

Functional role of the Theory of Mind Network in integrating mentalistic prior information with action kinematics during action observation

Azzurra Cristiano^{1,2}, Alessandra Finisguerra³, Cosimo Urgesi^{4,5}, Alessio Avenanti^{6,7*},
Emmanuele Tidoni^{8,9*}

1. Department of Psychology, Sapienza University of Rome and CLN²S@Sapienza, Italian Institute of Technology, Rome, IT.
2. IRCCS Santa Lucia Foundation, Rome, IT.
3. Scientific Institute, IRCCS E. Medea, Pesian di Prato, Udine, IT.
4. Scientific Institute, IRCCS E. Medea, Neuropsychiatry and Neurorehabilitation Unit, Bosisio Parini, 23842 Lecco, IT.
5. Laboratory of Cognitive Neuroscience, Department of Languages and Literatures, Communication, Education and Society, University of Udine, Udine, IT.
6. Department of Psychology, Centro studi e ricerche in Neuroscienze Cognitive, Alma Mater Studiorum - University of Bologna, Cesena, IT.
7. Centro de Investigación en Neuropsicología y Neurociencias Cognitivas, Universidad Católica del Maule, Talca, CL.
8. Human Technology Laboratory, School of Psychology and Social Work, University of Hull, Hull, UK.
9. School of Psychology, University of Leeds

* Both authors share co-senior authorship and correspondence. Alessio Avenanti: alessio.avenanti@unibo.it; Emmanuele Tidoni: e.tidoni@hull.ac.uk

Abstract

Inferring intentions from verbal and nonverbal human behaviour is critical for everyday social life. Here, we combined Transcranial Magnetic Stimulation (TMS) with a behavioural priming paradigm to test whether key nodes of the Theory of Mind network (ToMn) contribute to understanding others' intentions by integrating prior **knowledge** about an agent with the observed action kinematics. We used a modified version of the Faked-Action Discrimination Task (FAD), a forced-choice paradigm in which participants watch videos of actors lifting a cube and judge whether the actors are trying to deceive them concerning the weight of the cube. Videos could be preceded or not by verbal **description** (prior) about the agent's truthful or deceitful intent. We applied single pulse TMS over

three key nodes of the ToMn, namely dorsomedial prefrontal cortex (dmPFC), right posterior superior temporal sulcus (pSTS) and right temporo-parietal junction (rTPJ). Sham-TMS served as a control (baseline) condition. Following sham or rTPJ stimulation, we observed no consistent influence of priors on FAD performance. In contrast, following dmPFC stimulation, and to a lesser extent pSTS stimulation, truthful and deceitful actions were perceived as more deceptive only when the prior suggested a dishonest intention. These findings highlight a functional role of dmPFC and pSTS in coupling prior **knowledge** about deceptive intents with observed action kinematics in order to judge faked actions. Our study provides causal evidence that fronto-temporal nodes of the ToMn are functionally relevant to mental state inference during action observation.

Keywords

Action perception; Theory of Mind; Deception; Intention understanding; Transcranial Magnetic Stimulation

1. Introduction

Understanding others' actions is an essential ability for successful social interactions. This requires monitoring behaviours of other people and interpret intentions and motivations. Decades of research in cognitive neuroscience have highlighted two distinct brain systems that are engaged in these processes, namely the Action Observation Network (AON) and the Theory of Mind network (ToMn; Brass et al., 2007; De Lange et al., 2008; Van Overwalle, 2009; Van Overwalle & Baetens, 2009; Caspers et al., 2010; Wurm et al., 2011; Schurz et al., 2014; Catmur, 2015; Tidoni & Candidi, 2016; Paracampo et al., 2018; Thompson et al., 2019).

The AON is a widespread sensorimotor cortical network including frontal, parietal and temporal regions (Caspers et al., 2010; Hardwick et al., 2018) that are critical for action recognition (Avenanti et al., 2013a; Urgesi et al., 2014). The ToMn comprises a set of brain regions, including the ventral and dorsal sectors of the medial prefrontal cortex (mPFC) and the right temporoparietal junction (rTPJ) - active during mental state inferences (Frith & Frith, 2006; Van Overwalle & Baetens, 2009; Adolphs, 2009). Interestingly, these two networks partially overlap in the posterior Superior Temporal Sulcus (pSTS) - a key multisensory brain region sharing perceptual information about others' body movements and expressions with both the AON and the ToMn (Yang et al., 2015).

In a series of functional Magnetic Resonance Imaging (fMRI) studies, Spunt and colleagues (2010, 2011, 2012) highlighted that the core regions of AON are mainly recruited by identifying *how* an action is performed and *what* is being done (i.e., recognition of action kinematics and goals), whereas the ToMn is engaged when subjects deliberately reflect about the *why* of an observed action (i.e., the recognition of underlying intentions and motivations), with the dorsal mPFC (dmPFC) showing the most consistent involvement across studies (Spunt et al., 2010, 2011, 2012) and functional coupling with AON nodes (Spunt et al., 2012). These and other findings supported the hypothesis that the AON and the ToMn play distinct but complementary roles while inferring intentions during action observation (Brass et al., 2007; Keysers & Gazzola, 2007; De Lange et al., 2008; Ramsey and Hamilton, 2012). In this vein, the AON would be responsible for action recognition and provide the dmPFC and other ToMn nodes (e.g., pSTS; De Lange et al., 2008; Brosh et al., 2013) with information about action means and goals, supporting and constraining the inferential processes underlying intention understanding (Grèzes et al., 2004a, 2004b; Brass et al., 2007; De Lange et al., 2008; Thioux et al., 2008; Van Overwalle & Baetens, 2009; Catmur, 2015; Wurm & Schubotz 2018).

While the critical role of AON areas in action recognition has been widely demonstrated by means of non-invasive brain stimulation (Pobric & Hamilton, 2006; Tidoni et al., 2013; Avenanti et

al., 2013b, 2018; Jacquet & Avenanti, 2015; Makris & Urgesi, 2015; Valchev et al., 2017; Finisguerra et al., 2018), direct evidence supporting the functional relevance of ToMn regions during action observation is meagre (Schuwerk et al., 2014; Catmur, 2015; Bach & Schenke, 2017; Tamir & Thornton, 2018).

A substantial body of evidence shows that dmPFC is particularly active during social tasks requiring to integrate different sources of information (Amodio and Frith, 2006; Rilling et al., 2004; Volz et al., 2009), particularly, when intentional explicit judgements are required (Van Overwalle & Vandekerckhove, 2013). This research line highlights the role of dmPFC in providing flexible and context-independent metacognitive representations of the social world (Bzdok et al., 2013; Spunt & Adolphs, 2015). In line with this, a recent TMS investigation has shown that the dmPFC contributes to forming ‘long term’ impressions of another individual (e.g., how much he/she is trustworthy) by integrating visual information about her/his face and verbal descriptions of her/his social behaviour (e.g., ‘He offered to help a neighbour to fix a fence’ or ‘He criticized an old woman for being too slow’; Ferrari et al., 2016b). Authors showed that suppressing dmPFC resulted in more positive evaluations of other individuals when inconsistent information was provided. That is, when both positive and negative information about an agent was presented, suppression of dmPFC made participants more sensitive to positive rather than to negative information, thus biasing toward a positive evaluation.

The pSTS has been widely implicated in social perception (Pitcher and Ungerleider, 2021), particularly the perception of animacy and agency (Gao et al., 2012) and biological motion (Grossman and Blake, 2001) and it is thought to provide higher-order perceptual representations to other nodes of the AON and ToMn (Yang et al., 2015; Moessnang et al., 2017). In keeping, interference with the pSTS affects action recognition (Saygin 2007; Urgesi et al., 2014) and social judgments based on perceptual representations (Paracampo et al., 2018; Ferrari et al., 2018).

The rTPJ region contributes to mental states inference during judgments of agents’ behavior from textual scenarios (e.g., Young et al., 2010), whereas it is less clear whether temporo-parietal areas are relevant to inferring mental states from observed body movements. Previous studies on action observation found no influence of online repetitive TMS (rTMS) of rTPJ when participants were presented with smiling actors and were asked to discriminate between fake vs. true expressions of amusement (Paracampo et al., 2017); similarly, we found no influence of rTMS over the left TPJ when participants were presented with an agent lifting a box and had to discriminate between true vs. faked actions (i.e., actions in which the agent attempted to provide deceptive information concerning the weight of the lifted box; Tidoni et al., 2013). However, in a preprint study Thioux et al. (2018)

reported preliminary evidence that 1-Hz inhibitory rTMS over the rTPJ selectively impairs participants' performance in judging the level of hesitation of an actor executing a grasping action. This suggests that the rTPJ could contribute to deciphering mental states underlying hand actions.

Building on these works, we tested the contributions of the dmPFC, the right pSTS and the right TPJ in making sense of observed actions using TMS during a modified version of the Faked-Action Discrimination Task (FAD; Tidoni et al., 2013). This task requires participants to watch videos of an actor lifting a cube and judge whether the actor tried to deceive them about the cube's actual weight. In our new version of the FAD task, before each video, participants could read a short sentence describing the actor's intentional purpose (henceforth *prior*). Crucially, this text-based description could convey a positive, negative, or neutral impression of the actor's current intentional state (e.g., 'the actor is authentic' or 'the actor is fake'). This allowed us to test to what extent prior knowledge about an agent influenced the way her/his actions were perceived. We tested the influence of priors on FAD performance following sham TMS – serving as baseline and control condition – and active TMS of key fronto-temporal nodes of the ToMn. Specifically, we targeted the dmPFC, pSTS and rTPJ to test their functional relevance to coupling prior knowledge about truthful or deceptive intentions of an actor with the observed action kinematics during the FAD task.

While the different regions of the ToMn have been shown to work in strict connection (Van Overwalle, 2009; Schurz et al., 2014) and could thus contribute to accurately perform the task, based on previous studies we could attempt to differentiate the anticipated effects according to the stimulated areas. Overall, we expected that if an area is involved in integrating prior knowledge about the agent's intention with observed action kinematics, its stimulation would lead to better (or worse) performance levels depending on the type of prior (positive vs negative) and the type of intention inferred from the observed action (i.e., negative priors would facilitate identification of faked actions to a greater extent). Building on the evidence of integrative functions of dmPFC in social cognitive tasks reviewed above (e.g., Spunt et al., 2012), we expected that targeting this region would influence the coupling of prior knowledge about agent's intention with the observed action kinematics, thus affecting FAD performance depending on the type of prior and videos. Moreover, based on the valence-specific results of Ferrari et al., (2016b) we could expect that TMS over dmPFC would differentially affect the weight of negative vs. positive priors.

