Little is known about whether human beings' automatic mindreading is computationally restricted to processing a limited kind of content, and what exactly the nature of that signature limit might be. We developed a novel object-detection paradigm to test adults' automatic processing in a Level 1 perspective-taking (L1PT) context (where an agent's belief, but not his visuospatial perspective, is relevantly different) and in a Level 2 perspective-taking (L2PT) context (where both the agent's belief and visuospatial perspective are relevantly different). Experiment 1 uncovered that adults' reaction times in the L1PT task were helpfully speeded by a bystander's irrelevant belief when tracking two homogenous objects but not in the L2PT task when tracking a single heterogeneous object. The limitation is especially striking given that the heterogeneous nature of the single object was fully revealed to participants as well as the bystander. The results were replicated in two further experiments, which confirmed that the selective modulation of adults' reaction times was maintained when tracking the location of a single object (Experiment 2) and when attention checks were removed (Experiment 3). Our findings suggest that automatic mindreading draws upon a distinctively minimalist model of the mental that underspecifies representation of differences in perspective relative to an agent's position in space.

KEYWORDS: Level 1 perspective-taking; Level 2 perspective-taking, belief-tracking; automaticity; dual processing

1 Introduction

According to standard philosophical accounts (Davidson, 1980, 1990), beliefs have distinctive features that make inferences about mental states relatively demanding and made only when necessary. Beliefs carry propositional content (i.e., the referents of that clauses) and indicate the psychological relation between an individual and the world. Grasping that propositions can be evaluated in different ways by different people helps us appreciate that false beliefs are possible. Belief reasoning also has logical affinities with visual perspective-taking in the sense that both involve representing as well as integrating how the particular way an object, scene or state of affairs is experienced can give rise to different impressions, such as, "I see it as [the turtle standing on its feet], but he sees it as [the turtle lying on its back]." And analogously, "I know that [the chocolate is in the cupboard], but Maxi believes that [the chocolate is in the drawer]" (Apperly, 2011; Hamilton, Brindley, & Frith, 2009; Moll, Meltzoff, Merzsch, & Tomasello, 2012; Zeman, 2017). Appreciating beliefs and visual perspectives supports our inferences of others' actions, and yet the very characteristic that make such processes cognitively flexible—simultaneously acknowledging contrasting models of a particular thing to different people—is the same characteristic that makes mindreading slow and effortful. On the other hand, it is also commonly supposed that mindreading must be cognitively efficient to play a role during fast-moving social interaction. Given that these tensions tend not to co-occur in cognitive systems, a mindreading process is computationally efficient if there are signature limits on the kinds of input that can be automatically processed. We report converging data from three experiments revealing that adults automatically track an agent's belief in a task where differences in perspectives are not relevant, but do not show typical signs of automatic processing when both beliefs and perspectives differ in a relevant manner.

It is puzzling that there are seemingly conflicting sets of findings regarding the automaticity of belief inferences. On the one hand, studies measuring response times to unpredictable probe questions in incidental false-belief tasks show that adult humans can work out what someone is thinking, but this is not something that is performed automatically (Apperly, Riggs, Simpson, Chiavarino, & Samson, 2006; Back & Apperly, 2010). Adults take longer to respond to probes enquiring about an agent's belief of where an object is located than they take to respond to probes concerning the object's actual location (Apperly et al., 2006). Adults are only just as fast to respond to belief questions as they are to reality questions when explicitly instructed to keep track of an agent's belief of a target's whereabouts (Back & Apperly, 2010). There is also converging evidence suggesting that adults find it difficult to overcome egocentric biases when making judgements about others' beliefs (Birch & Bloom, 2007; Keysar, Barr, Balin, & Brauner, 2000; Keysar, Lin, & Barr, 2003) and their reasoning is impeded by increased cognitive load or decreased executive functioning (Apperly, Back, Samson, & France, 2008; Bull, Phillips, & Conway, 2008; McKinnon & Moscovitch, 2007; Rowe, Bullock, Polkey, & Morris, 2001). On the other hand, there is also evidence suggesting that belief inferences can be made automatically (Schneider, Bayliss, Becker, & Dux, 2012; Schneider, Nott, & Dux, 2014; Schneider, Slaughter, & Dux, 2017), to the extent that people's own action selections may be influenced by others' beliefs (van der Wel, Sebanz, & Knoblich, 2014), even when participants are explicitly instructed to prioritize their own beliefs (Meert, Wang, & Samson, 2017). Even in a simple object-detection task, where the goal is just to press a button to detect the presence of a ball, adults' reaction times are speeded when only a bystander happens to believe the object is present, compared to when neither the participant nor the bystander believes the object is present (Bardi, Desmet, & Brass, 2018; Deschrijver, Bardi, Wiersema, & Brass, 2016; El Kaddouri, Bardi, De Bremaeker, Brass, & Wiersema, 2019; Kovács, Téglás, & Endress, 2010; Nijhof, Brass, Bardi, & Wiersema, 2016; Nijhof, Brass, & Wiersema, 2017). In Kovács and colleagues' object-detection task, adults watched animated movies in which a Smurf character observed a ball move around a table. In the outcome phase a barrier fell away and the participant had to respond if the ball was present. The critical finding was that, compared to a baseline condition, in which neither the participant nor Smurf expected the ball to be present, participants were faster to respond when only the Smurf expected the ball to be present, implying that the Smurf's belief regarding the ball's location was automatically encoded.

