TPJ, our data also speak to the somewhat fragile nature of iToM anticipatory looking findings. To wit, while the first fixation measure revealed anticipatory looking behaviour consistent with implicit false belief processing, this was not the case when the fixation duration measure was employed - a failure to replicate previous work (Schneider et al., 2012). This discrepancy in the anticipatory eye movement measures may reflect that first fixations are a more sensitive measure of iToM, relative to fixation duration. Indeed, this fits with the idea that iToM is a rapid and unconscious process (Apperly and Butterfill, 2009) that perhaps does not sustain within a trial. Nevertheless, future work needs to further determine the best conditions and measures under which to assess iToM.

Another notable aspect of the current findings was the absence of anticipatory looking behaviour, indicative of iToM processing, during the first half of the sham trials. Past studies have consistently demonstrated sustained iToM looking behaviour across trials (Schneider et al., 2012; Schneider et al., 2013). In contrast, here, such an effect only emerged in the last half of the task. This pattern of results might appear to provide evidence for a learning account (Perner and Ruffman, 2005; Ruffman, 2014). However, the offline tDCS aspect of this study provides an alternate hypothesis. Before completing the task, subjects sat still for the entire stimulation period. This may have induced boredom/fatigue impacting passive viewing of the true- and false-belief movies that only reduced after subjects engaged in the tone task, which have taken some time. As the stimulation period is the only methodological difference this and past studies (Schneider et al., 2012; Schneider et al., 2013), it appears to be a key factor.

Taken together, the present results provide the first evidence that right hemisphere TPJ is causally involved in implicit belief processing. This helps to characterise the extent to which iToM and eToM overlap, as TPJ has been shown to be a key cortical node involved in the explicit analysis of others' mental states. Thus, the "two" iToM streams may not be as distinct as has been previous argued (Naughtin et al., 2017). Also, as noted above, iToM appears to be particularly impaired in Autism Spectrum Disorder (Schneider et al., 2014a,b), and the present results hint at a node where interventions could be targeted. Here offline tDCS impaired iToM processing, however employing online approaches often leads to enhancement of training outcomes (e.g., Filmer et al., 2014; Filmer et al., 2017). Thus, social processing my also respond to pairing training and stimulation concurrently. The present work represents the first attempt to casually link brain regions and implicit belief processing, providing an important finding, but also illustrating challenges when assessing this key social process. Future research should extend on this work to characterise further what the brain extracts implicitly from the social environment, its neural basis and the best way to reliably measure this operation.

Author contributions

HLF, AF and PED designed the study. HLF and AF collected the data. HLF and AF analysed the data. HLF, AF and PED interpreted the data. PED and AF wrote the paper, with edits from HLF.

Acknowledgements

This research was supported by the Australian Research Council Discovery grants to PED (DP140100266; DP180101885). HLF was supported by a Developmental Fellowship from The University of Queensland.

References

Apperly, I.A., Butterfill, S.A., 2009. Do humans have two systems to track beliefs and belief-like states? Psychol. Rev. 116 (4), 953–970. https://doi.org/10.1037/a0016923.

Bargh, J.A., Chartrand, T.L., 2000. The mind in the middle: a practical guide to priming and automaticity research. In: Reis, H.T., Judd, C.M. (Eds.), Handbook of Research Methods in Social and Personality Psychology. Cambridge University Press, New York, NY, pp. 253–285.

Baron-Cohen, S., Leslie, A.M., Frith, U., 1985. Does the autistic child have a "theory of mind"? Cognition 21 (1), 37–46. https://doi.org/10.1016/0010-0277(85)90022-8.

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., Clubley, E., 2001. The autism-spectrum quotient (AQ): evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. J. Autism Dev. Disord. 31 (1), 5–17. https://doi.org/10.1023/A:1005653411471.

Bikson, M., Datta, A., Elwassif, M., 2009. Establishing safety limits for transcranial direct current stimulation. Clin. Neurophysiol. 120 (6), 1033–1034. https://doi.org/10.1016/j.clinph.2009.03.018.

Bikson, M., Datta, A., Rahman, A., Scaturro, J., 2010. Electrode montages for tDCS and weak transcranial electrical stimulation: role of "return" electrode's position and size.

