Consciousness and Cognition 117 (2024) 103625

Contents lists available at ScienceDirect

Consciousness and Cognition

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

Principles of belief acquisition. How we read other minds

M.T. Pascarelli [a,b], D. Quarona [a,b], G. Barchiesi [a,b], G. Riva [c,d], S.A. Butterfill [e], C. Sinigaglia [a,b,f,*]

[a] *Cognition in Action (CIA) Unit, PHILAB, 20122 Milan, Italy*
[b] *Department of Philosophy, Università degli Studi di Milano, Milan, Italy*
[c] *Applied Technology for Neuro-Psychology Lab, IRCCS Istituto Auxologico Italiano, Milan, Italy*
[d] *Department of Psychology, Università Cattolica del Sacro Cuore, Milan, Italy*
[e] *Department of Philosophy, University of Warwick, Coventry, UK*
[f] *Department of Philosophy, Stanford University, CA, USA*

A R T I C L E I N F O

Keywords:
Social Cognition
Mindreading
Belief-tracking
Automatic processes

A B S T R A C T

Reading other minds is a pervasive feature of human social life. A decade of research indicates that people can automatically track an agent's beliefs regardless of whether this is required. But little is known about the principles t guide automatic belief tracking. In six experiments adapting a false belief task introduced by Kovács et al. (2010), we tested whether belief tracking is interrupted by either an agent's lack of perceptual access or else by an agent's constrained action possibilities. We also tested whether such manipulations create interruptions when participants were instructed to track beliefs. Our main finding: the agent's lack of perceptual access did not interrupt belief tracking when participants were not instructed to track beliefs. Overall, our findings raise a challenge: some of the phenomena that have been labelled mindreading are perhaps not mindreading at all, or—more likely—they are mindreading but not as we know it.

1. Introduction

Reading other minds is a hallmark of human sociality. Successful interaction very often depends on our ability to know the minds of others and to identify which beliefs predict and explain their behaviour. A standard way of assessing this ability involves tasks requiring a disparity between the participant's and target agent's mental states. For instance, in a classic false belief task, participants see a character named Sally looking at an object (e.g., a ball) being placed in a box before Sally leaves the room. During Sally's absence, another character, Anne, switches the ball to another box. When Sally comes back, participants are asked to indicate in which box they think she will search for the ball. Adults and children around four tend to correctly indicate the box Sally was looking at before leaving the room, and not the box containing the ball (Wimmer & Perner, 1983; Baron-Cohen et al., 1985).

Strikingly, a decade of research indicates that some mindreading may be automatic in the sense that whether it occurs is relatively unaffected by participants' tasks or intentions (Kovács, Téglás, & Endress, 2010; Schneider, Bayliss, Becker, & Dux, 2012; Schneider, Slaughter, & Dux, 2017; van der Wel, Sebanz, & Knoblich, 2014). Indeed, in a seminal paper, Kovács and colleagues (2010) adapted an object-location task by presenting adult participants with movies in which a character (a smurf) placed a ball on a table and looked at it rolling behind an occluder or away from it, just before leaving the scene. In the absence of the smurf, the ball could change its location. When the smurf came back, the occluder was removed. Participants were instructed to press a button as soon as they detected the ball's

* Corresponding author at: Department of Philosophy, Università degli Studi di Milano, Via Festa del Perdono 7, 20122 Milan, Italy.
E-mail address: corrado.sinigaglia@unimi.it (C. Sinigaglia).

https://doi.org/10.1016/j.concog.2023.103625
Received 14 June 2023; Received in revised form 28 November 2023; Accepted 9 December 2023
Available online 29 December 2023
1053-8100/

presence on the removal of the occluder. As expected, they were faster when they believed the ball was behind the occluder ('P+' conditions) than when they believed it wasn't there ('P–' conditions). Surprisingly, however, they turned out to be faster in pressing the button when the smurf supposedly believed that the ball was present ('P–A+' condition) compared to when both they and the smurf supposedly believed the ball was absent ('P–A–' condition); this is the critical '(P–A+) < (P–A–)' finding. If some mindreading is automatic, this effect of smurf's belief on responses may be a consequence of automatic mindreading.

There is intense debate about this finding. An early view was that it may be an artefact of extraneous features of the stimuli used (Phillips et al., 2015). A direct test of that view did not support it, however (Al Kaddouri et al., 2020), and a number of successful replications with a variety of stimuli would make finding extraneous features challenging (Nijhof et al., 2016; Bardi et al., 2017a,b; Low et al., 2020). We, therefore, started by assuming that in Kovács and colleagues' paradigm, it really is the agent's belief that influences participants' responses.

A question yet to be considered is which principles characterise automatic belief tracking. To illustrate, processes for tracking physical objects do not necessarily operate in accordance with all and only true physical principles, nor even in accordance with what we know to be the case. This is nicely illustrated by the tunnel effect, in which good continuity of trajectory requires us to see as one object what we actually know to be two (Scholl, 2007). Similarly, we reasoned that the principles which characterise automatic mindreading may not coincide with principles that do, or are widely held to, govern how beliefs are acquired in non-automatic mindreading. It would, therefore, be valuable to develop ways of testing which principles guide which mindreading processes. Doing so is crucial for understanding how people read other minds.

One basic principle concerns perceptual access. A natural thought is that, in order to first acquire a belief about the location of a ball (say), an agent has to see or otherwise perceive the ball (unless, of course, they have other sources of knowledge). Certainly, in hundreds of false belief tasks the agent initially sees a target object and—so the story—thereby acquires a belief about it; a subsequent failure to see the target is what causes false belief. This is why it may seem bizarre, or even incoherent, to create a false belief task in which an agent was looking away and lacking any perceptual access to the object right from the start. Yet we know that, while adults can be sensitive to an agent's visual access rather than their mere presence or orientation (e.g., Furlanetto et al., 2016), they also sometimes overestimate the informativeness of visual access (Wang et al., 2014) and are prone to neglect lack of visual access (Keysar et al., 2003). It is, therefore, coherent to ask, from the point of view of one or another mindreading process, whether, in the absence of other sources of knowledge, an agent's failure to perceive an object's movements affects what they believe about its location.

The question we are concerned with can be made clearer by reflection on Träuble et al. (2010)′s finding that infant mindreading