Because of mixed evidence in the literature, we had less straightforward hypotheses regarding the effect of pSTS and rTPJ stimulation. Based on the perceptual role of pSTS we anticipated that its stimulation could alter the ability to infer the actor's mental states independently of priors, thus affecting the overall accuracy of FAD performance. However, because STS may act as an interface

between AON and ToMn (Yang et al., 2015), its stimulation could affect integration of prior mentalistic knowledge with observed kinematics in a manner similar to dmPFC stimulation. Conversely, in keeping with the hypothesis that rTPJ may be involved in deciphering the actor's hesitation during action performance (Thioux et al., 2018), the effects of its stimulation could be specific for conditions of altered kinematics (i.e., when the actor is trying to deceive the observer), when the observer needed to attribute these alterations to an intention to deceive or to erroneous motor programming (see also Finisguerra et al., 2018).

2. Materials and Methods

We report how we determined our sample size, all data exclusions, all inclusion/exclusion criteria, all manipulations, and all measures in the study.

2.1 Participants

A total of 47 healthy volunteers from various Departments at the University of Bologna (e.g., Psychology, Engineering, Agriculture sciences) were recruited for two pilot experiments (28 subjects, 15 females and 13 males; age range 20-28; mean 23.8 ± 2.0 years old) and a TMS experiment (19 participants, 10 women and 9 men; age range 18-35; mean 23.11 ± 1.76 years old) through advertisement and word of mouth in the Cesena Campus community. Participants took part only in one of the experiments.

The design and sample size of the TMS experiment was similar to that of Ferrari et al. (2016a, 2016b). They collected between 12 (e.g., Ferrari et al., 2016a, experiment 1) and 20 (e.g., Ferrari et al., 2016b, experiment 1) participants, and observed medium-to-large effect size in their significant interactions. Based on our 4 x 3 x 2 experimental design (see Data Handling section), sample size for the main TMS experiment was determined using MorePower (version 6.0.4; Campbell & Thompson, 2012). The analysis indicated that 19 participants would be required given a power of 0.85, an alpha level of 0.05 and a large effect size (partial-eta set at 0.14; Cohen, 1992).

All participants provided written informed consent to participation and were naïve to the purpose of the experiment. None of the participants spontaneously reported having understood the goals of the study during the post-experimental debrief session. Participants had normal or corrected to normal vision and they were right-handed, as assessed by verbal reports of their manual preference. None of them had neurological, psychiatric, or other medical problems or demonstrated any contraindication to TMS (Rossi et al., 2009; 2021). The protocol was approved by the Bioethics committee of the University of Bologna and was carried out in accordance with the ethical standards

of the 2013 Declaration of Helsinki (World Medical Association, 2013). No discomfort or adverse effects during TMS were reported or noticed.

2.2 Visual Stimuli and Task

The employed video clips depicted from a lateral view the right hand of two non-professional actors (one man, one female) reaching, grasping, lifting, and placing a cube on a shelf (Finisguerra et al., 2018). In each clip, the actor lifted one of two cubes with identical visual appearance (6 x 6 x 6 cm) but different weights (100 g or 500 g) in two conditions. In the “truthful condition”, actors received truthful information about the cube’s weight and were instructed to provide truthful cues to the observers. In the “deceptive condition”, after receiving truthful information about the object weight, actors were asked to provide deceptive cues to the observers, pretending to lift the heavier cube (500 g) as if it weighted less (100 g) and vice-versa.

From an initial sample of 24 clips, we selected a subset of 12 action clips through two pilot experiments where participants completed the FAD Task. In Pilot 1, each of the 24 videos was repeated six times for a total of 144 trials. Based on participants’ performance, we selected the two most recognizable stimuli for each condition and actor for a total of 16 clips. In Pilot 2, each of these 16 clips were repeated 8 times for a total of 128 trials. Based on participants’ performance, we selected 7 clips for the truthful condition and 5 clips for the deceptive condition with an overall accuracy of 59%. Since for the TMS experiment we aimed at having 2 clips per condition for each actor, in the final set of stimuli we duplicated one truthful clip and three deceptive clips to reach a total of 16 clips (8 for the truthful condition, 8 for the deceptive condition).

In the TMS experiment participants completed a modified version of the FAD Task. Participants were informed that they had to observe an actor lifting a cube and that the actor, in half of the videos, was instructed to try and deceive the observer by lifting the cube as if it was heavier or lighter than it was. After the video, participants were instructed to indicate whether the observed actor tried or not to deceive them. Importantly, before the videos, participants were presented with a short sentence describing the actor’s intentional state (e.g., positive prior: ‘the actor is authentic’; negative prior: ‘the actor is fake’) or not (neutral prior: ‘the actor is #####’). Subjects were told they could use this information, together with the observed kinematics, to decide whether the actor was trying to deceive them (e.g., ‘natural or false?’).

Synonyms were used for composing the priors and the experimental question. Specifically, four Italian words have been chosen from the “Corpus and Frequency Lexicon of Written Italian” (COLFIS, see <https://www.istc.cnr.it/en/grouppage/colfis>) to guarantee the highest frequency of

occurrence and to maintain the same number of syllables. We selected the following words: natural (“naturale”; 4 syllables, 8 letters, absolute frequency 452) and authentic (“autentico”; 4 syllables, 9 letters, absolute frequency 197) as positive words, false (“falso”; 2 syllables, 5 letters, absolute frequency 199) and fake (“finto”; 2 syllables, 5 letters, absolute frequency 93) as negative. Then, we created two pairs of opposite words (authentic-false; natural-fake) and when a pair was administered as prior (e.g., ‘the actor is natural’, ‘the actor is fake’), the other was presented in the experimental question (e.g., ‘authentic or false?’) to reduce the risk participants based their answers exclusively on the **just read prior information**. The presentation of the couple of words was counterbalanced across participants. We also introduced catch trials where the prior contained an orthography error (e.g., ‘the atcor is authentic’, ‘the actor is fasle’). These trials were uncommon (8% of the total stimulus set) and were introduced to assess subjects’ attention to the prior during the whole experiment (see below). Of note, participants were instructed about the presence of these catch trials.

Each trial started with the appearance of a red fixation cross (duration of 750 ms; Fig. 1) presented simultaneously with a computer beep lasting 70 ms. Immediately after, the prior preceding the video clip was displayed for 1200 ms. **We chose this duration as the average speed reading for Italian language is about 188 ± 28 words per minute (about 3 words per seconds; Trauzettel-Klosinski et al., 2012) and our sentences were composed of four words (e.g., “the actor is false”, “the actor is sincere”)**. Finally, the clip appeared at the centre of the screen, for 1400 ms, on a uniform black background on a LED display (27 inch) on resolution of 1920x1080 pixels at 60 frames per second. At the end of each clip, participants answered as quickly and accurately as possible by pressing one of two keys (‘a’ or ‘d’) on a keyboard with their left ring and index finger respectively. The order of the button press was counterbalanced across participants. For catch trials, all subjects were instructed to signal them by pressing the ‘s’ key with their middle finger.

Presentation and randomization of the visual stimuli and response recording (Accuracy and Reaction Times) were controlled by custom scripts using Matlab (The MathWorks. Inc., Natick, MA, version R2015b) and Psychtoolbox (Kleiner et al., 2007).

2.3 TMS and Neuronavigation

The experiment comprised a preparatory procedure, three active TMS sessions and a sham TMS session. Stimulation was administered with a 70 mm figure-eight coil connected to a Magstim Rapid2 (The Magstim Company, Carmarthenshire, Wales, UK).

During the preparatory session, we selected the intensity of TMS to be used while participants completing the FAD task. We recorded Electromyographic (EMG) activity using a Biopac MP-35

system (Biopac System, Inc, CA, USA) by placing a pair of surface electrodes over the left first dorsal interosseous (FDI, active electrode) and the associated joint of the index finger (reference electrode) with the ground electrode located on the ventral surface of the left wrist. The TMS coil was placed tangentially to the motor cortex (M1) with the handle pointing backward and laterally at a 45° angle away from the midline (Rossini et al., 2015).

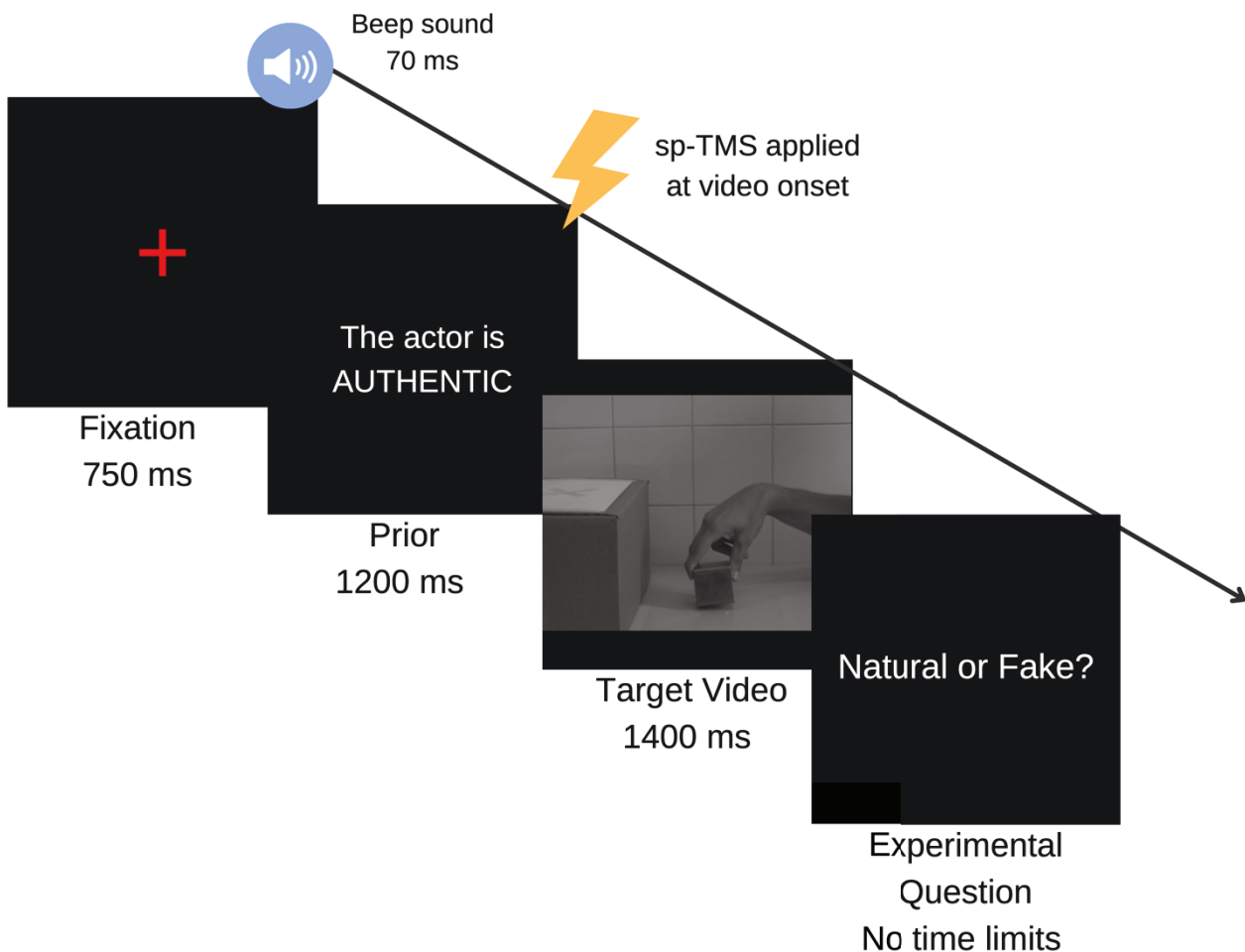


Figure 1. Example of an experimental trial. Each trial started with a computer beep lasting 70 ms presented simultaneously with the appearance of a red fixation cross (duration of 750ms). Then, a sentence (e.g., “the actor is AUTHENTIC”) appeared for 1200 ms providing participants prior knowledge about the actor’s intentional state. This was followed by the clip of an actor/actress lifting a cube (duration 1400 ms). At the end of each clip, the experimental question appeared at the centre of the screen (e.g., “Natural or Fake?”, note that when a word was administered in the prior – e.g., AUTHENTIC - the respective selected synonym was used in the experimental question – e.g., Natural).

