

Implicit false-belief processing in the human brain



Dana Schneider^{a,b,*}, Virginia P. Slaughter^a, Stefanie I. Becker^a, Paul E. Dux^{a,**}

^a School of Psychology, The University of Queensland, Australia

^b Institute of Psychology, Friedrich Schiller University Jena, Germany

ARTICLE INFO

Article history:

Accepted 9 July 2014

Available online 17 July 2014

Keywords:

Eye-movements

fMRI

Implicit theory of mind

Social cognition

ABSTRACT

Eye-movement patterns in 'Sally–Anne' tasks reflect humans' ability to implicitly process the mental states of others, particularly false-beliefs – a key theory of mind (ToM) operation. It has recently been proposed that an efficient ToM system, which operates in the absence of awareness (implicit ToM, iToM), subserves the analysis of belief-like states. This contrasts to consciously available belief processing, performed by the explicit ToM system (eToM). The frontal, temporal and parietal cortices are engaged when humans explicitly 'mentalize' about others' beliefs. However, the neural underpinnings of *implicit false-belief* processing and the extent to which they draw on networks involved in *explicit general-belief* processing are unknown. Here, participants watched 'Sally–Anne' movies while fMRI and eye-tracking measures were acquired simultaneously. Participants displayed eye-movements consistent with implicit false-belief processing. After independently localizing the brain areas involved in explicit general-belief processing, only the left anterior superior temporal sulcus and precuneus revealed greater blood-oxygen-level-dependent activity for false- relative to true-belief trials in our iToM paradigm. No such difference was found for the right temporal–parietal junction despite significant activity in this area. These findings fractionate brain regions that are associated with explicit general ToM reasoning and false-belief processing in the absence of awareness.

© 2014 Elsevier Inc. All rights reserved.

Introduction

'Theory of mind' (ToM) refers to the processes involved in representing another individual's cognitions by ascribing mental states (e.g., beliefs) to them and is thought to be crucial for social development (Hughes, 2011; Razza and Blair, 2009). Indeed, its impairment appears to play a key role in a range of psychiatric and developmental disorders (e.g., autism and schizophrenia; Frith, 2004; Frith, 2001). Of particular interest is false-belief processing, which is a prominent ToM operation that is considered a uniquely human ability (Call and Tomasello, 2008; Martin and Santos, 2014; Wimmer and Perner, 1983). Such processing is typically assessed using 'Sally–Anne' tasks, in which participants view a scenario in which 'Sally' places a ball in a basket and then leaves the room before 'Anne' transfers it to a box. Sally then returns to the room. False-belief processing is reflected by a participant being able to indicate that Sally expects the ball to be in the basket, which is contrary to the real state of events (i.e., that the ball is in the box).

Until recently, the ability to process false-beliefs was thought to reflect a developmental milestone, with children generally unable to

pass these Sally–Anne type tasks before 4 years of age (Perner and Lang, 1999). However, it has been proposed that humans actually possess two systems that can undertake belief processing: an explicit ToM (eToM) system, which is involved in overt responding and develops across the first few years of life, and an implicit ToM (iToM) system, possibly present from birth. The former is thought to act in a deliberate, controlled and conscious manner whereas the latter has been hypothesized to act efficiently and in the absence of awareness (Apperly and Butterfill, 2009). Implicit false-belief processing cannot be accessed using explicit responses (e.g., verbal report), because once participants have been questioned regarding others' cognitions one cannot guarantee that they are unconscious of engaging in ToM reasoning (Clements and Perner, 1994). Thus, eye-movements in Sally–Anne tasks have predominantly been employed to study implicit false-belief processing. The key finding is that children as young as 7 months will often look at a given location, without a stimulus, when it is consistent with another's false-belief about the stimulus being at that location (e.g., Kovács et al., 2010; Southgate et al., 2007).

To date, imaging studies have focused on the neural correlates of explicit general-belief processing, by having participants overtly respond to others' mental states (e.g., Saxe and Kanwisher, 2003), i.e., asking participants if an agent in a scenario has a false belief. Under these conditions, greater blood-oxygen-level-dependent (BOLD) activity is typically found for belief- relative to no-belief control trials (e.g., false-belief vs. false-photo) in the anterior cingulate cortex (ACC), medial prefrontal cortex (mPFC), the temporal–parietal junction (TPJ), the

* Correspondence to: D. Schneider, Institute of Psychology, Department of General Psychology and Cognitive Neuroscience, Friedrich Schiller University Jena, Am Steiger 3, Haus 1, 07743 Jena, Germany.

** Correspondence to: P.E. Dux, School of Psychology, The University of Queensland, McElwain Building, St Lucia, QLD 4072, Australia.

E-mail addresses: msdanaschneider@gmail.com (D. Schneider), paul.e.dux@gmail.com (P.E. Dux).

superior temporal sulcus (STS), temporal pole (TempP) and the precuneus (PC) suggesting these regions play an important role in explicit general-belief processing (Dodell-Feder et al., 2011). Importantly, the same network also displays greater activity for explicit false-belief processing relative to true-belief processing (Aichhorn et al., 2009; Sommer et al., 2007; but see, Döhnel et al., 2012 for evidence that not all eToM regions show false- vs. true-belief activation differences), indicating that these brain areas can represent belief states that contradict the true or current state of affairs.

The few previous studies that have aimed to assess the neural substrates of implicit ToM processing have attempted to do so by having participants view natural scenes, movies of everyday social events or of animate shapes (Iacoboni et al., 2004; Wagner et al., 2011; Wheatley et al., 2007). For example, Wagner et al. (2011) had participants categorize natural images into animal, vegetable or mineral categories without mentioning any social cognitive concepts so as to keep participants naïve to the experimental manipulation. The authors found that the dorsal medial prefrontal cortex (dmPFC), the PC and temporal poles showed increased activity for human versus nonhuman scenes. Furthermore, trait empathy correlated with dmPFC and temporal pole activities, especially for human social scenes. As these results were found in brain regions of the eToM network, it was inferred that an implicit ToM or mentalizing processes must have taken place. In this and related studies, however, it remained unclear whether the processing of another's cognitions indeed occurred as no measure of mentalizing was acquired. In short, no work to date has systematically examined the neural substrates of implicit false-belief processing. This is an important issue as false-belief tasks are unique in the way they require observers to make mental state inferences. Specifically, the observer cannot predict actions based on a representation of reality, but instead must establish a representation of another individual's knowledge in order to identify that this belief is false.

Here we examined the extent to which neural regions associated with explicit general-belief processing are involved in implicit false-belief analysis. Specifically, we aimed to identify regions engaged during overall general explicit ToM processing, those tapped when the beliefs of an agent are overtly analyzed, and then assessed which of these areas are specifically involved in implicit false-belief processing (see below). Participants were scanned using functional magnetic resonance imaging (fMRI) while they watched 'Sally-Anne' movies (Fig. 1, see the [Materials and methods](#) section). Implicit false-belief processing was assessed by simultaneously measuring participants' eye-movements. To examine the regions involved in explicit general-belief reasoning we employed a well-established explicit ToM localizer (Dodell-Feder et al., 2011).