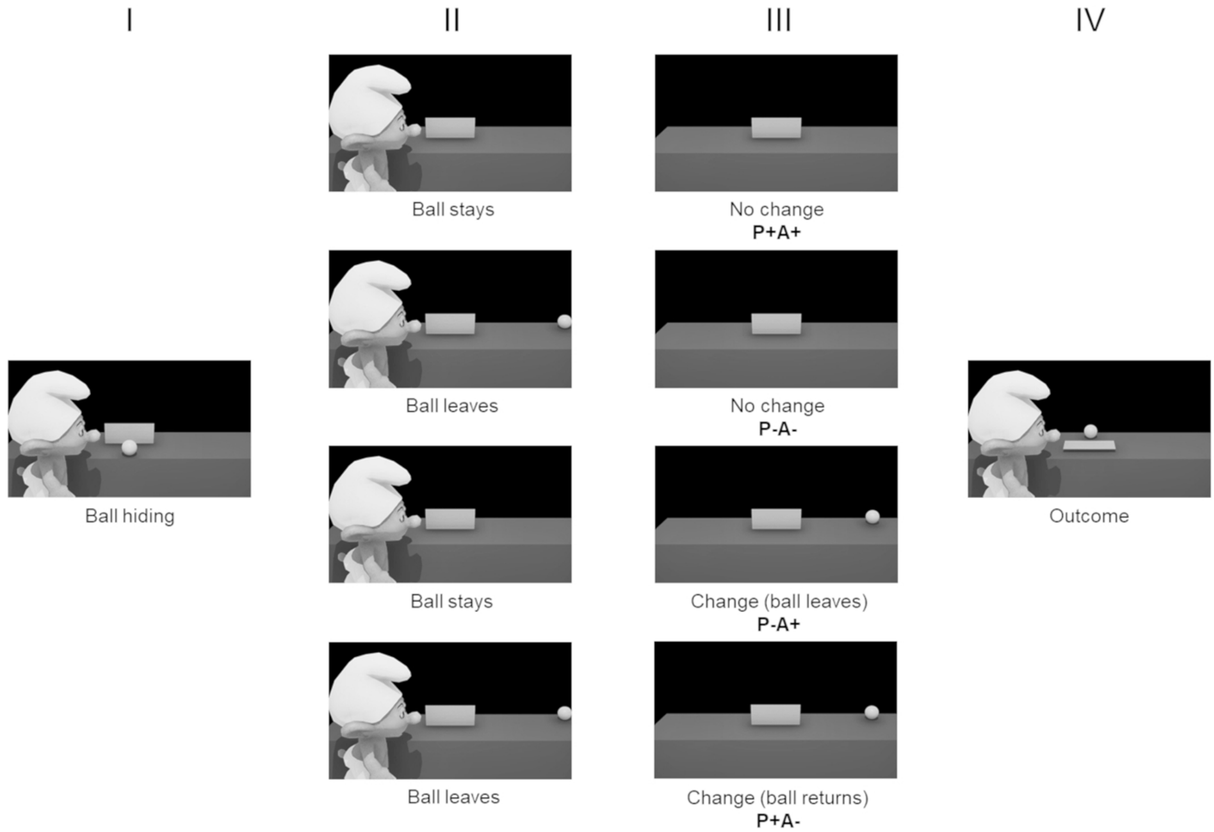

Fig. 1. Structure of events in the movies used in Experiment 1. For the actual timing of events see Table S1. For an example of this series of video clips see Movie S1.

does not appear to require visual access in all cases. In their paradigm, an agent who lacks visual access to a ball tilts a ramp to control its location. This is a wonderful way to test belief ascription when an agent has no visual access but does have another source of information about the ball's location. By contrast, our question was whether complete failure of perceptual access, in the absence of any other source of information, would influence mindreading processes.

Some readers may insist that, normatively, failure to perceive must affect what is believed (in the absence of other sources of knowledge). But even if they are right, it does not follow that all mindreading processes operate in accordance with that principle.

Another potential principle concerns acting. Does an agent have to be in a position to act on the ball in order to acquire a belief about it? If you share our pre-theoretical intuitions, you will be tempted to deny this immediately and without reflection. Whether or not you can act should not generally affect your beliefs about clearly visible events around you. But if we are open to the idea that mindreading processes are not necessarily characterised by true principles, nor by widely agreed principles, then we might be curious to know whether one or another mindreading process will only track beliefs when agents are in a position to act. We were motivated to consider this possibility by evidence that automatically taking the perspective of another individual might depend on perceiving her as being able to act (Cardellicchio et al., 2013; Costantini et al., 2011). Further, automatic (but not non-automatic) mindreading appears to be associated with a superior parietal network, including the superior marginal gyrus, typically related to sensory-motor processes and representations (Grosse Wiesmann et al., 2020; but see Bardi et al., 2017a,b). Consistent with this, Zani et al. (2020) demonstrated that automatic mindreading can modulate how actions are motorically processed in social interaction; and Low et al. (2020) provided evidence that automatic mindreading enables people to track a belief except when the agent was visibly constrained from acting. It is. therefore, coherent to ask whether, from the point of view of one or another mindreading process, an agent's having limited possibilities to act on an object might affect what they believe about that object.

In order to understand whether the principles which characterise automatic mindreading coincide with principles that do, or are widely held to, govern how beliefs are acquired in non-automatic mindreading, we ran three experiments where participants were never instructed to report beliefs and none of the experiments mentioned beliefs. In the first experiment, we implemented the same experimental setting and design as Kovács et al. (2010). Participants were presented with different movies with four distinct scenes: (i) an agent (the smurf) placing a ball on a table in front of an occluder; (ii) the agent seeing the ball rolling behind the occluder and then staying there or moving away; (iii) the agent leaving the scene and the ball remaining in position or changing position by either moving away from or returning to behind, the occluder; and (iv) the agent coming back and the occluder lowering (Fig. 1). As in Kovács et al. (2010), participants were asked to press a button as soon as they saw the ball behind the lowered occluder, with the ball being

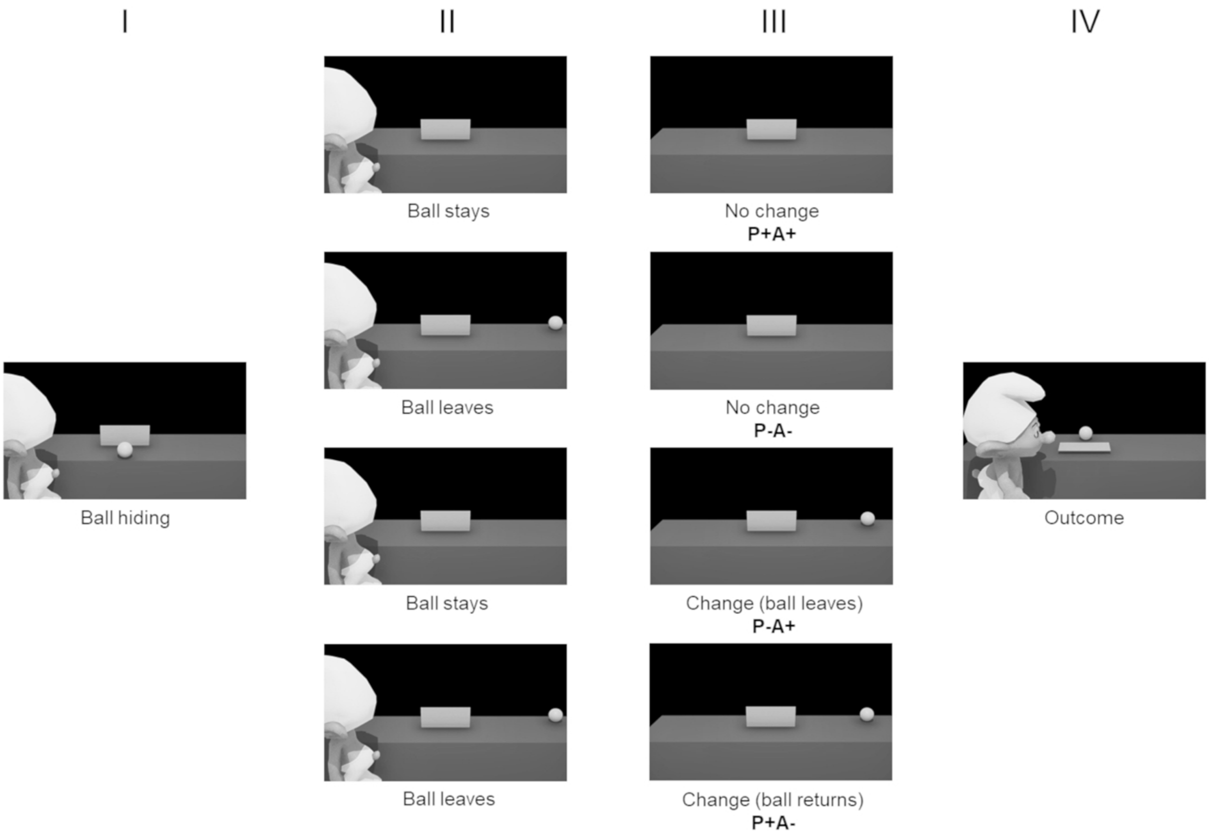

Fig. 2. Structure of events in the movies used in Experiment 2. For the actual timing of events see Table S2. For an example of this series of video clips see Movie S2.

randomly present in half of the trials in all the conditions. Finding the critical effect, (P–A+) < (P–A–), in this experiment would indicate that our materials can be used to further corroborate Kovács et al. (2010).