Research also shows that calculating others' visual perspectives is sometimes, but not always, automatic. In Samson and colleagues' dot-counting task (Samson, Apperly, Braithwaite, Andrews, & Bodley Scott, 2010) adults are instructed to indicate how many dots they themselves can see inside a room. Studies show that participants experience an altercentric interference effect whereby they respond more slowly and with more errors when an avatar in the room sees a different number of dots, compared to when he or she saw the same number as them (Furlanetto, Becchio, Samson, & Apperly, 2016;

Qureshi, Apperly, & Samson, 2010; Samson et al., 2010). Such findings suggest that the mental content of the avatar's visual perspective can be automatically computed, which results in interference during on-line judgements about self-perspective (though the interpretation of such work has been challenged on the grounds that altercentric interference may also be the result of experimental artefacts such as attentional cueing (Cole, Atkinson, Le, & Smith, 2016; Conway, Lee, Ojaghi, Catmur, & Bird, 2017; Heyes, 2014; Santiesteban, Catmur, Hopkins, Bird, & Heyes, 2014).

There are, however, different forms of visuospatial perspective processing. Pursuant to Flavell's model (1978, 1992), a simple case, referred to as Level 1 perspective-taking (L1PT), involves calculating the content of what is seen when someone gazes, and can be processed using line-of-sight information. A higher-level visuospatial perspective problem, termed Level 2 perspective-taking (L2PT), requires an understanding of *how* an entity is appreciated. The latter is regarded as the more representationally complex of the two, evidenced by later ontogenetic development and phylogenetic differences (Flavell, Everett, Croft, & Flavell, 1981; Karg, Schmelz, Call, & Tomasello, 2016; Masangkay et al., 1974; Moll & Meltzoff, 2011). The later-developing L2PT ability has been characterized as involving perspective-confrontation, which entails integrating in a single representation how two people looking at the self-same object from different viewpoints can arrive at different and contradictory descriptions (Moll et al., 2012; Perner, Stummer, Sprung, & Doherty, 2002). Confrontation of perspectives can come about not just when they are mutually exclusive (e.g., that the turtle is perceived as standing on its feet as opposed to lying on its back (Masangkay et al., 1974); or the object is believed to be in one location and not the other) but also arise when the alternatives are compatible. For example, a particular animal can be given two sortals (e.g., bunny, rabbit) allowing individuation of the self-same thing in distinct but synonymous ways (Doherty & Perner, 1998). Nonetheless, young children still treat alternative names as being somehow mutually exclusive. Overall, L2PT involves more than just tracking what someone else sees, but constructing and holding in mind a meta-relation that integrates alternative representations of one and the same thing held by two different people at the same time under a superordinate viewpoint. Several studies show that humans do not automatically compute how an object might appear differently to people with different perspectives (Hamilton & Ramsey, 2013; Surtees, Butterfill, & Apperly, 2012). In Surtees and colleagues' digit-appearance task, for example, adults were instructed to indicate the numeral that was shown on a table (the stimulus was a rotationally asymmetrical digit such as a ‘6’ or a ‘9’, and there was an avatar positioned behind the table such that he or she saw the digit from the opposite point of view from participants). In contrast to findings from the dot-counting task, there was no evidence of altercentric interference on the self-trials of the digit-appearance task: adults were no slower to respond when the avatar’s perspective of the digit was different from their own than when it was the same.

