Accepted in Quarterly Journal of Experimental Psychology, Oct, 2019

Spontaneous attribution of false beliefs in adults examined using a signal detection approach

Tian Ye[1], Stephen M. Fleming[2], Antonia Hamilton[1]

ICN, UCL

[1] Institution of Cognitive Neuroscience, University College London, 17 Queen Square, WC1N 3AZ, London, UK

[2] Wellcome Trust Centre for Neuroimaging, University College London, 12 Queen Square, WC1N 3BG, London, UK

Abstract

Understanding other people have beliefs different from ours or different from reality is critical to social interaction. Previous studies suggest that healthy adults possess an implicit mentalising system, but alternative explanations for data from reaction time false belief tasks have also been given. In this study, we combined signal detection theory (SDT) with a false belief task. Since application of SDT allows us to separate perceptual sensitivity from criteria, we are able to investigate how another person's beliefs change the participant's perception of near-threshold stimuli. Participants (n=55) watched four different videos in which an actor saw (or didn't see) a Gabor cube hidden (or not hidden) behind an occluder. At the end of each video, the occluder vanished revealing a cube either with or without Gabor pattern, and participants needed to report whether they saw the Gabor pattern or not. A pre-registered analysis with classical statistics weakly suggests an effect of the actor's belief on participant's perceptions. An exploratory Bayesian analysis supports the idea that when the actor believed the cube was present, participants made slower and more liberal judgements. Though these data are not definitive, these current results indicate the value of new measures for understanding implicit false belief processing.

Keywords: theory-of-mind, false belief task, signal detection theory, Bayesian hierarchical model

Introduction

In everyday social interaction, people often need to track other's mental states (eg. beliefs) either to understand their behaviours or generate social responses (Carruthers, 2015; Frith, 1999; Frith & Frith, 2005; Koster-Hale & Saxe, 2013; Samson, 2013). Recently, attention has focused on the question of the cognitive mechanisms needed to engage in mentalising, and it has been proposed that people may be able to engage implicit mentalising processes without the need for language or executive processing. Such implicit mentalising would make mindreading easier and more efficient (Apperly, 2009; Onishi & Baillargeon, 2005; Low & Perner, 2012; Baillargeon, Scott & He, 2010; Clements & Perner, 1994). However, the question of whether this mechanism really exists is still hotly debated (Apperly, 2009; Apperly, Riggs, Simpson, Chiavarino, & Samson, 2006; Back & Apperly, 2010; Low & Perner, 2012; Heyes, 2014; Leslie, Friedman & German, 2004).

The strongest evidence for spontaneous and implicit processing of other people's mental states comes from studies of altercentric intrusion effects, that is, cases where the mental state of another person impacts on the ability to do an individual task. Such effects were first demonstrated in the case of visual perspective taking (VPT) (Samson, Apperly, Braithwaite, Andrews, & Bodley Scott, 2010). In this study, participants see a room with an avatar in the centre looking towards one wall, and between one and three red discs on the walls either in front of or behind the avatar. They must report how many discs they can see, and results showed that when participant's and the avatar's perspective are incongruent (e.g. when a participant can see three discs but the avatar can only see two), participants responded slower and make more errors. Since participants were not required to take the avatar's perspective, such altercentric intrusion suggests that participants spontaneously considered other's perspective. However, the interpretation of these results remains controversial, with similar effects found for arrows (Santiesteban, Catmur, Hopkins, & Bird, 2014), implying a non-social explanation of the effect. Conversely, if participants see an avatar who wears opaque or transparent goggles, the altercentric intrusion is found only for the transparent goggles which implies a genuine mentalising explanation of the effect (Furlanetto, Becchio, Samson, Apperly, 2015; but see Conway, Lee, Ojaghi, Catmur & Bird, 2017).

A recent study builds on these results by showing that altercentric intrusion can also improve performance on a perceptual task. Seow and Fleming (preprint) combined a level-1VPT task with a signal detection paradigm in which participants must detect a near-threshold Gabor pattern on a grey disc, presented on the wall of Samson's room. On some trials, the avatar could also see the disc and other times he could not, either because he is facing the wrong way or is wearing a blindfold. Results showed that when participants knew that the avatar could see the disc, their perceptual sensitivities (for judgements from the participant's perspective) were enhanced as the *d'* in this condition were higher than in which the avatar could not see (Seow & Fleming, preprint). This study indicates that

processing another person's visual experience can occur spontaneously and can influence low-level perceptual processes.

However, understanding what another person can see is only a starting point for full mentalising, a more important building block is how we interpret other's beliefs. Recent studies suggested that people could implicitly predict other's behaviours based on their current beliefs, even on those conflicting with reality. A widely-used task here is the false belief task, where a protagonist is sometimes misled to believe that the target is hidden in a wrong location (or not), and participants at some points are asked to predict where the protagonist would look for the target (Wimmer & Perner 1983). Recent studies adapted this task into a non-verbal version and found participants tended to look to where the protagonist believed the target was hidden (Southgate, Senju & Csibra, 2007; Senju, Southgate, White & Frith, 2009). Later Schneider and colleagues (Schneider, Lam, Bayliss & Dux, 2012; Schneider, Nott & Dux, 2014) replicated these results and claimed that anticipatory looking could be observed even after participants have performed the task for 1hr. Neuroimaging studies further suggested that implicit behaviours share large overlap in neural substrates with those involved in explicit mentalising, but the former might be more sensitive to belief contents (Naughtin, Horne, Schneider, Venini, York & Dux, 2017; Kovács, Kühn, Gergely, Csibra, Brass, 2014; Schneider, Slaughter, Becker & Dux, 2014).

Nonetheless, there are also studies questioned if anticipatory looking can be taken as a robust behavioural sign for implicit mentalising as such behaviours were found to be influenced by task instructions and cognitive load (Cane, Ferguson & Apperly, 2017; Schneider et al., 2012b; Schneider et al., 2014a). An alternative approach was taken by Kovács and colleagues (2010), who designed a novel reaction-time based false belief task, where implicit mentalising was examined when participants were doing a belief-irrelevant task. In this study, participants were required to judge as fast as possible if a ball was present or not when an occluder vanished. Before they made this judgement, they saw a video clip in which an agent saw (or didn't see) a ball hide behind the occluder (or not). This design resulted in the agent having a true or false belief about the ball location, which was not relevant to the participant's task of deciding if the ball was present. Despite this, participants did respond faster if the agent believed the ball was behind the occluder, which was taken as evidence for spontaneous, implicit theory of mind.

Again, there has been controversy surrounding this finding. A recent paper from Phillips et al. (2015) presented 13 experiments based on Kovács' method and claimed that the reaction time pattern in Kovács' study cannot only be interpreted in terms of belief-relevant factors. Instead, these researchers concluded that the results from Kovács' study were due to refractory movements rather than belief attribution. With these ambiguous results, it remains hard to draw a strong conclusion about whether adults spontaneously and implicitly ascribe beliefs to others or not. One key reason for

these confusing results is that there are many factors influencing performance on reaction time tasks, including interference from previous actions and the physical features of the stimuli (Herman & Kantowitz, 1970; Niemi & Näätänen, 1981). Therefore, it may be useful to find alternative tasks which do not rely solely on reaction time measures. In this respect, Seow and Fleming's study offers a new approach to investigate belief ascription in the context of signal detection theory (SDT).