Materials and methods

Overview

False- and true-belief ToM movies were segmented into two sequences (26 s per sequence): sequence 1 – belief set-up and sequence 2 – belief test. For each trial, the sequences were separated by 14 s of fixation in order for us to ensure that the BOLD response to each returned to baseline and was not contaminated by the previous sequence. The set-up sequence displayed an actor watching a hand puppet moving a ball between different boxes. At different time points during the movies, an actor watching the puppet's actions left the room and the timing of this exit created either the context for a false- or true-belief trial in the test sequences. The test sequences displayed the puppet moving the ball one more time, now establishing the false- or true-belief for the actor. Then the actor re-entered the room, took a seat and looked down in between the boxes (the actor's eyes were hidden from view by wearing of a visor). Importantly, the test sequences were physically identical across false- and true-belief conditions and differed only with respect to whether a false- or true-belief context had been created for the returning actor by the paired set-up sequence. This allowed us to contrast conditions that varied only in the extent to which they placed demands on the resources involved in belief processing: In the false-belief trials the belief state of the actor and that for the participant, represented through the puppet (and ball location), were incongruent. Thus, both needed to be monitored to ensure that they did not interfere with one another. By contrast, in the true-belief trials the participant's belief state, represented through the puppet (and the ball location), was congruent with the belief state of the actor. Thus only one of these needed to be monitored and there was no opportunity for interference/confusion between the belief states.

At the end of each test sequence, the movie froze and participant's gaze behavior was assessed (i.e., first box fixations in the time period from when the actor re-entered the room) as an indicator for implicit false-belief processing. Throughout the sequences, to distract participants from actively engaging in belief processing, they were asked to discriminate, as quickly and accurately as possible, high versus low tones that were pseudo-randomly presented during the movies. In addition, to assess the extent to which participants were aware of our belief manipulation a 10-item funneled debriefing was administered at the end of the iTOM trials (see Supplementary material). Participants were included in the data analyses only when they reported being unaware of the belief manipulation. This approach has proven sensitive to explicit ToM processing in our previous multiple trial implicit false-belief processing studies (Schneider et al., 2012a,b, 2013, 2014). Finally,

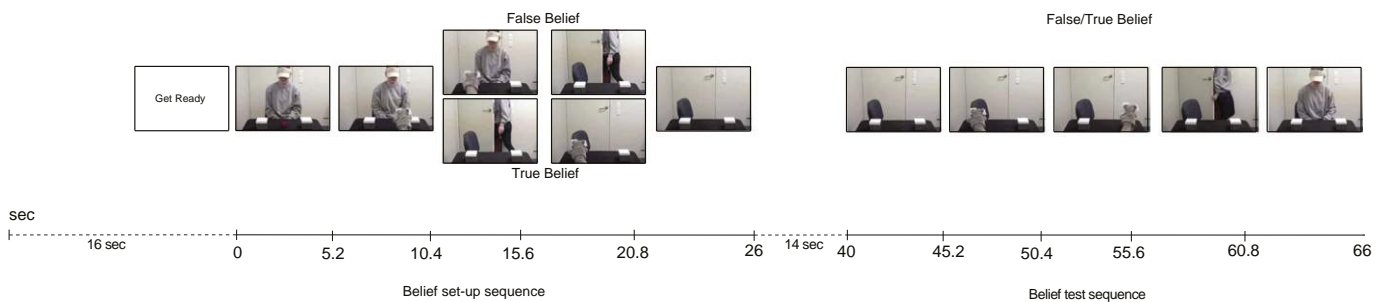


Fig. 1. Trial structure and conditions in the implicit false-belief task. A slow-event design was employed with a belief set-up sequence differing between false- and true-belief scenarios and a belief test sequence that was identical between these conditions and the focus of our analyses. Inter-stimulus-intervals were presented before and after each movie sequence to allow for the hemodynamic response to return to baseline. Set-up sequence: Upper stream: False-belief scenario (with ball finishing in the box to the right of the actor) – the hand puppet transfers a ball from one box to another, then the actor leaves the room. Lower stream: True-belief scenario (with ball finishing in the box to the left side of the actor) – once the ball is placed in one box the actor leaves the room then a hand puppet transfers the ball to the other box. Test sequence: The hand puppet appears and transfers the ball back to the other box. Thus, the returning actor has now either a false or true belief regarding the location of the ball. Crucially, the only difference for the test sequence between the conditions is the belief state of the actor created in the set-up sequence. During all the scenarios 1 of 2 tones were pseudo-randomly presented to participants through headphones as a distraction task from the belief manipulation. Participants were required to make high vs. low pitch discriminations as quickly and accurately as possible.

to isolate regions of interest involved in explicit general-belief processing participants completed a well-established story localizer task in the scanner (Dodell-Feder et al., 2011), consisting of false-belief and false-photo/map (control) stories for which they made false- and true-responses (see details below). Having exited the scanner all participants completed auxiliary questionnaire measures, to assess general social processing, administered in a counterbalanced order.

Participants

Twenty participants (11 females, mean age = 22.70 years, SD = 4.44 years) with normal or corrected-to-normal vision and no history of psychiatric illness or neurological injury or illness took part in the study. All participants gave informed written consent and received payment for taking part in the study. The protocol was approved by the ethics committee at The University of Queensland. Data from four participants were excluded from analyses due to a scanner malfunction ($n = 2$) or explicit belief processing during our implicit ToM task ($n = 2$; see below); hence the final number of participants came to $n = 18$ in the behavioral data analysis and $n = 16$ in the MRI analysis.

Implicit false-belief event-related task

Participants lay supine in the scanner with their head secured to minimize movement artifacts. They viewed the stimuli via a rear-projected mirror display mounted on the head coil. The eye-tracker was placed at the end of the bore and recorded one of the participants' eyes via the mirror. The stimuli were controlled with Presentation software (Neurobehavioral Systems, Albany, CA) and presented on a screen with dimensions of 30×24 cm and a resolution of 1280×1024 .

Using a slow event-related design, false- and true-belief movies were presented that were adapted from our previous work (Schneider et al., 2012a). The two belief trial types were generated by combining a unique set-up movie sequence for each condition (false- versus true-belief) with the same test movie sequence. In both trial types an actor was sitting behind a desk and watched a puppet move a ball between two boxes on the desk. In the false-belief set-up sequence the puppet hid the ball in one of the boxes and then moved it into the other box. The actor was present and watched these actions and then left the room (Fig. 1 upper stream; www.youtube.com/watch?v=2tG75nwk1E). The set-up sequences for the true-belief trials involved the actor leaving the room immediately after the first ball movement. Thus, the second ball movement occurred unbeknownst to the actor, while the actor was outside the room (Fig. 1 lower stream; www.youtube.com/watch?v=fXLI8PiiMxc). The false- and true-belief test sequences were identical (Fig. 1; www.youtube.com/watch?v=evFRneJxxYE), and started with the puppet moving the ball to the other box. Following this, the actor re-entered the room, took a seat behind the desk and then the movie froze for 6 s, in which the gaze behavior of the participants was monitored. Two versions of the experimental trials were employed, in which the initial and final locations of the ball and the actor's gaze were counterbalanced (false-belief: right and left ball location; and true-belief: right and left ball location). Note that our actor wore a visor that occluded the eyes to avoid gaze-cueing effects (Frischen et al., 2007). Previously, Schneider et al. (2012a) have ruled out gaze-cueing accounts for implicit false-belief anticipatory looking effects.