According to Apperly and Butterfill’s dual-process account (Apperly & Butterfill, 2009; Butterfill & Apperly, 2013), the apparently conflicting findings suggest that two relatively distinct processes are involved in arriving at others’ mental states: an efficient mindreading process and a flexible mindreading process. Efficient mindreading employs simple relational attitudes like registration instead of complex propositional attitudes like belief for predicting others’ behaviors, and in so doing makes minimal demands on central resources, is fast and automatic. A registration is an encountering relationship that persists even when the object is no longer in the agent’s field: “one stands in the registering relation to the object and location if one encountered it at that location and if one has not encountered it somewhere else” (p.962, Apperly & Butterfill). Registration is therefore belief-like in that it has a correctness condition which may or may not obtain but it falls short of being a proper propositional attitude in that it does not consider how a particular state of affairs is represented to the other. Fortunately, adults also have a flexible mindreading process which is able to compute belief as such, but such attributions make heavy demands on executive functioning, are slow and non-automatic.

A cornerstone prediction of the dual-process account is that signature limits on the efficient mindreading process arise from the fact that only objects and their relations to agents can be automatically computed to predict others’ behavior, which in turn means that false belief involving identity in the numerical sense cannot be ascribed by representing registrations. There is supportive evidence showing that adults automatically compute people’s false beliefs about an object’s location but not its numerical identity (Edwards & Low, 2017; Fizke, Butterfill, van de Loo, Reindl, & Rakoczy, 2017; Low, Drummond, Walmsley, & Wang, 2014; Low & Watts, 2013; Mozuraitis, Chambers, & Daneman, 2015; Oktay-Gür, Schulz, & Rakoczy, 2018). For example, Low and Watts found that adults’ efficient mindreading, as indicated by certain eye movements, allowed participants to make accurate search anticipations when the agent had a false belief about an object’s location but not when the agent’s false belief about object identity led him to think that there were two objects present when, in fact, there was only one. However, the dual-process account has yet to fully articulate the boundaries of the signature limit that distinguishes the automatic but rigid process of efficient mindreading. Representing mistakes over how objects are represented in the numerical sense may not be the elemental or primary marker that distinguishes efficient from flexible mindreading processes. In Low and Watts' task, confronting the truth of the agent's belief certainly requires making attributions of the agent's belief about there being multiple objects versus the reality that there is only one object. However, the absence of an altercentric interference effect on adults' performance on the self-trials of the digit-appearance task (Surtees et al., 2012) is also treated as converging evidence of a signature limit on adults' efficient mindreading, and yet that task does not involve tracking mistakes over numerical identity per se (i.e., the participant and the avatar are both aware there is a single digit on the table and there really is a single digit on the table). Instead of object identity per se, the commonality between such tasks and their constellations is that they require a meta-representational understanding of perspective, evaluating how people's epistemic states are relativized to the specific perspective by which others regard the world. L2PT, involving perspective-confrontation, may be the core signature limit operating on the automaticity of the efficient mindreading process whilst L1PT (e.g., tracking relational attitudes in object-location false-belief tasks or visibility in the dot-counting task) is potentially stimulus-driven and goal independent.

We created a new paradigm that weaved together belief-attribution and perspectivization to delineate the boundary of the signature limit operating on automatic mindreading. We used an object-detection paradigm to measure the extent to which adults were automatically influenced by the belief of a passive bystander in tasks that did and did not necessitate integrating contrasting perspectives. Using a within-subjects design, Experiment 1 profiled adults' reaction times in two closely-matched tasks. In the L1PT task, the participants and the bystander-agent observed a homogenous blue ball and a homogenous red ball moving around a table. At the end of each trial, one of the balls was hidden behind two screens so that neither the participant nor the agent could see it. In the L2PT task, the scene was identical except that a single heterogeneous object (a dog-robot) moved around the table, finishing its movements between the screens by the end of each trial. Both participant and agent were simultaneously shown that the object appeared blue from one viewing perspective and red from the opposite viewing perspective. Critically, the agent was irrelevant to both tasks; the participant was simply required to select the color (blue or red) that was revealed to himself or herself when the screen rapidly dropped away. The agent either witnessed all events (and so had beliefs consistent with the participant) or was absent for some of the events (so that the agent and participant had inconsistent beliefs). We adjusted Kovács et al.'s (2010) object-detection paradigm as follows. First, the agent was positioned so that he faced the participant, viewing events from the opposite (rather than same) perspective. Second, the opposing viewpoints necessitated the use of two screens (rather than one) to simultaneously mask the objects from the participant and the agent. Third, participants made forced-choice rather than Go/NoGo responses. All trials featured video clips of a human agent in a real-life setting rather than an avatar in an artificial environment.

