Journal of Experimental Social Psychology 121 (2025) 104785

Contents lists available at ScienceDirect

Journal of Experimental Social Psychology

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

Registered Report Stage 2: Full Article

Testing "quarantined" metarepresentational accounts of Theory of Mind: Are we biased by others' false beliefs?☆

Steven Samuel [a,*], Robert Lurz [b], Daizi Davies [a], Harry Axtell [a], Sarah K. Salo [c]

[a] *Department of Psychology, School of Health and Medical Sciences, City St. George's, University of London, UK*
[b] *Department of Philosophy, Brooklyn College, CUNY, USA*
[c] *School of Psychological Science, University of Bristol, UK*

ARTICLE INFO

Keywords:
Theory of Mind
Belief representation
Belief simulation
Sandbox task
False belief task

ABSTRACT

An important component of Theory of Mind is the ability to understand the beliefs (true or false) of others. Arguably the most widely-held view is that this is performed by a detached *belief-representation* process (e.g., metarepresenting that another agent has a belief about the world which one does not share). The standard belief-representation account posits a separation between one's own first-order representations of the physical environment and one's second-order representations of another agent's mental states, preventing the latter from infecting the former. An alternative process is engaged *belief-simulation* (e.g., imaginatively adopting another agent's belief about the world) which, in contrast to standard belief-representation, posits a correspondence in the mental states shared by oneself and another agent and predicts an influence of the other agent's beliefs on one's own first-order representations and egocentric actions. In the first two of three studies, a participant and an agent watched an object buried in a continuous space (sandbox). The participant then watched the same object moved from the first location to a new location. When participants were asked to search for the object, they demonstrated a bias towards the first location when the agent falsely believed the object to be there but not when the agent knew, like the participant, that the object was in the new location. Reasoning that the strength of this bias may have been limited by participants' knowledge of the object's true whereabouts, in a third and final study we hid the movement of the object so that participants did not know its true location. We also recruited a greater number of participants to increase statistical power. Contrary to expectations, there was now no evidence of belief infection. Overall, these results are more consistent with a belief-representation account for (human) adults' understanding of others' belief states.

1. Introduction

In the classic false-belief paradigm, a child watches an agent hide an object in one location before it is moved to a new location without the agent's knowledge. The child is then asked to say where the agent thinks the object is located, with the correct answer being the original location, not the new location where the participant knows the object to be hidden (Baron-Cohen et al., 1985; Wimmer & Perner, 1983). Children typically pass this task around the age of four (Wellman et al., 2001); although, in the last decade or so, evidence has emerged that children as young as seven months, as well as nonhuman primates, demonstrate an implicit understanding of others' (false) beliefs (Hayashi et al., 2020; Kovács et al., 2010; Krupenye et al., 2016; Lurz et al., 2022; Onishi & Baillargeon, 2005). This ability to understand the unobservable mental states of others is usually referred to as 'Theory of Mind' (Premack & Woodruff, 1978) or ToM and is considered a bedrock of our ability to interact and communicate with others (Baron-Cohen, 1997; Clark & Brennan, 1991; Sperber & Wilson, 2002; Tomasello et al., 2005). Precisely how humans process others' beliefs, however, is not yet clear.

1.1. Standard belief-representation

Arguably the most widespread view of how humans process others' beliefs is that of belief-representation, which concerns the

☆ This paper has been recommended for acceptance by Lile Jia

* Corresponding author at: City St. George's, University of London, Dept. of Psychology, School of Health and Medical Sciences, Northampton Square, London EC1v 0HB, UK.

E-mail address: steven.samuel@citystgeorges.ac.uk (S. Samuel).

https://doi.org/10.1016/j.jesp.2025.104785
Received 1 February 2024; Received in revised form 23 June 2025; Accepted 23 June 2025
Available online 16 July 2025

representation of another agent's belief as such – for example, *Agent believes that [the gold is buried in the west side of the yard]* (Cosmides & Tooby, 2000; Kovács et al., 2021; Perner, 1991). By being about another agent's mental state of believing, belief representation is different from a first-order belief about the physical state of the world (e.g., that the gold is buried in the west side of the yard), which is what guides one's own egocentrically driven actions towards the environment. Instead, a representation of another's belief is a second-order representation – a representation of a representation, or metarepresentation (Perner, 1991; Pylyshyn, 1978). Crucially, according to the standard belief-representation account, since the belief (mental state) that the gold is in the west side of the yard is attributed to the agent, it should not influence one's own first-order representations about the physical state of the world (Lurz et al., 2022). For example, if one wants to find the gold and believes (or knows) that it is buried in the *east* side of the yard, one's first-order representational state that the gold is in the east side of the yard guides one's egocentric search for the gold in the east side of the yard. If, on the other hand, one also knows that another agent mistakenly believes the gold is buried in the west side of the yard, then one's second-order representational state about the other agent's belief (mental state) may allow one to predict that the agent will search in the west side of the yard, but it does not cause oneself to search in the west side of the yard (see Fig. 1, left panel). In effect, metarepresentations keep separate ("quarantine") the attitudes and content attributed to others from one's own first-order beliefs about the environment (Cosmides & Tooby, 2000; Frith et al., 1991). Thus, on the standard meta-representational theory, belief attributions are not predicted to influence a subject's own egocentrically driven actions towards the environment.

1.2. Belief-simulation

The principal alternative to standard belief-representation is the belief-simulation account (Gallese & Goldman, 1998; Goldman, 2006; Gordon, 1986). According to this view, the mental states of other agents are not represented as such but are vicariously experienced or impersonated. In "low-level" simulation, this occurs automatically and implicitly without any metarepresentational processes involved. In "high-level" simulation, it can be deliberative and may generate a second-order representation by running the pretend state through one's own cognitive processing, then taking the outcome state back "off line" and attributing it to the agent in question to predict the agent's behaviour (Goldman, 2006; Goldman & Jordan, 2013; Shanton & Goldman, 2010). In both cases, simulating believing that the gold is in the west side of yard is a first-order representational state about the gold, just like believing that gold is in the west side of the yard. As a result, simulating believing that the gold is in the west side of the yard may have effects on one's own egocentric behaviour towards the gold similar to believing that the gold is in the west side of the yard (see Fig. 1, right panel). This is in contradistinction to what is predicted on the standard, "quarantined" belief-representation account. Belief-simulation therefore predicts the altercentric "belief infection" or blending of the "true" and simulated first-order representations in a way that standard belief-representation does not (Lurz et al., 2022). Gallese and Goldman (1998) were some of the first researchers to recognize this distinctive altercentric prediction of belief-simulation vis-a-vis the standard belief-representation account (here, Theory-Theory). In explaining why similar muscle groups are activated in both a mind-reader and a target agent (Fadiga et al., 1995), Gallese and Goldman write:

> ST [Simulation Theory] postulates mental occurrences in the mind-reader that are analogous to mental occurrences in the target, so it

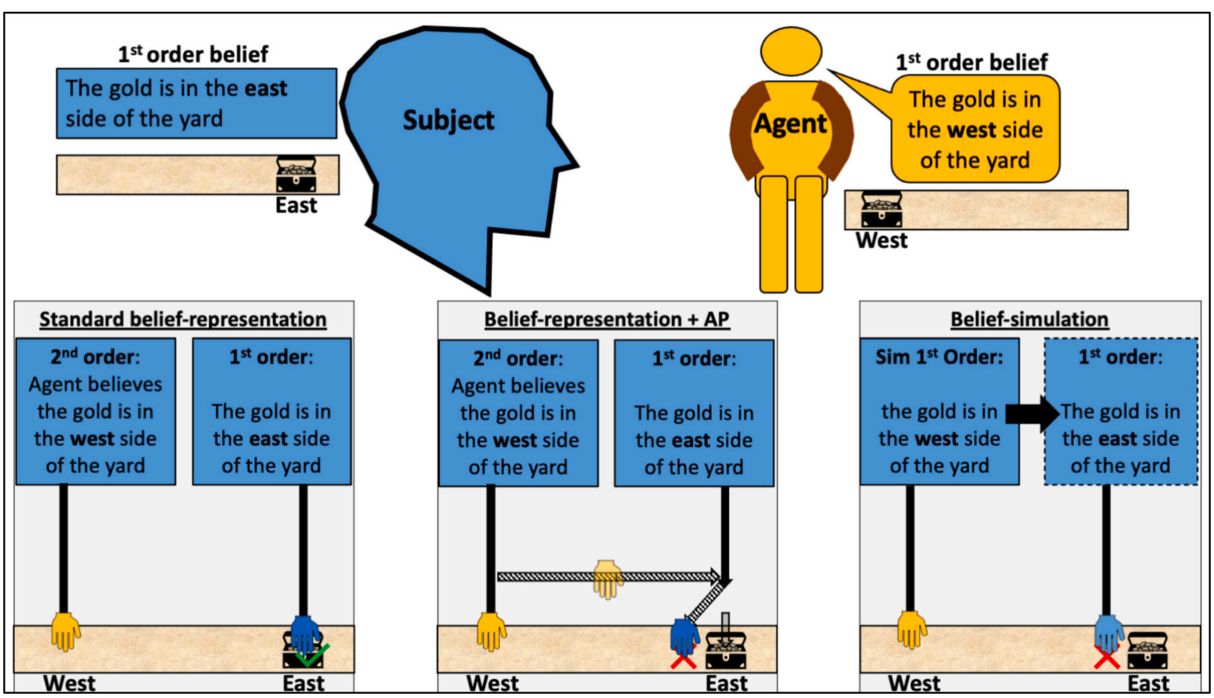

Fig. 1. AP = Action Prediction. When a subject is asked to search for an object (e.g., gold), another agent's belief can influence the subject's search behaviour under belief-representation + AP and under belief-simulation but not under standard belief-representation. Under standard belief-representation, metarepresentations about others' beliefs are kept separate from one's own first-order representations about the environment. Belief-simulation, on the other hand, posits that the agent's first-order belief about the location of the gold is simulated (imaginatively adopted) by the subject, producing a simulated first-order belief in the subject's mind that matches the content of the agent's first-order belief about the location of the gold, and thereby creating the opportunity for the subject's simulated first-order belief to bias the subject's own search behaviour for the gold. Under belief-representation + AP, first- and second-order beliefs are kept separate as per standard belief-representation, but the action prediction generated by the agent's belief can potentially interact with egocentric action planning and execution and subsequently bias the subject's own search behaviour. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

is not surprising that downstream motor activity is not entirely inhibited. If TT [Theory-Theory] were correct, and an observer represents a target's behaviour in purely theoretical fashion, it would not be predicted that the same muscle groups would be facilitated in the observer as in the target. But if ST were correct, and a mind-reader represents an actor's behaviour by recreating in himself the plans or movement intentions of the actor, then it is reasonable to predict that the same muscular activation will occur in the mind-reader. As matching muscular activation is actually observed in the observer, this lends support to ST as opposed to TT. (pp. 498, Gallese & Goldman, 1998).