Clin. Neurophysiol. 121 (12), 1976–1978. https://doi.org/10.1016/j.clinph.2010.05.020.
Brainard, D.H., 1997. The psychophysics toolbox. Spatial Vis. 10 (4), 433–436. https://doi.org/10.1163/156856897X00357.
Brüne, M., 2005. "Theory of Mind" in Schizophrenia: A Review of the Literature.
Butterfill, S.A., Apperly, I.A., 2013. How to construct a minimal theory of mind. Mind Lang. 28 (5), 606–637. https://doi.org/10.1111/mila.12036.
Clements, W.A., Perner, J., 1994. Implicit understanding of belief. Cogn. Dev. 9 (4), 377–395. https://doi.org/10.1016/0885-2014(94)90012-4.
Dodell-Feder, D., Koster-Hale, J., Bedny, M., Saxe, R., 2011. fMRI item analysis in a theory of mind task. Neuroimage 55 (2), 705–712. https://doi.org/10.1016/j.neuroimage.2010.12.040.
Donaldson, P.H., Kirkovski, M., Rinehart, N.J., Enticott, P.G., 2018. Autism-relevant traits interact with temporoparietal junction stimulation effects on social cognition: a high-definition transcranial direct current stimulation and electroencephalography study. Eur. J. Neurosci. 47 (6), 669–681. https://doi.org/10.1111/ejn.13675.
Filmer, H.L., Dux, P.E., Mattingley, J.B., 2014. Applications of transcranial direct current stimulation for understanding brain function. Trends Neurosci. 37 (12), 742–753. https://doi.org/10.1016/j.tins.2014.08.003.
Filmer, H.L., Fox, A., Dux, P.E., 2019. Modulating implicit theory of mind with transcranial direct current stimulation to the right temporal parietal junction. https://doi.org/10.14264/uql.2019.271.
Filmer, H.L., Varghese, E., Hawkins, G.E., Mattingley, J.B., Dux, P.E., 2017. Improvements in attention and decision-making following combined behavioural training and brain stimulation. Cerebr. Cortex 27, 3675–3682. https://doi.org/10.1093/cercor/bhw189.
Frischen, A., Tipper, S.P., 2006. Long-term gaze cueing effects: evidence for retrieval of prior states of attention from memory. Vis. Cognit. 14 (3), 351–364. https://doi.org/10.1080/13506280544000192.
Frischen, A., Bayliss, A.P., Tipper, S.P., 2007. Gaze cueing of attention: visual attention, social cognition, and individual differences. Psychol. Bull. 133 (4), 694–724. https://doi.org/10.1037/0033-2909.133.4.694.
Frith, C., Frith, U., 2005. Theory of mind. Curr. Biol. 15 (17), R644–R645. https://doi.org/10.1016/j.cub.2005.08.041.
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 (1), 11–21. https://doi.org/10.1016/S0028-3932(99)00053-6.
Gandiga, P.C., Hummel, F.C., Cohen, L.G., 2006. Transcranial DC stimulation (tDCS): a tool for double-blind sham-controlled clinical studies in brain stimulation. Clin. Neurophysiol. 117 (4), 845–850. https://doi.org/10.1016/j.clinph.2005.12.003.
Gelman, A., Hill, J., Yajima, M., 2012. Why we (usually) don't have to worry about multiple comparisons. Journal of Research on Educational Effectiveness 5 (2), 189–211. https://doi.org/10.1080/19345747.2011.618213.
Grainger, S.A., Henry, J.D., Naughtin, C.K., Comino, M.S., Dux, P.E., 2018. Implicit false belief tracking is preserved in late adulthood. Q. J. Exp. Psychol. 71 (9), 1980–1987. https://doi.org/10.1177/1747021817734690.
Heyes, C., 2014. False belief in infancy: a fresh look. Dev. Sci. 17 (5), 647–659. https://doi.org/10.1111/desc.12148.
Horvath, J.C., Carter, O., Forte, J.D., 2014. Transcranial direct current stimulation: five important issues we aren't discussing (but probably should be). Front. Syst. Neurosci. 8 (2), 1–8. https://doi.org/10.3389/fnsys.2014.00002.