In total, participants watched 20 false-belief and 20 true-belief trials presented in a pseudo-randomized order across 8 runs. Each of these runs lasted 430 s and consisted of 5 trials each made up of a set-up and test sequence (26 s each). All sequences were followed by an inter-stimulus interval of 14 s. In addition, a trial onset signal ('Get ready', 2 s) was presented before the beginning of each trial (Fig. 1).

To distract participants from the experimental belief manipulation, throughout the trials they were asked to identify, as quickly and accurately as possible, one of two tones (high [2 kHz] versus low [1.1 kHz]

judgment) via a button box response (upper button for high tones, lower button for low tones). Four tones were played during the set-up sequences and two tones during the test sequences (displayed always before the actress re-entered the room) and their temporal position and identity was pseudo-randomly determined. The auditory stimuli were presented using MRI-compatible headphones (MRI Confon, Leibniz Institute for Neurobiology, Magdeburg, Germany). To assess belief processing in the implicit task an MRI-compatible remote EyeLink 1000 (SR Research, Mississauga, Ontario, Canada) was employed, which sampled the gaze position monocularly at 500 Hz throughout the entire time course of the set-up and test sequence scans. The analysis of gaze behavior focused on the time frame from when the actor re-entered the room in the test sequence, which was divided into five areas of interest (face; left arm; right arm; left box; right box). This allowed us to examine whether participants were more inclined to look at the empty box (no-ball location) when the actor falsely believed that the ball was at that location (false-belief condition), compared with when she correctly believed that it was not at that location (true-belief condition). Participants were simply told to watch the movies and complete the tone task. They were aware that we were monitoring their eyes, thus it was unlikely they would ignore the movies (e.g., close their eyes) and just complete the tone task.

To ensure that we were indeed examining implicit false-belief processing, at the end of the implicit belief scanning-session participants were taken through a 10-item funneled debriefing protocol (see Supplementary material). Here the experimenter stood beside the MRI scanner and participants verbally answered the questions while still lying in the scanner. This debriefing procedure (Bargh and Chartrand, 2000), adjusted to our study content (Schneider et al., 2012a), probed with increasing specificity whether participants engaged in conscious processing of the actor's belief states. Participants were excluded as soon as they gave any indication of having consciously processed the actor's mental state (i.e., reported thinking about the puppet's or actor's mental states or made utterances regarding the expected actions of the characters). A typical answer, which would be coded as conscious processing of others' mental states would be: 'the puppet tricked the actor some of the time'. Participants were not excluded when they reported physical reasoning during debriefing. Typical answers during the debriefing, which were categorized as participants having not consciously processed the actor's mental state were, for example: 'this was a sequence of random events'; 'this was about multitasking' or 'this was about establishing eye-movement and tone interactions'. The implicit belief conditions, including the debriefing procedure, were always administered first, to avoid participants becoming aware of the experimental manipulations. It lasted approximately 75 min.

Explicit ToM localizer task

A well-established explicit general-belief localizer task was also administered to participants (<http://saxelab.mit.edu/superloc.php>; Dodell-Feder et al., 2011). The stories from the localizer were presented using a slow event-related approach. Stimuli consisted of short stories and questions, centrally presented in 30-point white text on a gray background. Stories were constructed to fit two categories: either to describe a false-belief or false-photograph/map scenario. The stories consisted of 2–4 short sentences. Immediately after the stories, a question was displayed, which consisted of a single sentence with a two-alternative true/false response prompt. An example false-belief story/question would be: 'When Lisa left Jacob he was deep asleep on the beach. A few minutes later a wave woke him. Seeing Lisa was gone Jacob decided to go swimming.' Question: 'Lisa now believes that Jacob is sleeping. True or False?' The correct answer for this example would be true, reflecting that participants had correctly represented Lisa's false/outdated belief. In a false-photograph/map example participants would be exposed to exactly the same story/question set-up, except they would be encouraged to represent an outdated/false photograph or map description. For

example, a false photograph story/question would be: ‘A large oak tree stood in front of City Hall from the time the building was built. Last year the tree fell down and was replaced by a stone fountain.’ Question: ‘An antique drawing of City Hall shows a fountain in front. True or False?’ The correct answer for this example would be false, reflecting that participants had correctly represented the false/outdated photograph. Participants pressed the upper button to choose the true option and the lower button for the false option.

Fifty percent of the belief and photograph/map questions required a ‘true’ response and the remaining trials a ‘false’ response. In total 2 runs were undertaken by each participant and each of these lasted for 340 s. Runs contained 5 false-belief and 5 false-photograph/map stories (presented for 12 s) followed by the true/false choice question (presented for 6 s). Fixation periods of 14 s were interleaved between the trials and the condition order was counterbalanced. Reaction time and accuracy data were collected, however only accuracy judgments were emphasized to participants. This explicit belief task component took approximately 15 min to administer.

Auxiliary measures

After exiting the MRI scanner participants completed three additional self-report questionnaires in a counterbalanced order to examine if they had any social processing difficulties that could influence the pattern of results. These included the Autism Quotient (AQ; Baron-Cohen et al., 2001), assessing autistic traits, the Empathy Quotient (EQ; Baron-Cohen and Wheelwright, 2004), which provides a score on an individual's capacity to recognize other's feelings and emotions and the Systemizing Quotient (SQ; Baron-Cohen et al., 2003), which assesses an individual's capacity to analyze and construct rules.

fMRI data acquisition

Anatomical and functional images were obtained using a 3 T Siemens Trio MRI scanner (Erlangen, Germany) with a 12-channel head coil. The high-resolution anatomical scan (T1) was acquired in the middle or at the end of the iTOM scanning session using a rapid gradient echo sequence (TR = 1900 ms, TE = 2.32 ms, FA = 9°, field of view (FOV) = 230 × 230, matrix = 256 × 256, resolution = 0.9 mm³). All functional (T2* weighted) images were acquired parallel to the AC–PC plane using an echo-planar imaging sequence with the following parameters: TR = 2000 ms, TE = 30 ms, FA = 90°, FOV = 190 × 190, matrix = 64 × 64 with 31 slices (3.0 mm thick, 0.3 mm gap). An interleaved axial slice order was used. Stimulus presentation was synchronized with each fMRI scan acquisition.

fMRI data preprocessing

The first 2 (i.e., for the implicit belief task) or 3 (i.e., for the explicit belief localizer) scans from each run were discarded from the analysis to allow for T1 equilibration. Data were preprocessed and analyzed using BrainVoyager QX (Brain Innovation, Maastricht, The Netherlands). Pre-processing comprised 3D-motion correction, slice-scan time correction, and the application of a high-pass temporal filter (3 cycles/run). Functional images were aligned to the image from the first run and transformed into Talairach space (Talairach and Tournoux, 1988).

fMRI data analysis

Explicit ToM localizer task

To isolate areas involved in explicit ToM/general-belief processing we conducted a whole-brain random effects general linear model (GLM) analysis to localize ROIs that reflected greater BOLD activation for false-belief relative to false-photo/map stories. We defined belief and photo/map events as regressors, which were then convolved with a double-gamma hemodynamic response function. Each event corresponded to

the 9 volumes that were acquired during the story and question presentation as recommended by Dodell-Feder et al. (2011). To identify voxels with significant activation we applied a cluster threshold of 108 mm³ with an alpha value of $p \leq .005$ ($t(15) = 3.286$). We employed this threshold to ensure that all relevant eToM areas were isolated (Bayliss et al., 2012; Todorov et al., 2007). ROIs were defined at the group level around the peak voxel of activation by including all voxels above statistical threshold up to a maximum voxel cluster range 10 × 10 × 10 mm³.

