Journal of Experimental Psychology: General

Rethinking Egocentric Bias: A Computer Mouse-Tracking Study of Adult Belief Processing

Richard J. O'Connor, Andrew P. Lucas, and Kevin J. Riggs

Online First Publication, October 26, 2023. https://dx.doi.org/10.1037/xge0001485

AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Experimental Psychology: General

ISSN: 0096-3445

https://doi.org/10.1037/xge0001485

Rethinking Egocentric Bias: A Computer Mouse-Tracking Study of Adult Belief Processing

Richard J. O'Connor, Andrew P. Lucas, and Kevin J. Riggs
Department of Psychology, University of Hull

Several theories of belief processing assume that processing another's false belief requires overcoming an egocentric bias toward one's current knowledge. The current evidence in support of this claim, however, is limited. In order to investigate the presence of egocentric bias in adult belief processing, computer mouse tracking was used across three experiments to measure attraction toward response options reflecting one's current knowledge while reporting a false belief. Participants viewed scenarios in which an agent either had a true belief or a false belief about the location of a set of keys. Participants used a mouse to answer reality questions "where are the keys currently hidden?" and belief questions "where does she think the keys are?" Mouse-tracking measures indexing attraction toward response options during decision making were measured, along with time taken to make a response and accuracy. Experiment 1 found no evidence, in any measures, that participants showed a bias toward their own knowledge when reporting another's false belief. Experiment 2 investigated whether differences in event timings between true belief and false belief scenarios in Experiment 1 masked an egocentric bias. Experiment 3 investigated whether the lack of egocentric bias could be explained by participants prioritizing encoding the other's beliefs. Neither follow-up experiment found evidence supporting the presence of an egocentric bias. Overall, contrary to many theories of belief processing, our results suggest that adults are readily able to process other people's beliefs without having to overcome a default bias toward their own knowledge.

Public Significance Statement

Older children and adults are able to infer what other people believe. A number of theoretical accounts of this ability claim that to infer the false belief of another person (e.g., she thinks the keys are in the red cup, when I know that they are in the blue cup) requires overcoming an egocentric bias toward one's own knowledge. We tested this claim across three experiments and found no evidence that young adults are egocentrically biased when reporting another's false belief. Instead of best being described as "egocentric creatures," with a default bias toward our own knowledge, our results suggest that as adults we are just as sensitive to the differing beliefs of other people as we are to our own.

Keywords: theory of mind, false belief task, egocentric bias, automaticity, mouse tracking

Supplemental materials: https://doi.org/10.1037/xge0001485.supp

The egocentrism of young children on theory of mind tasks is well-known: Until around 4 years of age, when asked to report on the belief of an agent that differs from their own, children respond based on their own belief instead. It has also been claimed that adults, despite having a fully developed theory of mind, show evidence of being egocentrically biased, often assuming that other people share our perceptions and knowledge about the world, sometimes incorrectly (e.g., Apperly et al., 2010; Fischhoff, 1975; Ghrear et al., 2016; Keysar et al.,

Richard J. O'Connor https://orcid.org/0000-0003-4326-7955

The results reported here have been presented at meetings of the Experimental Psychology Society and at the Society for Philosophy and Psychology in 2022. This work was supported by the Economic Social Research Council (Grant ES/T012528/1, awarded to Richard O'Connor and Kevin J. Riggs). The authors are grateful to Sara Incera and Naohide Yamamoto for their advice regarding the time bin analyses reported in the online supplemental materials. The authors have no conflicts of interest to declare that are relevant to the content of this article. All study materials are available from the corresponding author upon request. All experiments were preregistered, and data and analysis code for all experiments are publicly available as follows: Experiment 1 preregistered: https://aspredicted.org/er5k9.pdf. Experiment 2 preregistered: https://aspredicted.org/9wi54.pdf. Experiment 3 preregistered: https://aspredicted.org/eq98n.pdf. Experiments 1–3 data and analysis script: https://researchbox.org/1024.

Richard J. O'Connor served in a supporting role for formal analysis. Andrew P. Lucas served as lead for formal analysis, investigation, and software and served in a supporting role for writing–original draft. Richard J. O'Connor and Kevin J. Riggs contributed equally to conceptualization, writing–original draft, and funding acquisition.

Correspondence concerning this article should be addressed to Richard J. O'Connor, Department of Psychology, University of Hull, Cottingham Road, Hull HU6 7RX, United Kingdom. Email: richard.oconnor@hull.ac.uk

2003). This appears to be the case with adults even on very simple theory of mind tasks, such as those modeled on the classic developmental false belief task (e.g., Birch & Bloom, 2007; Sommerville et al., 2013). The view that humans are "egocentric creatures," demonstrating egocentric bias across the lifespan, is one held by many theory of mind researchers and is prominent in many cognitive theories of belief processing (e.g., Epley et al., 2004; Farrar & Ostojić, 2018; Friedman & Leslie, 2004; Leslie & Polizzi, 1998; Tamir & Mitchell, 2013; L. Wang & Leslie, 2016).

However, the data and interpretations supporting egocentric bias in adults deserve reappraisal. First, clarity is needed on what exactly explains "egocentric" errors. Second, many of the studies claiming to show egocentric bias in adult belief processing do not provide sufficient evidence, either due to nonreplicable results, available alternative explanations, or methods that do not directly measure egocentric bias. In the studies reported here we use a novel method—computer mouse tracking—that is capable of directly measuring egocentric bias while participants report another agent's false belief.

Causes of Egocentric Bias in Adults

Although adults are capable of inferring that another agent has a perspective or a belief that is different from their own, they often fail to make use of this ability in ongoing tasks, such as when inferring the referent of another's statement (e.g., in the widely used director task: Apperly et al., 2010; Keysar et al., 2003; Lin et al., 2010; J. J. Wang et al., 2020). Such behavior suggests a type of egocentrism: Adults may simply fail to appreciate the relevance of another person's mental state to the task at hand.

This suggestion can be contrasted with the claim that even when we deliberately employ theory of mind reasoning, we still need to overcome an egocentric bias (L. Wang & Leslie, 2016). For example, "anchor and adjustment" accounts (Epley et al., 2004; Farrar & Ostojić, 2018; Tamir & Mitchell, 2013) suggest that when adopting another person's perspective adults first use their own mental state as an anchor and then adjust away from this egocentric position. Where adults fail to make sufficient adjustment away from this default attribution of one's own mental state, this results in an egocentric bias. In a different account of belief processing, Leslie and colleagues propose an innate Theory-of-Mind module that automatically generates candidate beliefs (e.g., Friedman & Leslie, 2004; Leslie & Polizzi, 1998; L. Wang & Leslie, 2016). Among the possible beliefs generated is one's own salient true belief, which requires suppression when attributing a false belief to another.

There are, however, some accounts of belief processing do not assume an egocentric bias. For example, Rubio-Fernández and colleagues (e.g., Rubio-Fernández, 2017; Rubio-Fernández et al., 2019) have suggested that adults rapidly and accurately attribute beliefs to others without one's own knowledge necessarily interfering in this process (see also Kampis & Southgate, 2020). It is worth noting that both Rubio-Fernández and Leslie claim there is automaticity of belief processing. Thus, the question of whether there is an egocentric bias in adult belief processing is not, therefore, one of automaticity. The key difference between accounts such as that of Leslie and Rubio-Fernández—and the focus of the present study—is whether one's own knowledge necessarily interferes when processing another's false belief. If overcoming egocentric bias is inherent to false belief processing, then when adults deliberately employ theory of mind reasoning to report another's false belief, their responses should show evidence of egocentric bias. Yet, despite the centrality of egocentric bias to several theories of belief attribution, it is not clear that this prediction is supported by current data.

Measuring Egocentric Bias in Adults: The Sandbox Task and the Curse of Knowledge Task

Two highly cited and widely used tasks have been claimed to demonstrate evidence of egocentric bias: the sandbox task (Sommerville et al., 2013) and the curse of knowledge task (Birch & Bloom, 2007). Within each of these, adults are presented with an unexpected transfer false belief task. An agent first hides an object in one location and then in the agent's absence the object is moved to a second location. In the sandbox task, participants are asked to indicate where the protagonist will look for the object in a continuous array of undifferentiated space, following a delay of up to 45 s between participants seeing the events and giving their response. In the curse of knowledge task, participants judge the probability as to where the agent will look for the object first across four containers.

In the sandbox task, adults display a tendency to indicate that the agent will search at a location that is closer to where the participant knows the object is currently located, relative to a control memory question in which participants indicate where the object was first hidden (Bernstein, 2021; Bernstein et al., 2017; Coburn et al., 2015; Sommerville et al., 2013). In the curse of knowledge task, adults attribute a higher probability that the agent will search for the object in the container where the participant knows the object is hidden, relative to a control "ignorance" condition in which the participant knows that the object has been moved, but does not know its final location (Birch & Bloom, 2007; Converse et al., 2008; Farrar & Ostojić, 2018; Sassenrath et al., 2013; Todd et al., 2011).

The reliability of both of these tasks, however, has been called into question. Ryskin and Brown-Schmidt (2014) and Samuel (2023) both failed to replicate the original findings of the curse of knowledge task across several highly powered experiments. The original sandbox task effects have also failed to replicate (Samuel et al., 2018). Furthermore, it is not clear that performance on that task is best explained by egocentric bias. In Bernstein (2021), adults display a bimodal pattern of responding on both experimental (belief) and control (memory) trials (see Figure A1 of Bernstein, 2021). Responses cluster around both the original location of the object (the correct answer for both belief and memory questions) and the current location of the object (the incorrect answer for both questions). Responses close to the incorrect answer are typically not excluded—they are treated as extremely biased attempts to give the correct answer, rather than incorrect answers.[1] Participants, however, might be more likely to get the belief question wrong compared to the memory question for a reason other than egocentric bias: The belief question simply requires more information to be recalled (both where the object was first hidden and whether the agent saw it being moved) and an additional inference (predicting the agent's behavior). The memory question only requires the first piece of

[1] Removing any participant who gave an incorrect answer (14 in. or more away from the correct answer) and reanalysing the data from Bernstein (2021) in fact results in egocentric bias scores in the adult (18–64 year) group to no longer be significant (mean difference between memory and belief trials = .46 in.), $t(171) = 1.44$, $p = .15$, suggesting that incorrect answers drive the egocentric bias effect on this task.

information to be recalled and no further inference. These differing demands, and not egocentric bias, could result in more incorrect answers to the belief question.

Reaction Time (RT) Studies and Egocentric Bias

RT studies, however, have reliably produced evidence that is at least minimally consistent with an egocentric bias. Across a range of tasks adults take longer to report, or validate a statement about, an agent's false belief compared to statements about: current reality; one's own belief; or the agent's true belief. This is consistent with the claim that processing a false belief involves overcoming a bias toward one's own knowledge.

However, it is critical to note that a longer RT when reporting a false belief compared to either current reality, one's own belief or an agent's true belief is not sufficient evidence for egocentric bias. First, comparing responses to questions about reality or one's own belief to responses to questions about another agent's false belief confounds the presence of egocentric bias with a difference in question type (e.g., belief/reality). It has been found that even in situations where participants report on an agent's true belief—a response with no requirement to overcome an egocentric bias—participants are slower to report the belief of the agent compared to reality (Back & Apperly, 2010) or their own belief (Bradford et al., 2018; Meert et al., 2017). Second, comparing responses to questions about the belief of an agent where that belief is either true or false confounds egocentric bias with representational congruency. Scenarios in which an agent has a false belief necessarily involve an incongruency between reality and the agent's belief, but not so if the agent has a true belief. Simply processing two incongruent representations could increase RTs, independent of whether adults are egocentrically biased toward their own representation of events (Rubio-Fernández, 2017). Indeed, participants are often slower to report current reality or their own belief—responses that do not involve overcoming an egocentric bias—following scenarios in which an agent has a false belief compared to those in which the agent has a true belief (e.g., Back & Apperly, 2010; Schuwerk, Döhnel, et al., 2014; Schuwerk, Schecklmann, et al., 2014).

The minimum evidence for the presence of egocentric bias is therefore an interaction between question type (belief/reality or the participant's own belief) and belief scenario (true/false). In such a 2×2 design, egocentric bias should only be present when reporting the agent's belief on a false belief scenario. There is no egocentric bias to overcome when reporting the true belief of another (their belief will be consistent with one's own) or when reporting current reality (or one's own belief) in either true or false belief scenarios.

Interestingly, however, in the RT studies reported above that have used such a design, researchers have typically not found evidence for such an interaction. Back and Apperly (2010) report only a main effect of question (other belief > reality questions) and a main effect of scenario (false belief > true belief), with no evidence for an interaction in either RTs or error rates. Schuwerk and colleagues report only a main effect of scenario (false belief > true belief), and again no interaction, in either RTs or error rates (Schuwerk, Döhnel et al., 2014; Schuwerk, Schecklmann et al., 2014). Meert et al. (2017) did report an interaction between question and scenario in the expected direction for an egocentric bias effect, but this interaction was only present for error rates.

One possibility for why the tasks above do not reliably demonstrate evidence for egocentric bias is that while they measure accuracy and RT, they do not directly measure attraction toward one's own knowledge while the response is being made (see also Dale et al., 2018). Paradigms that do measure attraction while participants report a false belief may be better placed to reveal egocentric bias.

Measures of Attraction During Belief Processing: Eye Tracking and Mouse Tracking

A number of eye-tracking studies have measured where adults look while they watch unexpected-transfer false belief sequences to see if where an agent believes an object is attracts more attention relative to the object's actual location. Several studies demonstrate that prior to answering a question about the agent's belief, or in the absence of any questions or instructions to focus on the agent's belief, participants display anticipatory looking toward the location where the agent believes the object to be (e.g., Ferguson et al., 2010, 2015; Low & Watts, 2013; Rubio-Fernández, 2017; Schneider et al., 2012; Senju et al., 2009). While the aim of these studies is typically not to investigate egocentric bias per se, but rather to investigate the extent to which belief processing occurs spontaneously or even automatically, some authors have interpreted these anticipatory looking data as evidence against the presence of egocentric bias. Such anticipatory looking has been interpreted as evidence that adults can rapidly and accurately infer the belief of another agent without any default attentional bias toward where one knows the object to be (Rubio-Fernández, 2017).