In the L1PT task, the agent may hold a false belief about the final location of each ball because he was absent when the red ball and blue ball switched places. For example, before the reveal, the agent believes that there is a red ball between the screens and the participant believes that there is a blue ball between the screens. In this task, the agent's belief but not his visuospatial perspective is relevantly different, for when the screens drop both parties will see a blue ball. There is no confrontation of visuospatial perspective in the two-ball task because the two people looking at the object from different viewpoints will arrive at the same description. In the L2PT task, the agent may hold a false belief about the color that will be revealed when the occluders drop because he was absent during the object's final rotation. In this case, however, there is also confrontation of visuospatial perspective because at the reveal the two people looking at the self-same object from opposite viewpoints will arrive at different and contradictory descriptions. While both tasks involve tracking another's perspective of an object or objects (the content of what is seen when someone gazes), only the dog-robot task has the additional requirement of confronting perspectives: in this case the participant is required to evaluate how the self-same object is construed from one location, when that construal simultaneously represents the alternative viewpoint that the agent is instead expecting to only perceive from his opposite location. We can differentiate our L1PT task from our L2PT task in that only the latter involves simultaneously confronting two different visuospatial perspectives on the self-same object, which may require embodied self-rotations to imagine assuming others' positions in the world so as to reason about how an object in their environment is experienced by them (Kessler & Rutherford, 2010; Surtees, Apperly, & Samson, 2013).

For our L1PT task, we predicted that a bystander's belief about the presence of a specific object would helpfully modulate adults' own reaction times when detecting the presence of that object. However, for the closely matched L2PT task we expected to find that adults' reaction times would *not* be speeded when the bystander's belief about the presence of a specific object was dependent on his location in space. If, on the other hand, a facilitating influence of the bystander's belief extended to our L2PT task involving perspective-confrontation, then the dual-process account may be inaccurate and humans instead have a single mindreading process that is context sensitive.

In Experiment 2, we ran a single-ball version of the L1PT task to ascertain that our design modifications would still produce the critical finding seen in the typical version of the object-detection task (i.e., speeded reaction times to the appearance of an object when only the agent believed the object would be present). Finally, Experiment 3 sought to rule out the possibility that the evidence of automatic computations found in our L1PT task was merely an artefact of attention checks used in the object-detection paradigm to ensure participants' task compliance.

2 Experiment 1

2.1 Method

2.1.1 Participants

An a priori analysis using G*Power (Faul, Erdfelder, Lang, & Buchner, 2007) (input parameters: $\alpha = .05$, power =.8) determined that we required a sample size of at least 33 participants to detect the standardised effect size. While not a direct replication, the standardised effect size (r =.45) was calculated using the formula, $r^2 = t^2 / (t^2 + df)$, where t = reported t-test statistic of Kovács et al.'s (2010) critical effect = 2.42, and df = 23. A total of 54 adult participants, made available by the Victoria University of Wellington's Introduction to Psychology Research Programme (IPRP), signed up to take part in the study. Having a larger number of individuals safeguarded against participant dropout, and other factors affecting data collection such as experimenter error or computer malfunction. All participants signed informed consent forms prior to participation and were debriefed orally at the end of the session. One participant did not perform above chance level and was excluded. As a result, analysis was undertaken on the data of 53 participants. The ratio of females to males was 42/11 and the age mean was 18.36 years (Range 17 to 24). The study was approved by Victoria University of Wellington's Human Ethics Committee.

2.1.2 Materials

All stimuli and instructions to the participant were presented via E-Prime 2.0. Each individual watched a total of 80 videos in an object-detection paradigm. The on-screen video dimensions were 38cm x 21cm; all videos had a frame rate of 25 frames per second (fps) and a 720 x 576 resolution. There were 40 videos in the L1PT task and 40 videos in the L2PT task. Due to total experimental length considerations we reduced the duration of each video by speeding the footage by 120% using Adobe Premiere Pro. As a result, each L1PT video was 13.2 (from 15.8) seconds and each L2PT video was 17.8 (from 21.4) seconds in length. Sample videos used in the L1PT (S1 Movie and S2 Movie) and L2PT tasks (S3 Movie and S4 Movie) are available in supporting information.

L1PT videos: The L1PT videos began with an agent seated at a table facing the participant. On the table, visible to both agent and participant, were two stationary homogenous balls (one red, one blue) and two wooden screens. In the first movement, the two balls simultaneously moved between the two screens so that they could not be seen by either the participant or agent. Following this movement, the events in the videos varied to create four belief-induction conditions. These conditions differed according to whether the *participant* expected a particular color to be present (P+) or absent (P-) in the outcome phase and, further, whether the *agent* expected a particular color to be present (A+) or absent (A-) in the outcome phase.