1.3. A mediating process? Belief-representation plus action prediction

However, a further possibility is a belief-representation account that retains "quarantined" metarepresentations but allows that second-order belief representations might interact with one's own first-order representations about the environment via a mediating process. One candidate for this mediating process is action predictions (Southgate & Vernetti, 2014). Action predictions can be generated when someone understands that another agent has a goal (e.g. to find the gold), has a belief about where the gold is (e.g. that it is in the west side of the yard), and puts these two together to predict another's behaviour (the agent will search for the gold in the west side of the yard). This action prediction could then compete with one's own first-order motor representations, leading to a search that could tend towards the west side of the yard, even while the metarepresentation of the other agent's *belief* remains 100 % attributable to that agent (see Fig. 1 - center panel). Where this belief-representation plus action-prediction account (henceforth belief-representation +AP) differs from belief-simulation is therefore in how an influence of others' beliefs should arise. Specifically, belief-simulation implies and predicts the possibility of *belief infection* (i.e., the processing of another's belief as if it were one's own, perhaps unwittingly). No additional mechanism or processes need to be posited for an influence of others' beliefs to be explained. On the other hand, a belief-representation +AP allows for *action infection* and can arise in instances where egocentric action is planned or executed. In sum, both belief-simulation and belief-representation +AP predict that others' beliefs can influence one's own actions, but standard belief-representation does not. Standard belief-representation is therefore the only account that means that the subject's own search for the gold should not be affected by the subject's second-order representation of the agent's belief.

1.4. Evidence for an influence of others' perspectives on the self

To date, evidence for an influence of others' mental states on the self comes not from studies with beliefs but from studies on emotional empathy, whereby subjects share in another's (observed) emotional experience (Shamay-Tsoory, 2011; Singer et al., 2004), and from visuospatial perspective taking tasks, whereby subjects imaginatively adopt ("embody") the perspective of another person (Deroualle et al., 2015; Kessler & Thomson, 2010; Samuel, Legg, Manchester, Lurz, & Clayton, 2019; Surtees et al., 2013; Yu & Zacks, 2017). Being about cognitive rather than affective perspectives, the evidence from visuospatial perspective taking is likely to be more informative than emotional empathy about the possibility that others' (unemotional) beliefs about the world might also influence one's own egocentric behaviour. Here, a number of studies have reported results in favour of the possibility that other agents' perspectives can impair or facilitate the speed of self-perspective processing depending on whether the content of the alternative perspective is consistent (or not) with what one is attempting to process about the environment (Freundlieb et al., 2017; Freundlieb et al., 2018; Samson et al., 2010; Ward et al., 2019; Ward et al., 2020). However, some of the interpretations of these results about other agent's perspectives have been questioned, with alternatives such as domain-general processes proposed instead (Cole et al., 2020; Cole et al., 2022; Cole & Millett, 2019; Santiesteban et al., 2014). In addition, variations in response times reflect processing difficulty but do not tell us how beliefs are processed (i.e., whether they are represented or simulated). More relevant to such questions are perspective-taking studies that take as their principal measure accuracy and the nature of errors rather than the speed of responses. Here too the evidence favours the possibility that other agent's perspectives influence one's own egocentric behaviour. For example, subjects sometimes erroneously make manual responses consistent not with their own perspective but with that of a recently imagined alternative viewpoint of the same environment, even when the task is to respond entirely egocentrically (Samuel, Legg, Manchester, Lurz, & Clayton, 2019). In another study, it has been shown that taking the divergent visual perspective of another agent causes subjects to approximate the abstract beliefs of that agent and to report feeling more similar to the agent (Erle et al., 2018; Erle & Topolinski, 2017). Results like these suggest a degree of perspective blending that is predicted by mental simulation and, in the case of "infected" manual responses, also belief-representation + AP. However, neither is predicted by standard belief-representation.

Recently, evidence has emerged that others' beliefs can influence egocentric behaviour in chimpanzees. In a simple object-search task, Lurz et al. (2022) found that chimpanzees were biased in their own egocentric searches for hidden food by the false belief of another agent. In the false-belief condition of the task, in which the other agent mistakenly believed the food was still hidden in the original location but the chimpanzee knew it had been moved to a new location, chimpanzees were biased in their own search for the food towards the original location (where the agent falsely believed the food to be hidden), and a significant proportion of chimpanzees searched for the food in the original location area. This effect disappeared in the true-belief condition in which the other agent, like the chimpanzee, knew the food had been moved to the new location. The false belief of the observing agent appeared to influence the chimpanzees' own egocentric searches as the belief-simulation account would predict. If correct, this suggests that simulation is evolutionarily old, shared by humans' common ancestor with the chimpanzees.

Evidence for the permeability of one's own beliefs by those of other agents would be consistent with other theories that also propose altercentric influences on the self. For example, a number of researchers have suggested that others' visual perspectives can interfere with the efficient processing of one's own visual perspective (Elekes et al., 2016; Furlanetto et al., 2016; Samson et al., 2010), reduce the prominence of that perspective (Quesque et al., 2018; Tversky & Hard, 2009) and even complement that perspective with additional information gleaned from the other agent (Ward et al., 2019; Ward et al., 2020), though all these possibilities are debated (Cole et al., 2020; Cole & Millett, 2019; Samuel et al., 2021). Belief infection would also be broadly consistent with some social psychological theories, albeit with some important caveats. For example, deindividuation implies a reduction of one's sense of unique identity in certain group settings (Vilanova et al., 2017), a phenomenon that belief-simulation could contribute to, though not by itself explain. Theory of Collective Mind proposes the psychological amplification of representations which are shared ('we-representations') relative to egocentric representations ("I-representations") which are not (Shteynberg et al., 2023). Amplification could potentially be explained by simulation, whereby one's own belief and another's combine (overlap) to increasing effect. However, as Theory of Collective Mind appears to follow a standard belief-representation model rather than simulative one, evidence for belief infection may contradict this aspect of it. It is also unclear whether first-person 'plural' we-representations should be quarantined from first-order 'singular' representations in the same way that representations about other people are (i.e., a priori), given the overlapping role of the self in each.

The studies reported here are designed to adjudicate between the standard belief-representation account on the one hand, and either the belief-simulation *or* belief-representation + AP account on the other. To do so, we utilize the different predictions that the standard belief-representation account makes compared to these two alternatives. We gave young adult participants (henceforth capitalized *Participants*) an object-search task in which a ball is first hidden in one location (Location 1) in a continuous space (sandbox) but then moved to a second location (Location 2) before the Participant is instructed to search for it. We also manipulated the knowledge of a second participant, henceforth the *Agent*, by varying whether the Agent witnessed the ball's change of location. A positive detection of belief infection would occur if Participants search for the ball in a section of the sandbox closer to Location 1 on trials in which the Agent falsely believes the ball to be located there than on true-belief trials in which both the Participant and Agent know the ball to be in Location 2. A standard belief-representation account would not predict such results; the crucial assumption that second-order representations about others' unobservable mental states are detached would be compromised if their contents could so easily (and erroneously) "leak" this way into first-order representations about the physical environment. Instead, either the belief-simulation or belief-representation + AP account could explain this finding. On the other hand, an absence of evidence of belief infection would instead be consistent with belief-representation, though it would not constitute direct support for it.

The principal advantage of such a design is that it relies not on variation in response times, which reflect fluctuations in processing demands, but on accuracy, which can reveal crucial information about the nature of processing. Indeed, the sandbox task was adapted by Daniel Bernstein and his colleagues in order to detect the reverse of what we are investigating, i.e. an influence of the self-perspective in tasks where participants are asked to indicate where someone else will search in the trough (Mahy et al., 2017; Sommerville et al., 2013), also known as egocentric bias. Through our task we should therefore be able to identify how Participants understand others' beliefs. We used adults rather than children because their Theory of Mind abilities (whether simulative or metarepresentational) are more likely to be fully developed, less likely to be susceptible to interference from non-social processes such as poorer executive functioning, and because even healthy adults are susceptible to errors and biases on Theory of Mind tasks with measures sensitive enough to detect them (e.g., Apperly et al., 2010; Dietze & Knowles, 2020; Keysar et al., 2003; Samuel, 2023; Samuel, Roehr-Brackin, et al., 2019; Sommerville et al., 2013; Wardlow Lane & Ferreira, 2008).

In Experiment 1, we added an extra condition, based on research described in the Introduction that suggests that embodying another agent's divergent perspective might facilitate the blending of first- and second-order perspectives (Erle et al., 2018; Erle & Topolinski, 2017; Samuel, Legg, Manchester, Lurz, & Clayton, 2019). Specifically, the Participant and Agent each wore a different-colored glove on their dominant hand. In the Embodiment condition, Participants and Agents faced opposite directions and Participants were asked on which hand the Agent wore his or her glove ("left" or "right"). Facing in opposite directions was necessary in order to ensure divergent left/right perspective, which provides the best conditions for embodied perspective taking to occur (Kessler & Thomson, 2010) and promote self-other merging at other levels, such as trust, liking, and estimates (Erle et al., 2018; Erle & Topolinski, 2017). The embodiment induction question occurred immediately prior to responses on the critical false-belief trials. In the Control condition, Participants and Agents again faced in opposite directions but Participants were simply asked to name the colour of the glove on the Agent's hand. The Control question required participant to considered something about the other Agent but unlike the Embodiment question it did not require Participants to consider the other Agent's divergent perspective. We hypothesized that the potential for the Agent's belief to influence egocentric behaviour could be stronger in the Embodiment condition than in the Control condition because previous research had shown that embodied perspective-taking promotes the blending of perspectives. This itself might arise through a simulative process, thereby increasing activation of the agent's perspective and any resulting bias towards Location 1. Additionally, this contrast would help establish whether any such influence might in fact require prior embodiment to arise.