The optimal scalp position (OSP) and the resting motor threshold (rMT) were identified by recording motor-evoked potentials (MEPs) induced by single-pulse TMS. The OSP was selected to produce maximum amplitude of MEPs from the FDI muscle using suprathreshold TMS pulses. Then, the rMT was identified as the minimum intensity to produce a peak-to-peak MEP of $\geq 50\text{-}\mu\text{V}$ amplitude in 5 out of 10 consecutive trials (Rossini et al., 2015).

The coil positions for dmPFC, right pSTS and right TPJ were identified on each participant's scalp through the SofTactic Navigator System (Electro Medical Systems, E.M.S. s.r.l., Bologna, Italy). Four skull landmarks (nasion,inion and 2 preauricular points) and ~ 100 scalp points, providing a uniform representation of the scalp, were digitized by means of a Polaris Vicra Optical Tracking System (Northern Digital Inc, Ontario, Canada). Coordinates in Talairach space (Talairach and Tournoux, 1988) were automatically estimated by the SoftTactic Navigator from the resulting MRI-constructed stereotaxic template. The scalp locations that best corresponded to the selected target sites were marked on a bathing cap with a pen.

We identified the stimulation targets (Figure 2) based on previous experimental studies, reviews and meta-analysis (see Table S1), and targeted at Talairach coordinates the dmPFC ($x = 0, y = 25, z = 42$), pSTS ($x = 50, y = -40, z = 7$), and TPJ ($x = 51, y = -52, z = 21$). Finally, we targeted the vertex during sham stimulation. The vertex was manually localized in each participant's brain at the crossing of the midline between the inion and the nasion and the midline between the left and the right preauricular points.

For active stimulation of the three target areas, the stimulating coil was applied tangentially to the scalp above the searched Talairach coordinates. For dmPFC stimulation, the handle of the coil pointed backward, parallel to the midsagittal line (e.g., Ferrari et al., 2016a; Gamond and Cattaneo, 2016; Gamond et al., 2017), whereas for pSTS and rTPJ stimulation it pointed backward and laterally at a 45° angle away from the midsagittal line (e.g., Candidi et al., 2011; Thioux et al., 2018).

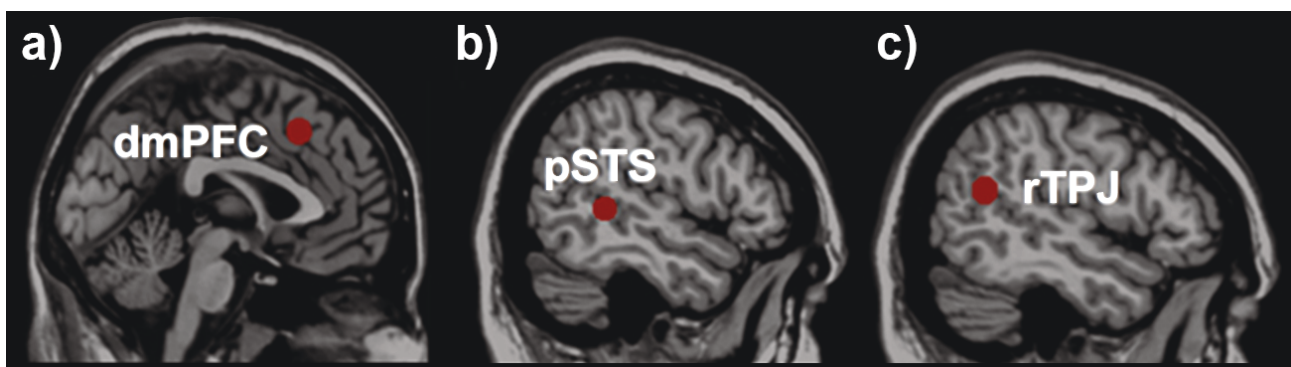


Figure 1. Targeted TMS sites. (a) dmPFC, (b) right pSTS, and (c) rTPJ.

During the active sessions, stimulation was block-wise administered over the target site at the onset of each test movie, in order to affect the potential influence that prior knowledge may have on the subsequent FAD videos judgements. The output strength of the TMS machine was set at 110% of the individual rMT. During the sham stimulation, we held the coil tangentially over the vertex, so that any cortical effect was unlikely to occur. This allowed to control nonspecific effects of TMS, such as those induced by the coil click or the coil pressure over the head.

2.4 Procedure

Before the TMS preparatory session, participants were trained to grasp and place on a shelf (located 30cm from a designed starting point) 2 cubes with similar size and weights to those depicted in the video clips (5x5x5cm, weighing 50gr and 500gr). Throughout the training, we asked participants to pay attention to their movements and the different weights to facilitate the use of their own motor experience of lifting the cubes when observing agents' movements in the FAD task (e.g., Wolpert et al., 2003; Tidoni et al., 2013; Finisguerra et al., 2018).

After the TMS preparatory session, participants were seated comfortably 60 cm away from the computer display, their right hand was placed on the computer keyboard, and the left hand rested on the table. They first familiarized themselves with the experimental task by performing a practice session of 56 trials. Then, they completed a total of four experimental sessions (i.e., one for each TMS condition) whose order was counterbalanced according to a Latin square. Each session comprised 96 trials and was divided into two blocks composed of 48 trials each. For each block, 16 trials consisted of presenting the actions with the neutral prior (half trials depicting the male actor, half depicting the female actress; with both actors performing half truthful and half deceptive actions, either with the light or the heavy cube) and 32 trials preceded by mentalistic priors. Specifically, 8 truthful and 8 deceptive clips were preceded by the positive prior, while the negative prior preceded the remaining 8 truthful and 8 deceptive actions. Finally, we included 4 catch trials for each block (see Visual Stimuli and Task section).

2.5 Data Handling

We collected task accuracy and reaction times (RTs, expressed in seconds) as performance measures. We excluded trials in which individual RTs (as recorded from the offset of the action-videos) were < 150 ms and >3000 ms (corresponding to 2.4% of the entire sample). Then, for each participant, we also removed trials in which individual RTs were higher than 3 standard deviations

(SDs) from the participant's median performance in each block (corresponding to 1.5% of the remaining sample).

First, the accuracy for the catch trials was analysed with the Friedman test, to check participants' attention to the prior words. Finally, both the accuracy and the RTs data of the remaining trials were separately entered into a 3-way repeated-measures ANOVA with Prior (neutral prior, negative prior, positive prior), Video (true, fake) and TMS session (Sham, dmPFC, pSTS, rTPJ) as within-subject factors.

The main statistics were run using Statistica 12 (StatSoft. Inc. Tulsa, USA). The significance level was set at $p \leq 0.05$ for all analyses and the Duncan correction was applied for post-hoc comparisons. To further facilitate the reader in assessing the strength of the evidence we also report the absolute value of the Cohen's d ($|d|$)¹, calculated by the software JASP version 0.14.1 (JASP Team, 2020).

No part of the study analyses was pre-registered prior to the research being conducted. The data that support the findings of this study are openly available in Open Science Framework (OSF) at https://osf.io/r8g2z/?view_only=aa4bdb1e588e42bd88a13df6bedeb30f

3. Results

3.1 Catch Trials Accuracy

Participants' accuracy scores on catch trials were analysed through the Friedman test. Results showed no significant effect of the stimulation sites ($\chi^2(3) = 3.19, p = 0.36$). Since the overall catch trials accuracy was consistently high (~94%), suggesting comparable levels of attention across the four TMS sessions. Catch trials were uncommon (about the 8% of the total stimulus set) and people can make sense of misspelled words (i.e., "the jumbled word effect"; Grainger and Whitney, 2004). Thus, the orthographic errors could only be detected by carefully reading the sentence. Our results thus suggest that participants paid attention to and read the priors in all TMS sessions as we observed no performance deterioration on these trials.

3.2 FAD Accuracy

The Prior*Video*TMS session ANOVA on accuracy scores revealed a significant main effect of Video ($F_{1,18} = 3.59, p = 0.003, \eta_p^2 = 0.39$), which was qualified by a Prior*Video interaction ($F_{2,36} = 6.01, p = 0.006, \eta_p^2 = 0.25$), and the three-way-interaction Prior*Video*TMS session ($F_{6,108} = 2.33,$

¹ Note that for main effect we computed the Cohen's d on mean-aggregated data.

$p = 0.04$, $\eta_p^2 = 0.11$). Hence, we performed separate ANOVAs with Prior and Video as within-subject factors, one for each TMS stimulation site.

In the sham session, we observed a main effect of Video ($F_{1,18} = 5.94$, $p = 0.03$, $\eta_p^2 = 0.25$), with higher accuracy for true actions (Mean % of accuracy \pm SD, [range across priors]; $64\% \pm 24$, [59-67%]) than for fake actions ($48\% \pm 19$, [47-49%], $|d| = 0.48$), but not main effect of Prior or Prior*Video interaction (all $F \leq 1.11$, $p \geq 0.34$, $\eta_p^2 \leq 0.06$).