SDT provides a framework for decomposing perceptual performance into two statistics: *d'*, the sensitivity of the system to the occurrence of the signal in signal-to-noise ratio units, and an overall bias to report signal presence, modelled as the criterion, *c*. This allows us to separate the perceptual decision-making criteria from individual differences in perceptual ability or reaction time (Rouder et al., 2005; Stanislaw & Todorov, 1999) . Seow's study revealed that knowing another person can see the same target as us can increase our perceptual sensitivity and change our decision criteria. In the present paper, we further predict that knowing the other person has seen something (or not) could also change perceptual sensitivity or decision criteria. Importantly, when either participants or an agent believed a target is present, participants should have a higher expectation that the target is present and should tend to report that they see the target. Hence, their perceptual sensitivities may increase while criteria may drop as participant's decisions becoming more liberal. In this way, SDT may provide a more sensitive measure of implicit mentalising.

In this current study, we thus applied a psychophysical approach that allows calculation of SDT measures to a version of Kovács' non-verbal false belief task. That is, we ask participants to detect a target feature (Gabor pattern) in contexts where another person either does or does not believe that the target feature is present. We then test if altercentric intrusion from another person's belief can influence participant's perceptual sensitivity (*d'*) or decision criterion (c). We term this paradigm a feature-detection false belief task. The methods and analysis of this study were pre-registered with the Open Science Framework (please see link https://osf.io/wxy2p/).

Method

Participants

Our target sample size was 40 participants. To achieve this, 66 participants were recruited to the experiment from the UCL-ICN participant database and were paid at a rate of £7.5 per hour. The study is under the ethical approval from the UCL Research Ethics Committee and conformed to the 1964 Declaration of Helsinki. Following the pre-registered data exclusion criteria, 26 participants were excluded from the original dataset, leaving a final sample of n=40 (16 males, age 24.6 ± 3.5).

The exclusion criteria are:

1. A participant scores 33 or above on the Autistic-Spectrum Quotient (AQ) test, which is suggested as a cut-off point between typical performance and autism (Baron-Cohen, Wheelwright, Skinner, Martin & Clubley, 2001);
2. A participant performs below 80% accuracy on the attention check trials.
3. Participant's overall accuracy on the belief tasks is below 55% or above 95%. Lower or higher accuracy will lead to extreme and unstable estimation of SDT measures such as *d'*.
4. A participant has 3 or more than 3 blocks in which biased responses are given. Chi-square test was used to examine whether participants' yes/no responses were significantly different from 50% yes and 50% no. With 24 trials in a block and a 0.05 significance level, if the number of 'Yes' responses was greater than 16 or less than 8, then participant's response profile for this block was significantly different from 50% 'Yes' and we considered this to be a biased block (Corder & Foreman, 2014). Participants with more than 3 biased blocks over the study were excluded.

Stimuli & Materials

Our stimuli were a set of 4 video clips in which a cube with a high-contrast grating on one side moves behind a barrier and then leaves the scene (or not) while an observer watches the cube move (or not) based on the logic of Kovács (2010) (Fig 1). We use the labels P (participant) and A (actor) to index who believes the moving cube is present (+) or absent (-). The event sequences were as follows:

1) P-A-: The cube moves behind the occluder and then moves out, all under the actor's gaze. The cube is last seen by the participant at 12.5s. Then the actor turns away from the table and starts to turn towards the table at 15s. The occluder disappears at 17s. In this condition, both the participant and actor hold a 'target absent' belief;

2) P-A+: The cube moves behind the occluder. Then the actor turns away from the table. While she is not looking, the cube moves out of the occluder and falls off the table (and out of the screen). The cube is last seen by the participant at 12.5s. Then the actor starts to turn towards the table at 15s. The occluder disappears at 17s. In this condition, the participant believes the target is absent but the actor has a false 'target present' belief;

3) P+A-: The cube moves behind the occluder and then moves out and falls off the table (and out of the screen), all under the actor's gaze. Then the actor turns away from the table. While she is not looking, a same cube moves in behind the occluder. The cube is last seen by the participant at 12.5s. Then the actor starts to turn towards the table at 15s. The occluder disappears at 17s. In this condition, the participant believes the target is present but the actor has a false 'target absent' belief;

4) P+A+: The cube moves behind the occluder. Then it moves out of the occluder to the right edge of the table and moves back behind the occluder. These all take place under the actor's gaze. The cube is last seen by the participant at 12.5s. Then the actor turns away from the table and starts to turn towards the table at 15s. The occluder disappears at 17s. In this condition, both the participant and the actor hold a 'target present' belief.

To create these video clips, we recorded the cube movements and the actor movements separately then superimposed them together using Adobe Premiere (Adobe, USA). In each video clip, the 3D environment and the cube's movement trajectories were generated in Vizard 5.7 (WorldViz, USA). All the movements took place against a dark grey [51, 51, 51] background. Each video clip involved a wooden table, a white occluder and a grey [128, 128, 128] cube with Gabor pattern on its front side. The Gabor stimuli on the moving cube contain sinusoidal gratings (contrast 0.25, spatial frequency of 6 cycles per degree and orientation 30 degrees.), superimposed with 10% white noise modulated by a Gaussian envelope.

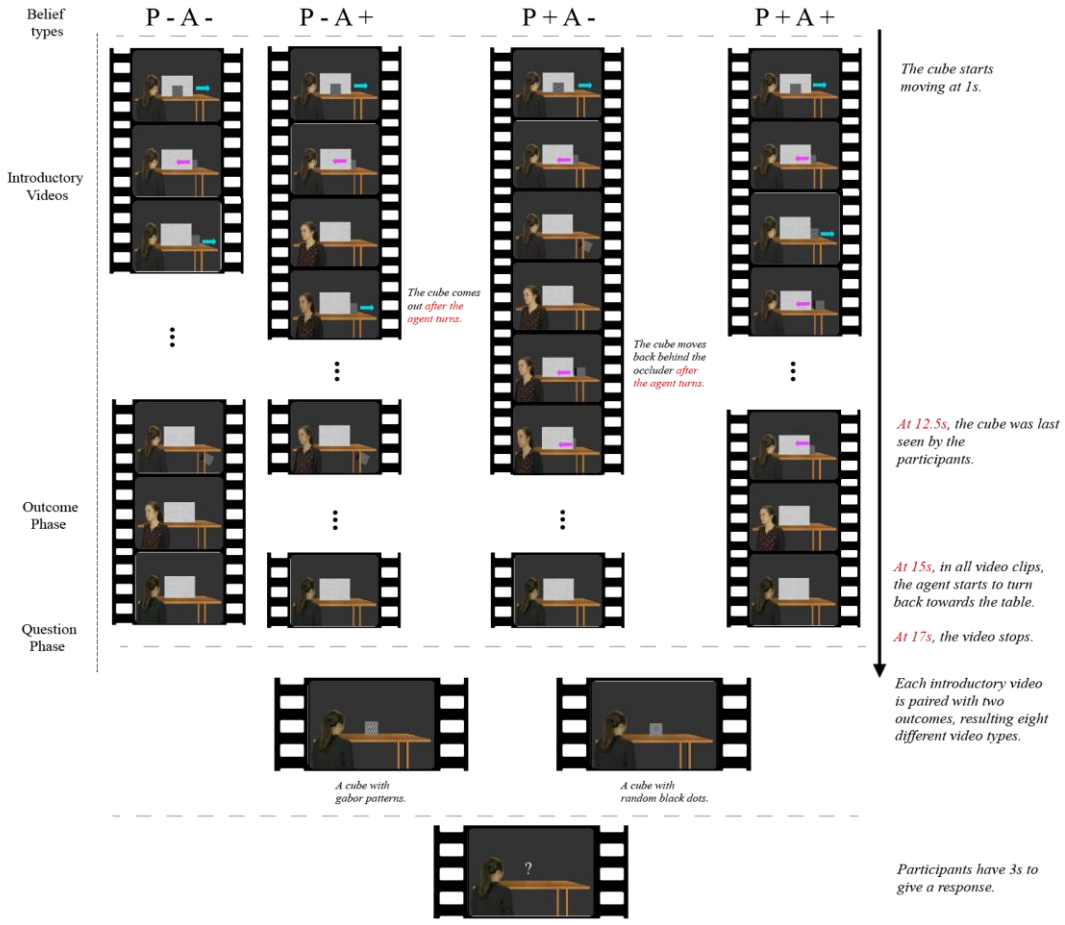

Fig 1. **A schematic illustration of all conditions in the belief task**. The four different starting events are illustrated as four downward film strips. Pink and blue arrows were not present in the videos but are included here to illustrate ball motion. All four clips ended with one of two outcome phases and then a single question screen.