There are, however, problems with such an interpretation of eye-tracking data, and with the use of eye tracking more broadly as a measure of egocentric bias. First, there have been a number of failures to replicate key anticipatory-looking findings in adults (Burnside et al., 2018; Kulke et al., 2018; Poulin-Dubois et al., 2018). Indeed, some studies (e.g., Ferguson & Breheny, 2012; L. Wang & Leslie, 2016) have claimed that looking behavior during false belief sequences reveals an egocentric bias, with the actual location of the object competing for attention as much as the "false belief" location.

Second, a fundamental challenge against using eye-tracking data to investigate egocentric bias is that participants' eye movements may not reflect the cognitive processes involved in explicitly reporting the agent's false belief. Young children's looking behavior demonstrates a sensitivity to the false beliefs of others, despite these children displaying "egocentric" errors when asked to verbally report that agent's false belief (Baillargeon et al., 2010). Furthermore, adults take longer and make more errors to report a false belief compared to a true belief on the exact same tasks where their eye movements show correct anticipation of another's false belief (Ferguson et al., 2015; Rubio-Fernández, 2017). Such dissociations between participants' eye movements and the responses they give when reporting another's false belief suggest that eye-tracking captures an implicit belief processing that is not subject to inhibitory demands (Baillargeon et al., 2010), and may reflect a different type of social processing from that involved in reporting another's belief (e.g., Apperly & Butterfill, 2009; Low & Watts, 2013).

What is needed, therefore, is a measure of attraction to different response options within the response itself when participants report on another's false belief. Computer mouse tracking provides such a measure. Mouse tracking has been widely used in social cognition research to investigate response competition across a diverse range of domains (for reviews, see Freeman, 2018; Stillman et al., 2018), including competition between egocentric and allocentric perspectives during spatial perspective-taking tasks (Duran et al., 2011; Galati et al., 2019). In a typical mouse-tracking task, participants respond by moving a computer mouse from the bottom center of the screen to select one of two responses located in the top-left and top-right corners of the screen. The path the mouse takes as the participant responds is recorded, along with temporal data such as the time taken to initiate mouse movement and then make the response. The greater the attraction of the nonselected response on a given trial, the more the participant's mouse path deviates toward it as they make their decision. This greater attraction is indexed by measures reflecting the spatial characteristics of the mouse path, such as the maximum absolute deviation (MAD) toward the alternative response from the ideal straight line between the start position and the final position of the mouse (Freeman, 2018; Stillman et al., 2018).

The spatial measurements afforded by mouse tracking appear to be more sensitive to differences in competition between response options during a decision, relative to RT: It is not uncommon in mouse-tracking studies for differences in spatial characteristics of the mouse path to emerge between experimental conditions even in the absence of any such differences in RT (e.g., Freeman & Ambady, 2011; Golshaie & Incera, 2021; Wojnowicz et al., 2009; see discussion in Stillman et al., 2018). Mouse tracking may therefore be able to provide a direct and sensitive measure of attraction toward a response option reflecting one's own knowledge whilst reporting another agent's false belief. We are aware of only one study that used mouse tracking to investigate belief processing (van der Wel et al., 2014). That study investigated whether the presence of another agent with a different belief regarding the location of an object biased participants' responses when reporting the actual location of the object: It did not measure whether one's own knowledge biases responding when reporting another agent's belief. The present study therefore provides the first use of mouse tracking to investigate the processes underlying explicit belief processing.

The Present Study

In three experiments, we used mouse tracking to directly measure egocentric bias during adult belief processing. Participants watched short videos in which a woman watched a man hide a set of keys in one of two cups: a red cup or a blue cup. The man then transferred the keys to the other cup, either in the presence of the woman ("true belief scenario") or once she had left the room ("false belief scenario"). At the end of each video, participants received one of two experimental questions: a "reality" question ("Where are the keys currently hidden?") or a "belief" question ("Where does she think the keys are?"). Participants answered by moving the mouse to select one of two answers—"red" or "blue"—located in the top-left and top-right corners of the screen.

Critically, it is only on the belief question following false belief scenarios that the incorrect answer to the question reflects participant's own knowledge of where the keys are currently hidden. On reality questions following either scenario, and on belief questions following the true belief scenarios, the correct answer is congruent with the participant's own knowledge. Therefore, if participant's show an attraction to response options reflecting their own knowledge—an egocentric bias—while answering questions about another agent's belief, this would manifest as greater curvature toward the incorrect answer (e.g., a greater MAD value) specifically when answering belief questions following false belief scenarios. Accounts proposing an egocentric bias in belief processing (e.g., Epley et al., 2004; Farrar & Ostojić, 2018; Friedman & Leslie, 2004; Leslie & Polizzi, 1998; Tamir & Mitchell, 2013; L. Wang & Leslie, 2016) would therefore predict a question by scenario interaction effect upon measurements of mouse-path curvature, specifically a greater difference between belief and reality questions following false belief scenarios compared to true belief scenarios. Accounts denying the presence of such an egocentric bias (e.g., Rubio-Fernández, 2017) would not predict such an interaction.

Experiment 1

Participants watched videos of a man transfer keys from one colored cup to another. A woman either saw this transfer (true belief scenarios) or was out of the room when the transfer took place (false belief scenarios) but returned before the end of the video. At the end of each video participants were asked to respond to either a belief question, a reality question or a filler question (see below). Participants were not told that the task was about belief processing or that they should deliberately pay attention to where the woman thinks the keys are.

Method

Transparency and Openness

The preregistration of the design and analysis plan for this experiment can be found at https://aspredicted.org/er5k9.pdf. All data and analysis code are available at https://researchbox.org/1024. Research materials are available from the authors upon request.

Participants

Following the recommendations of Brysbaert (2019), a minimum sample size of 75 was determined to have at least 80% power to detect an interaction, with appropriate post hoc tests, on a 2×2 repeated-measures design where one factor has an effect of $d_z =$ 0.50 at one level of the second factor, but is absent at the second level.[2]

Eighty-three participants completed the study, with a final sample of 76 participants after exclusions (see below). Participants were asked to report their gender, choosing from the following response options: "prefer not to say," "nonbinary or other identity," "male," and "female." Of the final sample, 24 identified as male and 52 identified as female. Participants' mean age was 27.4 years ($SD = 9.6$ years). Sixty-nine were ordinarily right handed and seven were ordinarily left handed, though all confirmed they normally used the computer mouse with their right hand. All had normal or corrected to normal vision, and all were fluent speakers of English. Participants were all recruited from the local university community. All participants were tested in accordance with procedures approved by the Faculty of Health Sciences Ethics Committee at the University of

[2] That is, as would be predicted by an egocentric bias: as egocentric bias is present only on belief questions on the false belief scenarios, egocentric bias should generate a difference between belief and reality questions on false belief scenarios, but not on true belief scenarios. The use of $d_z = 0.50$ is based on $d_{av} = 0.40$, the recommended estimate of the smallest effect size of interest by Brysbaert (2019), when $r \geq .68$. Calculations of ICC2 from pilot data indicated that all measurements were above that minimum level of reliability, both across and within conditions.

Hull, and were compensated for their time with course credits, or a £12 Amazon voucher.

Stimuli and Apparatus

The experiment was run in MouseTracker (Freeman & Ambady, 2010), on a Windows 10 desktop. Stimuli were presented on a 24″ monitor at 1,920 × 1,080 pixels resolution, with a 60 Hz refresh rate. Participants gave their responses with a generic USB mouse, with mouse *x*- and *y*-coordinates polled at a minimum of 60 Hz. As is recommended for mouse tracking (e.g., Kieslich et al., 2020), the mouse cursor speed was reduced, set at 50% of the native Windows maximum speed with mouse acceleration disabled.

Video stimuli were created by filming untrained actors, one male, one female, perform unexpected transfer theory of mind sequences, modeled on those used by Back and Apperly (2010). In every video, the two actors sat across from each other at a table on which there were a set of keys, a red cup and a blue cup. The actors wore either a red or a blue polo shirt. In each video (see Figure 1 for a schematic of the video sequences) the male actor first put the keys in one cup and then the female actor left the room. In false belief scenarios, the male actor then moved the keys to the second cup, and then the female actor returned. In true-belief scenarios, the female actor returned and then the male actor moved the keys to the second cup.

Eight different videos were created per scenario (false belief/true belief) by manipulating which cup the keys were hid in first (red or blue), which cup was closest to the female actor (red or blue), and which color shirt the female actor wore (red or blue—the male actor always wore the opposite color shirt), generating 16 videos overall. In order to maximize the number of trials participants completed, videos were played at twice their original speed. All videos were 20 s in duration and were shown at 960 × 540 pixels at 25 fps.

Design

The experiment used a 2 × 2 repeated-measures design in which, on experimental trials, participants answered either "belief" or "reality" questions after watching either a true belief scenario or a false belief scenario. Belief questions always asked participants to report where the female actor believed the keys were located, while reality questions asked participants to report where the keys were currently hidden. In order to prevent participants from anticipating the response they needed to give while watching the video, and to reduce the likelihood that they would strategically prioritize tracking the female actor's belief, participants also answered filler questions (for similar approaches, see Back & Apperly, 2010; Schuwerk, Döhnel et al., 2014; Schuwerk, Schecklmann et al., 2014). Questions on these filler trials asked participants to report either the color of the shirt of the male or female actor (red or blue) or the color of the cup closest to the male or female actor. On these trials, the colors of the actors' shirts and the cups on the table changed to white for the final frame, so that participants were required to respond from memory.

Across 128 trials, delivered in two blocks of 64 trials, participants answered 32 belief questions (16 following false belief scenarios, 16 following true belief scenarios), 32 reality questions (16 following false belief scenarios, 16 following true belief scenarios) and 64 filler questions (32 following false belief scenarios, 32 following true belief scenarios). All 16 videos appeared four times in each block, once with a belief question, once with a reality question and twice with a filler question, with randomized presentation of videos and questions within each block. For each video, the correct answer was always 50% "red" and 50% "blue," and thus unpredictable for participants.

Procedure

Each trial began by presenting participants with a start button, sized 384 × 108 pixels and positioned flush with the bottom of the screen, centered around the midline, and two response boxes, sized 480 × 270 pixels and positioned flush with the top left and top right corners of the screen. One response box was colored red and the other box colored blue: which box was blue and which box was red was held constant for a participant across the experiment and counterbalanced across participants.

After participants clicked the start button a central fixation cross was presented for 750 ms, followed by a video. At the end of the video the final frame remained on-screen. Below the final frame, a question appeared in 18pt font (see Figure 1). The two experimental questions were "Where does she think the keys are?" and "Where are the keys currently hidden?" (belief and reality, respectively), controlling for similar character length (34 and 36 characters, respectively). Filler trial questions were: "Which cup is nearest to her/him?" and "What shirt is he/she wearing?" with the gender of the target actor changing from trial to trial. When the question appeared, the mouse cursor position was reset to the bottom center of the screen. Participants gave their answer on each trial by clicking on one of the two response boxes with the mouse.

Participants were instructed to start moving the mouse as soon as the question appeared on-screen, even if they were not yet certain of the answer they were going to give. This instruction is common and best practice in mouse-tracking methodology, as it is needed to ensure that participants are moving the mouse while they are making their decision (Freeman & Ambady, 2010; Kieslich et al., 2020). If participants did not initiate a mouse movement within 1,000 ms a warning message appeared on-screen, after they had given their response, instructing them that they needed to start moving earlier, even if they were not yet entirely sure of their response. The use of a 1,000 ms initiation time-cut was based on RTs for responses to belief and reality questions on previous studies with similar designs and stimuli (Apperly et al., 2006; Back & Apperly, 2010; Cohen et al., 2015; Cohen & German, 2009, 2010). In those studies, typical condition means were greater than 2,000ms, and no condition mean fell below 1,200 ms. It was therefore considered unlikely that in the present study participants would be able read the question and make their decision within 1,000 ms. Movements initiated within 1,000 ms were therefore likely to reflect ongoing decision making.

Participants first completed six practice trials, containing one of each question type (belief, reality, and filler questions about the shirt or cup closest to the male or female actor). Whether the belief question in the practice followed a false belief or true belief scenario was counterbalanced across participants. Participants received feedback on their accuracy, along with any initiation time (IT) warnings as described above. If a participant made errors or triggered the IT

Figure 1
Video Scenario Sequences and Question Screen

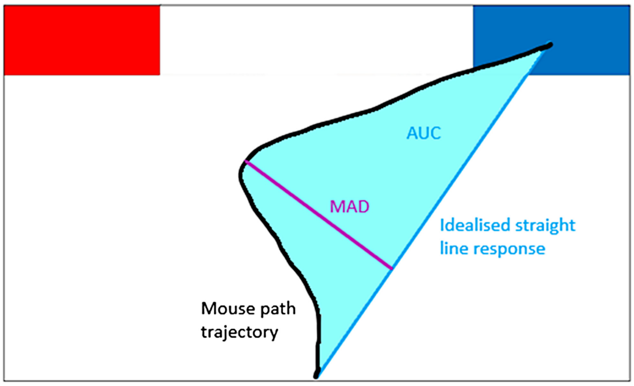
Note. Panel A: Event structure of video sequences used for false belief and true belief scenarios. The hiding location of the keys was counterbalanced between videos (see text for details). Panel B: Question screen (here using the belief question) as displayed to participants at the end of each trial. The location of the red (top left) and blue (top right) response boxes was counterbalanced between participants (see text for details). See the online article for the color version of this figure.

warning more than three times on the six practice trials they completed the practice a second time. Following the practice, participants then completed the main experiment. Participants did not receive feedback on their accuracy in the main experiment, but still received the IT warnings where necessary. Participants were given the opportunity for a break after the first block of 64 trials. In total, participants took around 60 min to complete the full study.