Following stimulation of the dmPFC we observed no effect of Prior ($F_{2,36} = 0.25$, $p = 0.78$, $\eta_p^2 = 0.01$), a main effect of Video ($F_{1,18} = 7.45$, $p = 0.01$, $\eta_p^2 = 0.29$), with larger accuracy for true actions ($66\% \pm 22$, [55-72%]) than for fake actions ($50\% \pm 20$, [46-58%], $|d| = 0.47$), and a Prior*Video interaction ($F_{2,36} = 9.96$, $p < 0.001$, $\eta_p^2 = 0.36$). Truthful actions were recognized with lower accuracy when preceded by the negative (55% \pm 23) relative to the positive (69% \pm 20, $p = 0.01$, $|d| = 0.52$) and the neutral prior (72% \pm 20, $p = 0.003$, $|d| = 0.72$) which in turn did not differ from one another ($p = 0.61$). A different pattern was observed for fake actions. In this case, accuracy was higher when fake actions were preceded by the negative (58% \pm 21) relative to the positive (46% \pm 19, $p = 0.03$, $|d| = 0.57$) and the neutral prior (46% \pm 17, $p = 0.03$, $|d| = 0.74$), which in turn did not differ from one another ($p = 0.93$).

Stimulation over pSTS revealed no effect of Prior ($F_{2,36} = 2.32$, $p = 0.11$, $\eta_p^2 = 0.11$), a main effect of Video ($F_{1,18} = 23.46$, $p < 0.001$, $\eta_p^2 = 0.57$), with larger accuracy for true actions (71% \pm 18, [62-74%]) than for fake actions (49% \pm 16, [46-50%], $|d| = 0.88$), and, importantly, a Prior*Video interaction ($F_{2,36} = 4.22$, $p = 0.02$, $\eta_p^2 = 0.19$). Truthful actions were recognized with lower accuracy when preceded by the negative (63% \pm 20) relative to the positive (75% \pm 17, $p = 0.008$, $|d| = 0.57$) and the neutral prior (75% \pm 15, $p = 0.008$, $|d| = 0.61$), which in turn did not differ from one another ($p = 0.93$). Fake actions were comparable across prior conditions (all $p \geq 0.37$).

Following rTPJ stimulation, we found no main effect of Prior or Prior*Video interaction (all $F \leq 3.13$, $p \geq 0.06$, $\eta_p^2 \leq 0.15$), while we observed a main effect of Video ($F_{1,18} = 8.83$, $p = 0.008$, $\eta_p^2 = 0.33$), with higher accuracy for true (66% \pm 21, [60-71%]) than for fake actions (49% \pm 19, [46-52%], $|d| = 0.570$).

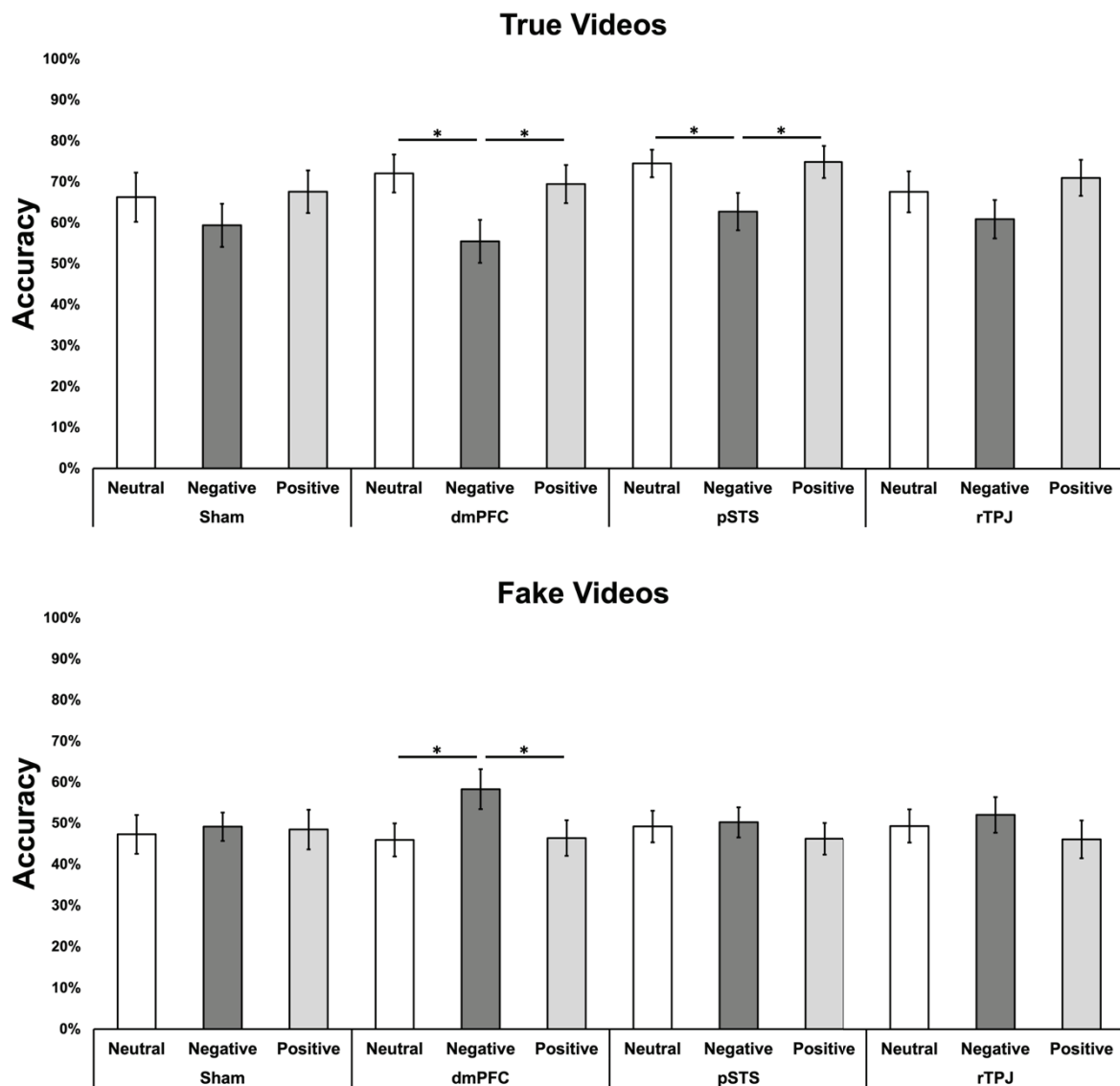


Figure 2. FAD Task Accuracy. The upper and lower panels display accuracy levels for true and fake videos respectively. No consistent modulation of task accuracy due to the priors (neutral, negative, positive) was observed in the sham condition or following stimulation of the rTPJ. Contrary, we observed a different pattern of results when TMS was delivered over the dmPFC and the pSTS (see main text for more details). Asterisks denote significant differences ($p < 0.05$).

Overall, we observed that TMS over the dmPFC and the pSTS increased participants' sensitivity to the negative prior: participants tended to commit more errors in judging truthful actions as fake actions in both TMS conditions; moreover, participants increased the ability to detect fake actions during dmPFC but not pSTS stimulation. In contrast, no consistent modulation of task accuracy due to the prior was observed in the sham control condition and following rTPJ stimulation.

We performed a further analysis to directly compare the effect of priors on FAD performance across the four TMS conditions. Since negative priors tended to induce opposite effects on the recognition of true and fake actions (Figure 3), we computed a global index of change in performance associated with negative priors. Specifically, for each TMS condition, we averaged two differences, the difference between the mean accuracy of positive and neutral conditions and the negative prior condition for true videos ($\text{avg}(\text{PositivePrior}, \text{NeutralPrior}) - \text{NegativePrior}$), and the specular difference for fake videos ($-1 * \text{avg}(\text{PositivePrior}, \text{NeutralPrior}) - \text{NegativePrior}$). Larger positive values indicate greater influence of negative priors on task accuracy (Figure 4a). A series of t-tests showed that priors affected recognition accuracy to a greater extent when TMS was applied over dmPFC (mean \pm s.e.m.: $13.73\% \pm 3.82\%$) relative to the Sham ($4.42\% \pm 3.81\%$; $p = 0.011$, $|d| = 0.65$), rTPJ ($6.37\% \pm 3.48\%$; $p = 0.001$, $|d| = 0.94$) or pSTS conditions ($7.22\% \pm 3.04\%$; $p = 0.062$, $|d| = 0.46$). No other differences were observed (all $p \geq 0.47$).

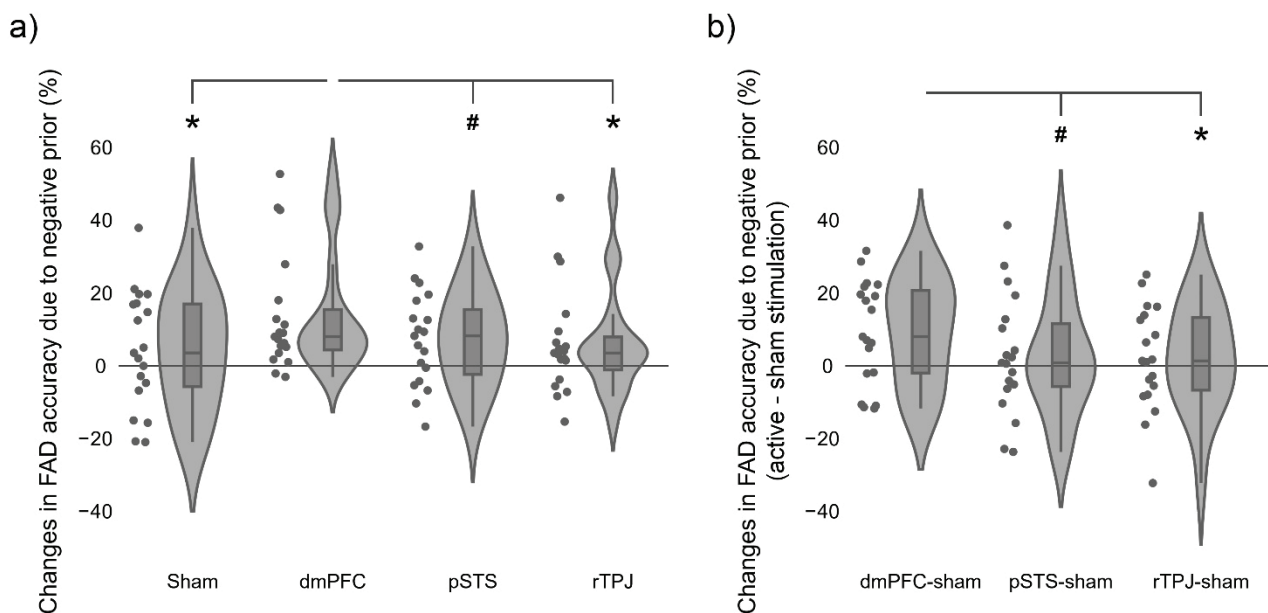


Figure 4. a) Influence of negative priors on FAD performance across TMS conditions. Violin plots display, for each stimulation conditions, the variation in the index of change in FAD performance associated with processing negative priors relative to processing neutral and positive priors. b) Individual variation in TMS-induced effects on changes in FAD performance due to negative priors. Violin display contrasts between active and sham TMS conditions in the index of change in performance due to negative priors. In both panels, asterisks (*) indicate significant t-tests ($p \leq 0.05$) whereas the symbol # denotes a marginally significant t-test ($p = 0.062$). We observed larger influences of negative priors following dmPFC stimulation relative to the other TMS conditions.