After the moving cube stimuli were generated, movements of a single actor were recorded in front of a blue background. The actor was required to stand still, look or turn following the director's orders, avoiding excessive movements or facial expressions. To achieve a close match between the actor's and the cube's movements for each condition, while filming the actor's movements, the corresponding cube's video was played aside to allow the director to give voice commands at the right time. Then the matched actor's and cube's videos were merged in Adobe Premiere Pro CC 2017, using Chromakey to remove the blue background. The position and sizes of the actor were carefully matched across conditions and each video was cut to 17 seconds duration and saved without sound and with a resolution of 1024*768 pixels. The final frame in each video was the one before the occluder disappeared. We also saved a test frame from each video, which was the frame immediately after the occluder vanished and depicted the actor facing an empty table. These four pictures were then appended seamlessly at the end of their parent videos in the experimental script to provide a background image for the outcome phase and response phase in the task.

In the experimental trials, a lower-contrast test cube was imposed on these background images, placed in the centre of the table. The test cube was the same grey colour as the moving cube but has either the Gabor pattern or a white noise pattern on its front side. The same background image was used during the thresholding task, where we determined the appropriate threshold for each participant to be able to detect the Gabor pattern (see below). This ensures that the physical environment for Gabor feature detection is identical throughout the experiment.

To monitor the participant's attention throughout the task, 5 'attention check' videos were created. These video clips were made by editing the belief videos to stop early, before the occluder vanishes. A blue question box appears immediately after the attention video stops, asking a question either about the cube's current location (on/off the table) or about the actor's orientation (looking towards / away from the table), but never about the actor's mental states. In summary, the final stimuli include 4 belief videos matched to 4 background pictures and 5 attention check videos. All the stimuli are presented by Cogent 2000 and Cogent Graphics (http://www.vislab.ucl.ac.uk/cogent.php) in MATLAB (The MathWorks, Natick, MA).

Procedure

When each participant arrived for the study, they read an information sheet about the study and then signed a consent form to take part. Instructions were given verbally by the experimenter, then the participant completed two computer-based tasks: threshold testing and the belief task. Finally, they completed the AQ (Baron-Cohen et al. 2001) and a post-experiment questionnaire. A detailed description of each task is given below.

Threshold testing

This phase of the study aims to measure each participant's Gabor detection threshold for use in the second part of the study. Participants were asked to detect Gabor patterns of varying contrast in a 2 interval forced choice task. The Gabors were 73 by 73 pixels (210 by 210 mm on the screen and viewed from approx. 60cm), with contrast varying from 0 to 1. As mentioned previously, test stimuli were presented on a grey cube in a scene drawn from the last frame of the video clips.

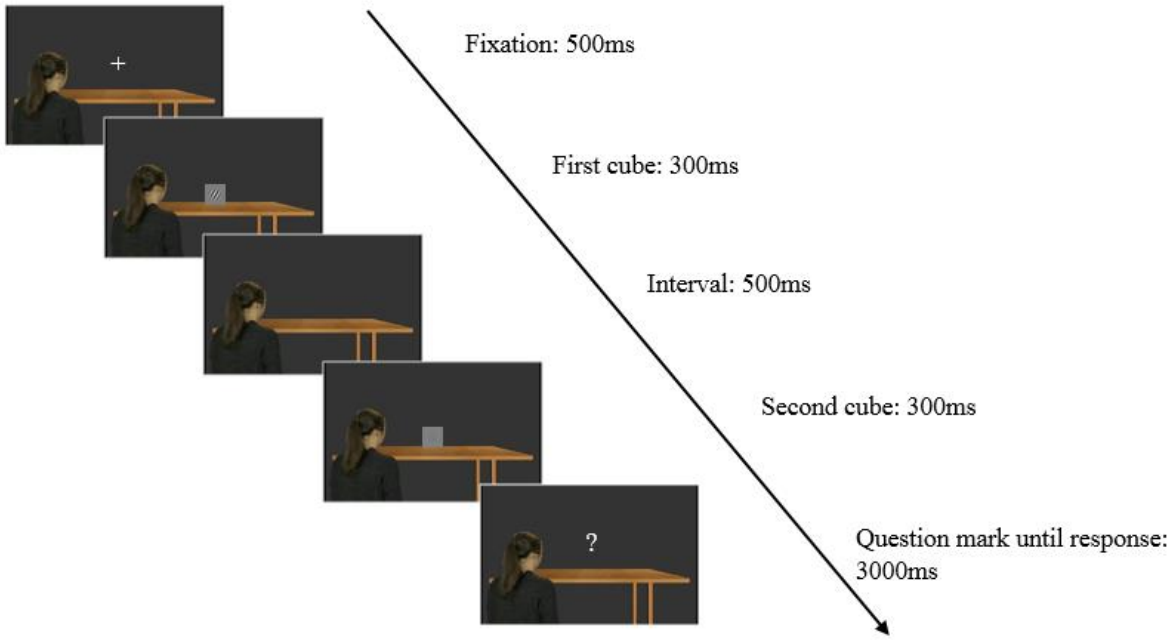

Fig 2. Procedure for the threshold testing stage

Participants were requested to detect which cube has the Gabor pattern, the first or the second. There were 10 practice trials before the formal threshold testing task, and participants were informed that the Gabor pattern would be difficult to detect during the formal session, so that they need to concentrate. In cases where they felt unable to identify the Gabor, they were instructed to make a best guess. There were 120 trials in the threshold testing session.

Each trial started with the onset of the 3D room picture, and then a white fixation appeared followed by the first grey cube, positioned in the middle on the table with a size of ~2 degrees of visual angle. The first cube appeared for 300ms then disappeared, after a short interval (500ms), the second cube appeared for another 300ms (Seow & Fleming, preprint). On half of the trials, the Gabor pattern was on the first cube and for the other half on the second, subtending ~1.84 degrees of visual angle to the centre of the front side of the cube. For the cube without the Gabor pattern, noise patches are drawn in the same area as the Gabor patch and consisted of uniformly random noise pixels at 10%

contrast, modulated by a Gaussian envelope. Right after the second cube disappeared a white question mark appeared and stayed on the screen until the participant made a key response. Accuracy was stressed but the participant was asked to give a key response in 3 seconds. The Matlab Quest Toolbox was used to adjust the contrast of the Gabor stimuli for each trial to identify a contrast value that leads to a performance level of 75% correct (Seow & Fleming, preprint; Watson & Pelli, 1983).