Implicit false-belief event-related task

Individual participant's time courses, for the implicit belief event-related task, were extracted from the ROIs defined via the localizer. Subsequently the average peak BOLD percent signal change was determined, relative to the volume prior to the onset of stimuli, for each participant for the false- and true-belief conditions, for both set-up and test sequences. The peak of the BOLD percent signal change was taken from volumes acquired between the 3rd scan from sequence onset (6 s from movie onset) and the 3rd scan after sequence offset (6 s from movie offset).

Results

Behavioral analysis

Collectively, the behavioral results confirmed that participants engaged in implicit false-belief processing during the ‘Sally–Anne’ movies. Specifically, upon the actor re-entering the room at the end of the test sequences, participants were significantly more likely to make their first fixation to the box that did not contain the ball (No ball location) when the actor had a false-belief that the ball was at that location as opposed to a true-belief that the ball was not at this location ($t(17) = 1.80$, $p = .045$, paired t -test, one-tailed; Fig. 2). There was a trend for the opposite first fixation difference at the box location that contained the ball ($p = .065$, Fig. 2). Notably, this eye-movement result was found even under conditions where participants had high accuracy ($M = 95\%$ correct, $SD = 7.96$) and fast reaction times ($M = 968$ ms, $SD = 207$ ms), with no differences between false- and true-belief conditions ($p > .27$, $p > .08$, respectively) on the tone discrimination task. This replicates our own and others' previous research that employed eye-movement measurements to investigate ToM in the absence of awareness (Schneider et al., 2012a,b, 2013; Senju et al., 2009; Southgate et al., 2007). This was important to show as it confirmed that implicit ToM processing occurred even though participants completed the low

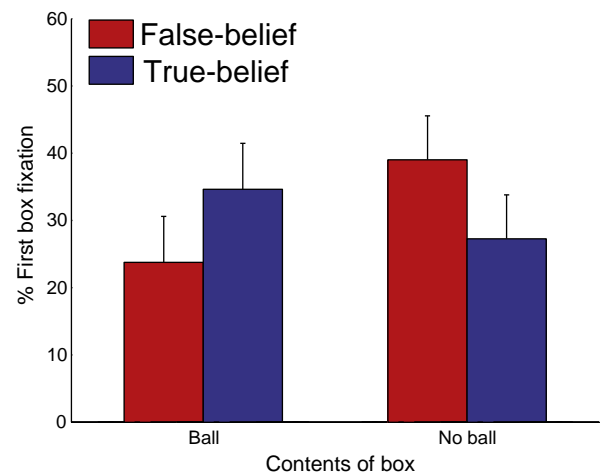


Fig. 2. Eye-movement results: implicit false-belief processing at the end of the belief test sequence. Percentage of first box fixations toward the boxes after the actor returned to the room as a function of her belief state. Error bars represent the standard error of within-subjects effects (false- vs. true-belief).

demand tone task and were in the noisy scanner environment (Schneider et al., 2012b).

Accuracy on the eToM task was high ($M = 85.7\%$, $SD = 12.24\%$) and reaction times were well within the 6-second response window ($M = 3.65$ s, $SD = 0.7$ s). There were no significant differences on accuracy or reaction time between the false-belief and false-photo conditions ($ps > .25$). This confirmed that participants engaged in explicit general-belief processing in this paradigm.

On the auxiliary measures none of the participants showed unusual empathic ($M = 38.19$, $SD = 9.98$ out of 80), systematic ($M = 64.19$, $SD = 20.45$ out of 150) or autistic traits ($M = 16.88$, $SD = 6.54$ out of 50, clinical range = 32 and above), which indicates that there were no social processing deficits in our participant group which could have influenced the imaging results.

MRI analysis: Explicit theory of mind

To localize the brain regions that have been consistently observed when participants engage in explicit belief information processing, we contrasted BOLD activity between false-belief and false-photo/map stories from the eToM task (Dodell-Feder et al., 2011). Thus, contrasting trials that differed in whether or not they engaged explicit general-belief processes. The resulting statistical parametric map (SPM) revealed a network of brain regions: right TPJ (BA 39, peak voxel Talarach coordinates 45, -58 , 25 [x,y,z]), left TPJ (BA 40, peak voxel Talarach coordinates -51 , -52 , 28), PC (BA 7, peak voxel Talarach coordinates 6, -55 , 31), right STS (BA 21, peak voxel Talarach coordinates 57, -19 , -5), left STS (BA 21, peak voxel Talarach coordinates -60 , -31 , -5), left TempP (BA 21, peak voxel Talarach coordinates -54 , -1 , -8) and

left middle frontal gyrus (left MFG; BA 6, peak voxel Talarach coordinates -33 , 2, 43). All these regions were significantly more activated in the belief compared to the photo/map trials ($t(15) = 3.286$, $p \leq .005$, cluster threshold 108 mm^3 random effects, Fig. 3A). This explicit general-belief processing network is highly consistent with that identified in previous eToM studies, with the exception of the mPFC not activating to this contrast (Aichhorn et al., 2009; Carrington and Bailey, 2009; Dodell-Feder et al., 2011; Sommer et al., 2007; van Overwalle, 2009; van Overwalle and Baetens, 2009).

MRI analysis: Implicit theory of mind

Analyses were focused on the test sequence, because as noted above, the implicit belief conditions in these movies did not physically differ. Consequently, any variations observed between implicit false- and true-belief conditions would reflect differences in the demands placed on belief processing operations induced by the context from the set-up sequences.