Inspection of Figure 4 shows that the effect of fake negative prior of FAD performance was variable across participants and stimulation conditions. Yet, following dmPFC stimulation, FAD performance was consistently influenced by the fake negative prior. Figure 4b shows individual variations in TMS effects computed as the difference between active and sham stimulation of the index of change in FAD performance due to negative priors. Relative to sham stimulation, the majority of participants (13 out of 19) showed enhanced influence of negative priors following dmPFC stimulation (mean: +17%; range: +5%, +32%), whereas the remaining 6 participants showed an opposite trend although smaller in size (mean: -8%; range: -2%, -12%).

3.3 FAD Reaction Times

The *Prior*Video*TMS session* ANOVA on median RTs (see Table 1) revealed a significant main effect of *Prior* ($F_{2,36} = 11.67, p < 0.001, \eta_p^2 = 0.39$). Post-hoc comparisons revealed that participants were slower when the neutral prior preceded the action clip (i.e., 742 ms \pm 315) compared to when the positive prior conditions (708 ms \pm 306, $p = 0.05, |d| = 0.32$), and faster the action clip was preceded by a negative prior (788 ms \pm 345, $p = 0.009, |d| = 0.29$). Accordingly, the negative and positive prior conditions differed from one another ($p < 0.0001, |d| = 0.54$). We also found a main effect of *Video* ($F_{1,18} = 4.56, p = 0.05, \eta_p^2 = 0.20$), with participants showing faster RTs in responding to true actions (771 ms \pm 348) than to fake actions (721 ms \pm 245, $|d| = 0.30$). No other main effects or interactions were found (all $F \leq 2.81$, all $p \geq 0.07$).

	True Video			Fake Video		
	Neutral Prior	Negative Prior	Positive Prior	Neutral Prior	Negative Prior	Positive Prior
dmPFC	757 \pm 78	832 \pm 95	751 \pm 82	760 \pm 88	745 \pm 71	685 \pm 67
pSTS	738 \pm 74	787 \pm 97	698 \pm 76	685 \pm 66	715 \pm 74	679 \pm 75
Sham	748 \pm 76	873 \pm 85	730 \pm 69	737 \pm 62	788 \pm 71	685 \pm 65
TPJ	779 \pm 88	822 \pm 94	736 \pm 81	734 \pm 70	745 \pm 67	697 \pm 70

Table 1. FAD Task RTs. Descriptive statistics (mean and s.e.m.) of the RTs for each experimental condition, reported in milliseconds.

4. Discussion

Social interaction depends upon the ability to infer beliefs and intentions in the minds of others. In the present study, we provide evidence that a core frontal region of the ToMn, namely the dmPFC (Frith & Frith, 2006; Van Overwalle & Baetens, 2009; Spunt & Adolphs, 2014; Catmur, 2015) and

an high-order multisensory area of convergence of both AON and ToMn, namely the pSTS (Yang et al., 2015) are called into play when inferring deceptive intents from body movements of other individuals and prior knowledge about their mental states.

Participants performed a modified version of the FAD task (Tidoni et al., 2013; Finisguerra et al., 2018). They were presented with short videos depicting an object-lifting action performed with truthful or deceptive intents and had to decide whether the actor was trying to deceive them about the objects' weight. The observed actions could be preceded by mentalistic priors referring to the observed agents (e.g., "the actor is authentic" or "the actor is false"). Participants received a single TMS pulse at the action onset over the dmPFC, pSTS, rTPJ, and a sham-TMS pulse. This way, we investigated how key nodes of the **ToMn contribute to participants' tendency to use prior knowledge about an actor intentional state when inferring their intentions from the observed action.**

In the sham session we did not observe any interaction between priors and type of videos, suggesting that FAD performance was not biased when brain stimulation was not active. In striking contrast, active TMS differently strengthened the integration of the prior with the observed action depending on the targeted site.

In line with the role of dmPFC in integrating social information from difference sources we found that dmPFC stimulation altered the way prior knowledge influenced FAD performance. This influence was prominent only when the prior conveyed information about deceptive intents (i.e., the agent is fake). In particular, we observed an increased sensitivity to the **negative** prior following dmPFC stimulation. **W**hen a prior suggesting deceptive intention was presented, participants were less accurate with True videos and more accurate with Fake videos, indicating that for both video types, they were more permeable to prior information. This indicates an increased bias in ascribing a deceptive intent to the actors. Hence dmPFC stimulation enhanced the influence of false (**negative**) prior on FAD judgements **as participants were more sensitive to the negative prior in both video types.** Our findings expand **previous** research that investigated the function of the dmPFC in social impression formation and updating (Mende-Siedlecki et al., 2013; Ferrari et al., 2016a, 2016b). Particularly relevant to our results is a study conducted by Ferrari et al. (2016a) where participants evaluated the trustworthiness of an observed human agent. In that study, participants simultaneously observed the face of the agent and read about his positive or negative social behaviours, then, they read a third behaviour that could be consistent or inconsistent with the first two. Ferrari et al. (2016b) applied an offline 1 Hz rTMS protocol for 15 minutes to suppress dmPFC activity before the beginning of the task. They found that dmPFC rTMS resulted in more positive evaluations of other individuals when inconsistent information was provided. That is, when both positive and negative

information about an agent was presented, suppression of dmPFC made participants more sensitive to positive rather than to negative information, thus biasing toward a positive evaluation.

In our study, single-pulse stimulation over dmPFC made participants more prone to ascribe a fake intent to others, creating a bias toward a “negative” evaluation of others’ intentionality. We note that in Ferrari et al. (2016a), participants judged stable personality attributes (i.e., trustworthiness), while in our study, participants judged the current mental state of the actor (i.e., motor intentions; Tamir and Thorton, 2018). Hence, these findings suggest that dmPFC may have a specific role in judging others’ ‘short-term’ mental states and stable ‘long-term’ personality dispositions by integrating their observable behaviour with negative or unexpected information. This would also be in line with previous studies showing greater mPFC response to the sight of deceptive (Grèzes et al., 2004b) and irrational actions (Brass et al., 2007; de Lange et al., 2008; Marsh et al., 2014) and support the view that the dmPFC is a key node of ToMn integrating mentalistic and action-related information to support social understanding (Zaki et al., 2009; Spunt & Liebermann, 2012).

An alternative account may suggest that the modulation observed following TMS over dmPFC may reflect an inability to monitor conflicting information (e.g., negative prior with a true action clip). It has been suggested that the posterior part of the medial frontal cortex (between 0 and 30mm on the y-coordinate) is recruited during reasoning regardless of mentalising content (Van Overwalle, 2011). Moreover, dmPFC is also active when people monitor their own cognitive processes (i.e., metacognition; Vaccaro & Fleming, 2018). However, were TMS over dmPFC affecting conflict monitoring we would have observed a selective influence for conflict information (e.g., when a fake video was preceded by a positive prior), whereas we observed reduced accuracy when negative priors preceded true videos but an increase in accuracy when negative priors preceded fake videos, indicating that dmPFC stimulation did not act on conflicting information, but rather facilitated the processing of the negative prior. Moreover, were TMS over dmPFC affecting general reasoning, then we should have expected reduced accuracy across the three priors. Therefore, a reduced ability of participants in reasoning and monitoring their own cognitive processes is unlikely to fully explain the current results.

Recent studies have shown that dmPFC is involved in constructing the mental model of a person’s mental state (Jiang et al., 2022) and in self-related belief updating (Kuzmanovic et al., 2018). In our task, participants were instructed to discriminate action where an actor may try (or not) to deceive them. It is then possible to assume that a negative prior was relevant for the participant. Being informed that a person may not be true to us is an important information to reduce the risk of being fooled (i.e., a positive self-related outcome). Hence, TMS over dmPFC may have biased participants

in interpreting the observed action as deceptive to reduce the risk of being fooled. This may also explain, why we did not find any effect with the positive prior as participants have the tendency to see other acting honestly (Levine et al., 1999; Vrij, 2000). This interpretation may also reconcile different views on the role of the dmPFC in conflict monitoring and mentalising. In other words, it might be possible that participants monitored the actor's intentions by updating their initial belief when a negative prior made them question about their truthfulness. Such explanation of our results may imply that the dmPFC is not only related to general reasoning (Van Overwalle, 2011) but also in constructing the mental model of a person's mental state (Jiang et al., 2022) and updating it based on new self-relevant information (Kuzmanovic et al., 2018).

Further, our hypothesis that stimulating the pSTS could influence the integration of prior mentalistic knowledge with observed kinematics has been in part corroborated by the results. We detected an enhanced influence of priors following right pSTS stimulation, with participants being less accurate on truthful action videos whenever they were told the actor had a deceptive intent. Vander Wyk et al. (2009; 2012) reported that the right pSTS is sensitive to the congruency between an observed action and the underlying predicted intention. In their fMRI studies, pSTS showed a great response when participants viewed an actress reaching an object after displaying an emotional expression incongruent with the subsequent reaching action (e.g., the actress expressed positive emotion towards one object and then grasped another one, or the actress expressed a negative emotion towards an object and then grasped that object). Other works demonstrated increased STS activation when participants viewed actors attempting to deceive them about the weight of a box (Grèzes et al., 2004b) or if the actors had a false belief about the weight of the box (Grèzes et al., 2004a). In both cases, since humans have a bias to judge other people's behaviour as truthful (Levine et al., 1999; Vrij, 2000) and to attribute true beliefs to others (Bartsch and Wellman, 1995) – a trend that we consistently observed with our videos in the pilot experiments and the sham condition of the main study – the researchers interpreted the increase of pSTS activity as resulting from the mismatch between the predicted (truthful) intention and the observed (e.g., deceptive) action.

However, in the present study, TMS over pSTS was not effective in strengthening the integration between the positive prior and the faked actions (i.e., we did not observe a drop in accuracy when detecting deceitful intents after suggesting the actor intention was truthful). We speculate that this result could be due to ambiguity of action kinematics in deceptive videos, as such these videos were not sufficiently informative about the actor's intentions. In fact, our participants recognized the fake actions at chance level even during the sham stimulation (i.e., the accuracy range was around 47-49% across all the three priors in the sham condition). Thus, the fake action may have not been

recognizable as such, and it remains unclear whether the positive prior information was perceived as congruent or incongruent. In other words, since the analyses of the fake action kinematics did not provide reliable cues to form an opinion about the actor's intent, the **prior knowledge** could not properly modulate participants' judgments (i.e., by further decreasing the accuracy).