Jasper, H., 1958. Report of the committee on methods of clinical examination in electroencephalography. Electroencephalogr. Clin. Neurophysiol. Suppl. 10 (2), 370–375. https://doi.org/10.1016/0013-4694(58)90053-1.
Jeffreys, H., 1961. Theory of Probability, third ed. Oxford University Press, New York.
Kulke, L., Reiß, M., Krist, H., Rakoczy, H., 2018a. How robust are anticipatory looking measures of Theory of Mind? Replication attempts across the life span. Cogn. Dev. 46, 97–111. https://doi.org/10.1016/j.cogdev.2017.09.001.
Kulke, L., von Duhn, B., Schneider, D., Rakoczy, H., 2018b. Is implicit theory of mind a real and robust phenomenon? Results from a systematic replication study. Psychol. Sci. 29 (6), 888–900. https://doi.org/10.1177/0956797617747090.
Moran, J.M., Young, L.L., Saxe, R., Lee, S.M., O'Young, D., Mavros, P.L., et al., 2011. Impaired theory of mind for moral judgment in high-functioning autism. Proc. Natl. Acad. Sci. U. S. A. 108 (7), 2688–2692. https://doi.org/10.1073/pnas.1011734108.
Naughtin, C.K., Horne, K., Schneider, D., Venini, D., York, A., Dux, P.E., 2017. Do implicit and explicit belief processing share neural substrates? Hum. Brain Mapp. 38 (9), 4760–4772. https://doi.org/10.1002/hbm.23700.
Nitsche, M.A., Paulus, W., 2001. Sustained excitability elevations induced by transcranial DC motor cortex stimulation in humans. Neurology 57 (10), 1899–1901. https://doi.org/10.1212/WNL.57.10.1899.
Nitsche, M.A., Fricke, K., Henschke, U., Schlitterlau, A., Liebetanz, D., Lang, N., et al., 2003. Pharmacological modulation of cortical excitability shifts induced by transcranial direct current stimulation in humans. J. Physiol. 553 (1), 293–301. https://doi.org/10.1113/jphysiol.2003.049916.
O'Connell, N.E., Cossar, J., Marston, L., Wand, B.M., Bunce, D., Moseley, G.L., De Souza, L.H., 2012. Rethinking clinical trials of transcranial direct current stimulation: participant and assessor blinding is inadequate at intensities of 2mA. PLoS One 7 (10), 1–5. https://doi.org/10.1371/journal.pone.0047514.
Pelli, D.G., 1997. The VideoToolbox software for visual psychophysics: transforming numbers into movies. Spatial Vis. 10 (4), 437–442. https://doi.org/10.1163/156856897x00366.
Perner, J., Ruffman, T., 2005. Infants' insight into the mind: how deep? Science 308 (5719), 214–216. https://doi.org/10.1126/science.1111656.
Premack, D., Woodruff, G., 1978. Does the chimpanzee have a theory of mind? Behav. Brain Sci. 1 (4), 515–526. https://doi.org/10.1017/S0140525X00076512.
Ruffman, T., 2014. To belief or not belief: children's theory of mind. Dev. Rev. 34 (3), 265–293. https://doi.org/10.1016/j.dr.2014.04.001.
Santiesteban, I., Banissy, M.J., Catmur, C., Bird, G., 2015. Functional lateralization of temporoparietal junction–imitation inhibition, visual perspective-taking and theory of mind. Eur. J. Neurosci. 42 (8), 2527–2533. https://doi.org/10.1111/ejn.13036.
Saxe, R., 2009. Theory of mind (neural basis). In: Banks, W.P. (Ed.), Encyclopedia of Consciousness, vol. 2. Academic Press, Oxford, UK, pp. 401–410.
Saxe, R., Kanwisher, N., 2003. People thinking about thinking people: the role of the temporo-parietal junction in "theory of mind". Neuroimage 19 (4), 1835–1842. https://doi.org/10.1016/S1053-8119(03)00230-1.