2. Experiment 1

2.1. Method

2.1.1. Participants

An a priori power analysis using G*Power (3.1.9.5) for a paired-sample *t*-test with 80 % power to detect a medium effect size ($d = 0.5$) found that 34 Participants were required. Therefore 34 pairs were recruited (data from the second of the pair, the Agent, were unimportant and discarded). The final sample was 34 young adults (8 identified as men; 26 as women) recruited through the University of Plymouth's online recruitment system. Participants averaged 20.6 years of age ($SD = 3.1$; range = 18–32) and were native English speakers with normal or correct-to-normal vision. All gave informed consent and were compensated financially for time spent in the lab. Three original Participants were replaced for mean biases beyond the preregistered limit of three times the SD of the mean for the group as a whole. Ethical approval was granted by the University of Plymouth's Ethics Committee. In this and all subsequent studies, all measures, manipulations and exclusions are reported either here or in Supplementary Materials.

2.1.2. Materials and procedure

One of each pair of recruits was randomly assigned to the role of either Participant or Agent. Neither knew to which role they had been assigned. Each received either a pink or blue rubber glove (counterbalanced across participants) that they were told to put on their dominant hand for the duration of the task.

At the start of every trial, the Participant and the Agent stood side by side on one side of a black trough measuring 100 cm (90 cm inside) x 22 cm × 18 cm and filled 15 cm high with play-safe sand, which was watered frequently so that marks made by fingers would be clearly visible. The experimenter stood on the opposite side of the trough facing the Participant and Agent. A tape measure in centimetres was secured along the top rim of the trough facing the experimenter, allowing only the experimenter to see it. The tape measure facilitated accurate ball placement and later measurement of Participants' responses from photos the experimenter took. Fig. 2 displays one such photo.

Trial procedure. The Participant and Agent were run through ten trials in total (see Table 1). Each trial *always* included the following, though other details varied according to trial type. First, both the Participant and the Agent watched the experimenter bury a ping-pong ball into the sand, called Location 1 (see Table 1). The experimenter then smoothed the sand over. Later in the trial the experimenter dug up the ball and moved it to the new location (Location 2) before smoothing the sand over again. Finally, either the Participant or the Agent was asked to search for the ball. This search instruction was the same across all trials: "*please make a mark in the sand where you think the ball is.*" The searcher then placed a finger in the sand where he or she believed the ball to be. The experimenter took a photo of the sand, using a smartphone and always from the same approximate angle (see Fig. 2 for an example), for later coding, and the trial ended. Note that Participants (and Agents) were only ever asked to indicate where they themselves believed the ball to be (i.e., they were never asked to consider the other's perspective).

Practice trials. The first two trials were always practice trials in which both the Participant and Agent witnessed the experimenter bury the ball in Location 1 and then Location 2. The Participant made the response on one trial, the Agent on the other. Although only data from trials where the Participant responded were analysed, having the Agent

Fig. 2. Example of images used for coding participant responses.

Table 1
Location and Trial list denotes both locations of the ball, before and after the location change in centimetres, and who is responding after location change. Trials in bold indicate the critical False Belief (FB) trials and their location-matched True Belief (TB) Baseline control trials (for half of Participants, trials 5 and 6 were False Belief and 9 and 10 were True Belief Baseline). For half the participants the first False Belief trial included the Embodiment question and the second the Control question, and for the other half the reverse was the case.

Trial	Location 1 (cm)	Location 2 (cm)	Distance (cm)	Searcher
1 Practice	50	70	20	Agent
2 Practice	60	20	40	Participant
3 TB Filler	70	30	40	Agent
4 TB Filler	30	50	20	Participant
5 TB Baseline	**80**	**30**	**50**	**Participant**
6 TB Baseline	**20**	**70**	**50**	**Participant**
7 TB Filler	20	50	30	Agent
8 TB Filler	40	10	30	Participant
9 FB	**80**	**30**	**50**	**Participant**
10 FB	**20**	**70**	**50**	**Participant**

respond in these early trials was important to demonstrate that the Agent could be asked to interact with the trough on any given trial, which could help to activate the Agent's belief (and any linked action prediction) even when the Participant was the searcher. If on practice trials either the Participant or Agent indicated a location greater than 10 cm off target (on either side), the practice trials were repeated once. If he or she failed again, the experiment would end, and the pair would be replaced. However, this never occurred. After the practice trials, there were three further trial types.

True Belief Filler trials. The 3rd, 4th, 7th and 8th trials were always True Belief Filler trials (see Fig. 3). These were not included in the final analyses, but served to maintain the Agent's interaction with the trough beyond the practice trials throughout the length of the experiment. True Belief Filler trials were conceptually identical to practice trials (albeit with different locations - see Table 1) except that following the burial of the ball in Location 2 the non-searcher was asked to leave for about ten seconds and then return. This brief absence did not create a false belief, as both Participant and Agent had already witnessed the ball being buried in Location 2. True Belief Filler trials were used to familiarise participants with the procedure before the experimental trials later in the sequence. After the non-searcher returned, the experimenter then asked whoever was designated as the searcher on that trial to search for the object.

True Belief Baseline trials. True Belief Baseline trials were included in the final analyses, and were the baseline from which any bias on False Belief trials were measured. There were two True Belief Baseline trials (see Fig. 3) that always occurred as a pair at either the 5th and 6th or 9th and 10th trial positions in the experiment. These trials were identical to True Belief Filler trials except that Locations 1 and 2 were matched with the False Belief trials to facilitate comparisons. Note that having the Agent leave and return after the ball was relocated to Location 2 in True Belief trials (both Filler and Baseline) served to eliminate a criticism that has been levelled at some ToM tasks in which the Agent (or equivalent) leaves and returns on False Belief trials but not on True Belief trials. It has been argued that the salience of the Agent's return in False Belief trials diminishes the Participant's memory of what happened before (i.e., the burial of the ball in Location 2) relative to True Belief trials where the agent stays (Heyes, 2014).

False Belief trials. False Belief trials were also a pair of trials, always at the 5th and 6th or 9th and 10th slots (counterbalanced order with True Belief Baseline trials across Participants). Crucially, on False Belief trials the Agent left *before* the ball was dug up and buried in Location 2, and he or she did not return until after the ball had been buried in Location 2 and the sand smoothed over (see Fig. 3). This also took about ten seconds to match the length of time the Agent was absent on True Belief Baseline trials. On False Belief trials alone, the Agent had a false belief that the ball was still in Location 1.

Additionally, upon returning, the Agent was asked to face briefly away from the trough. In one of the False Belief trials, the Embodiment question was asked ("*Which hand is the other participant wearing the glove on*?"); in the other False Belief trial, the Control question was asked ("*What is the colour of the glove the other participant is wearing*?"). The order of condition questions was counterbalanced across Participants, such that for half the participants the first False Belief trial included the Embodiment question and the second the Control question, and for the other half the revere was the case. Finally, the Agent was instructed to turn to face the trough again, and the Participant was then asked to indicate where the ball was. The task took approximately 10–15 min to complete.

2.1.3. Data analysis

Photos were labelled by Participant number and trial number but not by trial type (True Belief Baseline or False Belief) or condition (Embodiment or Control). As such, coding of response location was already part-blinded. As an additional layer of blinding, after the experiment was completed, the photos were saved again under randomly generated numbers before being coded by two researchers on the project, using the tape measure in the photograph alongside the hole made by the participants' fingers to judge where the search was made. The mean of these two judgments then re-associated with the original trial types by the first author by reversing the randomization process and using a trial type key. The first author adjudicated on any disagreements that exceeded 1 cm (three occasions). The only exception to this process concerned the photos from the three replacement participants, which were not saved under random numbers as they concerned only a small number of photos. The coding data is available here and in the Supplemental Materials. The bank of photographs is available here.

These data formed the dependent variable, measured in centimetres.

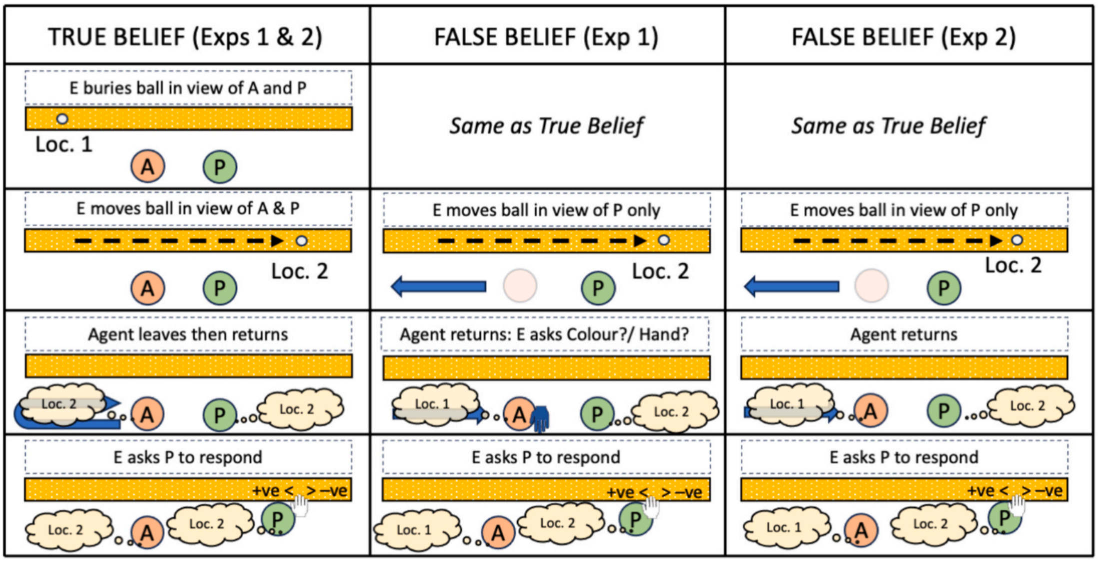

Fig. 3. Top-down schematic of lab for both True Belief type trials (Filler and Baseline) on the left and any changes made from these True Belief trials for the False Belief trials in Experiments 1 and 2, to the right. A = Agent, E = Experimenter, P = Participant. See text for full details of procedure.

It was then possible to compare the distance in centimetres towards Location 1 from Location 2 on False Belief trials with the same on True Belief Baseline trials. Support for an influence of the Agent's belief would come from bias towards Location 1 that is higher on False Belief trials than on True Belief Baseline trials. Additionally, a larger effect in the Embodiment condition than in the Control condition would suggest that taking someone's physical perspective serves to increase the influence of their false belief on one's own actions.

2.1.4. Transparency and openness

Details of the preregistration for the methodology and analyses of Experiment 1 can be found here: https://osf.io/dvna7. Data were analysed using JASP (version 0.16.4), and the data and JASP analysis file are available here: https://osf.io/5y96h/files/osfstorage.