Contrary, truthful actions kinematics provided enough information to recognise the observed behaviour as genuine (i.e., performance was above chance level in the sham session), and the negative prior may have been recognised as incongruent. For these reasons, the TMS pulse over pSTS was effective in reducing participants' ability to categorise a truthful action preceded by a negative prior.

So, while TMS over dmPFC may have increased the weight of the negative prior to impression formation irrespective of the observed movement, TMS over pSTS may have selectively biased the processing of truthful action preceded by an incongruent prior. Moreover, since we did not observe a decreased accuracy between the neutral and congruent prior when observing truthful actions, we can exclude that TMS over pSTS generally reduced the ability process biological motion. Thus, the present findings suggest that pSTS does more than simply identify biological motion and may be involved in representing the others' actions embedded in a social and mentalistic context—in the present case, a context established by (an incongruent) prior about others' intentionality.

Finally, regarding the role of rTPJ during the perception of altered kinematics, we found no evidence of prior effects on participants' accuracy following rTPJ stimulation. This result is in line with Tidoni et al. (2013) who showed no change in the FAD task performance while stimulating the left TPJ, suggesting that TPJ may not be critical for inferential and explicit mentalizing processes underlying the recognitions of others' (deceptive) intentions from action observation. Current evidence suggests that TPJ supports a basic mechanism of shifting or reorienting of attention, underlying spatial attention as well as social cognition (Corbetta et al., 2008; Mitchell, 2008, 2009; Van Overwalle, 2009; Hogeveen et al., 2015; Santiesteban et al., 2017). In social cognition tasks, for example, TPJ activity would be important to switch between potentially opposing representations—e.g., one's own and another person's mental state, perspective, or belief (Costa et al., 2008; Giardina et al., 2011; Santiesteban et al., 2012, 2017; Silani et al., 2013; Sowden et al., 2015; Krall et al., 2016). Building on **previous** works, one could expect that in the FAD task, TPJ stimulation would affect the shift between the kinematics representation of the observed action and the **knowledge of the actor's** intention, influencing the weight of **the** priors on FAD judgements. However, in the sham condition of the present study, we observed very little or no influence of priors on FAD task performance. This may suggest that priors were perceived as distinctive (participants paid attention to them as revealed by catch trials) but less important events with participants relying more on the ongoing action

kinematics cues rather than mentalistic information. Thus, the lack of consistent modulation on FAD task performance following rTPJ stimulation may be due to a reduced competition between priors and kinematics information.

In a recent study, Wurm & Schubotz (2018) showed increased TPJ bilateral activation during goal recognition from action observation. In particular, the authors showed that TPJ responded to the recognition of goal-directed actions in a naturalistic but not in a pixelized scene where detailed cues of the acting person were missing but the agent's goal was still recognisable. Based on this, Wurm & Schubotz (2018) proposed that TPJ is involved in ToM-related processes such as the detection of other agents and reasoning about their underlying mental states. Nonetheless, our findings showed no modulation following TPJ stimulation. Therefore, while we cannot rule out that TPJ is necessary to detect an intentional agent during action observation (Wurm & Schubotz, 2018), our results suggest that TPJ may not be crucial at integrating agent-related mentalistic and action kinematics information to discriminate different types of intentions. Future studies will need to address the relevance of TPJ when the competition between mentalistic and kinematics information is more distinctive.

Our results confirm previous findings and further extend the general contribution of TPJ in mentalizing tasks during action observation (Tidoni et al., 2013; Wurm & Schubotz, 2018). Moreover, the evidence that the stimulation over the rTPJ did not affect participants' performance while coupling mentalistic information with action kinematics further corroborates the results from the dmPFC and pSTS stimulation. That is, since TMS over TPJ had no effect, the results obtained by applying TMS over the dmPFC and pSTS may be a specific functional effect of TMS rather than a general change in participants' attentional focus.

4.1 Limits of the current Study

While current findings were specific to some of the target brain areas regions (i.e., anatomical specificity) when a positive or negative prior knowledge preceded the action video (i.e., no differences across priors were observed when TMS was delivered over rTPJ or sham), our study did not include a control task with similar task difficulty (Hartwigsen & Silvanto, 2022; Bergmann & Hartwigsen, 2020). For example, adding a task where the participant has to infer the real weight of the box may help disambiguate whether dmPFC and pSTS may play separate roles with dmPFC responsible for integrating prior knowledge with the observed kinematics and pSTS playing also a role in inferring the physical properties of the lifted box.

5. Conclusion

We provide evidence for the role of the dmPFC and the pSTS for accurate mental state inferences during action observation. Although more experimental evidence is required to complement our findings, using deceptive videos adequately informative about the actor's intentions, or employing – for example – advanced neurostimulation technique, such as the transcranial Ultrasound Stimulation (tFUS, Darmani et al., 2022) with its unprecedented spatial selectivity, or innovative approaches such as the High Definition - Transcranial Electrical Stimulation (HD-tES; Fusco et al., 2022), the current study suggests a differential contribution of human frontal and temporal brain areas during action observation. While dmPFC may have a specific function in biasing judgments about others' mental states towards negative or unexpected contents, pSTS seems to be sensitive to the congruency between the observed actions and their underlying intentions, having a role in representing not only biological motions, but how other persons' movements are coherently related to their intention.

Overall, this study reveals the relevance of both multisensory brain regions involved in processing the observed behaviour and the ToMn in forming and updating others' impression based on both visuo-motor and mentalistic information.

Competing interest

Authors have nothing to declare.

Acknowledgment

This work was financially supported by research grants from the Bial Foundation [347/18], the Ministero dell'Istruzione, dell'Università e della Ricerca [2017N7WCLP], and Fondazione del Monte di Bologna e Ravenna, Italy [1402bis/2021] awarded to AA.

Author's contributions

Study concept: AA; Study design: AA, CU, ET; Stimulus material: AF, CU; Fund acquisition: AA, CU; Data acquisition: AC, ET; Data analysis: AC, AA, ET; Supervision: AA, CU, ET; Manuscript drafting: AC, AA, ET; Manuscript revision: AC, AF, CU, AA, ET

REFERENCES

- Adolphs, R. (2009). The Social Brain: Neural Basis of Social Knowledge. *Annual Review of Psychology*, 60(1), 693–716. <https://doi.org/10.1146/annurev.psych.60.110707.163514>
- Amodio, D. M., & Frith, C. D. (2006). Meeting of minds: The medial frontal cortex and social cognition. *Nature Reviews Neuroscience*, 7(4), 268–277. <https://doi.org/10.1038/nrn1884>
- Avenanti, A., Annella, L., Candidi, M., Urgesi, C., & Aglioti, S. M. (2013). Compensatory Plasticity in the Action Observation Network: Virtual Lesions of STS Enhance Anticipatory Simulation of Seen Actions. *Cerebral Cortex*, 23(3), 570–580. <https://doi.org/10.1093/cercor/bhs040>
- Avenanti, A., Candidi, M., & Urgesi, C. (2013). Vicarious motor activation during action perception: Beyond correlational evidence. *Frontiers in Human Neuroscience*, 7. <https://doi.org/10.3389/fnhum.2013.00185>
- Avenanti, A., Paracampo, R., Annella, L., Tidoni, E., & Aglioti, S. M. (2018). Boosting and Decreasing Action Prediction Abilities Through Excitatory and Inhibitory tDCS of Inferior Frontal Cortex. *Cerebral Cortex*, 28(4), 1282–1296. <https://doi.org/10.1093/cercor/bhx041>
- Bach, P., & Schenke, K. C. (2017). Predictive social perception: Towards a unifying framework from action observation to person knowledge. *Social and Personality Psychology Compass*, 11(7), e12312. <https://doi.org/10.1111/spc3.12312>
- Bartsch, K., & Wellman, H. M. (1995). *Children Talk About the Mind*. Oxford University Press.
- Bergmann, T. O., & Hartwigsen, G. (2021). Inferring Causality from Noninvasive Brain Stimulation in Cognitive Neuroscience. *Journal of Cognitive Neuroscience*, 33(2), 195–225. https://doi.org/10.1162/jocn_a_01591
- Brass, M., Schmitt, R. M., Spengler, S., & Gergely, G. (2007). Investigating Action Understanding: Inferential Processes versus Action Simulation. *Current Biology*, 17(24), 2117–2121. <https://doi.org/10.1016/j.cub.2007.11.057>
- Brosch, T., Schiller, D., Mojdehbakhsh, R., Uleman, J. S., & Phelps, E. A. (2013). Neural mechanisms underlying the integration of situational information into attribution outcomes. *Social Cognitive and Affective Neuroscience*, 8(6), 640–646. <https://doi.org/10.1093/scan/nst019>
- Bzdok, D., Langner, R., Schilbach, L., Engemann, D. A., Laird, A. R., Fox, P. T., & Eickhoff, S. B. (2013). Segregation of the human medial prefrontal cortex in social cognition. *Frontiers in Human Neuroscience*, 7. <https://doi.org/10.3389/fnhum.2013.00232>
- Campbell, J. I. D., & Thompson, V. A. (2012). MorePower 6.0 for ANOVA with relational confidence intervals and Bayesian analysis. *Behavior Research Methods*, 44(4), 1255–1265. <https://doi.org/10.3758/s13428-012-0186-0>