A 2 (belief: false-belief vs. true-belief) \times 7 (region: right TPJ vs. left TPJ vs. PC vs. right STS vs. left STS vs. left TempP vs. left MFG) repeated measures ANOVA revealed a significant main effect of belief condition ($F(1, 15) = 7.93$, $p = .013$, $\eta_p^2 = .35$) with greater overall peak amplitude activity in this network for false-belief relative to true-belief trials (Fig. 3B). Crucially, however, there was also a significant interaction ($F(1, 15) = 2.91$, $p = .012$, $\eta_p^2 = .16$) which demonstrated that not all ROIs in the explicit general-belief processing network were equally sensitive to our implicit false- vs. true-belief contrast. Specifically, only left STS ($t(15) = 4.22$, $p = .001$, paired t -test) and PC ($t(15) = 2.23$, $p = .042$, paired t -test) peak amplitude activity revealed a significant

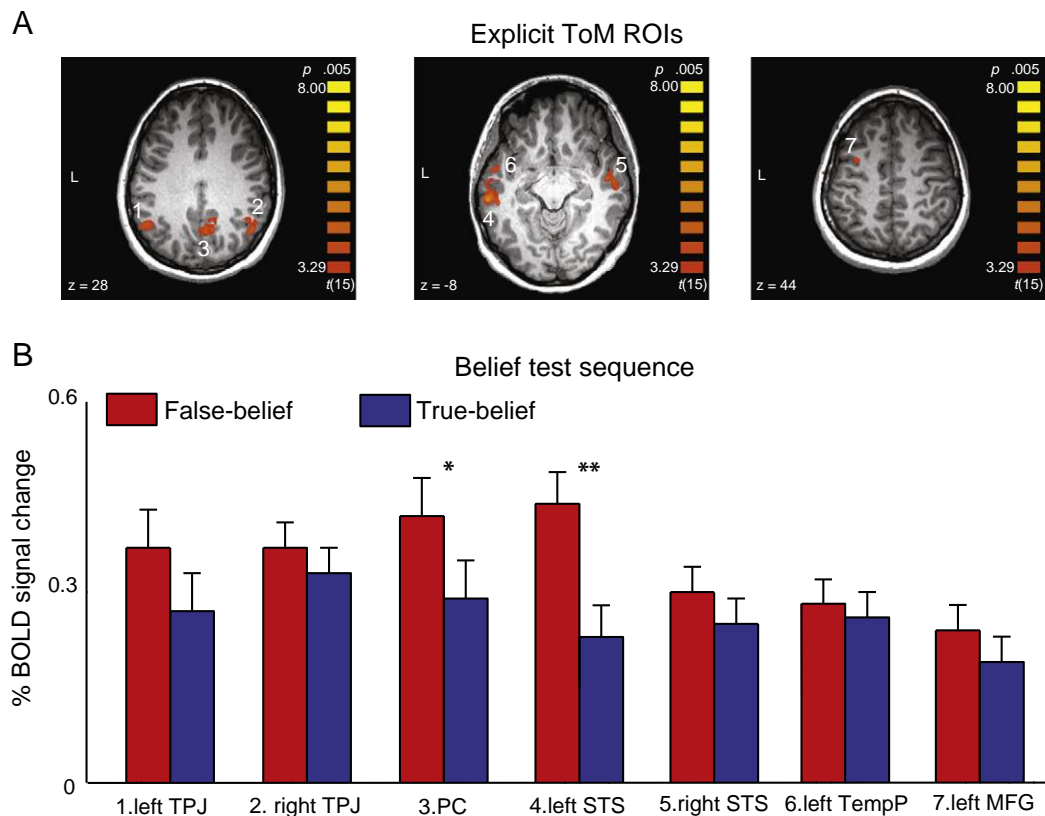


Fig. 3. Implicit false-belief processing in the explicit ToM network. A) eToM regions of interests (ROIs). Group SPM depicted on a single-subject T1-weighted image displaying significantly activated areas for whole-brain contrast of belief > photo/map stories (Dodell-Feder et al., 2011), $t(15) = 3.286$, $p \leq .005$, cluster threshold 108 mm^3 . eToM network included: Left panel: 1. left temporal-parietal junction – left TPJ (BA 40), 2. right temporal-parietal junction – right TPJ (BA 39) and 3. precuneus – PC (BA 7); middle panel: 4. left anterior superior temporal sulcus – left STS (BA 21), 5. right anterior superior temporal sulcus – right STS (BA 21), 6. left temporal pole – left TempP (BA 21); right panel: 7. left middle frontal gyrus – left MFG (BA 6). B) Implicit belief processing activity in the eToM ROIs for the belief test sequence: Mean BOLD % signal change peak amplitude as a function of the belief state of the actor. Error bars represent the standard error of within-subjects effects (false- vs. true-belief).

difference between the implicit false- and true-belief scenarios, with % signal change greater in the former relative to the latter condition (Fig. 3B; all other areas $p > .1$ for this contrast). Importantly, planned contrasts demonstrated that the difference in activity between the false- and true-belief conditions was significantly larger for the left STS ($t(15) = 3.12, p = .0035$, one-tailed, paired t -test) and PC ($t(15) = 1.95, p = .035$, one-tailed, paired t -test), relative to that for the right TPJ – the proposed key brain region for explicit general-belief processing (Saxe and Kanwisher, 2003; Scholz et al., 2009).

To confirm the above results we employed a Bayesian paired t -test approach (Rouder et al., 2009) in order to assess for a 'default' non-informative (JSZ) prior – the Bayes Factor – the ratio of evidence supporting the null hypothesis (i.e., no difference in BOLD activity between false- and true-belief trials) versus that supporting the alternative hypothesis (i.e., a meaningful difference in BOLD activity between false- and true-belief trials). We calculated the inverse Bayes Factor, values ≥ 3 indicate evidence for activation differences, those around 1 offer evidence neither in favor of, nor against, the null, while values $\leq 1/3$ indicate evidence for the null hypothesis (Tamber-Rosenau et al., 2013; Rouder et al., 2009). Evidence for meaningful differences, i.e., support for the alternative hypothesis, was observed for the left STS (inverse Bayes Factor = 50.9) no such evidence was observed for the PC, left TPJ, right STS and left MFG regions (inverse Bayes Factors: 1.5; 0.68; 0.44; 0.36, respectively) while support for the null was observed in right TPJ and left TempP (inverse Bayes Factors: 0.27; 0.2). Collectively the combined results demonstrate that the majority of areas subserving explicit general-belief analyses (i.e., right TPJ, left TPJ, right STS, left TempP and left MFG) are not sensitive to tracking the false- relative to true-belief states of others in the absence of awareness. In addition, SPMs generated using our iTOM event-related runs did not reveal any additional areas that significantly distinguished (either at Bonferroni or q [FDR] corrected levels) between implicit false- and true-belief conditions. Thus, it appears that only the left STS and, perhaps, PC are involved when false-belief information is tracked implicitly.

Discussion

Here we demonstrated for the first time the brain regions involved in processing false-beliefs of others in the absence of awareness. In addition, we assessed the extent to which these implicit false-belief processing areas overlapped with the now well-established network that underpins explicit general-belief processing. Of these 'typical' eToM regions, we found that only the left STS and PC displayed activity that differentiated between implicit false- and true-belief operations. This is an important finding as considerable past research has implicated the right hemisphere TPJ as the key region involved in explicitly reasoning about the mental states of others (e.g., Saxe and Kanwisher, 2003; however see Santiesteban et al., 2012). The current data suggest instead that other regions of the eToM network are most relevant for differentiating between belief states implicitly.