2.2. Results

All results are displayed in Fig. 4.

Control condition. On False Belief trials, Participants indicated a location that was closer to Location 1 ($M = -0.3$ cm, $SD = 2.1$) than on True Belief Baseline trials ($M = -1.4$ cm, $SD = 2.9$). This difference was significant, $t(33) = 2.595$, $p = .014$, 95 % $CI_{Diff} = [-0.5, -4.5]$, $BF_{10} = 1.254$, with a close-to-medium effect size, $d = 0.445$. A planned Wilcoxon signed-rank comparison, conducted in light of the non-normality of one cell in the data, supported this result, $W(34) = 158.5$, $z = 2.376$, $p = .018$, d[1] $= 0.467$. Of the 34 participants, a majority of 23 (68 %) displayed this pattern. This effect is consistent with the hypothesis that Participants are influenced in their own egocentric search behaviour by another person's false belief; though it is noteworthy that both *absolute* scores were negative (i.e., further away from Location 1 than Location 2 - see bottom of Fig. 3).

Embodiment condition. The mean location indicated by Participants was similar on both True Belief Baseline trials ($M = -1.3$, $SD = 4.1$) and False Belief trials ($M = -1.5$, $SD = 3.1$). This difference was not significant, $t(33) = 0.743$, $p = .46$, 95 % $CI_{Diff} = [-3.7, 1.7]$, $BF_{10} = 0.201$, with a small effect size, $d = 0.127$. Of the 34 participants, precisely half displayed this pattern. The Bayesian test confirmed that the data was at least three times more likely under the null hypothesis. A planned Wilcoxon signed-rank comparison, conducted in light of the non-normality of one cell in the data also supported this null result, $W(34) = 296.5$, $z = 0.017$, $d = 0.003$, $p = .99$. Again, both scores were negative (i.e., mean locations were further from Location 1 than Location 2).

Both conditions combined. The mean location indicated by Participants was similar on both True Belief Baseline trials ($M = -1.3$, $SD = 2.3$) and False Belief trials ($M = -0.9$, $SD = 1.9$). This difference was not significant, $t(33) = 1.198$, $p = .24$, 95 % $CI_{Diff} = [-1.1, 0.3]$, $BF_{10} = 0.355$, with a small effect size, $d = 0.206$. Of the 34 participants, a majority 20 (59 %) displayed this pattern. The Bayesian test were found that the data were almost three times more likely under the null hypothesis. Again, both scores were negative (i.e., mean locations were further from Location 1 than Location 2).

Comparing conditions. Results had shown a bias towards Location 1 on False Belief trials over True Belief Baseline trials, but only in the Control condition, not the Embodiment condition. Thus, contrary to one of our hypotheses Participants did not experience a greater influence of the Agent's false belief in the Embodiment condition; indeed, evidence of such an influence was in fact exclusive to the Control condition. A comparison of these two conditions found no difference between bias in the Control condition ($M = 1.1$ cm, $SD = 3.1$) compared to the Embodiment condition ($M = -0.3$ cm, $SD = 4.0$), $t(33) = 1.384$, $p = .176$, 95 % $CI_{Diff} = [-0.7, 3.5]$, $BF_{10} = 0.439$, with a small effect size, $d = 0.237$. The Bayesian test suggested the data were approximately twice as likely under the null than the alternative hypothesis. Though this did not meet our pre-registered threshold of three times as likely, it is important to note that the hypothesis was that influence would be greater in the Embodiment condition than the Control condition when in fact the data patterned in the opposite direction.

2.3. Discussion

Results yielded support for an influence of the Agent's false belief in the Control condition but not in the Embodiment condition. This suggests that another's false belief intrudes upon our ability to act according to our own reality, though surprisingly only in the Control condition and not when Participants had previously considered the physical perspective of the other individual. We had hypothesized that the search bias of Participants should be greater in the Embodiment condition than in the Control condition because the former condition required Participants to take the physical perspective of the Agent prior to making their

[1] The effect size (d) given for Wilcoxon tests here and throughout is the Hodges-Lehmann estimator.

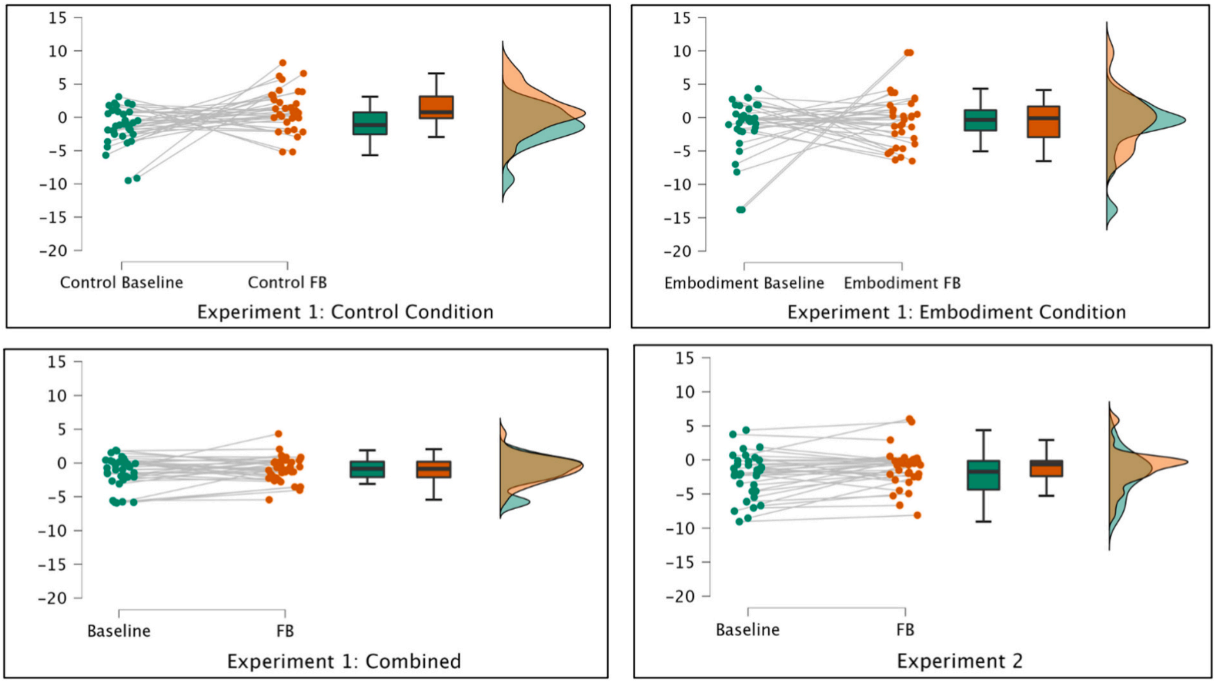

Fig. 4. Results of Experiments 1and 2. The y axis shows bias towards Location 1 (+ve) or away from it (−ve).

response, which should promote a stronger blending of perspectives (Erle & Topolinski, 2017). One reason for this divergence could be that the study by Erle and colleagues concerned an effect of physical perspective taking on the similarity of participants' numerical *guesses*, such as the height of the Eiffel Tower in feet, but in the present studies we employed belief states. Guesses are qualitatively different from belief states in that the former do not come with a commitment to their veracity. Guesses may therefore not fall within the purview of belief-simulation. A further reason for the discrepancy could be methodological. Erle et al. asked their participants to answer the left/right question about the other agent using their own left or right hand, but in our study, Participants responded verbally. This meant that Participants in our study did not have to "mirror" the Agents' perspectives in the same way, which may have diluted any effect of physical perspective taking (for a similar case and argument see Samuel, 2022). Alternatively, the spatial label "left"/"right" that Participants produced in the Embodiment Condition may have introduced spatial interference that limited the impact of the Agent's false belief about where the ball was. It may also be that our embodiment condition simply failed to activate a strong enough response; although our embodiment induction was conceptually consistent with previous research this particular form of embodiment induction has not been independently validated.

Critically, although the Embodiment condition failed to reveal a belief-infection effect it was present in the Control condition. This would support the belief-infection and belief-representation + AP accounts but would not appear to fit a standard belief-representation account. However, as this pattern of results was unexpected, we ran a similar experiment with a new sample to test the reliability of this effect. We therefore retained the original hypothesis that bias towards where the Agent believes the object to be should be greater on False Belief trials than on True Belief Baseline trials. However, to improve the statistical power of the experiment, the Embodiment and the Control condition questions were removed, and the two False Belief trials combined into a single analysis. Experiment 2 therefore resembled a lower-demand version of the Control Condition in Experiment 1.

3. Experiment 2

Experiment 2 was identical to Experiment 1 with the sole exception that no Embodiment or Control question was asked prior to responding. These questions were not replaced with other questions. As a result, Participants and Agents no longer wore gloves, and upon returning to the room on False Belief trials, the Agent always faced the trough and the instruction to the Participant to search for the ball followed immediately. The procedure for coding the photographs was identical to Experiment 1 except this time the two coders were one of the same researchers on the project as in Experiment 1 and the first author. Both scores were averaged for analysis and within 1 cm of each other with only one exception, adjudicated in favour of the first author.

3.1. Method

3.1.1. Participants

We retained the original power analysis (80 % power to detect a medium effect size $d = 0.5$ with a paired-sample *t*-test) requiring 34 Participants. The final sample was 34 young adults (8 identified as men; 25 as women, 1 non-binary) recruited through the University of Plymouth's online recruitment system. Participants averaged 19.5 years of age ($SD = 1.3$; range = 18–23). All were native English speakers with normal or correct-to-normal vision, and all gave informed consent and were compensated financially for time spent in the lab. One original Participant was replaced for mean biases beyond the preregistered limit of three times the SD of the mean for the group as a whole. Ethical approval was granted by the University of Plymouth's Ethics Committee.

3.1.2. Transparency and openness

Details of the preregistration for the methodology and analyses of Experiment 2 can be found here: https://osf.io/z8g6p. Data were analysed using JASP (version 0.16.4), and the data and JASP analysis file are available here: https://osf.io/5y96h/files/osfstorage.