Measures, Data Processing, and Exclusions

Response accuracy was recorded on each trial. The degree of attraction toward the nonselected response option on each trial was measured by the following three spatial characteristics of the mouse path (see Figure 2): "MAD," measuring the distance between the point of maximum deviation of the response trajectory away from the idealized straight line response; "area under curve" (AUC), measuring the area bounded by the trial's response trajectory and an idealized straight line response from start to target; and "path length" (PL), measuring the total length of the response trajectory, from trial start to trial end. Greater attraction to the unselected response is associated with greater values of MAD, AUC, and PL (Freeman & Ambady, 2010). While these different spatial measures are not independent of each other, we preregistered multiple spatial measures in order to check that any effects are not limited to the specific measure used. All three measures are widely used in mouse tracking research, with no current agreed best practice on which specific measure to report. For a discussion of this issue from our prior mouse-tracking research, in which we replicated the phonological cohort effect, see Lucas (2021).The following temporal characteristics of each response were also calculated: IT, measuring the time taken to initiate the first movement of the mouse, measured from the point at which the question appeared; and "response time" (RT), measuring the time taken

Figure 2
Spatial Measures Used to Record Attraction to the Incorrect Answer



Note. Here, the left response option is the incorrect answer. MAD measures the distance between the point of maximum deviation of the response trajectory away from the idealized straight-line response (here, the length of the magenta/dark gray line). AUC measures the area bounded by the trial's response trajectory and the idealized straight-line response from start to target (here, the turquoise/light gray shaded area). PL measures the total length of the response trajectory, from trial start to trial end (here, the total length of the black line). MAD = maximum absolute deviation; AUC = area under curve; PL = path length. See the online article for the color version of this figure.

from the point of this first movement to the time when participants clicked on a response box.[3]

Trial-by-trial mouse-tracking data were processed in the following preregistered steps, using the mousetrap package (Wulff et al., 2021) for R, Version 4.2.2 (R Core Team, 2022; see also analysis script at https://researchbox.org/1024). First, responses to filler trials and all trials with an incorrect response were removed. Mouse-path trajectories were spatially normalized (start points aligned to (0, 0) and end points to (8, 9)—appropriate for a 16 × 9 widescreen display), and remapped to the right-hand side of the screen. ITs were calculated, and any trial with an IT greater than 1,000 ms were excluded. The period prior to participants initiating movement, but subsequent to them clicking the start button (i.e., when the mouse was therefore stationary at the (0, 0) position), was removed from the trajectories data. Trajectories were time normalized (each trajectory divided into 101 bins of equal temporal duration) on a trial-by-trial basis, and MAD, AUC, PL, and RT calculated per trial. Any trial with a measurement of MAD, AUC, PL, IT,[4] or RT that was above or below three *SD* of the participant overall mean across experimental conditions was then excluded. Finally, participants ($n = 7$) with fewer than eight trials (i.e., 50%) remaining in any condition were excluded. For the final analyzed sample, 4,331 trials remained (out of a maximum of 4,864). On average, each participant contributed 56.9 trials, out of a maximum of 64 (89.0%).

Results

For ease of exposition, we report analyses for error rates, MAD, IT, and RT only. The data for these can be seen in Figures 3, 5, 6, and 7. Mouse-path trajectories, illustrating the overall shape of participant responses, can be seen in Figure 4. Other preregistered analyses (of AUC, PL, and analyses of *x*- and *y*-coordinates over time) are reported in the online supplemental materials. Our decision to report MAD specifically was guided by evidence that MAD may be more sensitive to competition effects involving an initial attraction to the incorrect response option (Maldonado et al., 2019), as predicted by the presence of egocentric bias. In any case, the analyses in the online supplemental materials report the same pattern of effects as those given below.

For temporal and spatial measures, 2 × 2 repeated-measures analyses of variance (ANOVAs) with the factors question (belief, reality) and scenario (false belief, true belief) were computed, followed where appropriate by post hoc paired-samples *t*-tests with Bonferroni correction ($\alpha = .0125$). Error data, however, were analyzed using mixed-effect logistic regression, instead of the preregistered ANOVA, given concerns raised by two anonymous reviewers regarding the suitability of ANOVA for such data (Jaeger, 2008). Models contained the fixed effects of question, scenario and their interaction, and a maximal random effects structure of random participant intercepts and random participant slopes for both main fixed effects and their interaction. We simplified the random effects only when the maximal structure failed to converge or resulted in singularity errors. Models were simplified by first removing random effect correlations and then the random slope with the lowest estimated variance, as necessary until the model converged (Barr et al., 2013). Final model equations can be found in the analysis script. Note that we did not include random item effects when building the maximal random effects structure. "Items" in our study correspond to the specific videos presented to participants. Other than the differences according to the fixed effect of scenario (false belief or true belief scenario videos), videos only otherwise varied on incidental factors such as the color of the shirt worn by each actor, as outlined above. There was no a priori reason to believe that these factors would result in interitem variability in the number of errors participants made on experimental questions. We confirmed this by attempting to add random item intercepts to the final models: These models failed to converge due to variability being too low for the item random effect. In all models we used deviation coding (−1, 1) for the effects of question and scenario, such that positive coefficients for question reflect more errors on belief compared to reality questions, and positive coefficients for scenario reflect more errors in false belief compared to true belief scenarios. Models were built using the glmer function in the lme4 package in R (Bates et al., 2015).

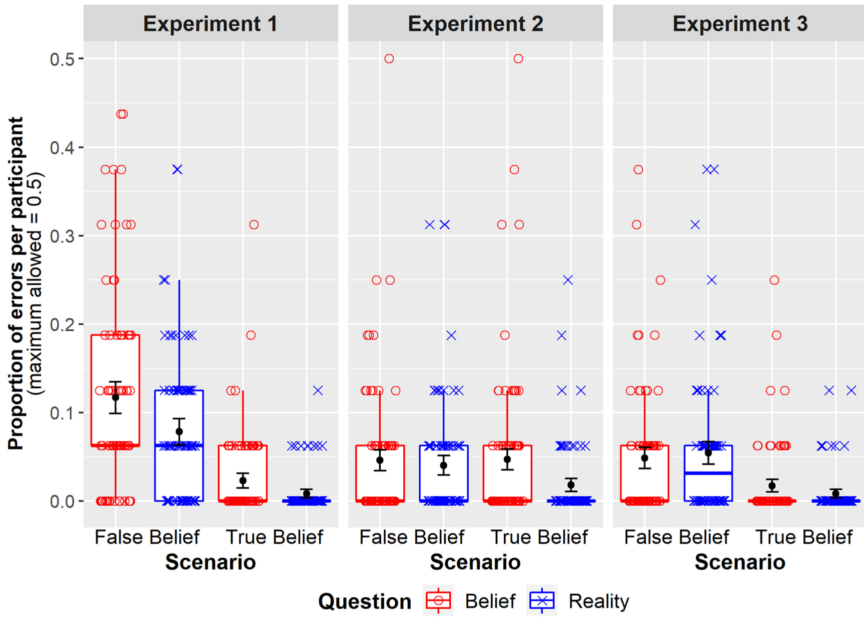
In all analyses, evidence for egocentric bias would have been indicated by a question × scenario interaction, driven by more erroneous responding, longer RTs, and/or greater curvature toward the incorrect answer (a greater MAD value) on belief questions compared to reality questions, specifically on the false belief scenario. Where we failed to find evidence for such an interaction, we further used Bayesian directional *t*-tests comparing mean differences (Morey & Rouder, 2011) to assess the extent to which our data supported the null hypothesis over the specific directional interaction

[3] It should be noted that the measurement of RT reported here is therefore different from typical reaction time measures. The overall reaction time on a trial is equal to the sum of IT and RT. While is it debatable how comparable mouse-tracking reaction times are to standard button-press reaction times (see Stillman et al., 2018), descriptive statistics and analysis of this overall reaction time measure, or "time to target," are reported in the online supplemental materials.

[4] It should be noted that due to an oversight in the preregistration, we did not originally list IT among the measures on which we would exclude based on three *SD* above/below the participant mean.

Figure 3

Error Data in Experiments 1–3



Note. Boxes indicate 25th–75th percentile (IQR), with the middle line indicating the median. The whiskers indicate ±1.5 IQR. Each ring/cross is an individual participant's mean. The black dot indicates the condition mean, with error bars representing 95% CI (i.e., 1.96 *SE*). IQR = interquartile range; CI = confidence interval. See the online article for the color version of this figure.

predicted by egocentric bias (i.e., that the difference between belief and reality questions would be greater in the false belief scenarios compared to true belief scenarios). For errors, in order to maintain comparability with the above logistic regression analyses, the mean number of errors per condition were first transformed to a logistic scale using the empirical logit transformation.

Bayes Factors reported below were calculated using the Bayes Factor package (Morey & Rouder, 2022) in R using a default one-sided (positive values only) Cauchy distributed prior centered at zero with a width of .707. This corresponds to a prior probability of 50% that the effect size lies between 0 and .707, with positive values reflecting a greater difference between belief and reality questions in false belief scenarios compared to true belief scenarios. We also provide sensitivity analyses in the online supplemental materials demonstrating that the calculated Bayes factors are robust: In all cases, our data continue to support the null hypothesis across a range of reasonable prior widths. We also calculated Bayes factors using a nonparametric Bayesian Wilcoxon signed-rank test in JASP (JASP Team, 2020); in all cases results from the parametric and nonparametric analyses converged. While these Bayesian analyses were not preregistered, their inclusion, as suggested by an anonymous reviewer, allow an appropriate interpretation of any null hypothesis significance testing failures to find evidence for egocentric bias.

Errors

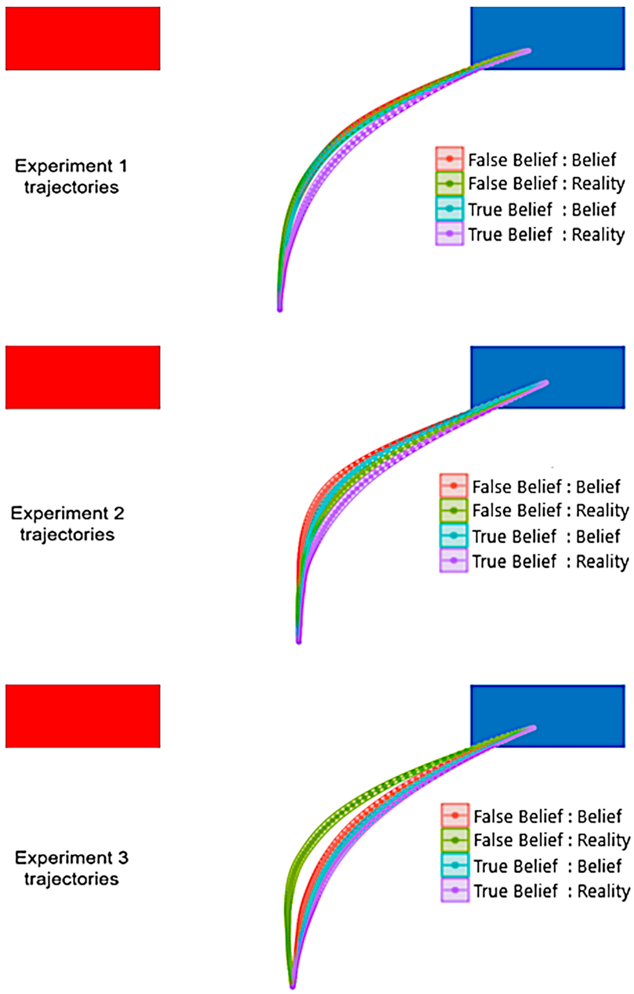
Mixed-model logistic regression revealed a significant main effect of question (coefficient = 0.40, $SE = 0.12$, $z = 3.38$, $p < .001$) and scenario (coefficient = 1.31, $SE = 0.19$, $z = 7.04$, $p < .001$), but no significant interaction (coefficient = −0.23, $SE = 0.13$, $z = -1.72$, $p = .085$).[5] More errors were made on belief questions compared to reality questions, and more errors were made on false belief scenarios compared to true belief scenarios. Bayesian analysis of empirical logit transformed mean errors found that the null hypothesis was 4.0 times more likely than the alternative hypothesis (that the difference in errors between belief and reality questions was greater in the false belief scenarios than in the true belief scenarios).

Spatial Data—MAD

ANOVA showed significant main effects of question, $F(1, 75) = 7.85$, $p = .006$, $\eta g^2 = .008$, and scenario, $F(1, 75) = 16.1$, $p < .001$, $\eta^2_g = .011$, as well as a significant interaction, $F(1, 75) = 4.56$, $p = .036$, $\eta^2_g = .003$. Post hoc *t*-tests revealed that this interaction was driven by a significant difference between belief and reality responses in the true belief scenarios, $t(75) = 5.11$, $p < .001$, $d = 0.586$, with greater attraction toward the incorrect answer on belief compared to reality questions. This pattern was not observed in the false belief scenarios, where there was no significant difference between question types, $t(75) = 0.650$, $p = .518$, $d = 0.075$ (see Figure 5). Therefore, the interaction was in the opposite direction to that predicted by the presence of an egocentric bias. Bayesian analysis of the mean differences found that the null hypothesis

[5] This matched the results of the preregistered ANOVA performed on untransformed mean errors, which also found main effects of question and scenario but no significant interaction.