- Candidi, M., Stienen, B. M. C., Aglioti, S. M., & de Gelder, B. (2011). Event-Related Repetitive Transcranial Magnetic Stimulation of Posterior Superior Temporal Sulcus Improves the Detection of Threatening Postural Changes in Human Bodies. *Journal of Neuroscience*, *31*(48), 17547–17554. <https://doi.org/10.1523/JNEUROSCI.0697-11.2011>
- Caspers, S., Zilles, K., Laird, A. R., & Eickhoff, S. B. (2010). ALE meta-analysis of action observation and imitation in the human brain. *NeuroImage*, *50*(3), 1148–1167. <https://doi.org/10.1016/j.neuroimage.2009.12.112>
- Catmur, C. (2015). Understanding intentions from actions: Direct perception, inference, and the roles of mirror and mentalizing systems. *Consciousness and Cognition*, *36*, 426–433. <https://doi.org/10.1016/j.concog.2015.03.012>
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155–159. <https://doi.org/10.1037/0033-2909.112.1.155>
- Corbetta, M., Patel, G., & Shulman, G. L. (2008). The Reorienting System of the Human Brain: From Environment to Theory of Mind. *Neuron*, *58*(3), 306–324. <https://doi.org/10.1016/j.neuron.2008.04.017>
- 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), 71–74. <https://doi.org/10.1155/2008/694632>
- Darmani, G., Bergmann, T. O., Butts Pauly, K., Caskey, C. F., de Lecea, L., Fomenko, A., Fouragnan, E., Legon, W., Murphy, K. R., Nandi, T., Phipps, M. A., Pinton, G., Ramezanpour, H., Sallet, J., Yaakub, S. N., Yoo, S. S., & Chen, R. (2022). Non-invasive transcranial ultrasound stimulation for neuromodulation. *Clinical Neurophysiology*, *135*, 51–73. <https://doi.org/10.1016/j.clinph.2021.12.010>
- de Lange, F. P., Spronk, M., Willems, R. M., Toni, I., & Bekkering, H. (2008). Complementary Systems for Understanding Action Intentions. *Current Biology*, *18*(6), 454–457. <https://doi.org/10.1016/j.cub.2008.02.057>
- Ferrari, C., Lega, C., Vernice, M., Tamietto, M., Mende-Siedlecki, P., Vecchi, T., Todorov, A., & Cattaneo, Z. (2016). The Dorsomedial Prefrontal Cortex Plays a Causal Role in Integrating Social Impressions from Faces and Verbal Descriptions. *Cerebral Cortex*, *26*(1), 156–165. <https://doi.org/10.1093/cercor/bhu186>
- Ferrari, C., Schiavi, S., & Cattaneo, Z. (2018). TMS over the superior temporal sulcus affects expressivity evaluation of portraits. *Cognitive, Affective, & Behavioral Neuroscience*, *18*(6), 1188–1197. <https://doi.org/10.3758/s13415-018-0630-4>

- Ferrari, C., Vecchi, T., Todorov, A., & Cattaneo, Z. (2016). Interfering with activity in the dorsomedial prefrontal cortex via TMS affects social impressions updating. *Cognitive, Affective, & Behavioral Neuroscience*, *16*(4), 626–634. <https://doi.org/10.3758/s13415-016-0419-2>
- Finisguerra, A., Amoroso, L., Makris, S., & Urgesi, C. (2018). Dissociated Representations of Deceptive Intentions and Kinematic Adaptations in the Observer's Motor System. *Cerebral Cortex*, *28*(1), 33–47. <https://doi.org/10.1093/cercor/bhw346>
- Frith, C. D., & Frith, U. (2006). The Neural Basis of Mentalizing. *Neuron*, *50*(4), 531–534. <https://doi.org/10.1016/j.neuron.2006.05.001>
- Fusco, G., Cristiano, A., Perazzini, A., & Aglioti, S. M. (2022). Neuromodulating the performance monitoring network during conflict and error processing in healthy populations: Insights from transcranial electric stimulation studies. *Frontiers in Integrative Neuroscience*, *16*, 953928. <https://doi.org/10.3389/fnint.2022.953928>
- Gamond, L., & Cattaneo, Z. (2016). The dorsomedial prefrontal cortex plays a causal role in mediating in-group advantage in emotion recognition: A TMS study. *Neuropsychologia*, *93*, 312–317. <https://doi.org/10.1016/j.neuropsychologia.2016.11.011>
- Gamond, L., Ferrari, C., La Rocca, S., & Cattaneo, Z. (2017). Dorsomedial prefrontal cortex and cerebellar contribution to in-group attitudes: A transcranial magnetic stimulation study. *European Journal of Neuroscience*, *45*(7), 932–939. <https://doi.org/10.1111/ejn.13529>
- Gao, T., Scholl, B. J., & McCarthy, G. (2012). Dissociating the Detection of Intentionality from Animacy in the Right Posterior Superior Temporal Sulcus. *The Journal of Neuroscience*, *32*(41), 14276–14280. <https://doi.org/10.1523/JNEUROSCI.0562-12.2012>
- Giardina, A., Caltagirone, C., & Oliveri, M. (2011). Temporo-parietal junction is involved in attribution of hostile intentionality in social interactions: An rTMS study. *Neuroscience Letters*, *495*(2), 150–154. <https://doi.org/10.1016/j.neulet.2011.03.059>
- Grainger, J., & Whitney, C. (2004). Does the huamn mnid raed wrods as a wlohe? *Trends in Cognitive Sciences*, *8*(2), 58–59. <https://doi.org/10.1016/j.tics.2003.11.006>
- Grezes, J. (2004). Brain Mechanisms for Inferring Deceit in the Actions of Others. *Journal of Neuroscience*, *24*(24), 5500–5505. <https://doi.org/10.1523/JNEUROSCI.0219-04.2004>
- Grèzes, J., Frith, C. D., & Passingham, R. E. (2004). Inferring false beliefs from the actions of oneself and others: An fMRI study. *NeuroImage*, *21*(2), 744–750. [https://doi.org/10.1016/S1053-8119\(03\)00665-7](https://doi.org/10.1016/S1053-8119(03)00665-7)

- Grosbras, M.-H., Beaton, S., & Eickhoff, S. B. (2012). Brain regions involved in human movement perception: A quantitative voxel-based meta-analysis. *Human Brain Mapping, 33*(2), 431–454. <https://doi.org/10.1002/hbm.21222>
- Grossman, E. D., & Blake, R. (2001). Brain activity evoked by inverted and imagined biological motion. *Vision Research, 41*(10–11), 1475–1482. [https://doi.org/10.1016/S0042-6989\(00\)00317-5](https://doi.org/10.1016/S0042-6989(00)00317-5)
- Hardwick, R. M., Caspers, S., Eickhoff, S. B., & Swinnen, S. P. (2018). Neural correlates of action: Comparing meta-analyses of imagery, observation, and execution. *Neuroscience & Biobehavioral Reviews, 94*, 31–44. <https://doi.org/10.1016/j.neubiorev.2018.08.003>
- Hogeveen, J., Obhi, S. S., Banissy, M. J., Santiesteban, I., Press, C., Catmur, C., & Bird, G. (2015). Task-dependent and distinct roles of the temporoparietal junction and inferior frontal cortex in the control of imitation. *Social Cognitive and Affective Neuroscience, 10*(7), 1003–1009. <https://doi.org/10.1093/scan/nsu148>
- Jacquet, P. O., & Avenanti, A. (2015). Perturbing the Action Observation Network During Perception and Categorization of Actions' Goals and Grips: State-Dependency and Virtual Lesion TMS Effects. *Cerebral Cortex, 25*(3), 598–608. <https://doi.org/10.1093/cercor/bht242>
- Jiang, S., Wang, S., & Wan, X. (2022). Metacognition and mentalizing are associated with distinct neural representations of decision uncertainty. *PLOS Biology, 20*(5), e3001301. <https://doi.org/10.1371/journal.pbio.3001301>
- Keysers, C., & Gazzola, V. (2007). Integrating simulation and theory of mind: From self to social cognition. *Trends in Cognitive Sciences, 11*(5), 194–196. <https://doi.org/10.1016/j.tics.2007.02.002>
- Kleiner, M., Brainard, D., & Pelli, D. (2007). What's new in Psychtoolbox-3?
- Koster-Hale, J., & Saxe, R. (2013). Theory of Mind: A Neural Prediction Problem. *Neuron, 79*(5), 836–848. <https://doi.org/10.1016/j.neuron.2013.08.020>
- 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: RTPJ-TMS in Attention and Social Interaction. *Human Brain Mapping, 37*(2), 796–807. <https://doi.org/10.1002/hbm.23068>
- Kuzmanovic, B., Bente, G., von Cramon, D. Y., Schilbach, L., Tittgemeyer, M., & Vogeley, K. (2012). Imaging first impressions: Distinct neural processing of verbal and nonverbal social information. *NeuroImage, 60*(1), 179–188. <https://doi.org/10.1016/j.neuroimage.2011.12.046>
- Levine, T. R., Park, H. S., & McCornack, S. A. (1999). Accuracy in detecting truths and lies: Documenting the “veracity effect”. *Communication Monographs, 66*(2), 125–144. <https://doi.org/10.1080/03637759909376468>

- Makris, S., & Urgesi, C. (2015). Neural underpinnings of superior action prediction abilities in soccer players. *Social Cognitive and Affective Neuroscience*, 10(3), 342–351. <https://doi.org/10.1093/scan/nsu052>
- Marsh, L. E., Mullett, T. L., Ropar, D., & Hamilton, A. F. de C. (2014). Responses to irrational actions in action observation and mentalising networks of the human brain. *NeuroImage*, 103, 81–90. <https://doi.org/10.1016/j.neuroimage.2014.09.020>
- Mende-Siedlecki, P., Cai, Y., & Todorov, A. (2013). The neural dynamics of updating person impressions. *Social Cognitive and Affective Neuroscience*, 8(6), 623–631. <https://doi.org/10.1093/scan/nss040>
- Mitchell, J. P. (2008). Activity in Right Temporo-Parietal Junction is Not Selective for Theory-of-Mind. *Cerebral Cortex*, 18(2), 262–271. <https://doi.org/10.1093/cercor/bhm051>
- Mitchell, J. P. (2009). Inferences about mental states. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1521), 1309–1316. <https://doi.org/10.1098/rstb.2008.0318>
- Moessnang, C., Otto, K., Bilek, E., Schäfer, A., Baumeister, S., Hohmann, S., Poustka, L., Brandeis, D., Banaschewski, T., Tost, H., & Meyer-Lindenberg, A. (2017). Differential responses of the dorsomedial prefrontal cortex and right posterior superior temporal sulcus to spontaneous mentalizing: DmPFC and pSTS Responses During Mentalizing. *Human Brain Mapping*, 38(8), 3791–3803. <https://doi.org/10.1002/hbm.23626>
- Paracampo, R., Pirruccio, M., Costa, M., Borgomaneri, S., & Avenanti, A. (2018). Visual, sensorimotor and cognitive routes to understanding others' enjoyment: An individual differences rTMS approach to empathic accuracy. *Neuropsychologia*, 116, 86–98. <https://doi.org/10.1016/j.neuropsychologia.2018.01.043>
- Paracampo, R., Tidoni, E., Borgomaneri, S., di Pellegrino, G., & Avenanti, A. (2016). Sensorimotor Network Crucial for Inferring Amusement from Smiles. *Cerebral Cortex*, cercor;bhw294v2. <https://doi.org/10.1093/cercor/bhw294>
- Pitcher, D., & Ungerleider, L. G. (2021). Evidence for a Third Visual Pathway Specialized for Social Perception. *Trends in Cognitive Sciences*, 25(2), 100–110. <https://doi.org/10.1016/j.tics.2020.11.006>
- Pobric, G., & de C. Hamilton, A. F. (2006). Action Understanding Requires the Left Inferior Frontal Cortex. *Current Biology*, 16(5), 524–529. <https://doi.org/10.1016/j.cub.2006.01.033>
- Ramsey, R., & Hamilton, A. F. de C. (2012). How does your own knowledge influence the perception of another person's action in the human brain? *Social Cognitive and Affective Neuroscience*, 7(2), 242–251. <https://doi.org/10.1093/scan/nsq102>