In a second experiment, the agent's (smurf's) *perceptual access* was manipulated. We created a new series of movies differing from those used in the first experiment in the first two scenes only. These are the scenes where, supposedly, the agent acquires their belief about the location of the ball. In each movie in this new series, in the first two scenes the agent turned their back on the table thus ensuring they were unable to see the ball's movements (Fig. 2). If the critical effect, (P–A+) < (P–A–), were found in this experiment, this would indicate that, from the perspective of automatic mindreading, an agent's failure to perceive an object, even in the absence of any other sources of knowledge, does not affect what they believe about its location.

In a third experiment, the agent's (smurf's) perceptual access was left unhindered while their *action possibilities* were constrained; this was achieved by encasing them in plexiglass. We created a third series of movies differing from those used in the earlier experiments in the first two scenes only. In each movie in this third series, in the first two scenes the agent appeared immobilised in their plexiglass cage and saw the ball rolling behind the occluder, staying there or moving away (Fig. 3). If the critical effect, (P–A+) < (P–A–), were found in this experiment, this would indicate that, from the perspective of automatic mindreading, an agent's having limited possibilities to act on an object does not affect what they believe about its location.

Motivated by the possibility of finding similarities and differences between automatic and non-automatic mindreading, we then switched our focus to non-automatic mindreading. We ran three further experiments with a modified design where participants were instructed to track the agent's beliefs. Taking a cue from previous studies (Nijhof et al., 2016; Bardi et al., 2017a), the modified design required participants not only to press the button as soon as they saw the ball but also to answer questions about the agent's belief about the ball's location. These questions were randomly presented at the end of the movies in one-sixth of the trials. This modified design was used with each of the series of videos from Experiments 1–3 (in Experiments 4–6, respectively). We assumed that explicit answers to the questions would be dominated by non-automatic mindreading processes. We also regarded it as possible that being instructed to answer a question about belief would increase the influence of non-automatic mindreading processes on response times; although Nijhof et al. (2016, p. 9) found no trace of any such effect, intriguingly Bardi et al. (2017a, p. 395) observed smaller differences in response time when participants were instructed about belief. The three new experiments thus allowed us to investigate whether, from the perspective of non-automatic mindreading, an agent's failure to perceive affects what they believe and also whether an agent's having limited action possibilities might affect what they believe.

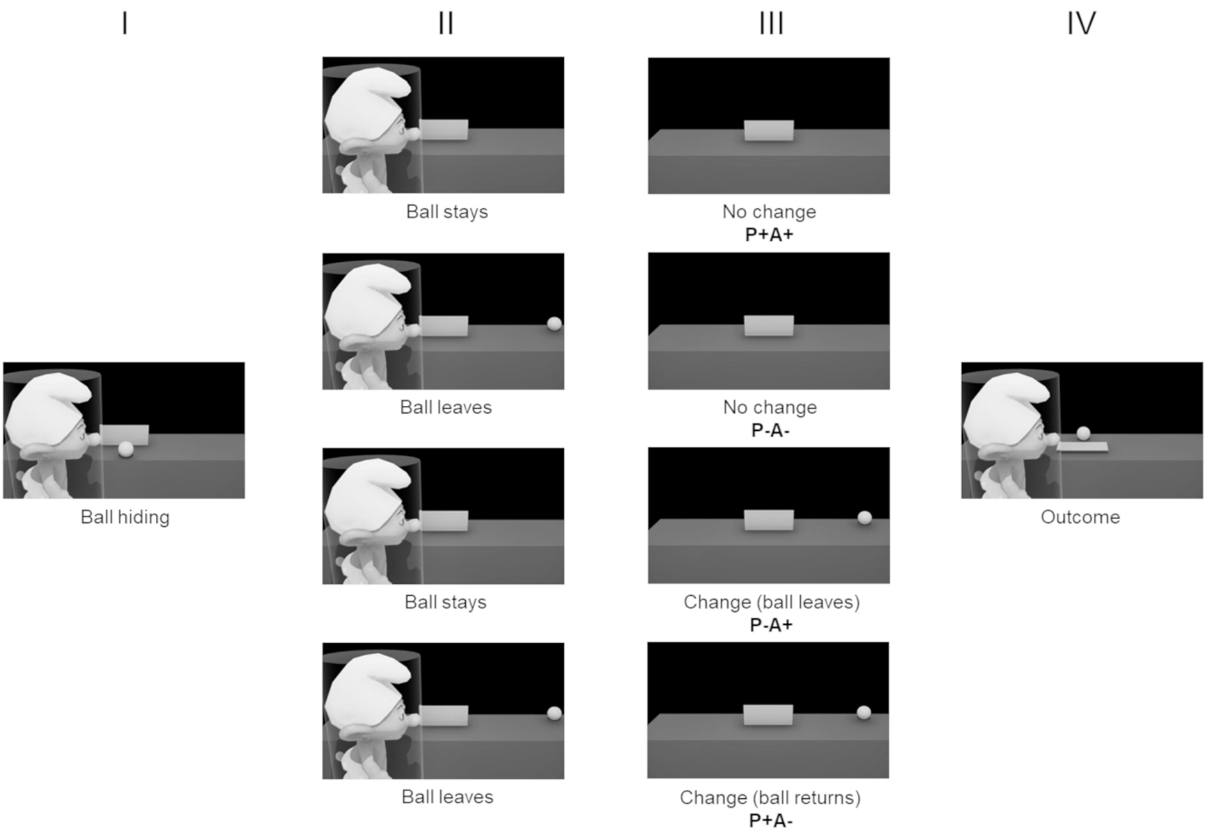

Fig. 3. Structure of events in the movie used in Experiment 3. For the actual time of events see Table S3. For an example of this series of video clips see Movie S3.

2. Methods

2.1. Sample size estimation

In order to reach standardised effect size, a sample size of at least 31 participants was determined with G*Power software (input parameters: $\alpha = 0.01$, Power (1- β err prob) = 0.8 and effect size dz = 0.657). Effect size dz was calculated from the difference of mean and SD of P–A– vs P–A+ comparison (Free-Agent experiment) in Low et al. (2020). A total of 210 participants took part in this study and they were randomly allocated to exactly one of each six experiments. All the participants were right-handed, with normal or corrected-to-normal vision, and with no history of either psychiatric or neurological disorders. All research methods were approved by the Local Ethics Committee and were carried out in accordance with the principles of the revised Helsinki Declaration (World Medical Association General Assembly, 2008). Written informed consent was obtained from all the participants.

3. Experiment 1 — Unistructed belief tracking, visual access, no plexiglass cage.

3.1. Participants

Thirty-five participants (19 females, age mean ± SE = 21.77 ± 0.47 years old) took part in this experiment 1.

3.2. Stimuli

The experimental stimuli consisted in 8 video clips (1920 x 1080 pixels,.mp4, 25 fps, 19 s) depicting a 3D room, with an agent (i.e., a smurf) taking a ball in his right hand, a red table, and a green occluder on the table (see). All video clips started with the smurf placing the ball in front of the occluder and the ball rolling behind it (Phase I). After that, the video clips presented four different possible scenarios:

1. The ball moved away from the occluder and returned behind it; then the smurf left the scene (Phase II). In absence of the smurf, the ball did not change its position (phase III).
2. The ball moved away from the occluder, returned behind it, and finally moved away again, leaving the scene; then the smurf left the scene (phase II). In absence of the smurf, the ball did not change its position (Phase III);
3. The ball moved away from the occluder and returned behind it; then the smurf left the scene (Phase II). In absence of the smurf, the ball changed its position, by moving away from the occluder (Phase III);
4. The ball moved away from the occluder, returned behind it, and finally moved away again, leaving the scene; then the smurf left the scene (Phase II). In the absence of the smurf, the ball changed its position, by returning behind the occluder (Phase III).