3.2. Results

Results are displayed in Fig. 4 (bottom right panel). Consistent with

Experiment 1, Participants indicated a location that was closer to Location 1 on False Belief trials ($M = -1.2$ cm, $SD = 2.8$) than on True Belief Baseline trials ($M = -2.2$ cm, $SD = 3.3$). This difference was significant, $t(33) = 2.469$, $p = .019$, 95 % $CI_{Diff} = [-1.9, -0.2]$, $BF_{10} = 2.526$, with a close-to-medium effect size, $d = 0.423$, though again it is noteworthy that both scores were negative. Of the 34 participants, a majority of 24 (71 %) displayed this pattern. This result replicates that found for Experiment 1 and is therefore also consistent with the hypothesis that Participants are influenced in their own egocentric search by the other person's false belief.

We also conducted an additional experiment in which we used the same setup as for Experiment 1 but instead of asking participants to search for the object we asked them to indicate where in the trough the Agent believed the object to be, i.e. a "classic" and explicit false belief version of the sandbox paradigm (e.g., Samuel et al., 2018a; Sommerville et al., 2013). We report the details in supplemental material to this article as it concerned a different research question, but for the present purposes it demonstrates that Participants in this paradigm have no problem accurately recalling Agents' false beliefs in this task. Results showed that they were very accurate in both conditions, with no statistically significant evidence of bias towards the location the Participant knows the object to be (i.e., no egocentric bias). These results confirm that any biases in this paradigm can indeed be explained by others' beliefs about the where the ball is.

3.3. Discussion

The results of Experiment 2 replicated those of Experiment 1. Participants' judgments about the location of the object were closer to Location 1 when the Agent falsely believed the object to be in Location 1 compared to when the Agent (correctly) believed the object to be in Location 2. In sum, both experiments showed a reliable bias towards where the Agent falsely believed the target to be when no extra embodiment question is asked. This bias carried a close-to-medium effect size in both experiments, with the smallest being $d = 0.423$. That this effect was generated by the Agent's belief rather than spatial interference from the first location was evidenced by there always being two locations in the trough where that the ball was buried on each trial, but bias was only reliable when the Agent falsely *believed* the target to be in the Location 1. Moreover, data from an additional experiment showed that Participants have no problem accurately recalling the location the Agent falsely believes the ball to be in this paradigm. This rules out the possibility that biases might reflect confusion about where Agents believe the ball to be on false belief trials. Together, the results of Experiments 1 and 2 suggest that one's egocentric actions can be influenced by others' mental states, and that a standard belief-representation account on its own is unlikely to explain how Agent's beliefs were processed.

An unexpected finding from Experiments 1 and 2 was that absolute bias scores were usually negative; that is, Participants indicated a location further away from Location 1 than Location 2. Though not predicted by any theory of mental state reasoning, negative scores in the sandbox task have been reported before (Haskaraca et al., 2023; Mahy et al., 2017; Samuel et al., 2018b) and are suggestive of an additional, spatial bias on responses. Indeed, early versions of the sandbox task were concerned not with others' mental states but with the potential to subdivide continuous spaces into segments that bias object location recall towards their centres (Huttenlocher et al., 1991; Huttenlocher et al., 1994; Huttenlocher et al., 2004). Mentally subdividing the trough into segments, or a general tendency to search closer to the edges of the trough than the centre, could explain why biases were often negative in the present studies. However, the important consideration here is that mean responses were reliably closer to Location 1 when the Agent falsely believed the ball to be in Location 1 than when the Agent (correctly) believed it to be in Location 2, despite the two locations being identical across both trial types. This eliminates the possibility that spatial biases can explain biases searching on False Belief trials.

The findings of Experiments 1 and 2 differ from that of a very recent study by Haskaraca et al. (2023). In their experiments, adults performed online versions of three sandbox tasks in which they were presented with short stories or video skits about two individuals, one of whom sometimes had a false belief about a target object's whereabouts. Participants were instructed to click on an area of the screen in response to questions about where the individual with a false belief would search for the object *or* where the object is now. Bias towards the second location in the former trials were taken as a measure of egocentric bias, and bias towards the first location in the latter trials were taken as a measure of altercentric bias. The latter of these instructions is the most similar to the instructions in the Experiments we report here. Haskaraca and colleagues found no reliable evidence of bias towards where the other agent believed the object to be in any of three experiments, including when both altercentric and egocentric question types were mixed (which could have made the other's belief more salient on trials where first-order representations were required). However, despite being conceptually similar tasks, there are considerable methodological differences between our studies and theirs that could explain our divergent findings. Probably the most important of these distinctions is that our versions were conducted in the lab with a real human other. Previous research has shown that people respond differently to present others compared to depicted others (Skarratt et al., 2012), and it is reasonable to assume that others' beliefs are more likely to be simulated or their actions more likely to be predicted in real social rather than virtual contexts – particularly, where the other person is not only real but also co-present at the time Participants made their judgments. Other differences include the performance of a distractor task between the hiding of the object in its final location and participants' responses, which was included to eliminate more perceptual response strategies. The extra cognitive load generated by this task could have made the simulation of belief or the generation of an action prediction less likely (indeed, perhaps the Embodiment question in our Experiment 1 eliminated bias for the same reason).

Nevertheless, we interpret the results of Experiments 1 and 2 so far as offering only preliminary evidence of an influence of others' beliefs on egocentric action, for the following reasons. Firstly, the absolute effect in terms of the size of the trough was small under these circumstances, around 1 % of its length. Secondly, Bayes Factor analyses did not provide convincing support (defined as $BF_{10} > 3$) for the experimental hypothesis in Experiment 1 or 2. Given these caveats, a modified third and final study was conducted. Firstly, because Participants knowledge that the ball is in Location 2 may be too salient and therefore mask any simulation concerning Location 1, we removed from Participants the knowledge of where the ball actually was. Participants now never saw the ball moved to Location 2, and therefore they could only guess its whereabouts when responding. However, Agents would either hold a false belief that the ball was still in Location 1 (False Belief trials) or, like the Participant, have no knowledge of its whereabouts (No Belief trials). An influence of others' beliefs would be supported if Participants' guesses are closer to Location 1 on False Belief trials than No Belief trials. This time, as there was no first-person knowledge to overcome when Participants made their guesses, belief infection effects could be larger than in Experiments 1 or 2. Additionally, as Participants had no knowledge of the object's true location any spatial biases that may have arisen out of attempts to recall Location 2 were now irrelevant; indeed, bias could not be positive or negative at all as there was no true location to measure from. Finally, Experiment 3 was powered to detect the lowest effect size previously found ($d = 0.423$ rather than a hypothesized $d = 0.5$) with a higher degree of confidence (90 % as opposed to 80 %). A replication of the results of Experiments 1 and 2 would therefore increase confidence in the underlying phenomenon, while an absence of an effect under these stricter conditions could indicate that the earlier findings were spurious. As before, evidence of bias would be interpreted as contrary to standard belief-representation but consistent with the belief-simulation or belief-representation + AP.

4. Experiment 3

4.1. Method

4.1.1. Participants

While Experiments 1 and 2 were pre-registered, Experiment 3 was conducted as part of a Registered Report extension to further test the reliability and robustness of belief infection.[2] We conducted an a priori power analysis using G*Power (3.1.9.7) for a two-tailed paired-sample *t*-test, assuming an alpha of .05 and 90 % power to detect the lowest effect size from a control (i.e., non-embodiment) condition ($d = 0.423$ from Experiment 2). The analysis found that 61 Participants were required. This increases to 64 if a Wilcoxon test is used instead of a *t*-test. Since the latter analysis would be applied if a Shapiro-Wilks test found the distribution of difference scores to be non-normal, data were collected until 64 pairs of participants are recruited (Participant *M* Age = 19.6 years, 5 men, the rest female). As previously, all were required to be native English speakers with normal or correct-to-normal vision. All gave informed consent and received course credit or were compensated financially for time spent in the lab. After data from 64 Participants were collected, we planned an outlier check. Participants whose mean biases exceed three times the SD of the mean for the group as a whole (higher or lower) would be replaced, and the analysis for outliers repeated again after each replacement to determine whether new outliers are produced after new data is added (as SDs change). This process would be repeated until all 64 participants fall within three SDs of the mean. No data needed to be discarded following this initial analysis. Ethical approval was granted by the City St. George's, University of London's Psychology Research Ethics Committee.

4.1.2. Materials and procedure

At the beginning of Experiment 3, the Experimenter told the Participant and the Agent that they must not speak during the task. This was to ensure no sharing of privileged information (e.g., if the Agent asks if the ball was moved while they were absent). At the beginning of every trial, the Experimenter buried the ball in Location 1 while both the Participant and the Agent watched, as previously. What happened next depended on which of three conditions that trial was part of.

Training trials. The first four trials were always "Training" trials, in which the Experimenter buried the ball in Location 1 in full view of both the Participant and Agent and then placed a barrier (drew a curtain) so that neither the Participant nor the Agent could see the sandbox or the Experimenter (See Fig. 5). The Experimenter then said the following: "*The ball has NOT moved. I will ask one of you to indicate where you think it is*." The Experimenter emphasised "NOT." The Agent was then instructed to leave the room. After 20 s, the Experimenter asked the Agent to return. Then, the Experimenter said, "*please make a mark in the sand where you think the ball is.*" For the first four Training trials the Agent fulfilled this request. After the response was made, the Experimenter dug up the ball to show where it was. If the response was within 10 cm of Location 1, the Experimenter said, "Thank you," and began the next trial. If the response missed the target by more than 10 cm in either direction, the Experimenter said, "*That is more than 10cm off target - let's try that one again.*" That trial was then repeated. Training trials create in the Agent the sense that when the Experimenter states that the ball has not moved it really has not moved, and that the Agent should try to indicate Location 1 as accurately as possible (or face a repetition of that trial).

No Belief trials. "No Belief" trials occurred at either trial positions 5 and 6, or trial positions 9 and 10 (counterbalanced across participants with False Belief trials). Different from Training trials, once the barrier was in place in No Belief trials, the Experimenter quietly removed the ball from the trough and waited ten seconds before saying the following to both the Agent and Participant, "*The ball has MOVED. I will ask one of you to indicate where you think it is*." The Experimenter emphasised "MOVED." Neither the Agent nor the Participant knew where the ball was. Next, the Agent left and then returned after 20 s as before. The Experimenter then said to the Participant, "*please make a mark in the sand where you think the ball is.*" This time the ball was not retrieved (it cannot be in any case as it is no longer in the trough), and no feedback was offered. The Experimenter took a photo of the response for later coding. No Belief trials formed a baseline measure of how close to Location 1 Participants might respond anyway, even in the absence of a false belief on the part of the Agent.