Belief task

After a short break, the participant performed the belief task. They were told that for each trial they were going to watch a short video, which involves an actor, a table, a white occluder and a grey cube with Gabor pattern on its front side. Each video began with the actor facing towards the table and the cube on the table in front of the occluder. After 1s, the cube started to move to a location behind the occluder. By manipulating the order of the events in the video (cube motion, actor motion), four belief conditions were created in a within-participant design (see Figure 1 and stimuli section). At the end of each video, right after the white occluder disappears and while the actor remains watching the table, a test cube appears in the middle on the table for 500ms. The test cube has the same colour (128, 128, 128) and size as the moving cube. On half trials, the test cube has a Gabor pattern on the front side and on the other half of the trials, the test cube has a white noise pattern on its front side. Unlike previous studies (Kovács et al., 2010), in the current study there was always a cube present at the end of the trial when the occluder was removed. This was because trials without a test cube cannot be analysed in our SDT approach, and so including these trials would add excess time to the study without adding useful data.

The contrast of the Gabor pattern on the test cube was determined by the previous thresholding test phase, and was selected so the participant should be able to achieve a 75% correct rate of response. The parameters of the noise dots were the same as the threshold testing stage. When the cube disappeared, a question mark showed up and stayed on the screen until response. Accuracy was stressed to the participants but they needed to respond in 3s. To enable SDT analysis, it is important that participants provide approximately 50% 'yes' and 50% 'no' responses in judging if the Gabor pattern is present on the test cube. To encourage this, participants were told explicitly at the beginning of the belief task that the proportions of Gabor pattern and noise patches were each 50/50. More importantly, if the script detected there was a response bias within a block (i.e. the number of YES responses is more than 16 or less than 8 in a block with 24 belief trials), the script would remind the participant during the next break to make approximately even numbers of yes/no judgements. Participants who persistently gave biased responses were excluded from the final analysis.

Overall, there were 4 belief conditions, each with 2 outcomes (either ending with a Gabor cube or with a noise cube), resulting in 8 types of trial. Each type of trial was repeated 24 times to allow enough trials for SDT analysis.

Attention check trials

To make sure that the participants were paying attention to the videos all the time rather than only to the last frame, we inserted another 48 trials as attention check trials. In the attention check trials, participants viewed shorter videos edited from the 4 belief conditions, and then answered a question either about the location of the cube or the orientation of the actor. There were 5 distinct versions of the attention check trials, each appearing randomly within each block. The accuracy of these questions was stressed to each participant and a low accuracy on these trials indicates participants were not paying enough attention to the videos so their data were excluded from further analysis.

In total, there were 240 trials for the whole belief task and 30 trials per block. Within each block, there were 6 trials from each belief condition (3 trials with a Gabor cube and 3 trials with a white noise cube) and 6 attention check trials. Belief trials and attention check trials were mixed randomly within each block. This manipulation helped participants to monitor their key responses and hence guarantee approximately equal numbers of 'yes/no' responses. Between blocks, there was a rest interval of at least 30s. The belief task took about 85min.

Post-experiment questionnaire

After completing all the computer-based tasks, the participants completed two questionnaires. The first one was related to the current task they performed, and mainly concerned their subjective report on their attention and knowledge about the actor, e.g. 'I paid a lot of attention to the actor,' and 'I wonder why she turns a lot.' Other questions are generally related to the participants' evaluation of the quality of the movie and the estimation of the task performance. The other questionnaire was the AQ (Baron-Cohen et al., 2001). The whole experiment took about 2hrs to be finished for each participant.

Data analysis

As specified in our pre-registration, we took the Gabor pattern detection data for the 40 valid participants and calculated *d'* and *criteria* values (Stanislaw & Todorov, 1999). The SDT *d'* was calculated by subtracting the z score for false alarm rate from the z score for hit rate (see formula 1) and criterion was calculated using the formula (2) listed below (Snodgrass & Corwin, 1988; Stanislaw & Todorov, 1999). These values were then submitted to a repeated measures ANOVA to test if the beliefs of the actor can bias the judgements of the participant, and thus could provide evidence for altercentric intrusion in this task.

$$d' = Z_{hit} - Z_{false\ alarm} \quad (1)$$

$$c = -0.5 * (Z_{hit} + Z_{false\ alarm}) \quad (2)$$

When preparing for this study, we did not anticipate having to exclude as many as 26 participants from our final data set[1], in 11 of these the exclusion was only because they gave unequal responses (e.g. all NO or all YES in a block) which cannot be used in a traditional SDT analysis. Therefore, we also decided to conduct exploratory analyses which could make use of more of the data which we collected, omitting items (3) and (4) in our original exclusion criteria and including all 55 participants who passed the attention check and scored within the typical range on the AQ. First, we analyzed reaction time data to test if there were differences related to either the participant's or the actor's belief, even though the task was not speeded. Here, we excluded trials where participants responded in less than 150ms, and those where participant's RT is below or above three standard deviations were also excluded from further analysis. Then the mean RT for each condition and each participant was submitted to a repeated-measures ANOVA.

Second, we adopted a Bayesian approach which is less influenced by extreme *d'* and *criteria* values. The BayesSDT model provides us a number of advantages:1) avoid edge corrections applied in standard SDT analyses that lead to biases in *d'* and *c* estimation when cell counts contain zeros; 2) allow group-level estimates of *d'* and *c* to mutually constrain extreme single-subject parameter estimates (Lee, 2008; Kruschke, 2010; Pleskac, Cesario, & Johnson, 2018). By performing this analysis we are able to obtain the posterior distributions of regression coefficients encoding the influence of our 2x2 factorial design on group-level sensitivity and criterion parameters. The SDT model was nested inside a regression model that encoded the two factors of our experimental design (participant beliefs × actor beliefs), plus their interaction. Thus each subject's *d'* parameter was specified as:

$$d' = d'_{base} + \beta_{pb} * I_{pb} + \beta_{ab} * I_{ab} + \beta_{i} * I_{pb} * I_{ab}$$

where I_{pb} and I_{ab} are indicator variables that are equal to 1 when participant/actor is holding a target present belief and -1 otherwise, and β_{pb} , β_{ab} and β_{I} are regression coefficients encoding the effects of participants beliefs, actor beliefs and their interaction, respectively. Uninformative (high variance) priors on these influences on *d'* were specified as follows (after JAGS convention, variances are written as precisions or the reciprocal of variance):

$$d'_{base} \sim N\ (0, 0.001)$$
$$\beta_{pb} \sim (0, 0.001)$$
$$\beta_{ab} \sim (0, 0.001)$$

[1] 11 participants failed on attention check criterion (item 2); 14 participants failed because of low overall accuracy (item 3); 11 participants failed because of giving too many biased responses (item 4). Some of the participants failed on more than one criteria.

$$\beta_i \sim (0, 0.001)$$

Analogous models and parameter estimation were also applied for the criterion, *c*. Markov Chain Monte Carlo (MCMC) implemented in JAGS in R was used to draw samples from the posterior distributions. When calling JAGS, we implemented 2000 adaptation steps, 5000 burn-in samples and 50000 effective samples. For each parameter we run 3 chains and convergence of all chains was assessed both visually and using Gelman & Rubin's potential scale-reduction statistic *R* for all parameters (Gelman, & Rubin, 1992). Our average *R* was 1.00 with a maximum value of 1.01, indicating good convergence.

The posterior distributions of each parameter returned by JAGS were then directly used for Bayesian inference. For each indicator variable, we calculated the probability that the coefficient is smaller than zero. For example, $P(\beta_{PB})$ for *d'* stands for the probability that regression coefficients for participant belief on *d'* is smaller than zero. To distinguish these probabilities from classical P-values, we denote them as P_θ.