Finally, in all the video clips the smurf came back and the occluder was lowered (Phase IV). The ball was present behind the occluder in half of the video clips (for the timing of events in each kind of video see Table S1; for an example of these videos see Movie S1 in Supplementary Materials).

We distinguished four different conditions. We adopted the same notation as Kovács et al. (2010) to characterise the four conditions. “P” stands for Participant, “A” for Agent (i.e., the smurf), “+” means that they supposedly believe that the ball was behind the occluder, and “–” means that they supposedly believe that the ball was not there. The conditions were:

‘P+A+’: both the participant and the agent supposedly believe that the ball is behind the occluder;

‘P–A–’: both supposedly believe that the ball was not behind the occluder;

‘P–A+’: the participant supposedly believes that the ball was not behind the occluder, while the agent supposedly believes that the ball was there;

‘P+A–’: the participant supposedly believes that the ball was behind the occluder, while the agent supposedly believes that the ball was not there.

3.3. Procedure

The experiment took place in a dimly lit room. Participants were instructed to observe a series of video clips on a monitor (1920 x 1080 pixels, refresh rate = 60 Hz). Each trial started with a fixation cross of 1000 ms, followed by one of the video clips. Each video clip was repeated 6 times in a random order for a total of 48 trials. Participants were instructed to observe the ball trajectory during each video clip and to press a button on a keyboard with their right index as soon as they detected the ball when the occluder was lowered. They were instructed not to respond when the ball was absent. The occluder fell down at 18 s in each video clip, so they had 1 s to respond. Reaction times were collected. To make sure that participants paid attention to the entire video clip, they were also asked to press a button with their left index finger as soon as the smurf left the scene (catch trials). Before the task, participants completed 8 practice trials, one for each video clip. The stimuli, timing, and randomization procedure were controlled by E-Prime Software 3.0 (Psychology Software Tools).

3.4. Data analysis

In this and the following experiments, RTs deviating more than ± 2.5 s.d. at the subject level were considered outliers and were eliminated from the analysis. Participants having a number of outliers/missing in the task, average RTs in the task, or a number of outliers/missing in the catch trials deviating more than ± 2.5 s.d. at the group level were considered outliers and were eliminated from the analysis. In Experiment 1, two participants were eliminated from the analysis. After that, average RTs for each condition were computed. Finally, the data were transformed into a logarithmic scale. As in Kovács et al. (2010), RTs were submitted to four pre-planned comparisons: (P–A–) vs (P+A+), (P–A–) vs (P–A+), (P–A–) vs (P+ A–), (P+A+) vs (P+A–). For all statistical tests, the alpha level of significance was set to 0.05, but corrected for multiple comparisons using Bonferroni's method (alpha/4 = 0.0125). This correction is arguably too stringent as, despite our stated plans, it may be thought that just two comparisons are truly relevant, (P–A–) vs (P+A–) to show that the participant's own belief matters and (P–A–) vs (P–A+) to test whether the agent's belief matters. But, as it turns out, the less stringent correction (alpha/2 = 0.025) would not alter the conclusions we draw (see Table S4). Further, we were also interested in the (P+A+) vs (P+ A–) comparison, which is potentially also an indicator of whether the agent's belief matters (even if this comparison did not yield significant results in published research to date). Traditional methods of data analysis (i.e. frequentist) were complemented by a Bayesian two-tailed *t*-test for all comparisons using JASP 0.14.1 software, using default Cauchy priors of 0.707.

4. Experiment 2 — Unistructed belief tracking, no visual access

4.1. Participants

Thirty-five participants (18 females, age mean ± SE = 22.17 ± 0.56 years old) took part in this experiment.

4.2. Stimuli

The experimental stimuli consisted of 8 new video clips. The clips differed from those used in Experiment 1 in that the agent's (smurf's) visual access to the ball was manipulated (see Fig. 2). In the new video clips, the smurf turned their back to the table after placing the ball in front of the occluder (phase I), which meant the smurf was manifestly unable to see all the ball's movements (phase II). Phases III and IV were the same as in Experiment 1. For the timing of events in each kind of video see Table S2; for an example of these videos see Movie S2 in Supplementary Materials.

4.3. Procedure

As in Experiment 1, participants were instructed to observe the ball's trajectory during each video clip and to press a button on a keyboard with their right index as soon as they detected the ball when the occluder was lowered. Reaction times were collected.

4.4. Data analysis

See Experiment 1. In Experiment 2, one participant was eliminated from the analysis.

5. Experiment 3 — Uninstructed belief tracking, plexiglass cage

5.1. Participants

Thirty-five participants (17 females, age mean ± SE = 22.09 ± 0.36 years old) took part in this experiment.

5.2. Stimuli

The experimental stimuli consisted of 8 new video clips. The clips differed from those used in Experiment 1 in that the smurf was immobilised in a plexiglass cage and entered the scene from the left, with the ball entering the scene from the right (Fig. 3). Phases II-IV of the videos were the same as in Experiment 1 except that the smurf always appeared in the cage. For the timing of events in each kind of video see Table S3; for an example of these videos see Movie S3 in Supplementary Materials.

5.3. Procedure

Because of the Covid-19 pandemic, the experiment was run online with E-Prime-Go (E-Prime Software 3.0, Psychology Software Tools). Participants ran the experiment on their own computer in a dimly lit room and were instructed to observe the video clips on their screen (refresh rate = 60 Hz). The experimenter monitored the video rendering with E-Prime-Go. Participants whose videos involved four or more dropped frames were to be excluded. No participants were excluded on these grounds. The experimenter also checked the correctness of all procedures by asking participants to share their screen and observing them through their webcam.

5.4. Data analysis

See Experiment 1. In Experiment 3, one participant was eliminated from the analysis.

6. Experiment 4 — Instructed belief tracking, visual access and no plexiglass cage

6.1. Participants

Thirty-five participants (18 females, age mean ± SE = 22.06 ± 0.38 years old) took part in the experiment.

6.2. Stimuli

Stimuli were the same as in Experiment 1.

6.3. Procedure

The procedure was almost the same as in Experiment 1. The only difference was that, following previous studies using an explicit version of Kovács et al (2010)'s paradigm (Nijhof,et al., 2016; Bardi et al., 2017a), in 8 out of the 48 experimental trials, after the end of the video clip a question was displayed on a black screen: "Did the smurf think that the ball was behind the occluder?" Participants had to respond "yes" or "no" by pressing "c" with their left middle finger or "v" with their left index finger. In half of the trials, the word "yes" was displayed below the question on the right side of the screen, and the word "no" was displayed on the left side of the screen. In the other half of the trials, it was vice versa. Participants had a maximum of 3000 ms to respond. Accuracies were collected. Before the experiment, participants completed six practice trials, five without the question and one with the question.

6.4. Data analysis

Data analysis was as in Experiment 1. Further, the accuracy of the question task was calculated and submitted to a one-sample *t*-test against the chance level (i.e., 0.5).

7. Experiment 5 — Instructed belief tracking, no visual access

7.1. Participants

Thirty-five participants (18 females, age mean ± SE = 22.03 ± 0.48 years old) took part in the experiment.

7.2. Stimuli

Stimuli were the same as in Experiment 2.

7.3. Procedure

The procedure was almost the same as in Experiment 2. The only difference was the addition of the question task from Experiment 4.

7.4. Data analysis

See Experiment 4.

8. Experiment 6 — Instructed belief tracking, plexiglass cage

8.1. Participants

Thirty-five participants (19 females, age mean ± SE = 23.34 ± 0.62 years old) took part in the experiment.

8.2. Stimuli

Stimuli were the same as in Experiment 3.