Figure 4
Mouse Path Trajectories in Experiments 1–3



Note. All trajectories are remapped such that the correct answer is on the right. In the case of the false belief : belief condition, the incorrect answer (the left response option) reflects participants' own knowledge of the location of the keys. For all other conditions, the correct answer (the right response option) reflects participants' own knowledge of the location of the keys. Ribbons indicate ± 1 *SE* in the *x*-position around a particular time bin. Lines indicate the condition mean trajectory, collapsed across subjects. See the online article for the color version of the figure.

was 23.7 times more likely than the alternative hypothesis (that the difference in MAD between belief and reality questions was greater in the false belief scenarios than in the true belief scenarios).

As is recommended practice for analyzing mouse-tracking data (Freeman & Dale, 2013), Hartigans' dip statistic (HDS), which tests for bimodality of responses, was computed for each spatial measure per condition. There was no evidence that the data were bimodally distributed (all $ps \geq .432$).

Temporal Data

ANOVA for IT showed no significant main effect of question, $F(1, 75) = 2.54$, $p = .115$, $\eta_g^2 < .001$, or scenario, $F(1, 75) = 2.06$, $p = .156$, $\eta g^2 < .001$. There was no significant interaction, $F(1, 75) = 0.451$, $p = .504$, $\eta_g^2 < .001$.

For RT there was a significant main effect of question, $F(1, 75) = 44.3$, $p < .001$, $\eta_g^2 = .011$, and scenario, $F(1, 75) = 58.0$, $p < .001$, $\eta_g^2 = .012$, but no significant interaction, $F(1, 75) = 0.798$, $p = .374$, $\eta_g^2 < .001$. Belief questions took longer to respond to than reality questions, and false belief scenarios took longer to respond to than true belief scenarios. Bayesian analysis of the mean differences found that the null hypothesis was 14.1 times more likely than the alternative hypothesis (that the difference in RT between belief and reality questions was greater in the false belief scenarios than in the true belief scenarios).

Exploratory Analyses: Performance on First Trials and First Half of the Experiment

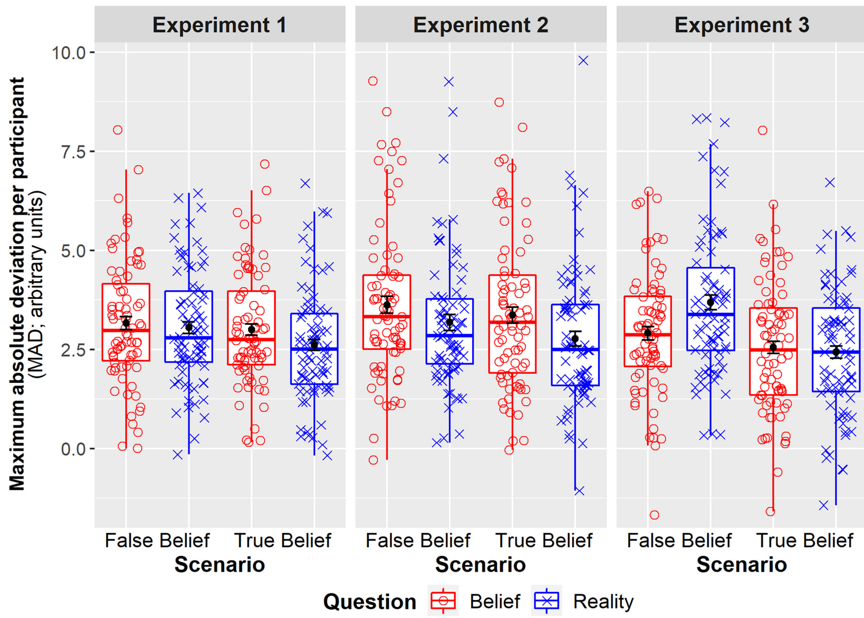
Our analyses of performance across all trials found no evidence of the interaction predicted by the presence of egocentric bias. However, it is possible, as suggested by an anonymous reviewer, that participants may have shown evidence of egocentric bias in earlier trials and then changed their performance as they adapted to the repeated presentation of the videos and questions. Additional exploratory analyses, reported in the online supplemental materials, did not support this suggestion. We conducted analyses of (a) the very first trial for each condition; (b) trials from the first half only; and (c) analyses of all trials with experiment half—whether trials came from the first or second half—included as an additional factor. There was no evidence of egocentric bias on participant's very first trial for each condition, although we also failed to find any consistent significant main effects (likely due to low power). Analysis of trials from the first half of the experiment only did not find evidence of egocentric bias either, but did find significant main effects of question and scenario. Finally, analysis of trials from both the first and second half together, with experiment half as an additional factor, did not find any evidence that experiment half interacted with either effects of question or scenario.

Discussion

If participants are egocentrically biased during belief processing, then answering false belief questions should involve overcoming an initial attraction toward the incorrect answer on our task. One would therefore predict an interaction between question and scenario, such that differences between belief and reality questions (greater MAD, and slower and less accurate responding) would be greater following false belief scenarios compared to true belief scenarios. Experiment 1 failed to find any evidence in support of this prediction in any of the recorded measures.

For errors and RT, we observed two main effects of question and scenario, but no interaction. Participants made more errors and responded slower to belief questions not only in false belief scenarios, but also in true belief scenarios—where the agent's belief was congruent with the participant's. Similarly, participants made more errors and responded slower to questions following false belief scenarios, but this was true when responding to questions about current reality as well as questions about beliefs. This pattern of two main effects for error and RT data is consistent with that seen in the RT study of Back and Apperly (2010). We will return to interpretations of these effects in the General Discussion section.

Figure 5
MAD Data in Experiments 1–3



Note. Boxes indicate 25th–75th percentile (IQR), with the middle line indicating the median. The whiskers indicate ±1.5 IQR. Each ring/cross is an individual participant's mean. The black dot indicates the condition mean, with error bars representing 95% CI (i.e., 1.96 *SE*). MAD = maximum absolute deviation; IQR = interquartile range; CI = confidence interval. See the online article for the color version of this figure.

In contrast to RT studies, the use of mouse tracking allowed us to directly measure attraction toward the incorrect answer during participant decision making. In our spatial measure of attraction (MAD) we observed an interaction between question and scenario, but in the opposite direction to that predicted by egocentric bias: While participants showed greater attraction toward the incorrect answer (greater MAD) on belief questions compared to reality questions following true belief scenarios, they did not show a significant difference between question types following false belief scenarios. Thus, there was no evidence that one's own knowledge was particularly salient for participants when reporting another agent's belief.

A possible concern, however, is that by the end of the false belief scenarios the participant's representation of where the keys are hidden may not have been particularly salient. Note that in response to reality questions, participants showed greater attraction toward the incorrect answer in false belief compared to true belief scenarios, along with more errors and longer RTs (see Figures 3, 5, and 7). One possibility is that participants were more uncertain as to where the keys were hidden in false belief scenarios relative to true belief scenarios. If so, this could explain why participants did not display any evidence of egocentric bias: Their knowledge of the final location of the keys in false belief scenarios may not have been salient enough to have biased belief processing.

Why might participants have been less certain regarding the final location of the keys in false belief compared to true belief scenarios? One possibility is that while false belief and true belief scenarios lasted the same overall length of time, they differed in when the question probe appeared relative to the final hiding of the keys. In true belief scenarios, the question probe occurred approximately 2 s after the switch of location, and that switch was the final event in the scenario. In false belief scenarios, however, the female agent returned to the room after the switch prior to the question probe appearing, leading to a delay of approximately 6 s between the switch and the question. This difference in timing, and the interruption of the female agent returning to the room, could have disrupted participants' representation of where the keys were hidden, to the extent that egocentric bias was not observed.

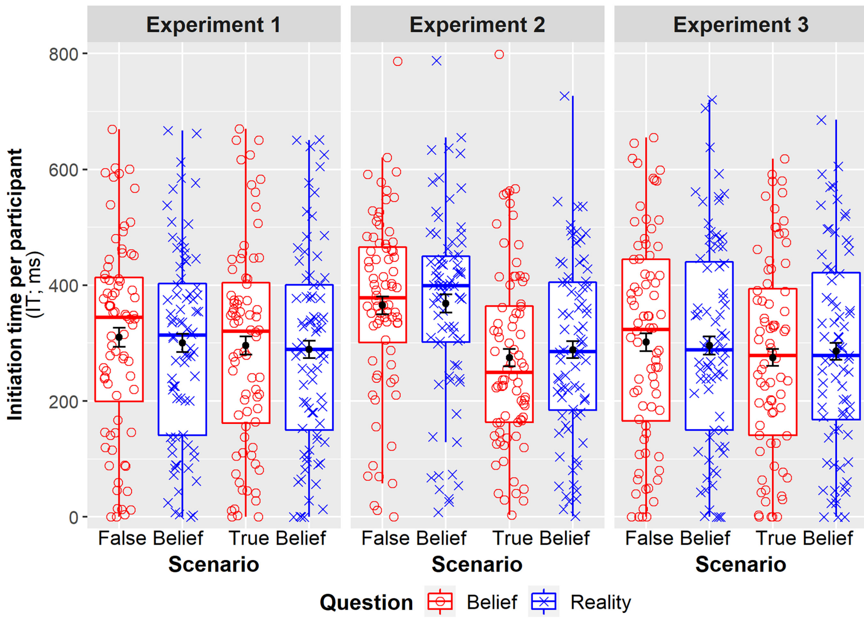
Furthermore, it is possible that in Experiment 1—and in Back and Apperly (2010), which also used similarly structured event sequences—the above difference in the relative timing of events between scenarios could explain the overall effects of scenario observed in those studies. The processing of events, particularly the location of the hidden object, could have been more disrupted in general in false belief compared to true belief scenarios, resulting in less accurate and more effortful responding. In Experiment 2, we addressed this possible explanation of both the failure to find evidence for egocentric bias and the observed effects of scenario in Experiment 1.

Experiment 2

In Experiment 2, participants completed the same task as Experiment 1 using the same stimuli. However, in the false belief scenarios the question probe now appeared once the keys were moved to the second cup, before the female actor came back into the room. Thus, the question probe occurred when the final location of the keys was maximally salient for participants—immediately

Figure 6

IT Data in Experiments 1–3



Note. Boxes indicate 25th–75th percentile (IQR), with the middle line indicating the median. The whiskers indicate ±1.5 IQR. Each ring/cross is an individual participant's mean. The black dot indicates the condition mean, with error bars representing 95% CI (i.e., 1.96 *SE*). IT = initiation time; IQR = interquartile range; CI = confidence interval. See the online article for the color version of this figure.

after they had just observed them being hidden. By probing at this moment in the false belief scenarios, Experiment 2 was arguably better placed than Experiment 1 to detect the presence of egocentric bias in belief processing. If in Experiment 2 we still found no evidence for the specific interaction predicted by egocentric bias—replicating the null findings of Experiment 1—then this would provide strong evidence against the claim that adults need to overcome an egocentric bias when reporting another's false belief.

In addition, the timing of the question probe relative to the final movement of the keys in Experiment 2 was now the same in the false belief and true belief scenarios. If the effects of scenario observed in Experiment 1 on error rate, RT and MAD were caused by a difference between scenarios in the relative timing of the question probes, then one would predict an absence of such scenario effects in Experiment 2.

Method

Transparency and Openness

The preregistration of the design and analysis plan for this experiment can be found at https://aspredicted.org/9wi54.pdf.[6] All data and analysis code are available at https://researchbox.org/1024. Research materials are available from the authors upon request.

Participants

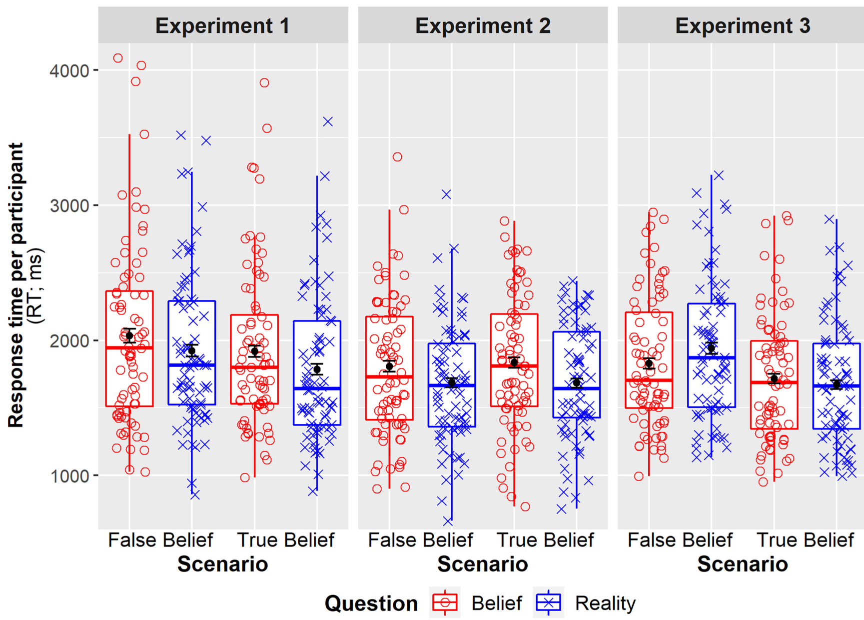
As reported above in Experiment 1, following the recommendations of Brysbaert (2019) a minimum sample size of 75 was determined to have at least 80% power to detect the relevant 2 × 2 interaction indicative of egocentric bias. Eighty-two participants completed the study, with a final sample of 76 participants after exclusions (see below). Participants were asked to report their gender with the same response options as in Experiment 1. Of the final sample, 23 identified as male and 53 identified as female. Participants' mean age was 26.3 years (*SD* = 7.9 years). Sixty-eight were ordinarily right-handed, and eight were ordinarily left handed, though all confirmed they normally used the computer mouse with their right hand. All had normal or corrected to normal vision and were fluent speakers of English. Participants were all recruited from the local University community. Ethical oversight and remuneration was the same as in Experiment 1.