By implementing Bayesian hierarchical analyses we are therefore able to use data from all 55 participants who pass our basic checks to evaluate whether the agent's belief impacts on participant's performance. In the 'Results' section, we clearly signpost our analysis as 'pre-registered' or 'exploratory' to facilitate understanding.

Results

Pre-registered analysis

Following the pre-registered document, the analyses below are based on data from 40 participants who met all our inclusion criteria.

d' Analysis

One critical prediction in this study is that an agent's belief may influence the participant's perceptual sensitivity. To test this hypothesis, *d'* for each participant from each belief condition was calculated by using formula (1) listed above (Snodgrass & Corwin, 1988; Stanislaw & Todorov, 1999), then we submitted *d'* from all four conditions to a repeated-measures analysis of variance (ANOVA) with Agent's belief (target present, target absent) and Participant's belief (target present, target absent) as within-subject factors. Before statistical analyses, the sphericity of the data was verified (Mauchly's test, all $p > .05$).

Results showed no main effect for agent's belief ($F(1, 39) = 0.034$, $p = 0.855$) and participant's belief ($F(1, 39) = 0.135$, $p = 0.715$). Also no interaction was found between these two factors ($F(1, 39) = 0.592$, $p = 0.446$) (Fig. 3A).

Criteria Analysis

The SDT criterion was calculated using the formula (2) listed above (Snodgrass & Corwin, 1988; Stanislaw & Todorov, 1999). Criteria for each participant from each condition were also submitted to a repeated-measurement ANOVA with Agent's belief (target present; target absent) and Participant's belief (target present, target absent) as within-participant factors. Interestingly, results revealed a trending main effect for Agent's belief (F (1, 39) = 3.985, p = 0.053, η^2 = 0.093).

Participant's criteria became less positive when the agent believed the cube was behind the occluder, indicating there is a weak trend to be more likely to report target presence in this condition. There was no main effect for participant's belief (F (1, 39) = 1.725, p = 0.197) and no interaction between Agent's belief and participant's belief (F (1, 39) = 0.109, p = 0.743) (Fig. 3B).

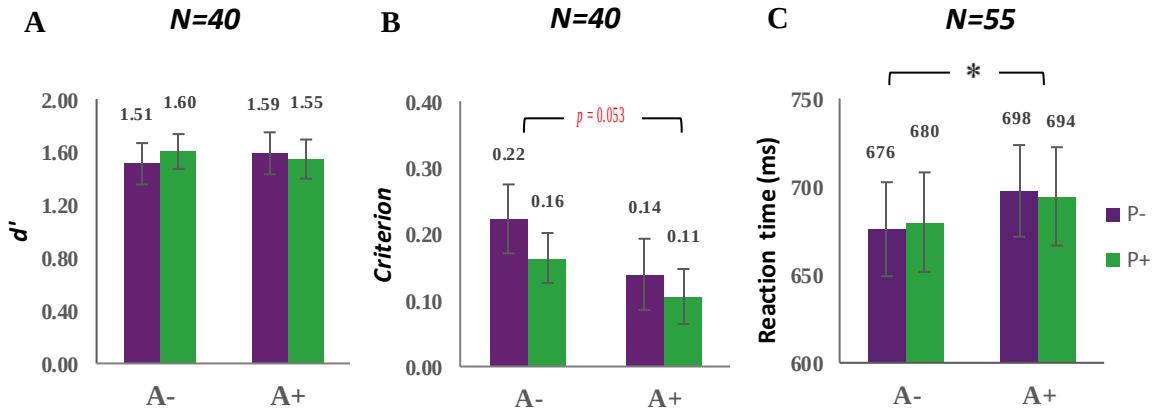

Fig. 3 Mean *d'*, *criterion* and reaction time between four belief conditions

Exploratory analysis

These analyses include all 55 participants who completed the task and passed the attention checks, and does not exclude those with biased responses.

RT analysis

RT data from all four conditions were submitted to a repeated-measures analysis (ANOVA). Results revealed a significant main effect for Agent's belief (F (1, 54) = 4.971, p = 0.030, η^2 = 0.084). Participants responded slower when the agent believed there was a target cube. No main effect was found for participant's belief (F (1, 54) = 0.003, p = 0.959) or for the interaction between Agent's belief and participant's belief (F (1, 54) = 0.319, p = 0.575) (Fig. 3C). It is worth noting that similar results were also found when analysing data from the 40 participants who met all criteria we set out in the pre-registration.

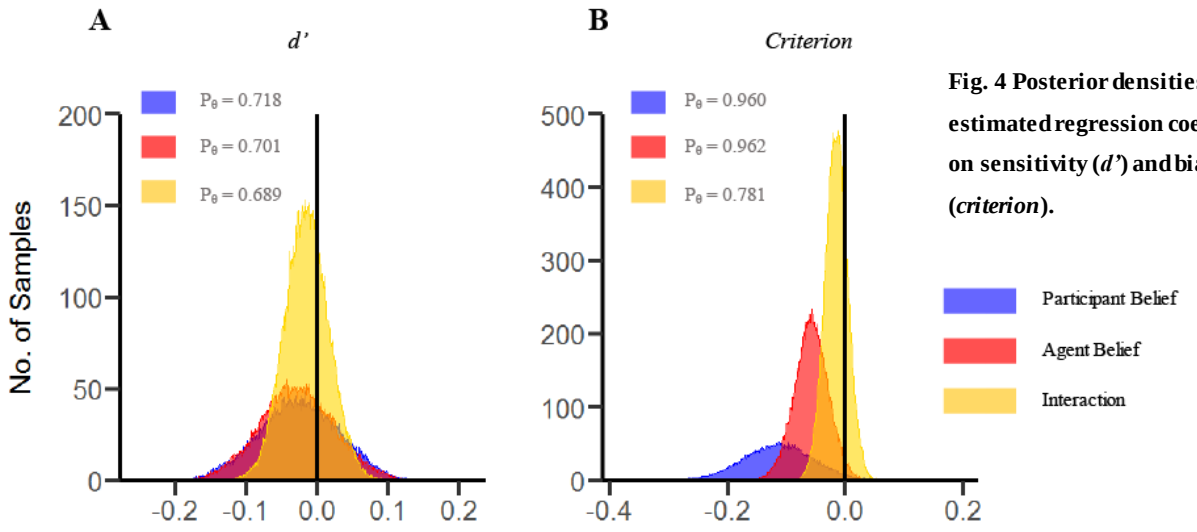

Fig. 4 Posterior densities of estimated regression coefficients on sensitivity (*d'*) and bias (*criterion*).

Bayesian SDT Analysis

The results from the Bayesian analysis engendered larger probabilities for impact both from participant beliefs and actor beliefs on the criteria but not on *d'*. Taking the *d'* analysis first (Fig. 4A). Bayesian analysis provides weaker support for positive influence (positive coefficients) either from participant beliefs (P_θ =0.718) or actor beliefs (P_θ =0.701) on performance with this measure, consistent with our classical pre-registered analysis. However, as Figure 4B shows, both factors are much likely to impact on criteria. When participants believe the cube is *present*, criteria are highly probable to decrease compare with when participants believe the cube is *absent* (P_θ =0.960) and similar trend is also observed when actor believes the cube is present (P_θ =0.962). But the probability for the interaction between these two factors is relatively smaller (P_θ =0.781).