8.3. Procedure

The procedure was almost the same as in Experiment 3 with the addition of the question task like in Experiment 4 and 5. Because of the Covid-19 pandemic, the experiment was run online with E-Prime-Go (E-Prime Software 3.0, Psychology Software Tools). Participants ran the experiment on their own computer in a dimly lit room and were instructed to observe the video clips on their screen (refresh rate = 60 Hz). The experimenter monitored the video rendering with E-Prime-Go. Participants whose videos involved four or more dropped frames were to be excluded. No participants were excluded on these grounds. The experimenter also checked the correctness of all procedures by asking participants to share their screens and observing them through their webcam.

8.4. Data analysis

See Experiment 4. In Experiment 6, one participant was eliminated from the analysis.

9. Results

9.1. Experiment 1 — Uninstructed belief tracking, visual access and no plexiglass cage.

The (P–A–) vs (P+A+) comparison revealed a significant difference ($t_{(34)} = 3.032$, $P = 0.005$, Cohen's d = 0.513). Indeed, participants were significantly faster at detecting the ball in the (P+A+) condition than in the (P–A–) condition (Log10 RTs mean difference ± SE = 0.021 ± 0.007 ms). The same held for the (P–A–) vs (P+A–) comparison, which revealed a significant difference ($t_{(34)} = 4.540$, $P < 0.001$, Cohen's d = 0.767), with participants being significantly faster at detecting the ball in the (P+A–) condition than in the (P–A–) condition (Log10 RTs mean difference ± SE = 0.034 ± 0.008 ms). The (P–A–) vs (P–A+) comparison, which has been held to be critical for testing automatic mindreading, revealed a significant difference ($t_{(34)} = 3.800$, $P < 0.001$, Cohen's d = 0.642), with participants being significantly faster at detecting the ball in the (P–A+) condition than in the (P–A–) condition (Log10 RTs mean difference ± SE = 0.027 ± 0.007 ms). Finally, the (P+A+) vs (P+A–) comparison revealed no significant difference ($t_{(34)} = 1.645$, $P = 0.109$) (see Fig. 4A; see also Table S4).

To complement traditional methods of data analysis (i.e., frequentist) and quantify the relative strength of our empirical data we ran Bayesian two-tailed *t*-test for all comparisons using JASP 0.14.1 software, using default Cauchy priors of 0.707. There was very strong evidence for the (P–A+)<(P–A–) effect ($BF_{10} = 52.168$) (see Fig. 4D). The other comparisons and the % error of the BF are shown in Table S5. All the row and processed data are uploaded to the repository of Open Science at the following link https://osf.io/7zvhy/?view_only=3e5d8212afbb40ceb9cfc7b8170be55f.

9.2. Experiment 2 — Uninstructed belief tracking, no visual access.

The (P–A–) vs (P+A+) comparison revealed a significant difference ($t_{(33)} = 3.929$, $P < 0.001$, Cohen's d = 0.674). As expected, participants were faster at detecting the ball in the (P+A+) condition than in the (P–A–) condition (Log10 RTs mean difference ± SE = 0.028 ± 0.007 ms). The same was held for the (P–A–) vs (P+A–) comparison, which revealed a significant difference ($t_{(33)} = 4.180$, $P < 0.001$, Cohen's d = 0.717). Participants were significantly faster at detecting the ball in the (P + A–) condition than in the (P–A–) condition (Log10 RTs mean difference ± SE = 0.048 ± 0.011 ms). Strikingly, the (P–A–) vs (P–A+) comparison, which has been held to be critical for testing automatic mindreading, revealed a significant difference ($t_{(33)} = 4.262$, $P < 0.001$, Cohen's d = 0.731). Participants were significantly faster at detecting the ball in the (P–A+) condition than in the (P–A–) condition (Log10 RTs mean difference ± SE = 0.044 ± 0.010 ms). Finally, the (P+A+) vs (P+A–) comparison did not reveal significant difference ($t_{(33)} = 2.150$, $P = 0.039$). (see Fig. 4B; see also Table S4).

The Bayesian two-tailed *t*-test comparisons revealed very strong evidence for the (P–A +)<(P–A–) effect ($BF_{10} = 166.228$) (see Fig. 4E). The other comparisons and the % error of the BF are shown in Table S5. All the row and processed data are uploaded to the repository of Open Science at the following link https://osf.io/7zvhy/?view_only=3e5d8212afbb40ceb9cfc7b8170be55f.

9.3. Experiment 3 — Uninstructed belief tracking, plexiglass cage.

The (P–A–) vs (P + A +) comparison revealed a significant difference ($t_{(33)} = 2.862$, $P = 0.007$, Cohen's d = 0.491). As expected, participants were significantly faster at detecting the ball in the (P + A +) condition than in the (P–A–) condition (Log10 RTs mean difference ± SE = 0.023 ± 0.008 ms). The same results were obtained with the (P–A–) vs (P+A–) comparison, which revealed a significant difference (t(33) = 6.211, P < 0.001, Cohen's d = 1.065). Participants were significantly faster at detecting the ball in the (P+A–) condition than in the (P–A–) condition (Log10 RTs mean difference ± SE = 0.056 ± 0.009 ms). The (P–A–) vs (P–A+) comparison, which has been held to be critical for testing mindreading, also revealed a significant difference ($t_{(33)} = 3.413$, $P = 0.002$, Cohen's d = 0.585). Participants were significantly faster in detecting the ball in the (P–A+) condition compared to the (P–A–) condition (Log10 RTs mean difference ± SE = 0.027 ± 0.008 ms). Finally, the (P+A+) vs (P+A–) comparison revealed a significant difference ($t_{(33)} = 3.410$, $P = 0.002$, Cohen's d = 0.585). Participants were significantly faster at detecting the ball in the (P+A–) condition than in the (P+A+) condition (Log10 RTs mean difference ± SE = 0.033 ± 0.010 ms) (see Fig. 4C; see also Table S4).

The Bayesian two-tailed *t*-test comparisons revealed strong evidence for the (P–A+)<(P–A–) effect ($BF_{10} = 19.884$) (see Fig. 4F). The other comparisons and the % error of the BF are shown in Table S5. All the row and processed data are uploaded to the repository of Open Science at the following link https://osf.io/7zvhy/?view_only=3e5d8212afbb40ceb9cfc7b8170be55f.