Stimuli and Apparatus

Stimuli and apparatus were the same as those in Experiment 1 but with one difference. Instead of controlling for overall length of the videos as in Experiment 1 (where videos in both false belief and true belief scenarios were 20 s long), in Experiment 2 we standardized the timing of the question probe. The probe now occurred immediately after the keys were transferred to the second cup in both false belief and true belief scenarios. The true belief scenario

[6] It should be noted that Experiments 2 and 3 were planned contemporaneously but completed in the opposite order to that reported here, as can be seen by the timestamps on the preregistrations: Experiment 3 was in fact completed before Experiment 2. While we believe that presenting these two experiments in this order provides a more logical structure for the reader, we provide this note for full transparency.

Figure 7
RT Data in Experiments 1–3



Note. Boxes indicate 25th–75th percentile (IQR), with the middle line indicating the median. The whiskers indicate ±1.5 IQR. Each ring/cross is an individual participant's mean. The black dot indicates the condition mean, with error bars representing 95% CI (i.e., 1.96 *SE*). RT = response time; IQR = interquartile range; CI = confidence interval. See the online article for the color version of this figure.

videos were thus the same as in Experiment 1, but the false belief scenario videos were trimmed to 14 s long, such that the video ended (and the question probed appeared) at the point when the male actor had completed placing the keys in the second cup.

Design and Procedure

The design and procedure were the same as in Experiment 1.

Mouse-Tracking Data Processing and Exclusions

Data processing procedures were the same as in Experiment 1. Raw data from 82 participants was collected but six participants were eliminated due to low trial counts (less than 50% in one or more conditions). This gave a final sample of 76 participants, who contributed 4,460 trials in total out of a maximum of 4,864. On average, each participant contributed 58.7 trials out of a maximum of 64 (91.7%).

Results

As in Experiment 1, we report analyses for error rates, MAD, IT, and RT only. Data for these can be seen in Figures 3, 5, 6, and 7, and mouse-path trajectories in Figure 4. Other preregistered analyses are reported in the online supplemental materials and show the same pattern of effects as below.

Errors

Mixed-model logistic regression revealed a significant main effect of scenario only (coefficient = .41, $SE = 0.18$, $z = 2.25$, $p = .024$), with more errors made on false belief scenarios compared to true belief scenarios. There was no significant main effect of question (coefficient = .27, $SE = 0.18$, $z = 1.48$, $p = .14$), and no significant interaction coefficient = −.27, $SE = 0.19$, $z = -1.40$, $p = .16$).[7] Bayesian analysis of empirical logit transformed mean errors found that the null hypothesis was 20.2 times more likely than the alternative hypothesis (that the difference in errors between belief and reality questions was greater in the false belief scenarios than in the true belief scenarios).

Spatial Data—MAD

ANOVA revealed main effects of question, $F(1, 75) = 31.3$, $p < .001$, $\eta_g^2 = .023$, and scenario, $F(1, 75) = 8.52$, $p = .005$, $\eta_g^2 = .007$, but no significant interaction, $F(1, 75) = 0.501$, $p = .481$, $\eta_g^2 < .001$. Participants showed greater attraction toward the incorrect answer on belief compared to reality questions, and on false belief compared to true belief scenarios. Bayesian analysis of the mean differences found that the null hypothesis was 12.7 times more likely than the alternative hypothesis (that the difference in MAD between belief and reality questions was greater in the false belief scenarios than in the true belief scenarios). HDS was computed for each spatial measure per condition and

[7] This differed from the results of the preregistered ANOVA performed on untransformed mean errors, which found a significant main effect of question, $F(1, 75) = 6.38$, $p = .014$, $\eta_g^2 = .014$, but not scenario, $F(1, 75) = 3.04$, $p = .085$, $\eta_g^2 = .005$. However, the ANOVA analysis also found no significant interaction, $F(1, 75) = 2.50$, $p = .118$, $\eta_g^2 = .006$: Both analyses were therefore consistent with regards to a lack of any evidence for egocentric bias.

found no evidence that the data were bimodally distributed (all $ps \geq .213$).

Temporal Data

ANOVA on IT revealed a significant main effect of scenario, $F(1, 75) = 92.7, p < .001, \eta_g^2 = .071$, but no significant main effect of question, $F(1, 75) = 1.79$, $p = .186$, $\eta_g^2 < .001$, or a significant interaction, $F(1, 75) = 1.30$, $p = .258$, $\eta_g^2 < .001$. Participants were slower to initiate their first mouse movement following the question probe on false belief scenarios compared to true belief scenarios.

For RT, ANOVA found a significant main effect of question, $F(1, 75) = 56.2, p < .001, \eta_g^2 = .021$, but no significant main effect of scenario, $F(1, 75) = 2.03$, $p = .159$, $\eta_g^2 < .001$, or a significant interaction, $F(1, 75) = 1.18$, $p = .281$, $\eta_g^2 < .001$. Participants were slower to respond to belief questions compared to reality questions. Bayesian analysis of the mean differences found that the null hypothesis was 15.5 times more likely than the alternative hypothesis (that the difference in RT between belief and reality questions was greater in the false belief scenarios than in the true belief scenarios).

Exploratory Analyses: Performance on First Trials and First Half of the Experiment

As for Experiment 1, in the online supplemental materials, we report additional exploratory analyses of (a) the first trial for each condition; (b) trials from the first half only; and (c) analyses of all trials with experiment half included as an additional factor. As in Experiment 1, there were no consistent main effects and no evidence of egocentric bias on participant's very first trial for each condition. Analysis of trials from the first half of the experiment only did not find any evidence of egocentric bias, but did find significant main effects of question and scenario. Finally, there was again no evidence that experiment half interacted with either effects of question or scenario.

Discussion

As in Experiment 1, in Experiment 2 we once again did not find any evidence in support of egocentric bias in belief processing. The critical scenario by question interaction was not observed in any measure. Even when probing at the time point when participants' representation of the actual location of the keys should have been maximally salient—immediately following their hiding—there was no evidence that participants' own knowledge biased their responses when reporting the false belief of another agent.

Attraction toward the incorrect answer during responding was affected by main effects of both question and scenario: Participants displayed greater MAD when responding to belief questions compared to reality questions, and greater MAD following false belief compared to true belief scenarios. There was no evidence that responding to a belief question, specifically following a false belief scenario, resulted in any greater attraction toward the incorrect answer beyond the sum of the two main effects. Furthermore, as well as the main effect of scenario on MAD, participants also showed a main effect of scenario on the temporal characteristics of their responses, albeit in their IT as opposed to RT as seen in Experiment 1. Therefore, the effects of scenario observed in Experiment 1 cannot be completely explained by differences in the relative timing of the question probe: Controlling for those differences in Experiment 2 did not completely remove the scenario effects on the spatial and temporal characteristics of participants' responses. We will consider possible explanations for the effects of scenario and question observed in Experiments 1 and 2 in the General Discussion section.

Overall, Experiment 2 replicated the null findings of Experiment 1: Across two well-powered experiments using the same paradigm we failed to observe evidence in any measures for the specific interaction predicted by the presence of egocentric bias. While this null result was consistent across both experiments, there were some differences in the patterns of significance of the main effects of question and scenario between the two experiments. One concern, raised by an anonymous reviewer, is that if our paradigm and measures cannot robustly elicit main effects of question and/or scenario, then we would be unlikely to observe an interaction of these two factors.

We address this concern with additional data from an online study using the same paradigm and stimuli as Experiment 1, in which we found consistent main effects of question and scenario across our key measures. This study is reported in full in online supplemental material 2. This study was powered to detect main effects ($n = 27$, providing 80% power to detect main effects with an effect size of $d_z = 0.40$), and demonstrated significant main effects for both question and scenario for errors, RT and MAD (all $p < .017$). In all cases, these effects were in the same direction as those seen in Experiments 1 and 2. Participants showed more errors, longer RTs and greater MAD for belief questions compared to reality questions, and for false belief scenarios compared to true belief scenarios. Our paradigm and measures are therefore capable of robustly eliciting main effects of question and scenario. The failure to observe the interaction predicted by egocentric bias in Experiments 1 and 2 cannot be explained by a lack of such robustness.

Instead, a different concern with our paradigm is that participants could have realized that the belief questions were particularly difficult to answer relative to the reality questions. This may have led them to strategically prioritize encoding the female actor's belief while watching the videos. If participants were deliberately prioritizing belief processing then they may have strategically controlled any egocentric bias while watching the videos themselves. This might explain why egocentric bias was not observed in their subsequent responses to the question probes. In Experiment 3 we addressed this explanation.

Experiment 3

In Experiment 3, participants completed the same task as Experiments 1 and 2 but were under the instruction to deliberately pay attention to the female actor's belief about the location of the keys throughout the videos. This instruction manipulation has been used in previous RT studies of belief processing (Apperly et al., 2006; Back & Apperly, 2010). In those studies, belief-instructed participants no longer showed any differences in RTs between belief and reality questions compared to participants without such instructions (Apperly et al., 2006), or those differences became reversed such that belief questions were easier to respond to than reality questions (Back & Apperly, 2010). These findings strongly suggest that participants do not deliberately prioritize encoding beliefs when watching similar events to those used in our experiments, unless under specific instruction to do so. If participants do spontaneously prioritize encoding beliefs when watching such events, then one would not expect the belief instruction manipulation to substantially change performance. While these findings speak against the suggestion that our participants had deliberately prioritized encoding beliefs in Experiments 1 and 2, we considered it important to demonstrate that the effects of manipulating belief instructions are seen with our mouse-tracking paradigm, given that our experiments used different stimuli, question probes and mode of responding to those previous studies.

Experiment 3 used the same video stimuli as in Experiment 1, and performance in Experiment 1 was compared against performance in Experiment 3. If participants in Experiment 1 (and, by extension, Experiment 2) were not deliberately prioritizing the encoding of the female actor's beliefs during the videos, then one would expect to observe different behavioral patterns in Experiment 3 compared to Experiment 1: Specifically, one would expect the effects of question to be reduced or reversed (cf., Apperly et al., 2006; Back & Apperly, 2010). Conversely, if participants in Experiment 1 had prioritized encoding beliefs while watching the videos, then one would expect the effects of question observed in Experiment 1 to persist in Experiment 3.

Method

Transparency and Openness

The preregistration of the design and analysis plan for this experiment can be found at https://aspredicted.org/eq98n.pdf. All data and analysis code are available at https://researchbox.org/1024. Research materials are available from the authors upon request.

Participants

In order to allow a balanced comparison between Experiment 1 and Experiment 3, the same sample size ($n = 76$) was collected in Experiment 3 as in Experiment 1. Given the effect sizes of the question effects reported in Experiment 1 for RT and MAD (a main effect of question on RT of $d_z \asymp 0.61$ and an effect of question on MAD within the true belief scenario of $d_z = 0.59$) and following the recommendations of Brysbaert (2019), this sample provided greater than 90% power to detect the presence of the same question effects within Experiment 3, and greater than 80% power to detect a significant experiment × question interaction in a comparison between Experiment 1 and Experiment 3 if those question effects were in fact absent in Experiment 3.

Eighty-five participants completed the study, with a final sample of 76 participants after exclusions (see below). Participants were asked to report their gender with the same response options as in Experiment 1. Of the final sample, 34 identified as male and 42 identified as female. Participants' mean age was 28.2 years ($SD = 10.9$ years). Sixty-nine were ordinarily right handed and seven were ordinarily left handed, though all confirmed they normally used the computer mouse with their right hand. All had normal or corrected to normal vision and all were fluent English speakers. Participants were all recruited from the local University community. Remuneration was exactly the same as in Experiment 1, as were the ethical procedures.

Stimuli and Apparatus

All stimuli and apparatus were the same as in Experiment 1. However, one change was made to the filler question trials. In Experiment 1 on filler question trials the shirts and cups were whitened-out on the final frame of the video when the question was asked. This was done to ensure that participants were responding from memory, with the aim that including questions on different elements of the video would reduce the likelihood that participants deliberately encoded the female actor's beliefs while watching the video. In Experiment 3, however, it was imperative that participants deliberately engaged in belief processing while watching the videos. Therefore we did not whiten these stimuli on filler trials—when the question was asked they appeared in full color as in the video. These filler trials were still included so that participants could not simply anticipate the response that they would be required to give on every video.

Design and Procedure

The design and procedure were the same as in Experiment 1, other than the following changes. First, participants were told that the task was about belief processing. They were instructed to keep in mind the belief of the agent with respect to the location of the keys as they watched the videos. On the six practice trials participants were asked two reality and two belief questions, one of each for each scenario, plus two filler questions. Participants' understanding of the task—that they had to pay attention to the female actor's belief on every trial, even though they were not always asked about her belief—was checked during the practice.

During the main experiment in each block of 64 trials participants received four randomly timed on-screen reminders instructing them that they needed to pay attention to the female actor's belief on every trial. After the first 64 trials, and again at the end of the experiment, participants were presented with the following on-screen manipulation-check question: "Did you pay attention to where the female character thinks the keys are hidden on every, or nearly every, video?" Participants could answer either "yes" or "no," using the computer mouse.

Mouse-Tracking Data Processing and Exclusions

Four participants who did not answer "yes" to the above manipulation checks were excluded from the final sample. One further participant was removed after visual inspection of their mouse paths revealed that they had understood the task differently to all other participants.[8] For the remaining participants, data processing and exclusions were the same as in Experiment 1.[9] Four participants were

[8] On every single trial this participant first moved the mouse to the response box reflecting the agent's belief, and then moved to the correct answer for the actual question probe (where necessary). This pattern was not seen in any other participant.