Discussion

In this study, we designed a feature-detection false belief task to test whether typical adults spontaneously attribute beliefs to a mere co-observer. Signal detection theory (SDT) was applied to test if there is altercentric intrusion from another person's belief influencing a participant's perceptual process, with perceptual discrimination (*d'*) and a decision criterion (c) symbolizing different perceptual components. Our pre-registered analysis (n=40) hinted at a small effect of the actor's belief on the participant's decision criteria but this was not significant. Our exploratory Bayesian analysis revealed that both the participant's belief and the agent's belief can change the decision criteria. When either participants or the actor believed the target was present, participants turn to give more liberal and slower responses. We discuss these results first in terms of the interpretation of the decision criterion value, and then consider the robustness of implicit theory of mind effects.

Measuring implicit ToM with signal detection

In a signal detection framework, two measures of perceptual discrimination performance can be obtained. *d'* gives a measure of a participant's perceptual sensitivity to signal occurrence in signal-to-noise units, and the *criterion* reflects an internal cut-off line above which a participant will regard the internal evidence to be strong enough to represent a signal (Stanislaw & Todorov, 1999; Wyart, Nobre, & Summerfield, 2012). Changes to both *d'* and criteria can be induced by cues preceding the stimulus. Perceptual sensitivity can be increased by location cues which drive participants to focus attention on a specific location or a relevance cue which reduces uncertainty about the upcoming stimulus. Changes in criteria can also be induced by location or relevance cues, but also by strategic factors such as increasing the reward available for one stimulus interpretation or by creating prior expectations (Downing, 1988; Fleming, Whiteley, Hulme, Sahani, & Dolan, 2010; Summerfield & de Lange, 2014; Summerfield & Egner, 2009; Summerfield & Koechlin, 2010; Whiteley & Sahani, 2008; Wyart et al., 2012). Criterion shifts are thought to reflect changes either at a perceptual or decisional stage of processing (Witt, Taylor, Sugovic & Wixted, 2015).

These differences between *d'* and *criteria* and how they can be manipulated may help us to understand the results we find in the present study, in relation to previous work using SDT in a social task. Seow & Fleming used SDT in the context of a visual perspective-taking task and found that both participant's *d'* and criterion changed as a function of the avatar's visual perspective. Specifically, participants were more sensitive and more liberal when the avatar could also see the target object. This altercentric intrusion occurs despite the fact participants were not asked about the avatar's point of view on those trials. In contrast, our study shows that participants may change their criterion under the influence of another person's belief, in the absence of changes in *d'*. That is, participants may use a more liberal threshold to decide if the Gabor pattern is present but they do not improve in sensitivity.

The discrepancy between results of the two studies suggests that visual perspective and belief may have different mechanisms in influencing an individual's decision making. In the VPT tasks, attention may be a critical factor affecting performance (Catmur, Santiesteban, Conway, Heyes, & Bird, 2016), and many SDT studies have revealed an enhancement in attention can boost signal detection sensitivity (Wyart, Nobre & Summerfield, 2012; Downing, 1988; Summerfield & Egner, 2009). In contrast, spatial attention is less relevant in the current Gabor detection task because the pattern always appeared at the same spatial location. This may explain why *d'* did not change in the current experiment.

Both our study and that of Seow et al. find evidence that altercentric intrusion may change the criterion people use to make their decision, with participants using a more liberal criterion when another person can see the same stimulus (Seow & Fleming) or when the other believes the Gabor is present (this study). One possible reason is that processing another's mental states changes our

expectations about what signals are present. A number of SDT studies have shown that expectations and rewards can bias perceptual decision-making on both behavioural and neural levels (Summerfield & Koechlin, 2010; Fleming, Whiteley, Hulme, Sahani & Dolan, 2010; Wyarta, Nobrea & Summerfield, 2012; Summerfield & de Lange, 2014), that is, when participants expect a stimulus to be present (e.g. it was frequently present on previous trials), then they have a more liberal criterion to judge that it is present in the future. Such expectations may be built up from previous experience in the current task (e.g. seeing the cube move behind the barrier leads to an expectation that it will be there later, a basic object permanency effect). This is reflected in the finding that, in our exploratory Bayesian analysis, participants are more likely to judge the pattern is present when they themselves believe the cube is present. This occurs despite the explicit instruction that the prior probability of signal/noise at the test phase 50/50.

Our data also gives hints of an altercentric intrusion effect in affecting criterion shifts. That is, the other's belief (or maybe expectation) that the cube is present also leads participants to be more liberal in their judgements and to report that the pattern was present. In this context, the actor's expectation that the cube is present seems to bias participant's judgements. This is a spontaneous altercentric intrusion because participants were never asked to make any judgements about the other person's beliefs nor to consider what the other knew during the task.

Reaction time measures

We also recorded reaction time data in our task, even though participants were not instructed to respond fast. There was some evidence that considering the other's beliefs may cause an increase in processing time. Reaction times on trials where the actor had a 'target present' belief were significantly slower compared to when the actor had a target absent belief, despite this being an unspeeded task. As we strictly controlled the timing of key events (when the cube was last seen by participant, when the actor turns towards the table), a plausible account to explain the increase in RT is that participants spontaneously processed actor's belief. When the actor is holding a target present belief, the onset of the final testing cube may trigger participants to compare the current cube with the actor's belief, which might delay the participant's response.

Effects of the participant's own belief

In our canonical analysis on 40 participants, there was no effect of the participant's own belief on *d'*, criteria or on RT. However, our Bayesian analysis with a larger sample size (N=55) did reveal a highly probable influence on criteria from participant's belief (P = 0.960). From the distribution of Beta coefficients for participant's belief, we can see the influence from participant's belief on criteria has a large variation across all participants, and this variation might explain why from canonical analyses we could not see a significant influence from participant's own beliefs.

Surprisingly, we did not find any influence from the participant's own belief on reaction times, despite this belief affecting the decision criterion. Such a result may be caused by our task design, as in previous studies the target object can be present or absent at the end of the trial, but in our task the object (cube) was always present in the end of each trial. Only the presence or absence of the Gabor pattern on the cube was varied. We kept the cube present on every trial because we could not collect any SDT judgements on cube-absent trials, and including such trials would make the experiment too long. However, the fact that the cube was always present meant that participants would not be surprised by the physical presence of the cube, and this may have reduced the self-belief effects on reaction time.

It is useful to consider here how our task relates to the idea of inherent processing limits in the capacity for implicit ToM. In particular, Low & Watts (2013) proposed that the iToM system can track the location of an object (is the cube behind the screen or not?) but cannot track the identity of the object (is this the Gabor cube or plain cube?). Framing the task in these terms makes it even more surprising that participants showed any evidence of taking the agent's belief into account, because it suggests that the participants go beyond the basic limitations of the iToM system. However, if we reframe the task in terms of 'tracking the location of the Gabor patch' and consider all other objects as distractors, then it might be possible to explain performance within a more limited location tracking iToM system. Further work will be needed to distinguish these possibilities.

Overall, our data provide hints that participant's perceptual judgements and their reaction times can be influenced by the beliefs of an actor, even when the actor is irrelevant to the task. This is manifested particularly in a change in criterion, which may be related to an unfolding expectation of events in the trial. However, these effects were weak and often marginal. We consider this issue next.