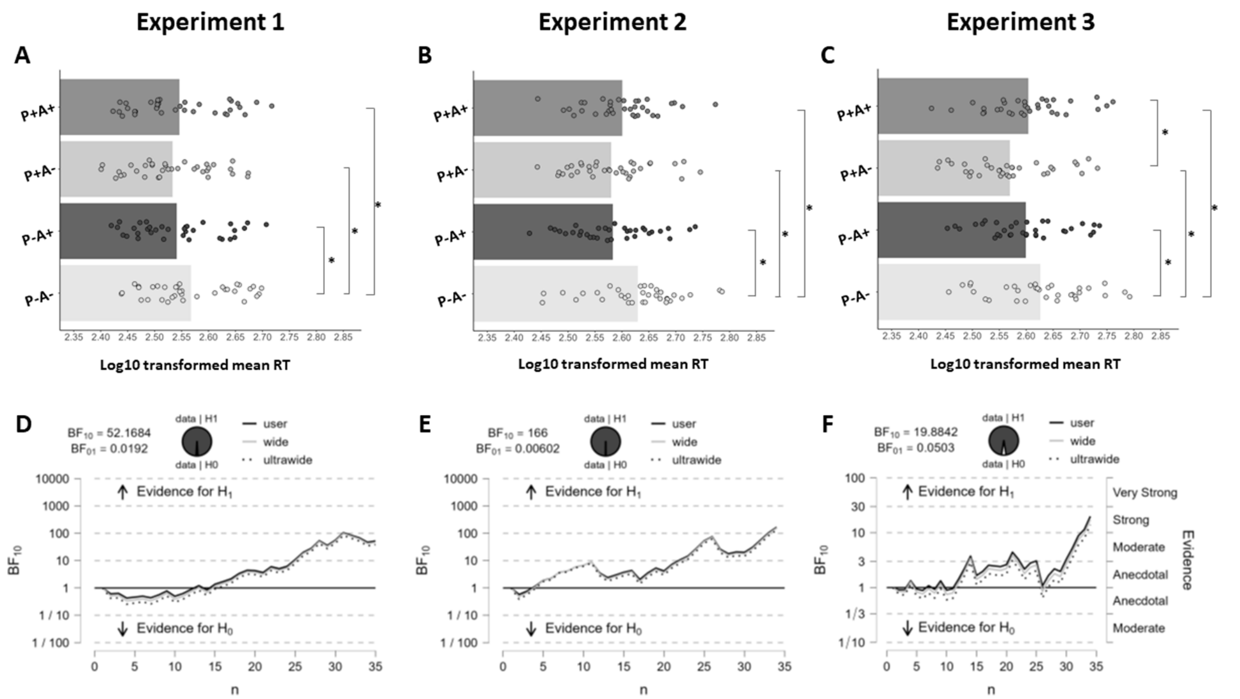

Fig. 4. Results of the Experiments 1–3. *Upper panels* Planned comparison of ball detection latencies (Reaction Times, or RTs) in Experiment 1 (A), Experiment 2 (B), and Experiment 3 (C). Bars represent the Log10 transformed of mean RTs in milliseconds, and error bars represent standard errors. * represents a p-value lower than 0.0125 (Bonferroni's correction is applied) of the planned comparisons. For a systematic overview of the pairwise comparisons in each condition of all the Experiments see Table S4. *Lower panels* Sequential analysis plots of the Bayes factor for the critical comparison (P–A– vs P–A+) in Experiments 1 (D), Experiment 2 (E), and Experiment 3 (F).

9.4. *Experiment 4 — Instructed belief tracking, visual access and no plexiglass cage.*

The (P–A–) vs (P + A +) comparison revealed a significant difference ($t_{(34)} = 3.319$, $P = 0.002$, Cohen's d = 0.561). Participants were significantly faster at detecting the ball in the (P+A+) condition than in the (P–A–) condition (Log10 RTs mean difference ± SE = 0.036 ± 0.011 ms). The (P–A–) vs (P+A–) comparison revealed a significant difference ($t_{(34)} = 5.630$, $P < 0.001$, Cohen's d = 0.952). Participants were significantly faster at detecting the ball in the (P+A–) condition than in the (P–A–) condition (Log10 RTs mean difference ± SE = 0.046 ± 0.008 ms). Further, the (P–A–) vs (P–A+) comparison, which has been held to be critical for testing mindreading, revealed a significant difference ($t_{(34)} = 2.848$, $P = 0.007$, Cohen's d = 0.481). Participants were significantly faster at detecting the ball in the (P–A+) condition than in the (P–A–) condition (Log10 RTs mean difference ± SE = 0.035 ± 0.012 ms). Finally, the (P+A+) vs (P+A–) comparison revealed no significant difference ($t_{(34)} = 1.558$, $P = 0.128$) (see Fig. 5A; see also Table S4). Participants completed the question task with an accuracy higher than 0.5 (mean ± SE = 0.76 ± 0.03, $t_{(34)} = 7.163$, $P < 0.001$).

The Bayesian two-tailed *t*-test comparisons revealed moderate evidence for the (P–A+)<(P–A–) effect ($BF_{10} = 5.517$) (see Fig. 5D). The other comparisons and the % error of the BF are shown in Table S5. All the row and processed data are uploaded to the repository of Open Science at the following link https://osf.io/7zvhy/?view_only=3e5d8212afbb40ceb9cfc7b8170be55f.

9.5. *Experiment 5 — Instructed belief tracking, no visual access.*

The (P–A–) vs (P+A+) comparison revealed a significant difference ($t_{(34)} = 4.176$, $P < 0.001$, Cohen's d = 0.706). Participants were faster at detecting the ball in the (P+A+) condition than in the (P–A–) condition (Log10 RTs mean difference ± SE = 0.031 ± 0.007 ms). The (P–A–) vs (P+A–) comparison revealed a significant difference ($t_{(34)} = 4.640$, $P < 0.001$, Cohen's d = 0.784). Participants were significantly faster at detecting the ball in the (P+A–) condition than in the (P–A–) condition (Log10 RTs mean difference ± SE = 0.050 ± 0.011 ms). The (P–A–) vs (P–A+) comparison, which has been held to be critical for testing mindreading, did not reveal a significant difference ($t_{(34)} = 2.084$, $P = 0.045$). Finally, also the (P+A+) vs (P+A–) comparison revealed no significant difference ($t_{(34)} = 1.973$, P = 0.057) (see Fig. 5B; see also Table S4). Participants completed the question task with an accuracy higher than chance (mean ± SE = 0.61 ± 0.06, $t_{(34)} = 3.060$, $P = 0.004$).

The Bayesian two-tailed *t*-test comparisons revealed anecdotal evidence for the (P–A+)<(P–A–) effect ($BF_{10} = 1.232$) (see Fig. 5E). The other comparisons and the % error of the BF are shown in Table S5. All the row and processed data are uploaded to the repository of Open Science at the following link https://osf.io/7zvhy/?view_only=3e5d8212afbb40ceb9cfc7b8170be55f.

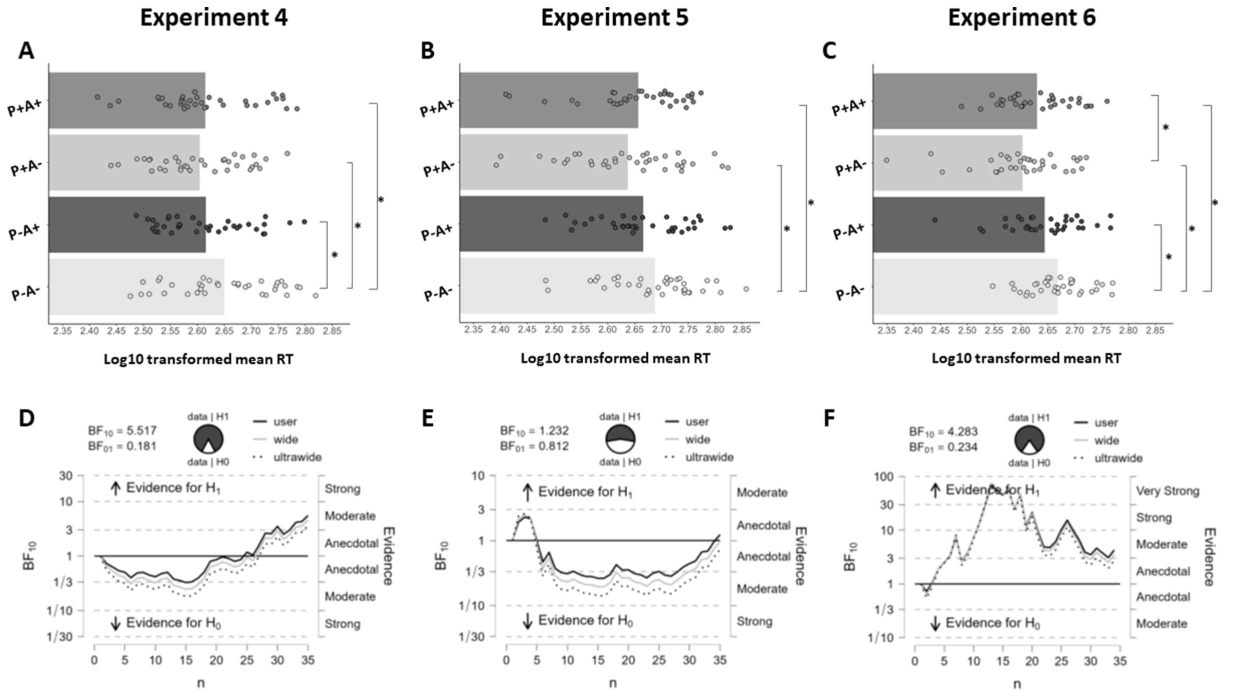

Fig. 5. Results of the Experiments 4–6. *Upper panels* Planned comparison of ball detection latencies (Reaction Times, or RTs) in Experiment 4 (A), Experiment 5 (B), and Experiment 6 (C). Bars represent the Log10 transformed of mean RTs in milliseconds, and error bars represent standard errors. * represents a p-value lower than 0.0125 (Bonferroni's correction is applied) of the planned comparisons. For a systematic overview of the pairwise comparisons in each condition of all the Experiments see Table S4. *Lower panels* Sequential analysis plots of the Bayes factor for the critical comparison (P–A– vs P–A+) in the Experiments 4 (D), Experiment 5 (E), and Experiment 6 (F).