False Belief trials. On "False Belief" trials, the Experimenter told the Agent and Participant that the ball had NOT moved, just as in Training trials. However, while the Agent was outside, the Experimenter removed the ball, waited ten seconds, and declared to the Participant *only*, "*The ball has MOVED.*" The Participant now knew that the Agent's belief that the ball was in Location 1 was false, although the Participant did not know where the ball was now. The Experimenter then instructed the Agent to return as usual, and the Participant was asked to make a mark in the sand according to where Participant thought the ball was. The Experimenter took a photo of the response for coding. Note that since the Agent had been told the ball has NOT moved, just as in their experience of Training trials, the Agent would have the same reason to believe that the ball is still in Location 1 on False Belief trials as the Agent did in Training trials. Crucially, if Participants respond closer to Location 1 on False Belief trials (when the Agent falsely believes it to be there) than on No Belief trials (when the Agent has no belief), then this would suggest a specific influence of the Agent's false belief.

4.1.3. Analysis plan

Four versions of the experiment were created to counterbalance for the position of No Belief and False Belief trials and their respective Locations in the trough (see Table 2). The influence of the Agent's belief was measured by calculating the average of bias towards Location 1 across both False Belief trials and the average of bias towards Location 1 across No Belief trials, and testing whether the average bias is larger in False Belief than No Belief. This will be done by means of a paired-sample *t*-test, or a Wilcoxon test (should the data fail to be normally distributed, ascertained by a Shapiro-Wilk test). Again, evidence of bias could support either a belief-simulation or belief-representation + AP account, but not a standard belief-representation account. The alpha for this test is .05, two-tailed, and the threshold for meaningful evidence (i.e., beyond statistical significance alone) is a Bayes Factor of 6 or greater from a Bayesian *t*-test (irrespective of the normality of the data). As before, each photo (256 in total) was independently coded by two coders with no knowledge of which trial they were coding. The average of the two coders' measurements (in cm) was taken as the final measure for each photo. Where the two coders differed in the measurement by more than 1 cm these were adjudicated between by the first author (13 occasions).

4.1.4. Transparency and openness

Experiment 3 was submitted as the final experiment in a Registered Report that began with Experiments 1 and 2 already recorded. Data were analysed using JASP (version 0.19.1), and the data file, photo coding details, and JASP analysis file are available here: https://osf.io/5y96h/files/osfstorage.

4.2. Results

Results are displayed in Fig. 6. The data were normally distributed (*p*

[2] Through an oversight on the part of the researchers a time-stamped pre-registration document was not generated prior to collecting data for Experiment 3. However, the methods were described in the manuscript which received in-principle acceptance prior to data collection (2nd June 2024). A post-hoc registration can be found here: doi: 10.17605/OSF.IO/45YW9.

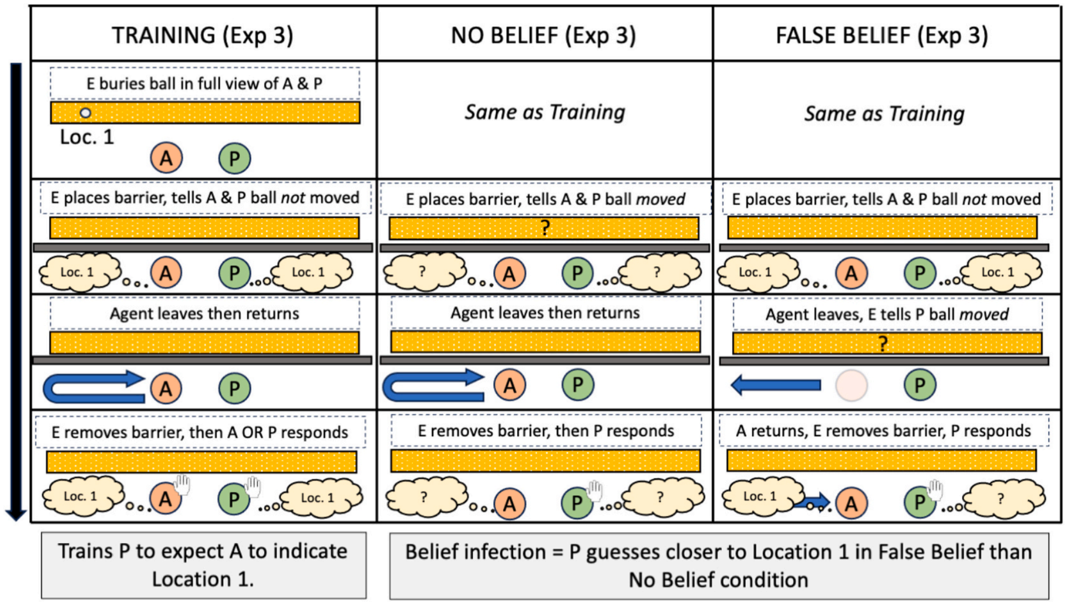

Fig. 5. *In Experiment 3, the Experimenter buried the ball in Location 1 and then placed a barrier (drew a curtain) in front of the sandbox. Now, when the Experimenter said that the ball had* moved, *this generated a new No Belief condition (centre panel), as neither the Participant nor the Agent knew where Location 2 is. When the Experimenter told the Agent that the ball had* not moved *but later told the Participant that it had moved, this formed a False Belief trial (right panel). On Training trials, the Experimenter said the ball had not moved, leading the Participant and Agent to believe the ball remains in Location 1. An influence of the Agent's belief would be evidenced by searches closer to Location 1 on False Belief trials than on No Belief trials. A = Agent, E = Experimenter, P = Participant. See text for full details of procedure.*

> .65 by Shapiro-Wilk test), so we proceeded with a paired-samples t-test. Contrary to the belief-simulation hypothesis, Participants showed no evidence of bias towards where the other agent falsely believed the object to be. Participants failed to indicate a location that was closer to Location 1 on False Belief trials ($M = 31.4$ cm from Location 1, $SD = 16.3$) than on No Belief Baseline trials ($M = 28.3$ cm from Location 1, $SD = 17.9$), $t(63) = 1.254$, $p = .21$, 95 % $CI_{Diff} = [-1.8, +8.0]$, $BF_{10} = 0.289$, with a small effect size, $d = 0.144$. The Bayes Factor analysis demonstrates meaningful support for the null ($BF_{10} < 0.3$) according to guidelines described by Dienes (2014).[3] Of the 64 participants, a minority 24 (37.5 %) patterned in a direction consistent with the hypothesis. In sum, this result fails to replicate that found for Experiments 1 and 2.

4.3. Discussion

The results of Experiment 3 failed to support the belief-simulation or the belief-simulation + AP account. Numerically at least, though not statistically, Participants in fact searched further from Location 1 when Location 1 was where the Agent falsely believed the object to be than when Location 1 was known to both to be a past location.

5. General discussion

Across three experiments, we tested the hypothesis that adults would be influenced in their own egocentric search behaviour by their understanding of another agent's belief about the object's whereabouts. Evidence in favour of such an influence would be most parsimoniously explained by belief-simulation, which posits the first-order simulation of others' beliefs (imagining believing something someone else believes), rather than a belief-representation account, which posits the separation between one's own first-order representations about the environment and one's metarepresentations of others' beliefs. The first two experiments showed a reliable bias towards where the Agent falsely believed the target to be when the Participant searched. This bias carried a close-to-medium effect size. That this effect was generated by the Agent's belief rather than spatial interference from the first location was evidenced by there always being two locations in the trough where that the ball was buried on each trial, but bias was only reliable when the Agent falsely *believed* the target to be in the Location 1. However, the absolute size of this bias was small, only around 1 % of the total length of the trough. This is within the diameter of the target ball itself, and roughly corresponds to a finger's width (the size of the hole made in the sand). One possibility is that the influence of belief-simulation was kept small because Participants were always aware of the true location of the object, and this knowledge was too influential for the simulation of the Agent's false belief to have a powerful impact. However, an alternative possibility is that the effect, being so small in absolute terms, was spurious and would not replicate under more rigorous testing. In Experiment 3, which formed the basis of a Registered Report building upon the first two studies, there were two principal differences from the previous studies. First, we removed Participants' awareness of the true location of the object by introducing No Belief trials. On these trials Participants knew the ball had moved, but not where to. Second, we increased the statistical power of the study from 80 % to 90 % and specified a slightly lower effect size than previously (the lowest effect size found in the data from the earlier studies). This required an almost 100 % increase in participant numbers from 34 to 64. If egocentric knowledge about the true location of the object had suppressed the impact of belief simulation, then Experiment 3 should replicate the effect and possibly increase it now that there was no competing egocentric knowledge. However, any absence of support in this better-powered and more sensitive study would suggest the earlier effects were potentially unreliable (false positives). Results favoured this latter interpretation; Participants failed to search closer to Location 1 when this was where the Agent falsely believed the object to be than when it was known to all that this was merely the previous location of the object. Bayesian analysis supported the null finding. This result fails to support the belief-

[3] This does not reach the higher threshold of six times as likely under the null as laid out in our methods prior to data collection, and this conclusion is therefore subject to a caveat. Although this expectation was not met, it appeared highly unlikely that further data collection would lead to any meaningful advance in our understanding of results.

Table 2
Locations and Trial list for Experiment 3. Trial in bold indicate the critical False Belief (FB) trials and their location-matched No Belief (NB) trials (for half of Participants, trials 5 and 6 were False Belief and 9 and 10 were No Belief). Within each of these two orders, half the time the first trial will involve the ball being buried 80 cm from the left of the trough and the second trial 20 cm from the left, and half the time the reverse order. This creates a total of four versions of the task. A total of 16 participants will perform each version.

Trial (Version 1)	Location 1 (cm)	Searcher
1 Training	50	Agent
2 Training	60	Agent
3 Training	70	Agent
4 Training	30	Agent
5 No Belief	**80**	**Participant**
6 No Belief	**20**	**Participant**
7 Training	20	Participant
8 Training	40	Agent
9 False Belief	**80**	**Participant**
10 False Belief	**20**	**Participant**

Trial (Version 2)	Location 1 (cm)	Searcher
1 Training	50	Agent
2 Training	60	Agent
3 Training	70	Agent
4 Training	30	Agent
5 False Belief	**80**	**Participant**
6 False Belief	**20**	**Participant**
7 Training	20	Participant
8 Training	40	Agent
9 No Belief	**80**	**Participant**
10 No Belief	**20**	**Participant**

Trial (Version 3)	Location 1 (cm)	Searcher
1 Training	50	Agent
2 Training	60	Agent
3 Training	70	Agent
4 Training	30	Agent
5 No Belief	**20**	**Participant**
6 No Belief	**80**	**Participant**
7 Training	20	Participant
8 Training	40	Agent
9 False Belief	**20**	**Participant**
10 False Belief	**80**	**Participant**

Trial (Version 4)	Location 1 (cm)	Searcher
1 Training	50	Agent
2 Training	60	Agent
3 Training	70	Agent
4 Training	30	Agent
5 False Belief	**20**	**Participant**
6 False Belief	**80**	**Participant**
7 Training	20	Participant
8 Training	40	Agent
9 No Belief	**20**	**Participant**
10 No Belief	**80**	**Participant**

simulation hypothesis (and by extension the belief-simulation + AP hypothesis). Given the greater power of this third experiment and the small real-world effects found in experiments 1 and 2, we interpret the totality of these results as more consistent with the belief-representation hypothesis than the belief-simulation hypothesis.