[9] It should be noted that we originally preregistered that participants who made more than one error on belief questions in either false belief or true belief scenarios would also be excluded. However, for the data reported here we have not included that exclusion criteria for the following reasons: (a) Using a different exclusion criteria to Experiment 1 could arguably invalidate the comparisons between the two experiment. (b) During testing, it was noticed that using that exclusion criteria was resulting in a high number of exclusions (23/85), with consequences for the resourcing of the study. (c) In any case, analyses based on that preregistered exclusion criteria produced the same overall pattern of results as those reported below.

excluded due to low trial counts (less than 50% of trials in one or more conditions) after trial exclusions. This gave a final sample of 76 participants who contributed 4,446 trials in total out of a maximum of 4,864. On average, each participant contributed 58.5 trials out of a maximum of 64 (91.4%).

Results

As in Experiment 1, we report analyses for error rates, MAD, IT, and RT only. Data for these can be seen in Figures 3, 5, 6, and 7, and mouse-path trajectories in Figure 4. Other preregistered analyses are reported in the online supplemental materials and show the same pattern of effects as below.

Errors

Mixed-model logistic regression revealed a significant main effect of scenario only (coefficient $= 0.79$, $SE = 0.12$, $z = 6.67$, $p < .001$), with more errors made on false belief scenarios compared to true belief scenarios. There was no significant main effect of question (coefficient $= 0.16$, $SE = 0.12$, $z = 1.29$, $p = .20$), and no significant interaction (coefficient $= -0.22$, $SE = 0.12$, $z = -1.86$, $p = .068$).[10]

Spatial Data—MAD

ANOVA revealed significant main effects of both question, $F(1, 75) = 8.41$, $p = .005$, $\eta^2_g = .009$, and scenario, $F(1, 75) = 44.9$, $p < .001$, $\eta^2_g = .052$, qualified by a significant interaction, $F(1, 75) = 17.7$, $p < .001$, $\eta^2_g = .017$. This interaction was due to responses to reality questions showing greater attraction toward the incorrect answer than responses to belief questions in the false belief scenarios only, $t(75) = 3.74$, $p < .001$, $d = 0.429$. In the true belief scenarios, there was no significant difference between questions, $t(75) = 1.72$, $p = .089$, $d = 0.198$. HDS was computed for each spatial measure per condition and found no evidence that the data were bimodally distributed (all $p\text{s} \geq .243$).

Temporal Data

ANOVA on IT revealed a significant main effect of scenario, $F(1, 75) = 14.5$, $p < .001$, $\eta^2_g = .004$. Participants took longer to initiate their first movement on false belief compared to true belief scenarios. There was no significant main effect of question, $F(1, 75) = 0.011$, $p = .918$, $\eta^2_g < .001$, and no significant interaction, $F(1, 75) = 1.78$, $p = .186$, $\eta^2_g < .001$.

For RT there was no significant main effect of question, $F(1, 75) = 3.09$, $p = .083$, $\eta^2_g < .001$. There was a significant main effect of scenario, $F(1, 75) = 97.7$, $p < .001$, $\eta^2_g = .035$, qualified by a significant interaction, $F(1, 75) = 20.2$, $p < .001$, $\eta^2_g = .005$. This interaction was due to participants taking longer to respond to reality questions compared to belief questions in the false belief scenarios only, $t(75) = 3.85$, $p < .001$, $d = 0.442$. In the true belief scenarios there was no significant difference, after Bonferroni correction ($\alpha = .025$), between questions, $t(75) = -2.02$, $p = .047$, $d = 0.231$.

Comparison of Experiments 1 and 3

Across the measures reported above, participants in Experiment 3 showed a different pattern of performance to those in Experiment 1. In order to assess whether these differences across experiments were themselves statistically significant, in the online supplemental materials we report full analyses of Experiment 1 and 3 together, with experiment (Experiment 1, Experiment 3) included as an additional factor. If experiment interacted with effects of question, this would indicate that the instruction to participants in Experiment 3 to attend to beliefs indeed changed participants' performance.

For both MAD and RT, these analyses found significant three-way interactions of experiment × scenario × question. Follow-up analyses of experiment and question effects within each scenario separately found significant experiment × question interactions in both false belief and true belief scenarios. For MAD, whereas in Experiment 1 in false belief scenarios participants showed no significant difference in attraction toward the incorrect answer between belief and reality questions, in Experiment 3 participants showed more attraction on reality compared to belief questions. In Experiment 1 in true belief scenarios participants showed greater attraction on belief compared to reality questions, but in Experiment 3 no significant difference in attraction between questions was observed. For RT, in Experiment 1 in both false belief and true belief scenarios participants took longer to respond to belief compared to reality questions. However, in Experiment 3 participants took longer to respond to reality compared to belief questions in false belief scenarios, and showed no significant difference in RT between question types in true belief scenarios. Analyses of errors and IT found no significant interactions of experiment with any other factor.

Discussion

Experiment 3 tested whether instructing participants to pay attention to the female actor's beliefs would result in different patterns of performance compared to Experiment 1. Recall that in Experiment 1, participants made more errors and took longer to respond to belief questions compared to reality questions. They also showed greater attraction toward the incorrect answer (greater MAD) when responding to belief questions compared to reality questions in true belief scenarios, but not in false belief scenarios.

In Experiment 3, however, those effects of question in true belief scenarios were no longer significant. In false belief scenarios, participants made more errors, took longer to respond, and showed more attraction toward the incorrect answer on reality questions. This pattern can be explained as follows: When attending to beliefs, the content of the female actor's belief (e.g., the keys are in the red cup) is incongruent with the required response only on reality questions following false belief scenarios (the keys are in the blue cup). Participants therefore experience greater competition in this particular condition, resulting in the observed pattern of effects. The reversal of the question effect matches data from Back and Apperly (2010) where participants also found answering reality questions more difficult than belief questions following false belief scenarios, when instructed to attend to beliefs.

Experiment 3 therefore rules out a possible explanation for why no effects of egocentric bias were observed in Experiments 1 and 2: That participants had been strategically prioritizing encoding the female agent's beliefs. Had they been, then one would

[10] This matched the results of the preregistered ANOVA performed on untransformed mean errors, which also found a significant main effect of scenario and no other effects.

have predicted the same pattern of effects in Experiment 3 as Experiment 1.

General Discussion

We failed to find any evidence for egocentric bias when participants ascribed false beliefs to another. In both Experiment 1 and Experiment 2 there was no evidence for an interaction between question and scenario, such that the cost of answering a belief versus a reality question was greater in false belief scenarios compared to true belief scenarios, as predicted by accounts that claim we have to overcome egocentric bias when reporting another's false belief (e.g., Epley et al., 2004; Farrar & Ostojić, 2018; Friedman & Leslie, 2004; Leslie & Polizzi, 1998; Tamir & Mitchell, 2013; L. Wang & Leslie, 2016). In both experiments the cost of answering a belief question compared to a reality question was also observed in true belief scenarios—where the agent shared the same belief as the participant—with no evidence that this cost was greater in false belief scenarios. Furthermore, our data consistently supported the null hypothesis over the prediction that any differences between question types would be greater in false belief compared to true belief scenarios. Experiment 3 ruled out one explanation for our failure to observe egocentric bias: That participants in Experiments 1 and 2 deliberately prioritized encoding beliefs while watching the videos. Participants under instruction to deliberately attend to the female actor's beliefs in Experiment 3 showed different patterns of performance to those in the other two experiments.

Another explanation of our failure to observe the interaction predicted by egocentric bias, raised by an anonymous reviewer, is that our mouse-tracking measures of attraction may not be reliable enough to consistently detect the competition effects necessary to detect egocentric bias. This is unlikely for several reasons. First, our measures clearly are able to detect an interaction resulting from a competition effect that occurs in one question by scenario combination but not the other three: We observed exactly such an interaction in Experiment 3. There, the instruction to attend to beliefs results in an incongruency between the correct answer and the ongoing belief processing on reality questions specifically following false belief scenarios. That we observed a question by scenario interaction on Experiment 3 provides a compelling positive control, or "sanity check," demonstrating that our methods are capable of detecting such effects.[11] Second, the results of the Bayesian analyses in Experiments 1 and 2 suggest our data are not simply too noisy to discriminate between null and alternate hypotheses: In all cases we found consistent evidence in favor of the null over the specific interaction predicted by egocentric bias.

Finally, across Experiment 1, Experiment 2 and the additional online study reported in the online supplemental materials we consistently observed main effects of question and scenario. Participants were generally less accurate, slower to respond and showed greater attraction to the incorrect answer when answering questions about another person's beliefs compared to questions about current reality. We also found main effects of scenario, with answers to questions in true belief scenarios generally more accurate, faster to produce and showing less attraction to the incorrect answer than answers to questions in false belief scenarios. Our paradigm and measures are capable of reliably demonstrating question and scenario effects, and both Experiment 1 and Experiment 2 were well-powered to detect their interaction—a lack of reliability per se therefore cannot explain the failure to observe the specific interaction predicted by egocentric bias. We now turn to explanations of these effects of question and scenario.

Explaining the Effects of Question

It is possible that incidental differences, such as linguistic complexity, between the belief and reality question probes could explain the main effects of question observed in Experiments 1 and 2. While question character length was controlled for, perhaps questions about belief states are just more difficult for participants to read and understand, compared to questions about reality. The findings of Experiment 3, however, make such an explanation unlikely. When participants attended to beliefs while watching the videos, they answered belief questions just as readily as reality questions. This suggests that belief probes are not simply harder to read and answer per se (see also this interpretation of the belief instruction manipulation in Apperly et al., 2006; Back & Apperly, 2010).

Furthermore, previous RT studies that have (a) equated linguistic complexity between belief and reality conditions (Apperly et al., 2006), or (b) used nonlinguistic question probes (Back & Apperly, 2010), or (c) have used a "reality" probe that refers to the participant's belief (i.e., "where do you think…" vs. " where does she think…," Bradford et al., 2018; Meert et al., 2017)," also report effects of question upon participant performance. Belief questions have therefore consistently been found to be harder than reality questions across different types of scenarios, question probes and modes of responding. Overall therefore, the question effects seen in Experiments 1 and 2 are most likely the result of differences in how belief and reality information is processed, rather than incidental differences in how participants respond to the specific question probes used.

How might the processing of belief information be different from reality processing? One explanation is that answering questions about reality requires accessing information that has already been encoded only, whereas answering belief questions also requires an inference to be made over that information (Apperly et al., 2006; Peterson & Riggs, 1999). In both our true belief and false belief scenarios, in order to report the agent's belief, participants had to recall: (a) the movement of the object; (b) whether the agent was present when the object was moved; and (c) combine these two pieces of information to infer what the agent believed. Responding to reality questions only requires the first of the above steps, therefore explaining their relative ease compared to belief questions. Note too that these steps in belief processing need not involve any egocentric bias toward assuming that others share one's own belief, consistent with the lack any evidence for egocentric bias in our data.

The question effects observed in our study further extend and replicate the findings of previous RT studies using belief and reality questions (Apperly, 2010; Apperly et al., 2006). Those authors

[11] The reviewer also suggested that replication of an unrelated, but reliable, mouse-tracking effect, such as the phonological cohort effect (Spivey et al., 2005) might serve as such a "sanity check." We have replicated that exact effect multiple times in our lab using the same underlying participant population as in the present studies (see Lucas, 2021). Given that with the same lab set-up and participant pool we are able to detect that robust effect, it is therefore not the case that our failure to detect an egocentric bias effect here can be explained by factors such as noisy or unreliable procedures, equipment, or participants.

also interpreted the cost of answering a belief question as reflecting the nonautomaticity of belief processing: When required to respond to belief questions, adults do not simply access a belief representation that was automatically encoded while watching the events unfold. Furthermore, our data rule out an objection by Cohen and German (2009) to that interpretation. Cohen and German objected that when there is both an instruction to attend to reality information (as in Apperly et al., 2006 and also Back & Apperly, 2010) and a delay between the agent forming a belief and the question probe appearing, automatically encoded belief information might decay faster compared to the deliberately maintained reality information. However, our participants in Experiments 1 and 2 were under no instruction to attend to any particular information, and in our true belief scenarios the question probe immediately followed the switch of object location. We still observed differences between belief and reality questions under these conditions on true belief scenarios—that difference cannot be explained by Cohen and German's account. Overall, the question effects observed across multiple studies using different stimuli, task instructions, and means of responding provide converging evidence supporting the claim that adults do not answer belief questions by accessing an automatically encoded belief representation.

Explaining the Effects of Scenario

It should be noted, however, that even if beliefs are not processed automatically, in general it still remains a possibility that participants may have engaged in some form of belief-relevant processing while watching our videos. Indeed, Experiments 1 and 2 provide evidence that participants did process the false belief and true belief scenarios differently. Across both experiments, effects of scenario were observed for both reality and belief questions.

That these effects of scenario were observed on reality questions implies that participants were processing some information relevant to the agent's beliefs while watching the scenarios. If participants did not engage in any such processing, then it is difficult to explain why responses to reality questions—that only involve recalling the final location of the object—were affected by whether the agent had a true or a false belief. We offer three possible explanations for this effect of scenario: (a) Participants were occasionally, but spontaneously, processing the beliefs of the agent (Back & Apperly, 2010); (b) they were processing the knowledge or ignorance of the agent (Phillips et al., 2021); and (c) they were processing something more akin to belief-like states (Apperly & Butterfill, 2009).

The first explanation, suggested by Back and Apperly (2010) to account for the scenario effects seen in their RT study, is that, on at least some trials, while watching the video participants spontaneously encoded the belief of the agent in parallel to encoding the location of the object. In the case of false belief scenarios, this would have resulted in participants maintaining two incongruent representations regarding the location of the object: the object's actual location and where the agent thinks it is (see also Rubio-Fernández, 2013, 2017). The processing of this incongruency per se may have affected participants' responses to both reality and belief questions, resulting in slower responding, and more competition between response options (as indexed by MAD) on false belief scenarios.