9.6. *Experiment 6 — Instructed belief tracking, plexiglass cage.*

The (P–A–) vs (P+A+) comparison revealed a significant difference ($t_{(33)} = 4.409$, $P < 0.001$, Cohen's d = 0.756). Participants were significantly faster at detecting the ball in the (P+A+) condition than in the (P–A–) condition (Log10 RTs mean difference ± SE = 0.038 ± 0.009 ms). The (P–A–) vs (P+A–) comparison revealed a significant difference ($t_{(33)} = 5.577$, $P < 0.001$, Cohen's d = 0.956). Participants were significantly faster at detecting the ball in the (P+A–) condition compared to the (P–A–) condition (Log10 RTs mean difference ± SE = 0.066 ± 0.012 ms). The (P–A–) vs (P–A+) comparison revealed a significant difference ($t_{(33)} = 2.730$, $P = 0.010$, Cohen's d = 0.468). Participants were significantly faster at detecting the ball in the (P–A+) condition than in the (P–A–) condition (Log10 RTs mean difference ± SE = 0.024 ± 0.009 ms). Finally, the (P+A+) vs (P+A–) comparison revealed a significant difference ($t_{(33)} = 2.893$, $P = 0.007$, Cohen's d = 0.496). Participants were significantly faster at detecting the ball in the (P+A–) condition than in the (P+A+) condition (Log10 RTs mean difference ± SE = 0.029 ± 0.010 ms). (see Fig. 5C; see also Table S4). Participants completed the question task with an accuracy higher than 0.5 (mean ± SE = 0.74 ± 0.05, $t_{(33)} = 5.145$, $P < 0.001$).

The Bayesian two-tailed *t*-test comparisons revealed moderate evidence for the (P–A+)<(P–A–) effect ($BF_{10} = 4.283$) (see Fig. 5F). The other comparisons and the % error of the BF are shown in Table S5. All the row and processed data are uploaded to the repository of Open Science at the following link https://osf.io/7zvhy/?view_only=3e5d8212afbb40ceb9cfc7b8170be55f.

10. Discussion

The main aim of the present study was to understand the principles on which automatic and non-automatic mindreading operate in tracking the initial acquisition of a belief. From the point of view of automatic mindreading, or of non-automatic mindreading (or both), does an agent's failure to perceive an object's movements affect what they believe about its location? And, from the same points of view, does an agent's inability to act on an object affect what they believe about its location?

There are two main findings and an interesting anomaly. First, in Experiment 1, we successfully replicated the key finding of Kovács et al., 2010. This adds to a now growing body of successful replications of that discovery (Nijhof,et al., 2016; Bardi et al., 2017a.b; Al Kaddouri et al., 2020; Low et al., 2020; but see also Phillips et al.,2015).

A second main finding concerns Experiment 2, where the agent does not perceive the ball and has no other source of information about its location. Participants' responses followed the same pattern as would be expected if their automatic mindreading processes were attributing false beliefs to the agent (that is, P–A+<P–A–). This suggests that from the point of view of automatic mindreading, failure to perceive does not prevent believing, even in the absence of any compensating information.

This suggestion presents a challenge to accounts of automatic mindreading processes which take for granted that belief tracking is influenced by lack of perceptual access. For instance, Butterfill & Apperly (2013) assume that registration (roughly belief) depends on encountering (roughly, perceptual access). While the theory does allow for exceptions (an agent's successful action is also permitted as a cue to what the agent has encountered), none of those can explain the pattern of findings we observed in Experiment 2. It seems we may therefore need alternative principles to characterise automatic mindreading. An alternative response might be to abandon the idea that the automatic processes responsible for the P–A+ < P–A– effect are genuinely mindreading processes. But any verdict on whether these responses are correct should also take into account our third main finding.

This third main finding comes from Experiment 5. In this experiment, the agent once again fails to perceive the ball, but participants knew they could be asked to answer questions about belief explicitly. Participants' answers revealed that they were generally attributing beliefs without regard to the agent's failure to perceive. Indeed, accuracy in answering the question about the agent's belief was significantly higher than chance. (By 'accurate', we mean an answer that attributes beliefs as if the agent were not looking away.) We do not infer, of course, that participants' non-automatic mindreading never takes failures to perceive into account when attributing beliefs. A more plausible possibility is that in mindreading, some of the details about what is perceived are sometimes ignored (as others also suggest; compare Keysar et al., 2003). For comparison with another domain: although adults can track physical objects' trajectories through momentary occlusion in simple cases, when tracking multiple objects simultaneously they disregard trajectory information and fall back to relying on a simple proximity heuristic (Franconeri et al., 2012). Similarly, although adults can track perceptual access in situations involving a static display (Furlanetto et al., 2016), when processing a more complex, changing scene their mindreading may fall back to relying on simpler heuristics. In the case of Experiment 5, it is possible that participants may have used simple heuristics such as people near an event often have correct beliefs about its most basic features.[1] Although researchers generally associate such heuristics with automatic mindreading, and with infants' or nonhumans' behavioural predictions, one might speculate that heuristics play a role when adult humans are answering explicit questions about belief.

There is a complication with Experiment 5, however. In contrast to Experiment 2, participants' response times in Experiment 5 provided no evidence that the participants' mindreading violated the idea that the agent's failure to perceive affects what they can believe in their response times (see Fig. 5B). We suggest that the lack of evidence should probably not carry any weight. Bardi et al. (2017a, p. 395) report smaller differences in response time when participants were instructed about belief than when uninstructed. Further, the Bayesian analysis provides anecdotal evidence for non-automatic mindreading in Experiment 5.

We note an interesting anomaly concerning results from Experiments 3 and 6, where the agent was encased in plexiglass. Participants' response times and their answers to questions about beliefs (in Experiment 6 only), both followed the pattern we would expect if their mindreading processes were attributing false beliefs to the agent. This appears to indicate that, from the point of view of those mindreading processes, an agent's having limited action possibilities does not affect what they believe; and, therefore, seems to conflict with Low et al. (2020)'s evidence that automatic mindreading was impaired when the agent was unable to act. However, we also note that when our participants believed the ball was present, they were significantly *slower* to detect the ball when the agent believed it was present than when the agent believed it was absent. So when participants themselves believe the ball is absent, they appear to be mindreading (that is, P–A+ < P–A–); when participants themselves believe the ball is present, they appear to be doing the inverse of mindreading (that is, P+A– < P+A+). This 'inverse mindreading' effect is in contrast to other studies in which either no significant differences are found when participants believe the ball to be present (for example, Kovács et al., 2010) or else significant differences in the opposite, expected direction are found (for example, Low et al., 2020). Indeed, others propose that these previous mixed results arise because an agent's belief is likely less impactful when participants have a true belief (Al Kaddouri et al., 2020, p. 1755). However, that explanation cannot account for a significant effect in the opposite direction. We, therefore, interpret this 'inverse mindreading' effect as removing, for now, any justification for concluding that the results of Experiments 3 and 6 provide evidence of mindreading. Of course, caution is needed because future discoveries may shed light on when and why 'inverse mindreading' is expected.