Robustness in the study of implicit theory of mind

Many previous studies of implicit forms of theory of mind in adults have given rather ambiguous results, with some papers showing evidence that adults spontaneously consider the mental states of others (Schneider, Bayliss, Becker, & Dux, 2012; Nijhof, Brass, Bardi, & Wiersema, 2016; Dumontheil, Apperly, & Blakemore, 2010; Schneider, Slaughter, Becker, & Dux, 2014; Van Der Wel, Sebanz, & Knoblich, 2014; Schneider, Nott, & Dux, 2014), while other papers argue against a pure iToM account (Catmur et al., 2016; Conway, Lee, Ojaghi, Catmur, & Bird, 2017; Phillips et al., 2015). Our results are also ambiguous, because our pre-registered analysis does not show an effect of the agent's belief (at $p = 0.053$ which is often considered marginal) and our exploratory Bayesian analysis does show an effect of agent's belief. Previous research suggests there are two plausible accounts for such a weak effect. One explanation is related with to which level ToM is measured in different tasks. As described above, it may be harder for an implicit mentalising system to track object identity, compared to object location (Low & Watts, 2013). As our task could be taken as an object identity task, this might explain why the results are weak. Another possible explanation is related with

participants' social motivation to be involved in others mental states. Cane (2017) and Elekes (2016) both found in VPT research that participant's performance can be influenced by social factors such as monetary rewards, task instructions or the partner's task (Cane, Ferguson & Apperly, 2017; Elekes, Varga & Király, 2016). It's also worth noting that both of them found such effect with level-2 VPT tasks, which is quite similar to our study where object identity is tested. Following this vein, if we adopt more socially-engaged scenarios, participants may invest more effort to the task, so it's possible that we will observe a larger difference.

Given this pattern, we do not wish to draw definitive conclusions here about whether people do or do not engage in implicit theory of mind. Rather, we believe that our data suggests that it is worth continuing to study this area and looking for factors which may influence this process. That is, we should not abandon the domain of adult implicit theory of mind as a dead-end. Rather, our study builds on the work of Seow & Fleming (preprint) in highlighting new ways to measure altercentric intrusions in adults. We believe that future studies using these methods may be able to provide definitive evidence in favour of the implicit mentalising hypothesis, and that such studies are worth pursuing. In particular, the signal-detection approach may provide a more sensitive measure than reaction times, and could be employed in a wider range of contexts in future.

Second, we suggest that it may be important to consider a wider range of factors when designing stimuli for studies of implicit mentalising. That is, it is not clear if implicit mentalising can be induced by any humanoid stimulus which moves in a human-like way, or if factors such as the animacy of the agent and the motivation of the observer to engage can change whether or not implicit mentalising is engaged. It may be that implicit mentalizing is more likely to occur when a richer social context is built as participants will be more motivated to be engaged in these tasks, but previous studies have not examined this. Both our study and Seow's study suggest that the application of SDT can help us decompose the perceptual process, and it will be worth exploring with SDT to ask how exactly these social factors influence the interaction between others' and our own mental contents.

Conclusion

In this study, we incorporated signal detection theory into a false belief task intending to explore in a minimal social context whether a co-observer's belief could influence our perception. We reveal that perceptual sensitivity was not influenced by other's belief contents; however, we find that the decision making process was influenced by whether another person believes the target it present or not. When a co-observer believes a target is present, our decisions are slower and more liberal, but this influence from another's belief represents a weak effect. Compared with previous paradigms used in investigating implicit ToM, the method of the current study is more directly related to the belief representations in mind and the results provide more details on how other beliefs influences perceptual judgements. Future studies could introduce more social cues to seek to create stronger contextual effects of others' beliefs.

Acknowledgement

This research was supported by ERC Starting Grant: 313398-INTERACT. This work is also partly supported by the scholarship from China Scholarship Council (CSC). We appreciate Roser Cañigueral for her help by being the actor in our videos.

References

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

Apperly, I. A., Riggs, K. J., Simpson, A., Chiavarino, C., & Samson, D. (2006). Is belief reasoning automatic? *Psychological Science*, *17*(10), 841–844. https://doi.org/10.1111/j.1467-9280.2006.01791.x

Back, E., & Apperly, I. A. (2010). Two sources of evidence on the non-automaticity of true and false belief ascription. *Cognition*, *115*(1), 54–70. https://doi.org/10.1016/j.cognition.2009.11.008

Baillargeon, R., Scott, R. M., & He, Z. (2010). False-belief understanding in infants. *Trends in cognitive sciences*, *14*(3), 110-118.

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The Autism Spectrum Quotient : Evidence from Asperger syndrome/high functioning autism, males and females, scientists and mathematicians. *Journal of Autism and Developmental Disorders*, *31*(1), 5–17. https://doi.org/10.1023/A:1005653411471

Cane, J. E., Ferguson, H. J., & Apperly, I. A. (2017). Using perspective to resolve reference: The impact of cognitive load and motivation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(4), 591.

Carruthers, P. (2015). Mindreading in adults: Evaluating two-systems views. *Synthese*, 1–16. https://doi.org/10.1007/s11229-015-0792-3

Catmur, C., Santiesteban, I., Conway, J. R., Heyes, C., & Bird, G. (2016). Avatars and arrows in the brain. *NeuroImage*, *132*, 8–10. https://doi.org/10.1016/j.neuroimage.2016.02.021

Clements, W. A., & Perner, J. (1994). Implicit understanding of belief. *Cognitive development*, *9*(4), 377-395.

Conway, J. R., Lee, D., Ojaghi, M., Catmur, C., & Bird, G. (2017). Submentalizing or mentalizing in a Level 1 perspective-taking task: A cloak and goggles test. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(3), 454–465. https://doi.org/10.1037/xhp0000319

Corder, G. W., & Foreman, D. I. (2014). *Nonparametric statistics: A step-by-step approach*. John Wiley & Sons.

Downing, C. J. (1988). Expectancy and Visual-Spatial Attention: Effects on Perceptual Quality. *Journal of Experimental Psychology: Human Perception and Performance*, *14*(2), 188–202. https://doi.org/10.1037/0096-1523.14.2.188

Dumontheil, I., Apperly, I. A., & Blakemore, S. J. (2010). Online usage of theory of mind continues to develop in late adolescence. *Developmental Science*, *13*(2), 331–338. https://doi.org/10.1111/j.1467-7687.2009.00888.x

Elekes, F., Varga, M., & Király, I. (2016). Evidence for spontaneous level-2 perspective taking in adults. *Consciousness and Cognition*, *41*, 93-103.

Fleming, S. M., Whiteley, L., Hulme, O. J., Sahani, M., & Dolan, R. J. (2010). Effects of Category-Specific Costs on Neural Systems for Perceptual Decision-Making. *Journal of Neurophysiology*, *103*(6), 3238–3247. https://doi.org/10.1152/jn.01084.2009

Frith, C. D. (1999). Interacting Minds--A Biological Basis. *Science*, *286*(5445), 1692–1695. https://doi.org/10.1126/science.286.5445.1692

Frith, C., & Frith, U. (2005). Theory of mind. *Current Biology*, *15*(17), R644–R645. https://doi.org/10.1016/j.cub.2005.08.041

Furlanetto T, Becchio C, Samson D, Apperly I. (2015). Altercentric interference in level 1 visual perspective taking reflects the ascription of mental states, not submentalizing. *Geologia Tecnica e Ambientale*, *19*(3), 55–79. https://doi.org/10.1016/j.jfms.2010.11.013.

Gelman, A., Rubin, D. B., Gelman, A., & Rubin, D. B. (1992). Inference from Iterative Simulation Using Multiple Sequences Linked references are available on JSTOR for this article : Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science*, *7*(4), 457–472.