Interestingly, it has been proposed that the PC and STS play different roles in ToM processing when stimuli are registered as being relevant for the 'self' or 'other's' mental states (Abu-Akel and Shamay-Tsoory, 2011). Specifically, it has been suggested that the PC is involved in computing stimulus-aspects relevant to the self through the dorsal attention pathway whereas the STS is important when stimuli are recognized as being relevant for another person through the ventral attention system. Thus, the current patterns of results suggest that both the dorsal and ventral attention systems are actively involved in processing false-belief information implicitly.

Did we fail to observe TPJ's involvement in implicit false-belief processing due to a lack of sensitivity? To identify the explicit general-belief processing network we used a well-established localizer task (Dodell-Feder et al., 2011), which isolated the commonly identified eToM regions. For all these brain areas implicit ToM processing led to significant activity compared to baseline, particularly for the right TPJ

across the set-up and test sequences ($t(15) > 5.60, p < .001$). However, only activity in the left STS and PC differentiated between implicit false- and true-belief conditions in the test processing stages. Thus, we had the sensitivity to find implicit belief processing activation differences with our approach. In addition, a SPM contrasting the test conditions (false-belief > true-belief), using a more liberal threshold ($p \leq .005$), did not implicate the right TPJ. The only other additional region, outside the eToM network, that was identified in this analysis was the right insula (Fig. 4); another area that has previously been suggested to play an important role in social information processing (Gobbini et al., 2007; Lissek et al., 2008; Sebastian et al., 2012). Further, performing an identical time course analysis, as mentioned above contrasting implicit false- and true-belief conditions, using an individual ROI approach also revealed no BOLD activity difference in the right TPJ ($t < 1.24, p > .24$). Thus, collectively, our findings demonstrate that implicit false-belief processing activates regions in the typical eToM network, however the majority of these brain areas do not distinguish between implicit false- and true-belief conditions (Fig. 3B). This is not to say that these areas are not involved in general *implicit* belief processing/ToM. Indeed, they may be tapped equally during false- and true-belief conditions. Of key relevance here is that there is still debate concerning whether or not false- and true-belief trials activate the eToM network differently when participants overtly mentalize (Aichhorn et al., 2009; Sommer et al., 2007; but see, Döhl et al., 2012).

Having said this, as argued in the Introduction section, it seems likely that our false-belief trials place greater demands on mentalizing than our true-belief trials, as in the former there is conflict between the actor and participant's belief state, whereas this is not the case in the true-belief trials. Indeed, this represents a parametric variation approach (commonly used across a variety of cognitive neuroscientific studies; see Marois and Ivanoff, 2005), because in the former the participants must monitor their own and the actor's belief but this is not the case in true-belief trials with the participant only having to track their own belief or the ball. Thus, one might still expect general iTOM areas to be sensitive to our implicit false- versus true-belief contrast. Further, as we only found activation differences between false- and true-belief trials during the test sequences our results cannot reflect low-level sensory differences between the conditions. That is because the implicit

Belief test sequence: FB>TB

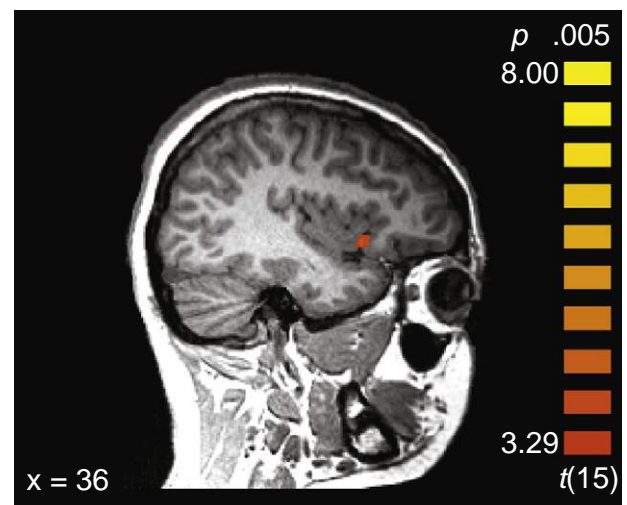


Fig. 4. Whole-brain analysis for implicit false-belief processing. Statistical parametric map (depicted on a single-subject T1-weighted image) displaying areas that showed greater peak BOLD amplitude for false- relative to true-belief scenarios in the test sequence, $t(15) = 3.286, p \leq .005$, cluster threshold 108 mm^3 .

belief conditions were physically identical and only differed in belief content created by the set-up sequence. In short, if not belief processing differences, what else could our imaging results reflect? Of course, an implicit false-belief versus no-belief contrast, as has been employed in explicit ToM studies, may implicate other areas in iToM. However, being unable to overtly instruct participants regarding the extent to which they should employ ToM in implicit studies renders it difficult to implement a no-belief control condition, particularly when conditions need to be physically identical. How can one prevent an individual from implicitly and automatically processing the beliefs of an actor in one condition (no-belief control) but not in another (belief trials) when test stimuli are identical and there are no task instructions? This is an important topic for future investigation.

The present findings isolate implicit false-belief operations in the healthy human brain, and converge with previous behavioral and theoretical work dissociating implicit and explicit ToM (Apperly and Butterfill, 2009; Clements and Perner, 1994; Low and Watts, 2013; Perner and Roessler, 2012). While no previous study has directly examined the neural substrates of implicit false-belief processing or the distinction between regions subserving eToM and iToM (Iacoboni et al., 2004; Wagner et al., 2011; Wheatley et al., 2007) some previous work has assessed related topics. For example, the brain regions activated when participants give ToM descriptions regarding the movement of animated shapes (Castelli et al., 2000; Gobbini et al., 2007); the areas tapped for ToM processing of verbal versus non-verbal tasks (Gallagher et al., 2000); brain structure correlates of spontaneously mentalizing about naturalistic movies (Rice and Redcay, *in press*); and the differential activity observed between conditions where participants simply read mentalizing stories as opposed to actively infer traits from them (Ma et al., 2010) have all been examined. In addition, Young et al. (2007) had participants read scenarios and make judgments on the moral permissibility of actor's actions that lead to either a negative (someone's death) or neutral outcome (no death). Unbeknownst to the participants (no task instructions to monitor beliefs of the actor) the scenarios also detailed whether these actions were either intended (based on the actor's beliefs) or incidental. Right TPJ activity was influenced by the interaction of beliefs and moral permissibility of the actor – i.e., showed greatest activity for intended harm.

Geangu et al. (2012) employed an approach similar to the present study in which participants passively watched (i.e., without having their attention drawn to any particular aspect of the stimuli) photographic sequences that portrayed false- and true-belief scenarios and at the same time recorded electroencephalography (EEG). These researchers found that a range of frontal, central and parietal electrodes could distinguish between false- and true-belief conditions from up to 200 ms after the onset of the last image. However, EEG, particularly in the absence of source localization analysis, has poor spatial resolution and consequently cannot provide stark insights into the neural substrates of implicit false-belief processing. In addition, no eToM task was used in the experiment and therefore a neural association/dissociation between regions involved in explicit and implicit belief processing could not be assessed. Of primary importance though, is that the study did not include an additional distraction task, follow-up debriefing and/or a measure of belief processing, thus it cannot be determined whether the beliefs of the actor in the photographs were processed implicitly or explicitly. Indeed, this is a potential shortcoming issue for most studies that examined ToM abilities with limited or no explicit instructions.