Our results are therefore consistent with recent studies by Haskaraca et al. (2023) and by Speiger and colleagues (2025), who also reported no evidence of altercentric biases in an online, computerised version of the sandbox task. There was one exception; in the latter study, when participants had first performed a block of trials where they were required to take the other agent's perspective explicitly there was evidence of searches closer to Location 1 on false belief than true belief trials on a second block when the task was now to search egocentrically (the erosion of the egocentric perspective after practice taking other perspectives has also been shown in a recent study using a spatial perspective-taking tasks; Samuel et al., 2025). This altercentric bias in block 2 could arise due to low-level practice effects from block 1 carrying over into block 2 (enhanced salience of the other agent or of the

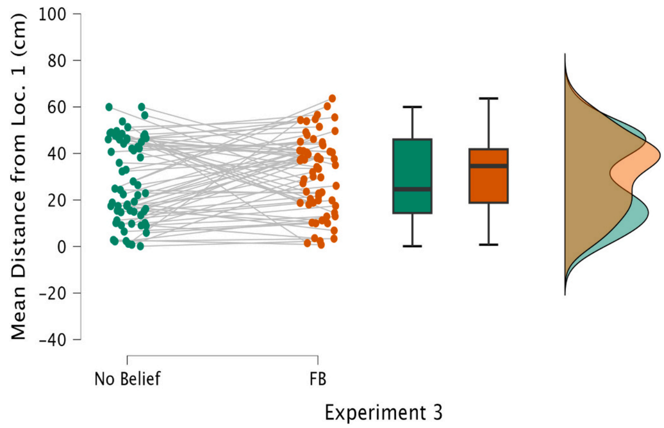

Fig. 6. Results of Experiment 3. The y axis shows distance Location 1 (which could only be positive).

first location, practice discarding one's own belief, etc) rather than a sudden emergence of simulation. Since in the experiment reported here Participants were never asked to take the Agent's perspective or search according to that perspective, our results are consistent with those experiments of Haskaraca et al. and Speiger et al. where explicit perspective taking did not precede egocentric searches, and extend their findings to a more ecologically valid context with real human agents.

Despite this interpretation, there were two methodological differences between experiment 3, where there was no evidence of infection, and experiments 1 and 2, where there was, which might account for the discrepancy. Firstly, the belief-representation induction cues were slightly different between the first two studies and the last study, and this may have influenced whether or how strongly the participant represented the other agent's false belief. In the first two studies, part of the induction cue consisted in Participants actually seeing the ball being move in the agent's absence; whereas, in the last study, the Participants were simply told that the ball was moved in the Agent's absence. Following the adage that seeing is believing, it is possible that in the first two studies, what the Participant saw convinced them that the ball had been moved in the Agent's absence, allowing them to represent that the Agent has a false belief about the location of the ball. However, Participants in the last study may not have found it as compelling to believe that the ball had been moved in the Agent's absence (perhaps, the experimenter was trying to mislead them for the sake of the experiment), resulting in no or a weak representation that the Agent has a false belief about the location of the ball. If Participants in the last study did not have or had a weaker representation of the Agent's false belief than participants in the first two studies, we would not expect the representation of the Agent's false belief to have the same effect on the Participants' own searches in the last study compared to the first two studies.

Secondly, it could be that the search mechanism used by Participants in the first two studies may have been different from the search mechanism used by Participants in the last study, and this difference may have affected whether or how biased Participants were in their search for the ball in the last study compared to the first two studies. In the first two studies, the Participants could have used a belief-input search mechanism that takes the Participant's belief about the location the ball was buried in the sandbox as an input and outputs a behaviour routine directed at searching in that location in the sandbox. Since this search mechanism takes a Participant's actual beliefs about the location of objects as inputs it is vulnerable to a biasing effect when a simulated belief about the location of the object is fed into it as an input. However, in the last study Participants do not have beliefs about the location of ball in the sandbox, and so they could not use a belief-input search mechanism guiding their pointing behaviour at a location in the sandbox when prompted. The Participants in the last study are, after all, simply guessing where the ball is buried compared to the Participants in the first two studies who clearly believe the ball is buried in a particular location in the sandbox. Since the Participants in the last study are not using a belief-input search mechanism, we perhaps should not expect that their putative simulated belief of the Agent's would bias their searches as they might if they were using a belief-input search mechanism.

Assuming methodological divergences between experiments 1 and 2 on the one hand and experiment 3 on the other did not lead to the absence of evidence for simulation in the latter, why might simulation be evident in chimpanzees (Lurz et al., 2022) but not so in humans? One possibility, albeit speculative, is that simulation is evolutionarily older, and the only kind of ToM available to chimpanzees. An ongoing debate in the ToM literature around belief-representation is whether language (syntax, complement structure) is necessary to generate meta-representations (e.g. de Villiers, 2021); this would naturally preclude nonhuman animals. Relatedly, why might humans *not* simulate? One possibility is that humans are capable of both simulation and meta-representation but might default to the latter for 'cognitive' belief attributions and to the former for spatial and affective perspective taking. If this is correct, it might suggest an evolution from simulation for spatial and affective perspective taking, followed (chronologically) in humans alone by belief-representation for cognitive perspective taking, possibly facilitated by the emergence of language. Human ontogeny may by coincidence follow a similar path, with children potentially solving perspective taking problems through belief-simulation until sufficiently advanced in their cognitive and linguistic development. To illustrate, belief-simulation could account for some recent developmental data. In a study by Kovács et al. (2021), 15-month-old infants watched as a first experimenter (E1) hid a toy in one of two boxes in view of a second experimenter (E2). The infant, however, did not see in which box the toy was hidden but knew that both experimenters had knowledge of its location. Next, E2, who had merely observed the hiding, showed the infant which box the toy was in but then proceeded to hide it again such that the infant no longer knew where it was. E1 was either present and witnessed this second hiding or was absent and thus held a (potentially) false belief. Either way, E1's first belief about the location of toy was now revealed to the infant. When the infants were allowed to search for the toy, they could not know which box it was in, but were more likely to search in this revealed location corresponding to E1's first belief when E1 had been absent during the second hiding. The researchers interpreted this finding within the framework of a standard belief-representation account and argued that the infants generated and sustained a metarepresentation of E1's (potentially) false belief because, unlike when E1 was present and her belief could be updated to reflect reality, it was still worth sustaining information that E1 thought was true. Kovács and colleagues were also interested in the incremental construction of the different components of a metarepresentation (i.e., that it was not possible to code all at once all the different components of the metarepresentation). They argued that metarepresentation allowed the infants to maintain "placeholder" components, each distinct from the other, and to populate these sections as more information was revealed throughout the task. This, they argued, makes metarepresentation a strong candidate for belief attributions even in infants. Additionally, Kovács and colleagues suggested that infant altercentrism could support the belief-priming effect (Southgate, 2020). However, an alternative explanation is that once E1's specific belief (that the toy was in box A) was revealed, the infants simulated E1's belief that the toy was hidden in box A, and this biased their own search behaviour as predicted by a belief-simulation account. It is beyond the scope of this paper to evaluate the potential for simulation accounts to also be constructed incrementally, as Kovacs and colleagues suggest metarepresentations can, but it is not impossible that infants first simulated E1's belief by imagining believing that the toy was in a box and then updating this simulated general belief (once the specific content of E1's belief was revealed) to simulate E1's specific belief that the toy was in box A.

There are other types of belief-reasoning that are harder to fit within a belief-simulation account. We know that human adults can not only reason about others' mental states, but also reason about others' mental states *about* others' mental states. Belief-simulation should struggle, relatively speaking, to account for "recursive" ToM, whereby one person's belief is embedded in another person's belief, such as in *John believes that Mary believes that [the treasure is in the chest]* (Oesch & Dunbar, 2017; Wilson et al., 2023). Attempting to simulate John's more complex belief in this case, while not impossible, would also appear to require the simulation of Mary's. Given that simulation allows for belief infection, generating a detached representation would seem better able to avoid confusion.

Overall, our results suggest that adult humans do not automatically simulate the (false) beliefs of others, or are not impacted by any such simulation in their own egocentrically-directed action. While a null result cannot usually be considered powerful evidence in favour of a hypothesis, these data are consistent with a theory by which the beliefs of another agent can indeed be quarantined from our own egocentric representations such that our own search behaviour can proceed unimpaired.

CRediT authorship contribution statement

Steven Samuel: Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Robert Lurz:** Writing – review & editing, Methodology, Investigation, Funding acquisition, Conceptualization. **Daizi Davies:** Investigation, Data curation. **Harry Axtell:** Investigation, Data curation. **Sarah K. Salo:** Writing – review & editing, Investigation, Data curation.

Declaration of competing interest

None.

Acknowledgments

This research was funded by a grant from the Experimental Psychology Society (EPS) awarded to SS and RL. We are very grateful to Amanda Rama for her assistance with coding photographs for Experiment 3.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jesp.2025.104785.

References

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.

Baron-Cohen, S. (1997). *Mindblindness: An essay on autism and theory of mind*. MIT press.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind"? *Cognition, 21*(1), 37–46.

Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In *, 13. Perspectives on socially shared cognition* (pp. 127–149).

Cole, G. G., & Millett, A. C. (2019). The closing of the theory of mind: A critique of perspective-taking. *Psychonomic Bulletin & Review*, 1–16.

Cole, G. G., Millett, A. C., Samuel, S., & Eacott, M. J. (2020). Perspective taking: In search of a theory. *Vision, 4*(2), 30.

Cole, G. G., Samuel, S., & Eacott, M. J. (2022). A return of mental imagery: The pictorial theory of visual perspective-taking. *Consciousness and Cognition, 102*, Article 103352.