Expectations were induced by manipulating the movements of the balls and by varying the time that the agent left the scene. The agent's return to the scene signalled the onset of the final phase. There were two possible outcomes in the final phase: either a blue ball or a red ball was revealed when the screens rapidly fell away. As such, participants experienced 8 trial types, comprised of four belief-induction conditions paired with one of two possible outcomes (see Table 1a for an overview of conditions). For clarity and efficiency, we detail the four conditions (P+A+, P-A-, P+A-, P-A+) when paired with the blue outcome only (trials 1, 3, 5 and 7, as shaded in Table 1a).

Table 1. Belief-induction Conditions in the L1PT and L2PT Tasks

(a) L1PT task			
Condition	**Trial**	**Outcome P ≡ Outcome A**	**Expectations based on belief-induction phase**
P+A+	1	Blue	Both P and A expect the outcome.
	2	Red	Both P and A expect the outcome.
P-A-	3	Blue	Neither P or A expect the outcome.
	4	Red	Neither P or A expect the outcome.
P+A-	5	Blue	P, but not A, expects the outcome.
	6	Red	P, but not A, expects the outcome.
P-A+	7	Blue	A, but not P, expects the outcome.
	8	Red	A, but not P, expects the outcome.

(b) L2PT task				
Condition	**Trial**	**Outcome P**	**Outcome A**	**Expectations based on belief-induction phase**
P+A+	1	Blue	Red	Both P and A expect the outcome.
	2	Red	Blue	Both P and A expect the outcome.
P-A-	3	Blue	Red	Neither P or A expect the outcome.
	4	Red	Blue	Neither P or A expect the outcome.
P+A-	5	Blue	Red	P, but not A, expects the outcome.
	6	Red	Blue	P, but not A, expects the outcome.
P-A+	7	Blue	Red	A, but not P, expects the outcome.
	8	Red	Blue	A, but not P, expects the outcome.

Each condition is described following the first movement (in which both balls moved between the screens). Let us first consider the P+A+ and P-A- conditions which resulted in expectations that were consistent between the participant and agent. As illustrated in Fig 1, events in the P+A+ condition led both the participant and the agent to expect the presence of the blue ball in the outcome phase; in the final movement, both saw the red ball exit the scene, inducing a belief that the blue ball remained between the screens. Likewise, in the P-A- condition, both participant and agent witnessed the blue ball ultimately exit the scene, so that neither were led to believe that a blue ball would be revealed in the outcome phase (i.e., both were expecting the presence of the red ball). The P+A- and P-A+ conditions induced inconsistent expectations. In the P+A- condition, the participant and agent saw the blue ball leave the scene. However, the agent was absent when the red ball exited and the blue ball returned to rest between the screens. In this case, the participant was led to expect the outcome but agent was not. Finally, in the P-A+ condition the agent was present when the red ball left the scene but did not witness the red ball's return after the blue ball's exit. Again, the agent's and participant's expectations were inconsistent as the eventual outcome was not expected by the participant, but it *was* expected by the agent.