Herman, L. M., & Kantowitz, B. H. (1970). The psychological refractory period effect: Only half the double-stimulation story? *Psychological Bulletin*, (May). https://doi.org/10.1037/h0028357

Heyes, C. (2014). Submentalizing. *Perspectives on Psychological Science*, *9*(2), 131–143. https://doi.org/10.1177/1745691613518076

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

Kovács, A. M., Teglas, E., & Endress, A. D. (2010). The Social Sense: Susceptibility to Others' Beliefs in Human Infants and Adults. *Science*, *330*(6012), 1830–1834. https://doi.org/10.1126/science.1190792

Kovács, Á. M., Kühn, S., Gergely, G., Csibra, G., & Brass, M. (2014). Are all beliefs equal? Implicit belief attributions recruiting core brain regions of theory of mind. *PloS one*, *9*(9).

Kruschke, J. K. (2010). What to believe: Bayesian methods for data analysis. *Trends in Cognitive Sciences*, *14*(7), 293–300. https://doi.org/10.1016/j.tics.2010.05.001

Lee, M. D. (2008). BayesSDT: Software for Bayesian inference with signal detection theory. *Behavior Research Methods, 40*(2), 450–456. https://doi.org/10.3758/BRM.40.2.450

Leslie, A. M., Friedman, O., & German, T. P. (2004). Core mechanisms in 'theory of mind'. *Trends in Cognitive Sciences, 8*(12), 528–533. https://doi.org/10.1016/j.tics.2004.10.001

Low, J., & Perner, J. (2012). Implicit and explicit theory of MD: State of the art. *British Journal of Developmental Psychology, 30*(1), 1–13. https://doi.org/10.1111/j.2044-835X.2011.02074.x

Low, J., & Watts, J. (2013). Attributing false beliefs about object identity reveals a signature blind spot in humans' efficient mind-reading system. *Psychological Science, 24*(3), 305-311.

Naughtin, C. K., Horne, K., Schneider, D., Venini, D., York, A., & Dux, P. E. (2017). Do implicit and explicit belief processing share neural substrates?. *Human brain mapping*, *38*(9), 4760-4772.

Niemi, P., & Näätänen, R. (1981). Foreperiod and simple reaction time. *Psychological Bulletin*, *89*(1), 133–162. https://doi.org/10.1037/0033-2909.89.1.133

Nijhof, A. D., Brass, M., Bardi, L., & Wiersema, J. R. (2016a). Measuring mentalizing ability: A within-subject comparison between an explicit and implicit version of a ball detection task. *PLoS ONE, 11*(10), 1–15. https://doi.org/10.1371/journal.pone.0164373

Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs?. *Science, 308(5719),* 255-258.

Phillips, J., Ong, D. C., Surtees, a. D. R., Xin, Y., Williams, S., Saxe, R., & Frank, M. C. (2015). A Second Look at Automatic Theory of Mind: Reconsidering Kovács, Teglas, and Endress (2010). *Psychological Science*, 1–15. https://doi.org/10.1177/0956797614558717

Pleskac, T. J., Cesario, J., & Johnson, D. J. (2018). How race affects evidence accumulation during the decision to shoot. *Psychonomic Bulletin and Review, 25*(4), 1301–1330. https://doi.org/10.3758/s13423-017-1369-6

Rouder, J. N., Lu, J., Speckman, P., Sun, D., & Jiang, Y. (2005). A hierarchical model for estimating response time distributions. *Psychonomic Bulletin & Review, 12(2),* 195-223.

Samson, D. (2013). *Theory of Mind. Encyclopedia of Identity*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195376746.013.0059

Samson, D., Apperly, I. A., Braithwaite, J. J., Andrews, B. J., & Bodley Scott, S. E. (2010). Seeing it their Way: Evidence for rapid and involuntary computation of what other people see. *Journal of Experimental Psychology: Human Perception and Performance, 36*(5), 1255–1266. https://doi.org/10.1037/a0018729

Santiesteban, I., Catmur, C., Hopkins, S. C., & Bird, G. (2014). Avatars and arrows: Implicit mentalizing or domain-general processing? *Journal of Experimental Psychology: Human Perception and Performance, 40*(3), 929–937. https://doi.org/10.1037/a0035175

Schneider, D., Bayliss, A. P., Becker, S. I., & Dux, P. E. (2012). Eye movements reveal sustained implicit processing of others' mental states. *Journal of Experimental Psychology: General, 141*(3), 433–438. https://doi.org/10.1037/a0025458

Schneider, D., Lam, R., Bayliss, A. P., & Dux, P. E. (2012). Cognitive load disrupts implicit theory-of-mind processing. *Psychological science, 23*(8), 842-847.

Schneider, D., Nott, Z. E., & Dux, P. E. (2014). Task instructions and implicit theory of mind. *Cognition, 133*(1), 43–47. https://doi.org/10.1016/j.cognition.2014.05.016

Schneider, D., Slaughter, V. P., Becker, S. I., & Dux, P. E. (2014). Implicit false-belief processing in the human brain. *NeuroImage, 101*, 268–275. https://doi.org/10.1016/j.neuroimage.2014.07.014

Senju, A., Southgate, V., White, S., & Frith, U. (2009). Mindblind eyes: an absence of spontaneous theory of mind in Asperger syndrome. *Science, 325*(5942), 883-885.

Seow, T. X. F., & Fleming, S. M. (2018). Perceptual sensitivity is modulated by what others can see.

Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General, 117*(1), 34–50. https://doi.org/10.1037/0096-3445.117.1.34

Southgate, V., Senju, A., & Csibra, G. (2007). Action anticipation through attribution of false belief by 2-year-olds. *Psychological Science, 18*(7), 587-592.

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers, 31*(1), 137–149. https://doi.org/10.3758/BF03207704

Summerfield, C., & De Lange, F. P. (2014). Expectation in perceptual decision making: Neural and computational mechanisms. *Nature Reviews Neuroscience, 15*(11), 745–756. https://doi.org/10.1038/nrn3838

Summerfield, C., & Egner, T. (2009). Expectation (and attention) in visual cognition. *Trends in Cognitive Sciences, 13*(9), 403–409. https://doi.org/10.1016/j.tics.2009.06.003

Summerfield, C., & Koechlin, E. (2010). Economic Value Biases Uncertain Perceptual Choices in the Parietal and Prefrontal Cortices. *Frontiers in Human Neuroscience, 4*(November), 1–12. https://doi.org/10.3389/fnhum.2010.00208

Van Der Wel, R. P. R. D., Sebanz, N., & Knoblich, G. (2014). Do people automatically track others' beliefs? Evidence from a continuous measure. *Cognition*, *130*(1), 128–133. https://doi.org/10.1016/j.cognition.2013.10.004

Watson, A. B., & Pelli, D. G. (1983). Quest: A Bayesian adaptive psychometric method. *Perception & Psychophysics*, *33*(2), 113–120. https://doi.org/10.3758/BF03202828

Whiteley, L., & Sahani, M. (2008). Implicit knowledge of visual uncertainty guides decisions with asymmetric outcomes. *Journal of Vision*, *8*(3), 2. https://doi.org/10.1167/8.3.2

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, *13*(1), 103-128.

Witt, J. K., Taylor, J. E. T., Sugovic, M., & Wixted, J. T. (2015). Signal detection measures cannot distinguish perceptual biases from response biases. *Perception, 44(3)*, 289-300.

Wyart, V., Nobre, A. C., & Summerfield, C. (2012). Dissociable prior influences of signal probability and relevance on visual contrast sensitivity. *Proceedings of the National Academy of Sciences*, *109*(9), 3593–3598. https://doi.org/10.1073/pnas.1120118109