We have been assuming that, in our paradigm, it really is the agent's belief that explains the differences in participants' responses. Perhaps this is wrong. As we noted in the Introduction, some researchers have held that differences in participants' responses are an artefact of one or another extraneous feature of the stimuli (Phillips et al., 2015; Heyes, 2014a,b). The difficulty for such views until now is that a direct test failed to support one leading contender (Al Kaddouri et al., 2020); and, as we noted above, a number of successful replications with a variety of stimuli make finding extraneous features challenging (Nijhof et al., 2016; Bardi et al., 2017a, b), particularly as some of these find significant effects of the agent's belief on participants' responses even when participants believe the ball to be present (for example, Low et al., 2020). However, our findings may give new hope to anyone aiming to identify extraneous factors. After all, observing a mindreading effect even in the absence of perceptual access (in Experiments 2 and 5) may suggest to some that this is not, after all, a mindreading effect. This perspective is not without problems. The fact that explicit responses to a question show the same pattern (in Experiment 5) is tricky to explain from this perspective and may require further investigation. On the other hand, the anomaly between the way the agent's belief affects participants' responses when they themselves believe the ball absent compared to when they themselves believe the ball present in Experiments 3 and 6 (discussed above) is consistent with the focus on finding an extraneous factor affecting only conditions where the participants believe the ball absent (Heyes, 2014a,b). One route to making convincing an 'extraneous factor' view would be to successfully explain and predict the mixed pattern of discrepancies between conditions where participants themselves believe the ball present and where they themselves believe it absent.

[1] We thank an anonymous reviewer for this suggestion.

To understand mindreading we must know which principles govern it. To date, researchers have mostly assumed that the principles are known *a priori*. We have taken a step forward by treating the principles as hypotheses and developing ways to test whether one or another mindreading process accords with the principles or not. Overall, our findings motivate considering whether principles of belief acquisition that researchers are used to taking for granted really do characterise how people actually read minds. If phenomena that have been labelled automatic mindreading are mindreading at all, they are perhaps unlikely to be mindreading as we know it.

CRediT authorship contribution statement

M.T. Pascarelli: Conceptualization, Data curation, Formal analysis, Writing – original draft, Writing – review & editing. **D. Quarona:** Conceptualization, Data curation, Formal analysis. **G. Barchiesi:** Formal analysis, Methodology, Writing – review & editing. **G. Riva:** Conceptualization, Funding acquisition, Resources, Supervision. **S. Butterfill:** Conceptualization, Supervision, Writing – original draft, Writing – review & editing. **C. Sinigaglia:** Conceptualization, Funding acquisition, Resources, Supervision, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

All the row and processed data are uploaded to the repository of Open Science at the following link https://osf.io/7zvhy/?view_only=3e5d8212afbb40ceb9cfc7b8170be55f.

Acknowledgements

This article was supported by the Department of Philosophy ‘Piero Martinetti’ of the University of Milan with the Project “Departments of Excellence 2018-2022” awarded by the Italian Ministry of Education, University and Research (MIUR) (to MTP, DQ, and CS) and by the PRIN 2017 project “The cognitive neuroscience of interpersonal coordination and cooperation: a motor approach in humans and non-human primates” (Cod. Prog. 201794KEER; to CS). It has also been supported by the UCSC D3.2 2020 project “Behavioural change: prospettive per la stabilizzazione di comportamenti virtuosi verso la sostenibilità” (to GR). We are grateful to Roberto Ranon for his critical contribution in making the different series of video clips.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.concog.2023.103625.

References

Al Kaddouri, R., Bardi, L., De Bremaeker, D., Brass, M., & Wiersema, J. R. (2020). Measuring spontaneous mentalizing with a ball detection task: Putting the attention-check hypothesis by Phillips and colleagues (2015) to the test. *Psychological Research, 84*, 1749–1757.
Bardi, L., Desmet, C., Nijhof, A., Wiersema, J. R., & Brass, M. (2017a). Brain activation for spontaneous and explicit false belief tasks overlaps: New fMRI evidence on belief processing and violation of expectation. *Social Cognitive and Affective Neuroscience, 1*(12), 391–400.
Bardi, L., Six, P., & Brass, M. (2017b). Repetitive TMS of the temporo-parietal junction disrupts participant’s expectations in a spontaneous Theory of Mind task. *Social Cognitive and Affective Neuroscience, 1*(12), 1775–1782.
Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a “theory of mind”? *Cognition, 21*, 37–46.
Butterfill, S. A., & Apperly, I. (2013). How to construct a minimal theory of mind. *Mind & Language, 28*, 606–637.
Cardellicchio, P., Sinigaglia, C., & Costantini, M. (2013). Grasping affordances with the other’s hand: A TMS study. *Social Cognitive and Affective Neuroscience, 8*, 455–459.
Costantini, M., Committeri, G., & Sinigaglia, C. (2011). Ready both to your and to my hands: Mapping the action space of others. *PLoS One, 4*(6), e17923.
Franconeri, S. L., Pylyshyn, Z. W., & Scholl, B. J. (2012). A simple proximity heuristic allows tracking of multiple objects through occlusion. *Attention, Perception, & Psychophysics, 74*(4), 691–702.
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*, 158–163.
Grosse Wiesmann, C., Friederici, A. D., Singer, T., & Steinbeis, N. (2020). Two systems for thinking about others’ thoughts in the developing brain. *Proceedings of the National Academy of Science USA, 24*(117), 6928–6935.
Heyes, C. (2014a). Submentalizing: I am not really reading your mind. *Perspective on Psychological Science, 9*, 131–143.
Heyes, C. (2014b). False Belief in Infancy: A Fresh Look. *Developmental Science, 17*(5), 647–659.
Keysar, B., Lin, S., & Barr, D. J. (2003). Limits on theory of mind use in adults. *Cognition, 89*(1), 25–41.
Kovács, A. M., Téglás, E., & Endress, A. D. (2010). The social sense: Susceptibility to others’ beliefs in human infants and adults. *Science, 330*, 1830–1834.
Low, J., Edwards, K., & Butterfill, S. A. (2020). Visibly constraining an agent modulates observers’ automatic false-belief tracking. *Scientific Report, 10*, 11311.
Nijhof, A. D., Brass, M., Bardi, L., & Wiersema, J. R. (2016). Measuring mentalizing ability: A within- subject comparison between an explicit and implicit version of a ball detection task. *PLoS ONE, 11*, e016437.

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

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

Schneider, D., Slaughter, V. P., & Dux, P. E. (2017). Current evidence for automatic Theory of Mind processing in adults. *Cognition, 162*, 27–31.

Scholl, B. J. (2007). Object Persistence in Philosophy and Psychology. *Mind & Language, 22*(5), 563–591.

Träuble, B., Marinovic, V., & Pauen, S. (2010). Early theory of mind competencies: Do infants understand others' beliefs? *Infancy, 1*(11), 1–11.

van der Wel, R. P. R. D., Sebanz, N., & Knoblich, G. (2014). Do people automatically track others' beliefs? Evidence from a continuous measure. *Cognition, 130*, 128–133.

Wang, J. J., Miletich, D. D., Ramsey, R., & Samson, D. (2014). Adults see vision to be more informative than it is. *Quarterly Journal of Experimental Psychology, 67*(12), 2279–2292.

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.

Zani, G., Butterfill, S. A., & Low, J. (2020). Mindreading in the balance: Adults' mediolateral leaning and anticipatory looking foretell others' action preparation in a false-belief interactive task. *Royal Society Open Science, 7*(1), Article 191167.