Cosmides, L., & Tooby, J. (2000). *Adaptations for decoupling and metarepresentation. Metarepresentations: A multidisciplinary perspective*.

Deroualle, D., Borel, L., Devèze, A., & Lopez, C. (2015). Changing perspective: The role of vestibular signals. *Neuropsychologia, 79*, 175–185.

Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psychology, 5*, 781.

Dietze, P., & Knowles, E. D. (2020). Social class predicts emotion perception and perspective-taking performance in adults. *Personality and Social Psychology Bulletin, 0146167220914116*.

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

Erle, T. M., Ruessmann, J. K., & Topolinski, S. (2018). The effects of visuo-spatial perspective-taking on trust. *Journal of Experimental Social Psychology, 79*, 34–41.

Erle, T. M., & Topolinski, S. (2017). The grounded nature of psychological perspective-taking. *Journal of Personality and Social Psychology, 112*(5), 683–695.

Fadiga, L., Fogassi, L., Pavesi, G., & Rizzolatti, G. (1995). Motor facilitation during action observation: A magnetic stimulation study. *Journal of Neurophysiology, 73*(6), 2608–2611.

Freundlieb, M., Kovács, Á. M., & Sebanz, N. (2018). Reading your mind while you are reading—Evidence for spontaneous visuospatial perspective taking during a semantic categorization task. *Psychological Science, 29*(4), 614–622.

Freundlieb, M., Sebanz, N., & Kovács, Á. M. (2017). Out of your sight, out of my mind: Knowledge about another person's visual access modulates spontaneous visuospatial perspective-taking. *Journal of Experimental Psychology: Human Perception and Performance, 43*(6), 1065.

Frith, U., Morton, J., & Leslie, A. M. (1991). The cognitive basis of a biological disorder: Autism. *Trends in Neurosciences, 14*(10), 433–438.

Furlanetto, T., Becchio, C., Samson, D., & Apperly, I. (2016). Altercentric interference in level 1 visual perspective taking reflects the ascription of mental states, not submentalizing. *Journal of Experimental Psychology: Human Perception and Performance, 42*(2), 158.

Gallese, V., & Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Sciences, 2*(12), 493–501.

Goldman, A. I. (2006). *Simulating minds: The philosophy, psychology, and neuroscience of mindreading*. Oxford University Press on Demand.

Goldman, A. I., & Jordan, L. C. (2013). *Mindreading by simulation: The roles of imagination and mirroring. Understanding other minds: Perspectives from developmental social neuroscience* (pp. 448–466).

Gordon, R. M. (1986). Folk psychology as simulation. *Mind & Language, 1*(2), 158–171.

Haskaraca, F. N., Proft, M., Liszkowski, U., & Rakoczy, H. (2023). How robust are egocentric and altercentric interference effects in social cognition? A test with explicit and implicit versions of a continuous false belief task. *Frontiers in Psychology, 14*.

Hayashi, T., Akikawa, R., Kawasaki, K., Egawa, J., Minamimoto, T., Kobayashi, K., … Iijima, A. (2020). Macaques exhibit implicit gaze bias anticipating others' false-belief-driven actions via medial prefrontal cortex. *Cell Reports, 30*(13), 4433–4444 (e4435).

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

Huttenlocher, J., Hedges, L. V., Corrigan, B., & Crawford, L. E. (2004). Spatial categories and the estimation of location. *Cognition, 93*(2), 75–97.

Huttenlocher, J., Hedges, L. V., & Duncan, S. (1991). Categories and particulars: Prototype effects in estimating spatial location. *Psychological Review, 98*(3), 352.

Huttenlocher, J., Newcombe, N., & Sandberg, E. H. (1994). The coding of spatial location in young children. *Cognitive Psychology, 27*(2), 115–147.

Kessler, K., & Thomson, L. A. (2010). The embodied nature of spatial perspective taking: Embodied transformation versus sensorimotor interference. *Cognition, 114*(1), 72–88.

Keysar, B., Lin, S., & Barr, D. J. (2003). Limits on theory of mind use in adults. *Cognition, 89*(1), 25–41.

Kovács, Á. M., Téglás, E., & Csibra, G. (2021). Can infants adopt underspecified contents into attributed beliefs? Representational prerequisites of theory of mind. *Cognition, 213*, Article 104640.

Kovács, Á. M., Téglás, E., & Endress, A. D. (2010). The social sense: Susceptibility to others' beliefs in human infants and adults. *Science, 330*(6012), 1830–1834.

Krupenye, C., Kano, F., Hirata, S., Call, J., & Tomasello, M. (2016). Great apes anticipate that other individuals will act according to false beliefs. *Science, 354*(6308), 110–114.

Lurz, R. W., Krachun, C., Mareno, M. C., & Hopkins, W. D. (2022). Do chimpanzees predict others' behavior by simulating their beliefs. *Animal Behavior and Cognition, 9*(2), 153–175.

Mahy, C. E., Bernstein, D. M., Gerrard, L. D., & Atance, C. M. (2017). Testing the validity of a continuous false belief task in 3-to 7-year-old children. *Journal of Experimental Child Psychology, 160*, 50–66.

Oesch, N., & Dunbar, R. I. (2017). The emergence of recursion in human language: Mentalising predicts recursive syntax task performance. *Journal of Neurolinguistics, 43*, 95–106.

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

Perner, J. (1991). *Understanding the representational mind*. The MIT Press.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences, 1*(4), 515–526.

Pylyshyn, Z. W. (1978). When is attribution of beliefs justified?[P&W]. *Behavioral and Brain Sciences, 1*(4), 592–593.

Quesque, F., Chabanat, E., & Rossetti, Y. (2018). Taking the point of view of the blind: Spontaneous level-2 perspective-taking in irrelevant conditions. *Journal of Experimental Social Psychology, 79*, 356–364.

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

Samuel, S. (2022). A curse of knowledge or a curse of uncertainty? Bilingualism, embodiment, and egocentric bias. *Quarterly Journal of Experimental Psychology, 76* (8), 1740–1759.

Samuel, S. (2023). A curse of knowledge or a curse of uncertainty? Bilingualism, embodiment, and egocentric bias. *Quarterly Journal of Experimental Psychology, 76* (8), 1740–1759.

Samuel, S., Cole, G. G., Eacott, M. J., & Edwardson, R. (2025). EXPRESS: How to eliminate (and even reverse) egocentric bias in perspective taking. *Quarterly Journal of Experimental Psychology, 0*(ja). https://doi.org/10.1177/17470218251341289

Samuel, S., Hagspiel, K., Eacott, M. J., & Cole, G. G. (2021). Visual perspective-taking and image-like representations: We don't see it. *Cognition, 210*. Article 104607.

Samuel, S., Legg, E. W., Lurz, R., & Clayton, N. S. (2018a). Egocentric bias across mental and non-mental representations in the sandbox task. *Quarterly Journal of Experimental Psychology, 71*(11), 2395–2410.

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

Samuel, S., Legg, E., Manchester, C., Lurz, R., & Clayton, N. (2019). Where was I? Taking alternative visual perspectives can make us (briefly) misplace our own. *Quarterly Journal of Experimental Psychology, 73*(3), 468–477.

Samuel, S., Roehr-Brackin, K., Jelbert, S., & Clayton, N. S. (2019). Flexible egocentricity: Asymmetric switch costs on a perspective-taking task. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 42*(2), 213–218.

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

Shamay-Tsoory, S. G. (2011). The neural bases for empathy. *The Neuroscientist, 17*(1), 18–24.

Shanton, K., & Goldman, A. (2010). Simulation theory. *Wiley Interdisciplinary Reviews: Cognitive Science, 1*(4), 527–538.

Shteynberg, G., Hirsh, J. B., Wolf, W., Bargh, J. A., Boothby, E. B., Colman, A. M., … Rossignac-Milon, M. (2023). Theory of collective mind. *Trends in Cognitive Sciences, 27*(11), 1019–1031.

Singer, T., Seymour, B., O'doherty, J., Kaube, H., Dolan, R. J., & Frith, C. D. (2004). Empathy for pain involves the affective but not sensory components of pain. *Science, 303*(5661), 1157–1162.

Skarratt, P. A., Cole, G. G., & Kuhn, G. (2012). Visual cognition during real social interaction. *Frontiers in Human Neuroscience, 6*, 196.

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.

Southgate, V. (2020). Are infants altercentric? The other and the self in early social cognition. *Psychological Review, 127*(4), 505.

Southgate, V., & Vernetti, A. (2014). Belief-based action prediction in preverbal infants. *Cognition, 130*(1), 1–10.

Sperber, D., & Wilson, D. (2002). Pragmatics, modularity and mind-reading. *Mind & Language, 17*(1–2), 3–23.

Surtees, A., Apperly, I., & Samson, D. (2013). The use of embodied self-rotation for visual and spatial perspective-taking. *Frontiers in Human Neuroscience, 7*, 698.

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). In search of the uniquely human. *Behavioral and Brain Sciences, 28*(5), 721–727.

Tversky, B., & Hard, B. M. (2009). Embodied and disembodied cognition: Spatial perspective-taking. *Cognition, 110*(1), 124–129.

Vilanova, F., Beria, F. M., Costa, Â. B., & Koller, S. H. (2017). Deindividuation: From Le Bon to the social identity model of deindividuation effects. *Cogent Psychology, 4*(1), 1308104.

de Villiers, J. (2021). With language in mind. *Language Learning and Development, 17*(2), 71–95.

Ward, E., Ganis, G., & Bach, P. (2019). Spontaneous vicarious perception of the content of another's visual perspective. *Current Biology, 29*(5), 874–880. e874.

Ward, E., Ganis, G., McDonough, K. L., & Bach, P. (2020). Perspective taking as virtual navigation? Perceptual simulation of what others see reflects their location in space but not their gaze. *Cognition, 199*, Article 104241.

Wardlow Lane, L., & Ferreira, V. S. (2008). Speaker-external versus speaker-internal forces on utterance form: Do cognitive demands override threats to referential success? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*(6), 1466.

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development, 72*(3), 655–684.

Wilson, R., Hruby, A., Perez-Zapata, D., van der Kleij, S. W., & Apperly, I. A. (2023). Is recursive "mindreading" really an exception to limitations on recursive thinking? *Journal of Experimental Psychology: General, 152*(5), 1454–1468.

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

Yu, A. B., & Zacks, J. M. (2017). Transformations and representations supporting spatial perspective taking. *Spatial Cognition & Computation, 17*(4), 304–337.