Evidence for such an effect of incongruency (i.e., the effect of scenario) should not, however, be confused with evidence for egocentric bias. Egocentric bias would predict that if participants were maintaining two incongruent representations, the representation reflecting participants' own knowledge should be the most salient. Therefore, the detriment encountered when answering questions following a false belief scenario should be greatest for belief questions compared to reality questions (because on belief questions the incorrect answer reflects participants' own knowledge, whereas on reality questions the incorrect answer reflects the agent's belief). Answering belief questions should therefore amplify any effect of incongruency (scenario) upon participants' responses. This is not what was observed in the spatial characteristics of participants' responses: There was no evidence that the magnitude of the scenario effect on MAD was larger for belief questions than for reality questions.

What was observed Experiments 1 and 2, therefore, could at best be described as a bidirectional effect of incongruency. If, on some trials, participants were maintaining a representation of both the agent's belief and the actual location of the object, then the representation of the agent's belief interfered with subsequent responding to reality questions to the same extent as the representation of the object's actual location interfered with responding to belief questions. That the ongoing processing of another agent's belief can affect one's own processing of reality—altercentric interference—is widely attested in the theory of mind literature (e.g., Apperly et al., 2008; Back & Apperly, 2010; Meert et al., 2017). The findings here are therefore consistent with recent suggestions that rather than best being characterized as "egocentric creatures," humans in fact regularly engage in altercentric cognition and display effects of altercentric interference across the lifespan (Kampis & Southgate, 2020).

The second explanation for the effects of scenario observed in Experiments 1 and 2 is that participants processed some other relevant aspect of the agent's mental state, but not her beliefs. For example, it has recently been argued that knowledge ascription is a more basic and fundamental theory of mind process than belief ascription and is likely to be automatic (Phillips et al., 2021). In false belief scenarios, the agent is ignorant of something that the participant knows—the final location of the object. Simply processing events in which another agent is ignorant of a relevant fact, without necessarily representing what they believe, could come with a cognitive cost compared to the true belief scenarios where both participant and agent know the same things.

The final explanation is provided by the two-system account of mindreading (Apperly & Butterfill, 2009; Butterfill & Apperly, 2013; Low et al., 2016). This account distinguishes between a flexible-but-cognitively-demanding system that can support deliberate, verbal reasoning about beliefs (System 2) and a limited-but-efficient system that cannot process beliefs, but instead is capable of tracking belief-like "registrations" (System 1). The question effects observed in Experiments 1 and 2 may reflect the need to engage System 2 in response to belief question probes, consistent with our suggestions above. The scenario effects, however, may reflect the automatic tracking of registrations—the encoding of where an agent last encountered an object—by System 1. In false belief scenarios, any such registration for the female actor's encounter with the keys will be incongruent with the reality of where the keys are located at the end of the video. It will also be incongruent with any registration for the male actor who hides the keys. This incongruency when tracking registrations, rather than an incongruency when processing a belief, or costs associated with processing ignorance, may have resulted in the effects of scenario.

The two-system account suggests that at least some elements of theory of mind processing are automatic, but denies that there is automatic processing of beliefs. Given that our task was designed to test for the presence of egocentric bias in explicit belief processing, our data are not able to distinguish whether the observed scenario effects are the result of automatic processes or not. Stronger evidence for automaticity would be obtained if the effects of scenario persisted even when participants did not have to answer questions about beliefs, and thus had no reason to engage in any form of theory of mind processing while watching the videos (for examples of such tasks, see Bardi et al., 2019; Kaddouri et al., 2020; Kovács et al., 2010; Low et al., 2020; van der Wel et al., 2014; for debate regarding the interpretation of such tasks see Heyes, 2014; Phillips et al., 2015, 2021). Whether effects of scenario continue to be seen in our task when belief questions are removed is open to future research.

Reconsidering Egocentric Bias

While the exact nature of the processes underlying the question and scenario effects observed in our studies is open to further debate and research, we found no evidence supporting the presence of egocentric bias in participants' belief processing. While our results are therefore inconsistent with theories of belief processing that emphasize the need to overcome an egocentric bias as discussed above, our failure to find evidence for egocentric bias is consistent with data from other studies of belief processing (e.g., Back & Apperly, 2010; Schuwerk, Döhnel, et al., 2014; Schuwerk, Schecklmann, et al., 2014), including those that failed to replicate previous findings claiming to show evidence of egocentric bias in adults (Ryskin & Brown-Schmidt, 2014; Samuel, 2023; Samuel et al., 2018). More broadly, our findings are consistent with the suggestion that adults are readily able to infer and make use of other people's mental states—for example, during interactive communication—and are able to keep in mind both their own and another's beliefs without a necessary, default bias either toward self or other (e.g., Kampis & Southgate, 2020; see also Brown-Schmidt et al., 2008; Dale et al., 2018; Hanna et al., 2003; Heller et al., 2008; Rubio-Fernández, 2013, 2017).

Constraints on Generality

We acknowledge, however, that our data are built upon a specific task and target population: explicit belief processing in healthy young adults on an unexpected-transfer false belief task. There is, of course, very good reason for that: It is with exactly that type of task and target population that others have claimed to demonstrate evidence of egocentric bias (e.g., Birch & Bloom, 2007; Sommerville et al., 2013). Nonetheless, it is plausible that different populations, such as elderly adults (Bernstein et al., 2011), may be more likely to demonstrate egocentric bias in belief processing. It is also plausible that tasks with less straightforward scenarios in which the other's belief is more difficult to infer may be more likely to result in egocentric belief attributions. However, such tasks may face a challenge in discriminating between effects due to egocentric bias in attempted belief processing and effects due to participants simply not engaging in belief processing at all due to the task difficulty (see the Causes of Egocentric Bias in Adults section above). Finally, it is possible that neuroimaging techniques might reveal processes indicative of egocentric bias in the absence of any such effects upon behavioral measures such as RT and mouse tracking (Bricker, 2020). Neuroimaging studies to date, however, have typically not used designs that cross question (belief v reality) with scenario (false belief v true belief) such that egocentric bias effects can be isolated within a given sample (e.g., Aichhorn et al., 2009; Liu et al., 2004; Meinhardt et al., 2011; Mossad et al., 2016; Sommer et al., 2007). Whether a lack of any egocentric bias effect continues to be observed across a wider range of techniques and paradigms, once appropriate designs are used, remains to be seen.

Conclusion

We used a mouse-tracking paradigm to directly measure the salience of one's own knowledge while reporting the false belief of another agent. We found no evidence that young adults are egocentrically biased on any measure of responding. Instead, participants' responses suggest that beliefs in general are more effortful for participants to report compared to reality, and that while participants are sensitive to the incongruence between beliefs and reality in false belief scenarios, they are not biased toward their representation of reality over their representation of another agents' false belief.

References

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. *Journal of Cognitive Neuroscience*, *21*(6), 1179–1192. https://doi.org/10.1162/jocn.2009.21082

Apperly, I. A. (2010). *Mindreaders: The cognitive basis of "theory of mind."* Psychology Press.

Apperly, I. A., Back, E., Samson, D., & France, L. (2008). The cost of thinking about false beliefs: Evidence from adults' performance on a non-inferential theory of mind task. *Cognition*, *106*(3), 1093–1108. https://doi.org/10.1016/j.cognition.2007.05.005

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

Apperly, I. A., Carroll, D. J., Samson, D., Humphreys, G. W., Qureshi, A., & Moffitt, G. (2010). Why are there limits on theory of mind use? Evidence from adults' ability to follow instructions from an ignorant speaker. *Quarterly Journal of Experimental Psychology*, *63*(6), 1201–1217. https://doi.org/10.1080/17470210903281582

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. https://doi.org/10.1016/j.tics.2009.12.006

Bardi, L., Desmet, C., & Brass, M. (2019). Spontaneous theory of mind is reduced for nonhuman-like agents as compared to human-like agents. *Psychological Research*, *83*(7), 1571–1580. https://doi.org/10.1007/s00426-018-1000-0

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bernstein, D. M. (2021). Hindsight bias and false-belief reasoning from preschool to old age. *Developmental Psychology*, *57*(8), 1387–1402. https://doi.org/10.1037/dev0001226

Bernstein, D. M., Coolin, A., Fischer, A. L., Thornton, W. L., & Sommerville, J. A. (2017). False-belief reasoning from 3 to 92 years of age. *PLoS ONE*, *12*(9), Article e0185345. https://doi.org/10.1371/journal.pone.0185345

Bernstein, D. M., Thornton, W. L., & Sommerville, J. A. (2011). Theory of mind through the ages: Older and middle-aged adults exhibit more errors than do younger adults on a continuous false belief task. *Experimental Aging Research*, *37*(5), 481–502. https://doi.org/10.1080/0361073X.2011.619466

Birch, S. A., & Bloom, P. (2007). The curse of knowledge in reasoning about false beliefs. *Psychological Science*, *18*(5), 382–386. https://doi.org/10.1111/j.1467-9280.2007.01909.x

Bradford, E. E., Hukker, V., Smith, L., & Ferguson, H. J. (2018). Belief-attribution in adults with and without autistic spectrum disorders. *Autism Research*, *11*(11), 1542–1553. https://doi.org/10.1002/aur.2032

Bricker, A. M. (2020). The neural and cognitive mechanisms of knowledge attribution: An EEG study. *Cognition*, *203*, Article 104412. https://doi.org/10.1016/j.cognition.2020.104412

Brown-Schmidt, S., Gunlogson, C., & Tanenhaus, M. K. (2008). Addressees distinguish shared from private information when interpreting questions during interactive conversation. *Cognition*, *107*(3), 1122–1134. https://doi.org/10.1016/j.cognition.2007.11.005

Brysbaert, M. (2019). How many participants do we have to include in properly powered experiments? A tutorial of power analysis with reference tables. *Journal of Cognition*, *2*(1), Article 16. https://doi.org/10.5334/joc.72

Burnside, K., Ruel, A., Azar, N., & Poulin-Dubois, D. (2018). Implicit false belief across the lifespan: Non-replication of an anticipatory looking task. *Cognitive Development*, *46*, 4–11. https://doi.org/10.1016/j.cogdev.2017.08.006

Butterfill, S. A., & Apperly, I. A. (2013). How to construct a minimal theory of mind. *Mind & Language*, *28*(5), 606–637. https://doi.org/10.1111/mila.12036

Coburn, P. I., Bernstein, D. M., & Begeer, S. (2015). A new paper and pencil task reveals adult false belief reasoning bias. *Psychological Research*, *79*(5), 739–749. https://doi.org/10.1007/s00426-014-0606-0

Cohen, A. S., & German, T. C. (2009). Encoding of others' beliefs without overt instruction. *Cognition*, *111*(3), 356–363. https://doi.org/10.1016/j.cognition.2009.03.004

Cohen, A. S., & German, T. C. (2010). A reaction time advantage for calculating beliefs over public representations signals domain specificity for 'theory of mind'. *Cognition*, *115*(3), 417–425. https://doi.org/10.1016/j.cognition.2010.03.001

Cohen, A. S., Sasaki, J. Y., & German, T. C. (2015). Specialized mechanisms for theory of mind: Are mental representations special because they are mental or because they are representations? *Cognition*, *136*, 49–63. https://doi.org/10.1016/j.cognition.2014.11.016

Converse, B. A., Lin, S., Keysar, B., & Epley, N. (2008). In the mood to get over yourself: Mood affects theory-of-mind use. *Emotion*, *8*(5), 725–730. https://doi.org/10.1037/a0013283

Dale, R., Galati, A., Alviar, C., Contreras Kallens, P., Ramirez-Aristizabal, A. G., Tabatabaeian, M., & Vinson, D. W. (2018). Interacting timescales in perspective-taking. *Frontiers in Psychology*, *9*, Article 1278. https://doi.org/10.3389/fpsyg.2018.01278

Duran, N. D., Dale, R., & Kreuz, R. J. (2011). Listeners invest in an assumed other's perspective despite cognitive cost. *Cognition*, *121*(1), 22–40. https://doi.org/10.1016/j.cognition.2011.06.009

Epley, N., Keysar, B., Van Boven, L., & Gilovich, T. (2004). Perspective taking as egocentric anchoring and adjustment. *Journal of Personality and Social Psychology*, *87*(3), 327–339. https://doi.org/10.1037/0022-3514.87.3.327

Farrar, B. G., & Ostojić, L. (2018). Does social distance modulate adults' egocentric biases when reasoning about false beliefs? *PLoS ONE*, *13*(6), Article e0198616. https://doi.org/10.1371/journal.pone.0198616

Ferguson, H. J., Apperly, I., Ahmad, J., Bindemann, M., & Cane, J. (2015). Task constraints distinguish perspective inferences from perspective use during discourse interpretation in a false belief task. *Cognition*, *139*, 50–70. https://doi.org/10.1016/j.cognition.2015.02.010

Ferguson, H. J., & Breheny, R. (2012). Listeners' eyes reveal spontaneous sensitivity to others' perspectives. *Journal of Experimental Social Psychology*, *48*(1), 257–263. https://doi.org/10.1016/j.jesp.2011.08.007

Ferguson, H. J., Scheepers, C., & Sanford, A. J. (2010). Expectations in counterfactual and theory of mind reasoning. *Language and Cognitive Processes*, *25*(3), 297–346. https://doi.org/10.1080/01690960903041174

Fischhoff, B. (1975). Hindsight is not equal to foresight: The effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human Perception and Performance*, *1*(3), 288–299. https://doi.org/10.1037/0096-1523.1.3.288

Freeman, J. B. (2018). Doing psychological science by hand. *Current Directions in Psychological Science*, *27*(5), 315–323. https://doi.org/10.1177/0963721417746793