- Rilling, J. K., Sanfey, A. G., Aronson, J. A., Nystrom, L. E., & Cohen, J. D. (2004). The neural correlates of theory of mind within interpersonal interactions. *NeuroImage*, 22(4), 1694–1703. <https://doi.org/10.1016/j.neuroimage.2004.04.015>
- Rossi, S., Antal, A., Bestmann, S., Bikson, M., Brewer, C., Brockmüller, J., Carpenter, L. L., Cincotta, M., Chen, R., Daskalakis, J. D., Di Lazzaro, V., Fox, M. D., George, M. S., Gilbert, D., Kimiskidis, V. K., Koch, G., Ilmoniemi, R. J., Lefaucheur, J. P., Leocani, L., ... Hallett, M. (2021). Safety and recommendations for TMS use in healthy subjects and patient populations, with updates on training, ethical and regulatory issues: Expert Guidelines. *Clinical Neurophysiology*, 132(1), 269–306. <https://doi.org/10.1016/j.clinph.2020.10.003>
- 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), 2008–2039. <https://doi.org/10.1016/j.clinph.2009.08.016>
- Rossini, P. M., Burke, D., Chen, R., Cohen, L. G., Daskalakis, Z., Di Iorio, R., Di Lazzaro, V., Ferreri, F., Fitzgerald, P. B., George, M. S., Hallett, M., Lefaucheur, J. P., Langguth, B., Matsumoto, H., Miniussi, C., Nitsche, M. A., Pascual-Leone, A., Paulus, W., Rossi, S., ... Ziemann, U. (2015). Non-invasive electrical and magnetic stimulation of the brain, spinal cord, roots and peripheral nerves: Basic principles and procedures for routine clinical and research application. An updated report from an I.F.C.N. Committee. *Clinical Neurophysiology*, 126(6), 1071–1107. <https://doi.org/10.1016/j.clinph.2015.02.001>
- Santiesteban, I., Banissy, M. J., Catmur, C., & Bird, G. (2012). Enhancing Social Ability by Stimulating Right Temporoparietal Junction. *Current Biology*, 22(23), 2274–2277. <https://doi.org/10.1016/j.cub.2012.10.018>
- Santiesteban, I., Kaur, S., Bird, G., & Catmur, C. (2017). Attentional processes, not implicit mentalizing, mediate performance in a perspective-taking task: Evidence from stimulation of the temporoparietal junction. *NeuroImage*, 155, 305–311. <https://doi.org/10.1016/j.neuroimage.2017.04.055>
- Saygin, A. P. (2007). Superior temporal and premotor brain areas necessary for biological motion perception. *Brain*, 130(9), 2452–2461. <https://doi.org/10.1093/brain/awm162>
- Schurz, M., Radua, J., Aichhorn, M., Richlan, F., & Perner, J. (2014). Fractionating theory of mind: A meta-analysis of functional brain imaging studies. *Neuroscience & Biobehavioral Reviews*, 42, 9–34. <https://doi.org/10.1016/j.neubiorev.2014.01.009>

- Schuwerk, T., Langguth, B., & Sommer, M. (2014). Modulating functional and dysfunctional mentalizing by transcranial magnetic stimulation. *Frontiers in Psychology*, 5. <https://doi.org/10.3389/fpsyg.2014.01309>
- Silani, G., Lamm, C., Ruff, C. C., & Singer, T. (2013). Right Supramarginal Gyrus Is Crucial to Overcome Emotional Egocentricity Bias in Social Judgments. *The Journal of Neuroscience*, 33(39), 15466–15476. <https://doi.org/10.1523/JNEUROSCI.1488-13.2013>
- Spunt, R. P., & Adolphs, R. (2014). Validating the Why/How contrast for functional MRI studies of Theory of Mind. *NeuroImage*, 99, 301–311. <https://doi.org/10.1016/j.neuroimage.2014.05.023>
- Spunt, R. P., & Adolphs, R. (2015). Folk Explanations of Behavior: A Specialized Use of a Domain-General Mechanism. *Psychological Science*, 26(6), 724–736. <https://doi.org/10.1177/0956797615569002>
- Spunt, R. P., Falk, E. B., & Lieberman, M. D. (2010). Dissociable Neural Systems Support Retrieval of *How* and *Why* Action Knowledge. *Psychological Science*, 21(11), 1593–1598. <https://doi.org/10.1177/0956797610386618>
- Spunt, R. P., & Lieberman, M. D. (2012). Dissociating Modality-Specific and Supramodal Neural Systems for Action Understanding. *Journal of Neuroscience*, 32(10), 3575–3583. <https://doi.org/10.1523/JNEUROSCI.5715-11.2012>
- Spunt, R. P., Satpute, A. B., & Lieberman, M. D. (2011). Identifying the What, Why, and How of an Observed Action: An fMRI Study of Mentalizing and Mechanizing during Action Observation. *Journal of Cognitive Neuroscience*, 23(1), 63–74. <https://doi.org/10.1162/jocn.2010.21446>
- Talairach, J. and Tournoux, P. (1988) Co-planar stereotaxic atlas of the human brain: 3-Dimensional proportional system: An approach to cerebral imaging. Thieme Medical Publishers, Inc., New York.
- Tamir, D. I., & Thornton, M. A. (2018). Modeling the Predictive Social Mind. *Trends in Cognitive Sciences*, 22(3), 201–212. <https://doi.org/10.1016/j.tics.2017.12.005>
- Thioux, M., Gazzola, V., & Keysers, C. (2008). Action Understanding: How, What and Why. *Current Biology*, 18(10), R431–R434. <https://doi.org/10.1016/j.cub.2008.03.018>
- Thioux, M., Suttrup, J., & Keysers, C. (2018). *Perceiving Hesitation requires both the Motor and Mentalizing systems* [Preprint]. Neuroscience. <https://doi.org/10.1101/454298>
- Thompson, E. L., Bird, G., & Catmur, C. (2019). Conceptualizing and testing action understanding. *Neuroscience & Biobehavioral Reviews*, 105, 106–114. <https://doi.org/10.1016/j.neubiorev.2019.08.002>

- Tidoni, E., Borgomaneri, S., di Pellegrino, G., & Avenanti, A. (2013). Action Simulation Plays a Critical Role in Deceptive Action Recognition. *Journal of Neuroscience*, 33(2), 611–623. <https://doi.org/10.1523/JNEUROSCI.2228-11.2013>
- Tidoni, E., & Candidi, M. (2016). Commentary: Understanding intentions from actions: Direct perception, inference, and the roles of mirror and mentalizing systems. *Frontiers in Behavioral Neuroscience*, 10. <https://doi.org/10.3389/fnbeh.2016.00013>
- Urgesi, C., Candidi, M., & Avenanti, A. (2014). Neuroanatomical substrates of action perception and understanding: An anatomic likelihood estimation meta-analysis of lesion-symptom mapping studies in brain injured patients. *Frontiers in Human Neuroscience*, 8. <https://doi.org/10.3389/fnhum.2014.00344>
- Vaccaro, A. G., & Fleming, S. M. (2018). Thinking about thinking: A coordinate-based meta-analysis of neuroimaging studies of metacognitive judgements. *Brain and Neuroscience Advances*, 2, 239821281881059. <https://doi.org/10.1177/2398212818810591>
- Valchev, N., Tidoni, E., Hamilton, A. F. de C., Gazzola, V., & Avenanti, A. (2017). Primary somatosensory cortex necessary for the perception of weight from other people's action: A continuous theta-burst TMS experiment. *NeuroImage*, 152, 195–206. <https://doi.org/10.1016/j.neuroimage.2017.02.075>
- Van Overwalle, F. (2009). Social cognition and the brain: A meta-analysis. *Human Brain Mapping*, 30(3), 829–858. <https://doi.org/10.1002/hbm.20547>
- Van Overwalle, F., & Baetens, K. (2009). Understanding others' actions and goals by mirror and mentalizing systems: A meta-analysis. *NeuroImage*, 48(3), 564–584. <https://doi.org/10.1016/j.neuroimage.2009.06.009>
- Van Overwalle, F., & Vandekerckhove, M. (2013). Implicit and explicit social mentalizing: Dual processes driven by a shared neural network. *Frontiers in Human Neuroscience*, 7. <https://doi.org/10.3389/fnhum.2013.00560>
- Vander Wyk, B. C., Hudac, C. M., Carter, E. J., Sobel, D. M., & Pelphrey, K. A. (2009). Action Understanding in the Superior Temporal Sulcus Region. *Psychological Science*, 20(6), 771–777. <https://doi.org/10.1111/j.1467-9280.2009.02359.x>
- Vander Wyk, B. C., Voos, A., & Pelphrey, K. A. (2012). Action representation in the superior temporal sulcus in children and adults: An fMRI study. *Developmental Cognitive Neuroscience*, 2(4), 409–416. <https://doi.org/10.1016/j.dcn.2012.04.004>
- Vrij, A. (2000). *Detecting lies and deceit: The psychology of lying and the implications for professional practice*. John Wiley.

- Wolpert, D. M., Doya, K., & Kawato, M. (2003). A unifying computational framework for motor control and social interaction. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 358(1431), 593–602. <https://doi.org/10.1098/rstb.2002.1238>
- Wurm, M. F., & Schubotz, R. I. (2018). The role of the temporoparietal junction (TPJ) in action observation: Agent detection rather than visuospatial transformation. *NeuroImage*, 165, 48–55. <https://doi.org/10.1016/j.neuroimage.2017.09.064>
- Wurm, M. F., von Cramon, D. Y., & Schubotz, R. I. (2011). Do we mind other minds when we mind other minds' actions? A functional magnetic resonance imaging study. *Human Brain Mapping*, 32(12), 2141–2150. <https://doi.org/10.1002/hbm.21176>
- Yang, D. Y.-J., Rosenblau, G., Keifer, C., & Pelphrey, K. A. (2015). An integrative neural model of social perception, action observation, and theory of mind. *Neuroscience & Biobehavioral Reviews*, 51, 263–275. <https://doi.org/10.1016/j.neubiorev.2015.01.020>
- 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), 6753–6758. <https://doi.org/10.1073/pnas.0914826107>
- Zaki, J., Weber, J., Bolger, N., & Ochsner, K. (2009). The neural bases of empathic accuracy. *Proceedings of the National Academy of Sciences*, 106(27), 11382–11387. <https://doi.org/10.1073/pnas.0902666106>