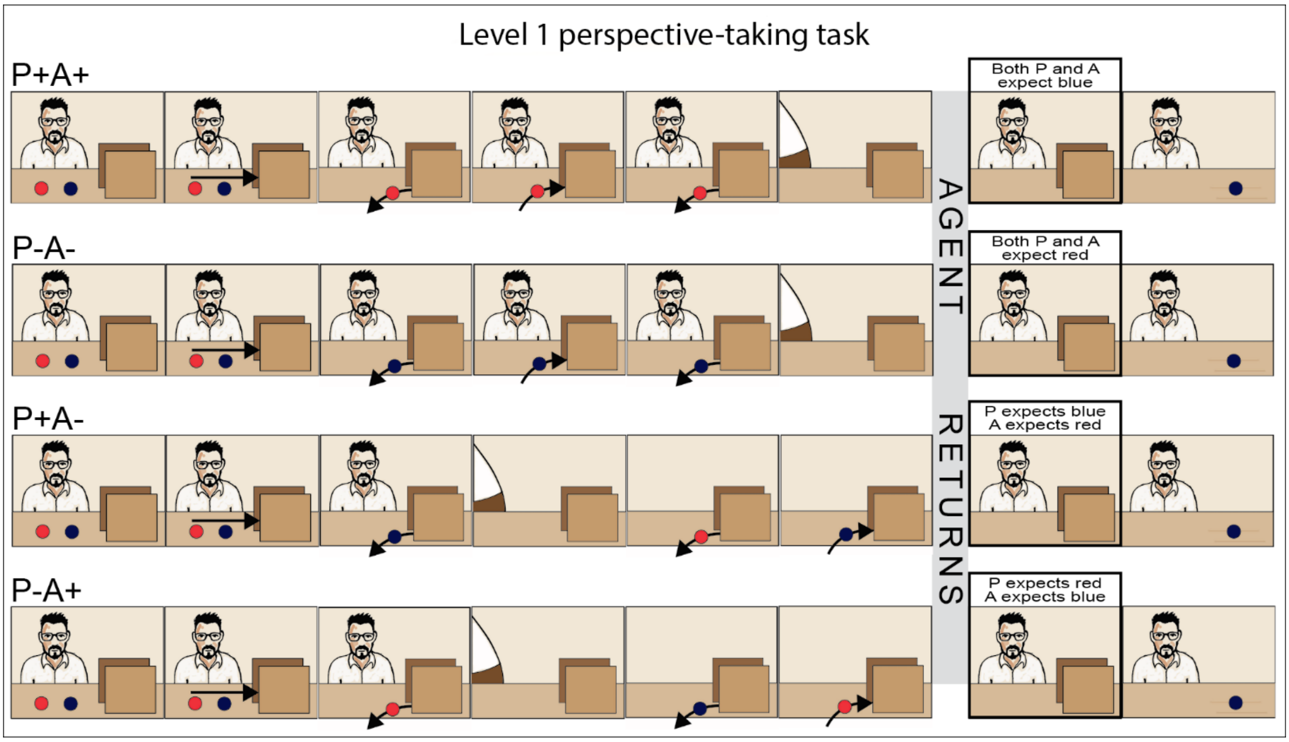

Fig 1. Schematic Storyboard Showing the Main Belief-inducing Events of the Four Conditions in the L1PT Task Movies. The main belief-inducing events represent conditions where there is a blue outcome. In the P+A+ condition (consistent), both participant and agent expected blue; in the P-A- condition (consistent) neither participant nor agent expected blue. In the P+A- condition (inconsistent) only the participant expected blue, and in the P-A+ condition (inconsistent) only the agent expected the blue.

L2PT videos: The L2PT videos were designed to match the L1PT videos as closely as possible. Each video began with the same agent seated at a table facing the participant. The screens were present but instead of there being two balls on the table there was a single object (a dog-robot) that was blue on one side and red on the other (see Fig 2). The dual nature of this object was revealed to the participant and agent at the beginning of each video; it twice turned 180˚ (anticlockwise) before making its initial move behind the screens.

Fig 2. Object Used in the L2PT Task. Turning through 180˚, the heterogeneous object presents its red and blue aspects to the participant and agent.

As in the L1PT task, the sequence of events leading up to the final phase varied (according to the object's movements and timing of agent's departure) to create four conditions culminating in one of two outcomes ('blue-facing-participant' or 'red-facing-participant'). This combination resulted in 8 trials types (see Table 1b for an overview of all conditions in the L2PT task). Here, we detail the four conditions when paired with blue-facing-participant outcomes (trials 1, 3, 5 and 7, as shaded in Table 1b). In Fig 3 we illustrate the critical belief-inducing events following the initial spinning motion of the dog-robot and its first movement between the screens (common to all conditions). Due to the dual nature of the object, the participant's and agent's beliefs were consistent when they expected *different* colors in the outcome phase. For example, in the P+A+ condition, where both the participant and agent expect the eventual outcome (blue-facing-participant, red-facing-agent) the dog-robot's blue aspect was presented to the participant in its last movement inducing a belief in the participant that the blue aspect would be revealed in the outcome. From the agent's viewpoint, the red aspect was presented when the dog-robot made its last move behind the screens so the agent was induced to believe he would see a red aspect when the screens dropped. Similarly, expectations were consistent in the P-A- condition. Before its final move between the screens, the dog-robot spun to reveal its red aspect to the participant and its blue aspect to agent. As a result, neither the participant nor agent expected the eventual outcome. In the P+A- condition the agent was induced to believe that the blue aspect would be revealed as he saw the object's blue aspect enter the screens before he left the scene. In the agent's absence, the participant then saw the dog-robot re-emerge and spin to reveal its blue aspect to the participant before returning behind the screens (with its red aspect facing the agent). In this case, both the participant and the agent last saw the object's blue aspect, but the outcome only met the participant's expectation. Finally, in the P-A+ condition the agent expected the eventual outcome (blue-facing-participant) because he last saw the dog-robot's red aspect enter the screens. The participant, however, saw (in the agent's absence) that the dog re-emerged, turned to present its red aspect to the participant, then retreated behind the screens. The events of the P-A+ condition induced the agent, but not the participant, to expect the outcome.