For our implicit belief task, we employed an anticipatory looking paradigm while participants were scanned with MRI, thus guaranteeing high spatial resolution. In addition, a distraction task and extensive post-experimental debriefing, which built on our previous work (Schneider et al., 2012a,b, 2013, 2014), ensured that participants were not aware of tracking the mental state of the actor in our belief movies. As this protocol was sensitive enough to identify two participants who had at least some awareness of engaging in belief processing, we feel confident that the other participants were unaware of processing the

belief state of the actor. Of course, it may still be possible that participants explicitly considered the mental states of the actor during the movies, but failed to report it during debriefing as they focused on the relationship between the tone task and movies. However, we specifically asked in our debriefing direct questions regarding the social aspects of the movies (e.g., Did you notice the occasional false- or true-belief state in the actress? Did you notice that the actress in the videos was occasionally tricked?). Thus, it seems likely that if participants engaged in explicit ToM processing, it would have been detected.

Recently, the temporal profile of processing another person's visual perspective has been investigated using event-related potentials (ERPs; McCleery et al., 2011). Specifically, in this study, participants were asked to actively take the perspective of an agent or simply their own regarding the number of displayed disks on a wall. This paradigm could be interpreted as a 'Sally-Anne' false-belief type task; with false-belief trials being those where participants must take the perspective of another person and true-belief trials being those where participants take their own perspective. Under these conditions, using EEG and source analysis, McCleery and colleagues concluded that early in visual perspective processing the temporal and parietal cortices distinguish between self and other perspectives and then, later, the frontal cortex resolves conflicts between these representations during response selection. Importantly, however, McCleery and colleagues' study design was focused on visual perspective taking and did not examine brain regions involved in a classic false-belief processing task, which is key as there is still considerable debate regarding the overlap of processes subserving visual perspective taking and belief analysis (Aichhorn et al., 2006; David et al., 2008; Santiesteban et al., 2012) and developmental mastery of visual perspective-taking proceeds that of false-belief reasoning by at least a year (Flavell, 1988). Nonetheless, as we found in the current study also the parietal and temporal cortices (i.e., left STS and PC) to be sensitive to differentiating between implicit false- and true-belief processes it may be the case that early visual perspective processing stages overlap with early mental state processing stages.

In conclusion, we isolated brain areas subserving explicit general-belief reasoning that are also sensitive to implicit false- vs. true-belief processing. Importantly, the right TPJ that previous studies have indicated as a key region in eToM (Saxe and Kanwisher, 2003; Scholz et al., 2009) was significantly activated under general implicit belief processing demands, however failed to differentiate between implicit false- and true-belief trials – conditions that draw on belief processing resources to a different extent. Instead the left STS and PC appear to be the key nodes for false-belief processing in the absence of awareness. These findings provide physiological evidence for the proposal that a system exists for efficiently and implicitly processing the beliefs of others (Apperly and Butterfill, 2009; Clements and Perner, 1994).

Acknowledgments

DS was supported by a The University of Queensland PhD Centennial Scholarship, SIB by an Australian Research Council Discovery Grant and APD Fellowship (DP120103721; DP110100588); and PED by an Australian Research Council Discovery Grant and APD fellowship (DP0986387) and an Australian Research Council Future Fellowship (FT120100033).

Conflict of interest

The authors declare no competing financial interests.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.neuroimage.2014.07.014>.