Schneider, D., Bayliss, A.P., Becker, S.I., Dux, P.E., 2012a. Eye movements reveal sustained implicit processing of others' mental states. J. Exp. Psychol. Gen. 141 (3), 433–438. https://doi.org/10.1037/a0025458.
Schneider, D., Lam, R., Bayliss, A.P., Dux, P.E., 2012b. Cognitive load disrupts implicit theory-of-mind processing. Psychol. Sci. 23 (8), 842–847. https://doi.org/10.1177/0956797612439070.
Schneider, D., Nott, Z.E., Dux, P.E., 2014a. Task instructions and implicit theory of mind. Cognition 133 (1), 43–47. https://doi.org/10.1016/j.cognition.2014.05.016.
Schneider, D., Slaughter, V.P., Dux, P.E., 2015. What do we know about implicit false-belief tracking? Psychonomic Bull. Rev. 22 (1), 1–12. https://doi.org/10.3758/s13423-014-0644-z.
Schneider, D., Slaughter, V.P., Bayliss, A.P., Dux, P.E., 2013. A temporally sustained implicit theory of mind deficit in autism spectrum disorders. Cognition 129 (2), 410–417. https://doi.org/10.1016/j.cognition.2013.08.004.
Schneider, D., Slaughter, V.P., Becker, S.I., Dux, P.E., 2014b. Implicit false-belief processing in the human brain. Neuroimage 101, 268–275. https://doi.org/10.1016/j.neuroimage.2014.07.014.
Schneider, D., Slaughter, V.P., Dux, P.E., 2017. Current evidence for automatic Theory of Mind processing in adults. Cognition 162, 27–31. https://doi.org/10.1016/j.cognition.2017.01.018.
Schurz, M., Radua, J., Aichhorn, M., Richlan, F., Perner, J., 2014. Fractionating theory of mind: a meta-analysis of functional brain imaging studies. Neurosci. Biobehav. Rev. 42, 9–34. https://doi.org/10.1016/j.neubiorev.2014.01.009.
Scott, R.M., Baillargeon, R., 2017. Early false-belief understanding. Trends Cognit. Sci. 21 (4), 237–249. https://doi.org/10.1016/j.tics.2017.01.012.
Senju, A., Southgate, V., Snape, C., Leonard, M., Csibra, G., 2011. Do 18-month-olds really attribute mental states to others? A critical test. Psychol. Sci. 22 (7), 878–880. https://doi.org/10.1177/0956797611411584.
Senju, A., Southgate, V., White, S., Frith, U., 2009. Mindblind eyes: an absence of spontaneous theory of mind in Asperger syndrome. Science 325 (5942), 883–885. https://doi.org/10.1126/science.1176170.
Southgate, V., Senju, A., Csibra, G., 2007. Action anticipation through attribution of false belief by 2-year-olds. Psychol. Sci. 18 (7), 587–592. https://doi.org/10.1111/j.1467-9280.2007.01944.x.
Van de Schoot, R., Kaplan, D., Denissen, J., Asendorpf, J.B., Neyer, F.J., van Aken, M.A., 2014. A gentle introduction to Bayesian analysis: applications to developmental research. Child Dev. 85 (3), 842–860. https://doi.org/10.1111/cdev.12169.
Wasserman, L., 2000. Bayesian model selection and model averaging. J. Math. Psychol. 44 (1), 92–107, 10.1006 jmps.1999.1278.
Wimmer, H., Perner, J., 1983. Beliefs about beliefs: representation and constraining function of wrong beliefs in young children's understanding of deception. Cognition 13 (1), 103–128. https://doi.org/10.1016/0010-0277(83)90004-5.
Woods, A.J., Antal, A., Bikson, M., Boggio, P.S., Brunoni, A.R., Celnik, P., et al., 2016. A technical guide to tDCS, and related non-invasive brain stimulation tools. Clin. Neurophysiol. 127 (2), 1031–1048. https://doi.org/10.1016/j.clinph.2015.11.012.
Zellner, A., Siow, A., 1980. Posterior odds ratios for selected regression hypotheses. In: Bernardo, J.M., DeGroot, M.H., Lindley, D.V., Smith, A.F.M. (Eds.), Bayesian Statistics: Proceedings of the First International Meeting. University of Valencia Press, Valencia, pp. 585–603.