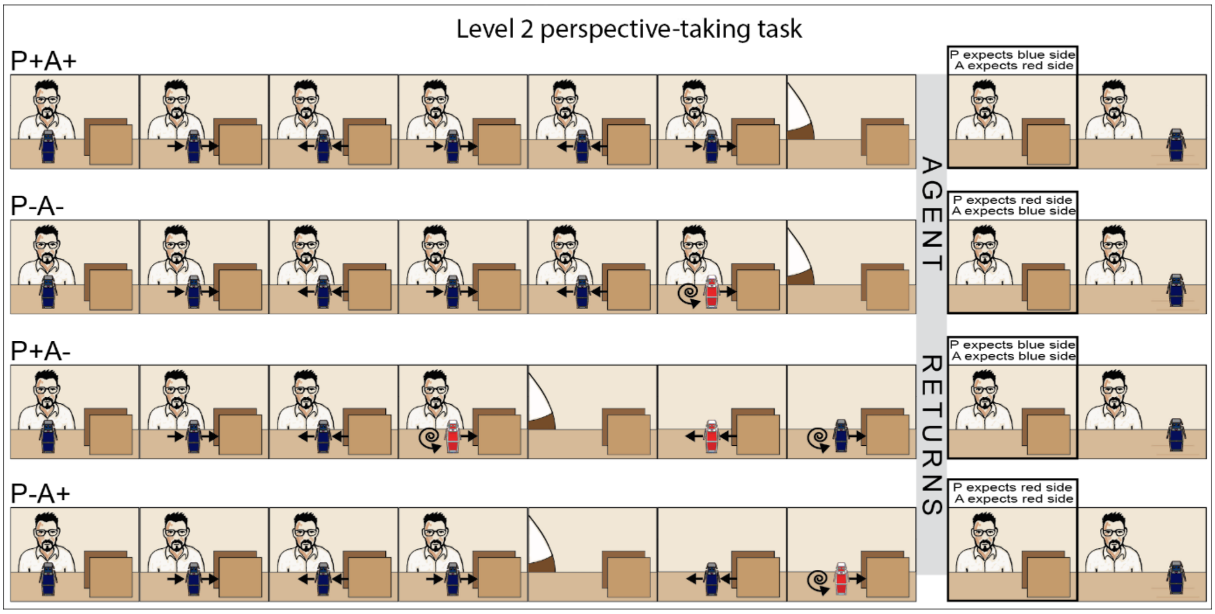

Fig 3. Schematic Storyboard Showing the Main Belief-inducing Events of the Four Conditions in the L2PT Task Movies. The main belief-inducing events represent conditions in blue-facing-participant outcomes. In the P+A+ condition (consistent), the participant expects blue and the agent expects red; in the P-A- condition (consistent) the participant expects red and the agent expects blue; in the P+A- condition (inconsistent) both participant and agent expect blue; and in the P-A+ condition (inconsistent) both participant and agent expect red.

2.1.3 Procedure

Participants were tested in a room in which there were two stand-alone workstation cubicles, so that one or two adults could separately and privately undertake the experiment in a single session. Each person sat at a Dell Optiplex 9020 desktop with a 23" screen (16:9 aspect ratio). Participants were guided through each task via on-screen directions which explained the format of the test trials and provided the correct procedure for responding. The initial screen stated, "This is an object-detection task. Your job is to press a key as quickly as you can when you see something appear behind a wall". Task order was counterbalanced; the L1PT task instructions were as follows (the L2PT task instructions were identical except for the information in brackets): "In the first half of the experiment you will see 40 videos, lasting a total of about 10 (15) minutes. They will look like this (relevant frame of video provided). In each video, the person will leave the scene, then return. Press the 'Q' key with your left hand as soon as the person has completely left the scene. When the walls disappear do one of the following with your right hand: Press the 'N' key if BLUE is revealed; Press the 'M' key if RED is revealed". The outcome response buttons, both depressed by fingers of the right hand, were not counterbalanced.

Each trial consisted of an initial fixation cross (1000ms), then a short video. During each video, the participant had to make two responses: an attention check (pressing a key within 2000ms of the agent leaving the scene), and a color detection (selecting blue or red when an object was revealed). The timings of each trial's events differed by task and condition (see S1 Fig for timings of critical events in the L1PT and L2PT tasks). For each task, 40 test trials were presented in a pseudorandom order in two blocks. The first block contained 24 trials comprising three cycles of four different conditions with a red or a blue outcome. After a student-led break the participants experienced another block of 16 trials (two cycles of four different conditions with either a red or a blue outcome). Thus, over the two tasks, participants experienced 80 trials in total. A training phase exposed participants to 4 practice trials with feedback. These were undertaken before the experimental trials of each task. No performance feedback was given during the test phase to minimize trial time and distraction. The entire experiment took approximately 30 minutes in total. On completion of the experiment participants were asked to complete a form purportedly surveying their experience of how easy it was to sign up for laboratory experiments in exchange for partial course credits (e.g., "Have you found it easy to find suitable timeslots?"). The final question, "What was the experimenter testing?" sought to determine whether the participants were primed to consider the bystander's belief. Although not a funneled debriefing protocol we were confident from survey responding that mental state attribution was not deemed to be the target of our research; all answers referred to the measuring of attention and/or reaction times in the pursuit of object detection.