References

- Abu-Akel, A., Shamay-Tsoory, S., 2011. Neuroanatomical and neurochemical bases of the theory of mind. *Neuropsychologia* 49, 2971–2984.
- Aichhorn, M., Perner, J., Kronbichler, M., Staffen, W., Ladurner, G., 2006. Do visual perspective tasks need theory of mind? *NeuroImage* 30, 1059–1068.
- Aichhorn, M., Perner, J., Weiss, B., Kronbichler, M., Staffen, W., Ladurner, G., 2009. Temporo-parietal junction activity in theory-of-mind tasks: falseness, beliefs, or attention. *J. Cogn. Neurosci.* 21, 1179–1192.
- Apperly, I.A., Butterfill, S., 2009. Do humans have two systems to track beliefs and belief-like states? *Psychol. Rev.* 116, 953–970.
- 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., Wheelwright, S., 2004. The empathy quotient: an investigation of adults with Asperger syndrome or high functioning autism, and normal sex differences. *J. Autism Dev. Disord.* 34, 163–175.
- 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, 5–17.
- Baron-Cohen, S., Richler, J., Bisarya, D., Gurunathan, N., Wheelwright, S., 2003. The systemizing quotient: an investigation of adults with Asperger syndrome or high-functioning autism, and normal sex differences. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 358, 361–374.
- Bayliss, A.P., Naughtin, C.K., Lipp, O.V., Kritikos, A., Dux, P.E., 2012. Make a lasting impression: the neural consequences of re-encountering people who emotive inappropriately. *Psychophysiology* 49, 1571–1578.
- Call, J., Tomasello, M., 2008. Does the chimpanzee have a theory of mind? 30 years later. *Trends Cogn. Sci.* 12, 187–192.
- Carrington, S., Bailey, A., 2009. Are there theory of mind regions in the brain? A review of the neuroimaging literature. *Hum. Brain Mapp.* 30, 2313–2335.
- Castelli, F., Happé, F., Frith, U., Frith, C., 2000. Movement and mind: a functional imaging study of perception and interpretation of complex intentional movement patterns. *NeuroImage* 12, 314–325.
- Clements, W.A., Perner, J., 1994. Implicit understanding of belief. *Cogn. Dev.* 9, 377–395.
- David, N., Aumann, C., Santos, N.S., Bewernick, B.H., Eickhoff, S.B., Newen, A., Shah, N.J., Fink, G.R., Vogetley, K., 2008. Differential involvement of the posterior temporal cortex in mentalizing but not perspective taking. *Soc. Cogn. Affect. Neurosci.* 3, 279–289.
- Dodell-Feder, D., Koster-Hale, J., Bedny, M., Saxe, R., 2011. fMRI item analysis in a theory of mind task. *NeuroImage* 55, 705–712.
- Döhnel, K., Schuwerk, T., Meinhardt, J., Sodian, B., Hajak, G., Sommer, M., 2012. Functional activity of the right temporo-parietal junction and of the medial prefrontal cortex associated with true and false belief reasoning. *NeuroImage* 60, 1652–1661.
- Flavell, J., 1988. The development of children's knowledge about the mind: from cognitive connections to mental representations. In: Astington, J., Harris, P., Olson, D. (Eds.), *Developing Theories of Mind*. Cambridge University Press, New York, pp. 244–270.
- Frischen, A., Bayliss, A., Tipper, S., 2007. Gaze cueing of attention: visual attention, social cognition and individual differences. *Psychol. Bull.* 133, 694–724.
- Frith, U., 2001. Mind blindness and the brain in autism. *Neuron* 32, 969–980.
- Frith, C.D., 2004. Schizophrenia and theory of mind. *Psychol. Med.* 34, 385–389.
- Gallagher, H., Happé, F., Brunswick, N., Fletcher, P., Frith, U., Frith, C., 2000. Reading the mind in cartoons and stories: an fMRI study of “theory of mind” in verbal and nonverbal tasks. *Neuropsychologia* 38, 11–21.
- Geangu, E., Gibson, A., Kaduk, K., Reid, V., 2012. The neural correlates of passive viewing sequences of true and false beliefs. *Soc. Cogn. Affect. Neurosci.* 8, 432–437.
- Gobbini, M., Koralek, A., Bryan, R., Montgomery, K., Haxby, J., 2007. Two takes on the social brain: a comparison of theory of mind tasks. *J. Cogn. Neurosci.* 19, 1803–1814.
- Hughes, C., 2011. *Social Understanding and Social Lives: From Toddlerhood Through to the Transition to School*. Psychology Press, Hove, East Sussex.
- Iacoboni, M., Lieberman, M.D., Knowlton, B.J., Molnar-Szakacs, I., Moritz, M., Throop, C.J., Fiske, A.P., 2004. Watching social interactions produces dorsomedial prefrontal and medial parietal BOLD fMRI signal increases compared to a resting baseline. *NeuroImage* 21, 1167–1173.
- Kovács, Á.M., Téglás, E., Endress, A.D., 2010. The social sense: susceptibility to others' beliefs in human infants and adults. *Science* 330, 1830–1834.
- Lissek, S., Peters, S., Fuchs, N., Witthaus, H., Nicolas, V., Tegenthoff, M., Juckel, G., Brüne, M., 2008. Cooperation and deception recruit different subsets of the theory-of-mind network. *PLoS ONE* 3, e2023.
- Low, J., Watts, J., 2013. Attributing false beliefs about object identity reveals a signature blind spot in humans' efficient mind-reading system. *Psychol. Sci.* 24, 305–311.
- Ma, N., Vandekerckhove, M., van Overwalle, F., Seurinck, R., Fias, W., 2010. Spontaneous and intentional trait inferences recruit a common mentalizing network to a different degree: spontaneous inferences activate only its core areas. *Soc. Neurosci.* 6, 123–138.
- Martin, A., Santos, L.R., 2014. The origins of belief representation: Monkeys fail to automatically represent others' beliefs. *Cognition* 130 (3), 300–308.
- Marois, R., Ivanoff, J., 2005. Capacity limits of information processing in the brain. *Trends Cogn. Sci.* 9, 296–305.
- McCleery, J., Surtees, A., Graham, K., Richards, J., Apperly, I.A., 2011. The neural and cognitive time-course of theory of mind. *J. Neurosci.* 31, 12849–12854.
- Perner, J., Lang, B., 1999. Development of theory of mind and executive control. *Trends Cogn. Sci.* 3, 337–344.
- Perner, J., Roessler, J., 2012. From infants' to children's appreciation of belief. *Trends Cogn. Sci.* 16, 519–525.
- Razza, R.A., Blair, C., 2009. Associations among false-belief understanding, executive function, and social competence: a longitudinal analysis. *J. Appl. Dev. Psychol.* 30, 332–343.
- Rice, K., Redcay, E., 2014. Spontaneous mentalizing captures variability in the cortical thickness of social brain regions. *Soc. Cogn. Affect. Neurosci.* <http://dx.doi.org/10.1093/scan/nsu081>, pii: nsu081 (Epub ahead of print).
- Rouder, J.N., Speckman, P.L., Sun, D., Morey, R.D., Iverson, G., 2009. Bayesian t tests for accepting and rejecting the null hypothesis. *Psychon. Bull. Rev.* 16, 225–237.
- Santesteban, I., Banissy, M.J., Catmur, C., Bird, G., 2012. Enhancing social ability by stimulating right temporoparietal junction. *Curr. Biol.* 22, 2274–2277.
- Saxe, R., Kanwisher, N., 2003. People thinking about thinking people: the role of the temporo-parietal junction in “theory of mind”. *NeuroImage* 19, 1835–1842.
- 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, 433–438.
- Schneider, D., Lam, R., Bayliss, A.P., Dux, P.E., 2012b. Cognitive load disrupts implicit theory-of-mind processing. *Psychol. Sci.* 23, 842–847.
- 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, 410–417.
- Schneider, D., Nott, Z.E., Dux, P.E., 2014. Task instructions and implicit theory of mind. *Cognition* 133, 43–47.
- Scholz, J., Triantafyllou, C., Whitfield-Gabrieli, S., Brown, E.N., Saxe, R., 2009. Distinct regions of right temporo-parietal junction are selective for theory of mind and exogenous attention. *PLoS ONE* 4, e4869.
- Sebastian, C.L., McCrory, E.J., Cecil, C.A., Lockwood, P.L., De Brito, S.A., Fontaine, N.M., Viding, E., 2012. Neural responses to affective and cognitive theory of mind in children with conduct problems and varying levels of callous-unemotional traits. *Arch. Gen. Psychiatry* 69, 814–822.
- Senju, A., Southgate, V., White, S., Frith, U., 2009. Mindblind eyes: an absence of spontaneous theory of mind in Asperger syndrome. *Science* 325, 883–885.
- Sommer, M., Döhnel, K., Sodian, B., Meinhardt, J., Thoermer, C., Hajak, G., 2007. Neural correlates of true and false belief reasoning. *NeuroImage* 35, 1378–1384.
- Southgate, V., Senju, A., Csibra, G., 2007. Action anticipation through attribution of false belief by 2-year-olds. *Psychol. Sci.* 18 (7), 587–592.
- Talairach, G., Tournoux, P., 1988. *Co-planar Stereotaxic Atlas of the Human Brain*. Thieme, New York (NY).
- Tamber-Rosenau, B.J., Dux, P.E., Tomblu, M.N., Asplund, C.L., Marois, R., 2013. Amodal processing in human prefrontal cortex. *J. Neurosci.* 33, 11573–11587.
- Todorov, A., Gobbini, M.I., Evans, K.K., Haxby, J.V., 2007. Spontaneous retrieval of affective person knowledge in face perception. *Neuropsychologia* 45, 163–173.
- van Overwalle, F., 2009. Social cognition and the brain: a meta-analysis. *Hum. Brain Mapp.* 30, 829–858.
- van Overwalle, F., Baetens, K., 2009. Understanding others' actions and goals by mirror and mentalizing systems: a meta-analysis. *NeuroImage* 15, 564–584.
- Wagner, D.D., Kelley, W.M., Heatherton, T.F., 2011. Individual differences in the spontaneous recruitment of brain regions supporting mental state understanding when viewing natural social scenes. *Cereb. Cortex* 21, 2788–2796.
- Wheatley, T., Milleville, S.C., Martin, A., 2007. Understanding animate agents distinct roles for the social network and mirror system. *Psychol. Sci.* 18, 469–474.
- Wimmer, H., Perner, J., 1983. Beliefs about beliefs: representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition* 13, 103–128.
- Young, L., Cushman, F., Hauser, M., Saxe, R., 2007. The neural basis of the interaction between theory of mind and moral judgment. *Proc. Natl. Acad. Sci. U. S. A.* 104, 8235–8240.