Freeman, J. B., & Ambady, N. (2010). Mousetracker: Software for studying real-time mental processing using a computer mouse-tracking method. *Behavior Research Methods*, *42*(1), 226–241. https://doi.org/10.3758/BRM.42.1.226

Freeman, J. B., & Ambady, N. (2011). Hand movements reveal the time-course of shape and pigmentation processing in face categorization. *Psychonomic Bulletin & Review*, *18*(4), 705–712. https://doi.org/10.3758/s13423-011-0097-6

Freeman, J. B., & Dale, R. (2013). Assessing bimodality to detect the presence of a dual cognitive process. *Behavior Research Methods*, *45*(1), 83–97. https://doi.org/10.3758/s13428-012-0225-x

Friedman, O., & Leslie, A. M. (2004). A developmental shift in processes underlying successful belief-desire reasoning. *Cognitive Science*, *28*(6), 963–977. https://doi.org/10.1207/s15516709cog2806_4

Galati, A., Dale, R., & Duran, N. D. (2019). Social and configural effects on the cognitive dynamics of perspective-taking. *Journal of Memory and Language*, *104*, 1–24. https://doi.org/10.1016/j.jml.2018.08.007

Ghrear, S. E., Birch, S. A., & Bernstein, D. M. (2016). Outcome knowledge and false belief. *Frontiers in Psychology*, *7*, Article 118. https://doi.org/10.3389/fpsyg.2016.00118

Golshaie, R., & Incera, S. (2021). Grammatical aspect and mental activation of implied instruments: A mouse-tracking study in Persian. *Journal of Psycholinguistic Research*, *50*(4), 737–755. https://doi.org/10.1007/s10936-020-09742-3

Hanna, J. E., Tanenhaus, M. K., & Trueswell, J. C. (2003). The effects of common ground and perspective on domains of referential interpretation. *Journal of Memory and Language*, *49*(1), 43–61. https://doi.org/10.1016/S0749-596X(03)00022-6

Heller, D., Grodner, D., & Tanenhaus, M. K. (2008). The role of perspective in identifying domains of reference. *Cognition*, *108*(3), 831–836. https://doi.org/10.1016/j.cognition.2008.04.008

Heyes, C. (2014). Submentalizing: I am not really reading your mind. *Perspectives on Psychological Science*, *9*(2), 131–143. https://doi.org/10.1177/1745691613518076

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and toward logit mixed models. *Journal of Memory and Language*, *59*(4), 434–446. https://doi.org/10.1016/j.jml.2007.11.007

JASP Team. (2020). *JASP* (Version 0.14) [Computer software]. https://jasp-stats.org/

Kaddouri, E., Bardi, R., De Bremaeker, L., Brass, D., & Wiersema, M., & R, J. (2020). Measuring spontaneous mentalizing with a ball detection task: Putting the attention-check hypothesis by Phillips and colleagues (2015) to the test. *Psychological Research*, *84*(6), 1749–1757. https://doi.org/10.1007/s00426-019-01181-7

Kampis, D., & Southgate, V. (2020). Altercentric cognition: How others influence our cognitive processing. *Trends in Cognitive Sciences*, *24*(11), 945–959. https://doi.org/10.1016/j.tics.2020.09.003

Keysar, B., Lin, S., & Barr, D. J. (2003). Limits on theory of mind use in adults. *Cognition*, *89*(1), 25–41. https://doi.org/10.1016/S0010-0277(03)00064-7

Kieslich, P. J., Schoemann, M., Grage, T., Hepp, J., & Scherbaum, S. (2020). Design factors in mouse-tracking: What makes a difference? *Behavior Research Methods*, *52*(1), 317–341. https://doi.org/10.3758/s13428-019-01228-y

Kovács, ÁM., Téglás, 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

Kulke, L., von Duhn, B., Schneider, D., & Rakoczy, H. (2018). Is implicit theory of mind a real and robust phenomenon? Results from a systematic replication study. *Psychological Science*, *29*(6), 888–900. https://doi.org/10.1177/0956797617747090

Leslie, A. M., & Polizzi, P. (1998). Inhibitory processing in the false belief task: Two conjectures. *Developmental Science*, *1*(2), 247–253. https://doi.org/10.1111/1467-7687.00038

Lin, S., Keysar, B., & Epley, N. (2010). Reflexively mindblind: Using theory of mind to interpret behavior requires effortful attention. *Journal of Experimental Social Psychology*, *46*(3), 551–556. https://doi.org/10.1016/j.jesp.2009.12.019

Liu, D., Sabbagh, M. A., Gehring, W. J., & Wellman, H. M. (2004). Decoupling beliefs from reality in the brain: An ERP study of theory of mind. *NeuroReport*, *15*(6), 991–995. https://doi.org/10.1097/00001756-200404290-00012

Low, J., Apperly, I. A., Butterfill, S. A., & Rakoczy, H. (2016). Cognitive architecture of belief reasoning in children and adults: A primer on the two-systems account. *Child Development Perspectives*, *10*(3), 184–189. https://doi.org/10.1111/cdep.12183

Low, J., Edwards, K., & Butterfill, S. A. (2020). Visibly constraining an agent modulates observers' automatic false-belief tracking. *Scientific Reports*, *10*(1), 1–12. https://doi.org/10.1038/s41598-019-56847-4

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. https://doi.org/10.1177/0956797612451469

Lucas, A. P. (2021). *The nature of novel word representations: Computer mouse tracking shows evidence of immediate lexical engagement effects in adults* [Thesis]. University of Hull. https://hull-repository.worktribe.com/output/4223534

Maldonado, M., Dunbar, E., & Chemla, E. (2019). Mouse tracking as a window into decision making. *Behavior Research Methods*, *51*(3), 1085–1101. https://doi.org/10.3758/s13428-018-01194-x

Meert, G., Wang, J., & Samson, D. (2017). Efficient belief tracking in adults: The role of task instruction, low-level associative processes and dispositional social functioning. *Cognition*, *168*, 91–98. https://doi.org/10.1016/j.cognition.2017.06.012

Meinhardt, J., Sodian, B., Thoermer, C., Döhnel, K., & Sommer, M. (2011). True-and false-belief reasoning in children and adults: An event-related potential study of theory of mind. *Developmental Cognitive Neuroscience*, *1*(1), 67–76. https://doi.org/10.1016/j.dcn.2010.08.001

Morey, R. D., & Rouder, J. N. (2011). Bayes Factor approaches for testing interval null hypotheses. *Psychological Methods*, *16*(4), 406–419. https://doi.org/10.1037/a0024377

Morey, R. D., & Rouder, J. N. (2022). *BayesFactor: Computation of Bayes factors for common designs* (R package Version 0.9.12-4.2) [Computer software]. https://cran.r-project.org/package=BayesFactor

Mossad, S. I., AuCoin-Power, M., Urbain, C., Smith, M. L., Pang, E. W., & Taylor, M. J. (2016). Thinking about the thoughts of others; temporal and spatial neural activation during false belief reasoning. *Neuroimage*, *134*, 320–327. https://doi.org/10.1016/j.neuroimage.2016.03.053

Peterson, D. M., & Riggs, K. J. (1999). Adaptive modelling and mindreading. *Mind & Language*, *14*(1), 80–112. https://doi.org/10.1111/1468-0017.00104

Phillips, J., Buckwalter, W., Cushman, F., Friedman, O., Martin, A., Turri, J., Santos, L., & Knobe, J. (2021). Knowledge before belief. *Behavioral and Brain Sciences*, *44*, Article e140. https://doi.org/10.1017/S0140525X20000618

Phillips, J., Ong, D. C., Surtees, A. D., Xin, Y., Williams, S., Saxe, R., & Frank, M. C. (2015). A second look at automatic theory of mind: Reconsidering Kovács, Téglás, and Endress (2010). *Psychological Science*, *26*(9), 1353–1367. https://doi.org/10.1177/0956797614558717

Poulin-Dubois, D., Rakoczy, H., Burnside, K., Crivello, C., Dörrenberg, S., Edwards, K., Krist, H., Kulke, L., Liszkowski, U., Low, J., Perner, J., Powell, L., Priewasser, B., Rafetseder, E., & Ruffman, T. (2018). Do infants understand false beliefs? We don't know yet—A commentary on Baillargeon, Buttelmann and Southgate's commentary. *Cognitive Development*, *48*, 302–315. https://doi.org/10.1016/j.cogdev.2018.09.005

R Core Team. (2022). *R: A language and environment for statistical computing* (Version 4.2.2) [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

Rubio-Fernández, P. (2013). Perspective tracking in progress: Do not disturb. *Cognition*, *129*(2), 264–272. https://doi.org/10.1016/j.cognition.2013.07.005

Rubio-Fernández, P. (2017). Can we forget what we know in a false-belief task? An investigation of the true-belief default. *Cognitive Science*, *41*(1), 218–241. https://doi.org/10.1111/cogs.12331

Rubio-Fernández, P., Mollica, F., Ali, M. O., & Gibson, E. (2019). How do you know that? Automatic belief inferences in passing conversation. *Cognition*, *193*, Article 104011. https://doi.org/10.1016/j.cognition.2019.104011

Ryskin, R. A., & Brown-Schmidt, S. (2014). Do adults show a curse of knowledge in false-belief reasoning? A robust estimate of the true effect size. *PLoS ONE*, *9*(3), Article e92406. https://doi.org/10.1371/journal.pone.0092406

Samuel, S. (2023). EXPRESS: A curse of knowledge or a curse of uncertainty? Bilingualism, embodiment, and egocentric bias. *Quarterly Journal of Experimental Psychology*, *76*(8), 1740–1759. https://doi.org/10.1177/17470218221132539

Samuel, S., Legg, E. W., Lurz, R., & Clayton, N. S. (2018). The unreliability of egocentric bias across self-other and memory-belief distinctions in the Sandbox Task. *Royal Society Open Science*, *5*(11), Article 181355. https://doi.org/10.1098/rsos.181355

Sassenrath, C., Sassenberg, K., & Semin, G. R. (2013). Cool, but understanding… experiencing cooler temperatures promotes perspective-taking performance. *Acta Psychologica*, *143*(2), 245–251. https://doi.org/10.1016/j.actpsy.2013.03.011

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

Schuwerk, T., Döhnel, K., Sodian, B., Keck, I. R., Rupprecht, R., & Sommer, M. (2014). Functional activity and effective connectivity of the posterior medial prefrontal cortex during processing of incongruent mental states. *Human Brain Mapping*, *35*(7), 2950–2965. https://doi.org/10.1002/hbm.22377

Schuwerk, T., Schecklmann, M., Langguth, B., Döhnel, K., Sodian, B., & Sommer, M. (2014). Inhibiting the posterior medial prefrontal cortex by rTMS decreases the discrepancy between self and other in theory of mind reasoning. *Behavioural Brain Research*, *274*, 312–318. https://doi.org/10.1016/j.bbr.2014.08.031

Senju, A., Southgate, V., White, S., & Frith, U. (2009). Mindblind eyes: An absence of spontaneous theory of mind in Asperger syndrome. *Science*, *325*(5942), 883–885. https://doi.org/10.1126/science.1176170

Sommer, M., Döhnel, K., Sodian, B., Meinhardt, J., Thoermer, C., & Hajak, G. (2007). Neural correlates of true and false belief reasoning. *Neuroimage*, *35*(3), 1378–1384. https://doi.org/10.1016/j.neuroimage.2007.01.042

Sommerville, J. A., Bernstein, D. M., & Meltzoff, A. N. (2013). Measuring beliefs in centimeters: Private knowledge biases preschoolers' and adults' representation of others' beliefs. *Child Development*, *84*(6), 1846–1854. https://doi.org/10.1111/cdev.12110

Spivey, M. J., Grosjean, M., & Knoblich, G. (2005). Continuous attraction toward phonological competitors. *Proceedings of the National Academy of Sciences*, *102*(29), 10393–10398. https://doi.org/10.1073/pnas.0503903102

Stillman, P. E., Shen, X., & Ferguson, M. J. (2018). How mouse-tracking can advance social cognitive theory. *Trends in Cognitive Sciences*, *22*(6), 531–543. https://doi.org/10.1016/j.tics.2018.03.012

Tamir, D. I., & Mitchell, J. P. (2013). Anchoring and adjustment during social inferences. *Journal of Experimental Psychology: General*, *142*(1), 151–162. https://doi.org/10.1037/a0028232

Todd, A. R., Hanko, K., Galinsky, A. D., & Mussweiler, T. (2011). When focusing on differences leads to similar perspectives. *Psychological Science*, *22*(1), 134–141. https://doi.org/10.1177/0956797610392929

van der Wel, R. P., 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

Wang, J. J., Ciranova, N., Woods, B., & Apperly, I. A. (2020). Why are listeners sometimes (but not always) egocentric? Making inferences about using others' perspective in referential communication. *PLoS One*, *15*(10), Article e0240521. https://doi.org/10.1371/journal.pone.0240521

Wang, L., & Leslie, A. M. (2016). Is implicit theory of mind the 'Real Deal'? The own-belief/true-belief default in adults and young preschoolers. *Mind & Language*, *31*(2), 147–176. https://doi.org/10.1111/mila.12099

Wojnowicz, M. T., Ferguson, M. J., Dale, R., & Spivey, M. J. (2009). The self-organization of explicit attitudes. *Psychological Science*, *20*(11), 1428–1435. https://doi.org/10.1111/j.1467-9280.2009.02448.x

Wulff, D. U., Kieslich, P. J., Henninger, F., Haslbeck, J., & Schulte-Mecklenbeck, M. (2021). *Movement tracking of cognitive processes: A tutorial using mousetrap*. PsyArXiv. https://doi.org/10.31234/osf.io/v685r

Received December 8, 2022
Revision received July 21, 2023
Accepted August 2, 2023 ■