2.2 Results and Discussion

All statistical analyses were conducted with IBM SPSS Statistics 23 (SPSS Inc., Chicago, IL, USA). Analysis was undertaken on correct responses, defined as those in which the participant detected a color that matched the revealed object. All statistical tests were two-tailed. We excluded reaction times for trials in which participants failed to respond to an attention check (1.5% of trials). Following an outlier analysis, we removed all data points greater than 3 standard deviations above or below the participant's overall mean in each task. As a result, 39 individual RTs were omitted (0.6% of individual responses in the L1PT task and 1.2% of individual responses in the L2PT task). Tests for normality revealed a positive skew in reaction times and error rates. We performed a logarithmic transformation of the reaction time data to fit the assumptions of an ANOVA before proceeding with further statistical analyses. As such, all means and standard deviations reported in the main text describe logarithmically transformed data. Mean response times for each condition are presented in S1 Table (transformed) and S2 Table (untransformed). The extent of the positive skew for the error data necessitated non-parametric testing. Mean error proportions are presented in S3 Table. Greenhouse Geisser corrections were used whenever the assumption of sphericity was violated.

2.2.1 Response times

In keeping with Kovács et al.'s (2010) analyses we initially compared responding between conditions. There was no theoretical basis to suggest that the color of the target in the outcome phase (blue or red) would influence responding, so we performed a 2 (Task: L1PT, L2PT) x 2 (Order: L1PT first, L2PT first) x 4 (Condition: P+A+, P+A-, P-A+, P-A-) mixed model ANOVA. We discovered a main effect of Task, F (1, 51) = 215.00, $p < .001$, $\eta p^2 = .81$; reaction times in the L1PT task (m = 2.61, sd = 0.13) were significantly faster than those in the L2PT task (m = 2.71, sd = 0.10). Planned comparisons between the corresponding conditions in each task (see S4 Table for an overview of analysis) revealed that reaction times were consistently slower in the L2PT task. We uncovered a main effect of Condition, F (1, 51) = 149.17, $p < .001$, $\eta p^2 = .75$, but no main effect of Order (p =.423. There was no 3-way interaction (p = .482), but we found a two-way Task x Order interaction, F (1, 52) = 20.19, $p < .001$, $\eta p^2 =$ .28. Post hoc independent samples t-tests found a single significant difference when comparing how participants performed in conditions depending on what order they completed the tasks; participants were faster in the P+A- condition if they completed the L1PT task first, t (52) = 2.17, $p < .036$, though this did not survive a Bonferroni correction. Finally, we found a two-way Task x Condition interaction, F (1, 52) = 50.06, $p < .001$, $\eta p^2 = .50$, which we explored further by task.

L1PT task: A one-way ANOVA revealed that response times differed significantly between conditions, F (2.54, 131.86) = 173.93, $p < .001$, $\eta p^2 = .78$. This was explored by performing Bonferroni-corrected pairwise comparisons. Our critical prediction was supported: response times were significantly faster in the P-A+ condition than in the P-A- condition, t (52) = 11.60, $p < .001$. We then compared response times for the other conditions (see S5 Table for an overview of pairwise comparisons). The pattern of responding is shown in Fig 4A: participants were fastest to respond in the P+A+ condition and slowest to respond in the P-A- condition; in addition, their reaction times in the P+A- condition were significantly faster than in the P-A+ condition. These findings suggest that, in the L1PT task, speed of response was modulated by both the participants' and the bystander's beliefs.

L2PT task: Participants' reaction times differed per condition, as revealed by a one-way ANOVA, F (2.35, 122.19) = 31.32, $p < .001$, $\eta p^2 = .38$. Bonferroni-corrected pairwise comparisons showed that there was no difference between response times in the P-A+ and P-A- conditions ($p = .689$) supporting our primary hypothesis for this task. As illustrated in Fig 4B, the pattern of responding diverged from the L1PT task. In the L2PT task there was no difference between the P+A+ and P+A- conditions, and no difference between the P-A- and P-A+ conditions, suggesting that participants were not influenced by the bystander's belief. A statistical overview of the pairwise comparisons for each condition is provided in S5 Table.

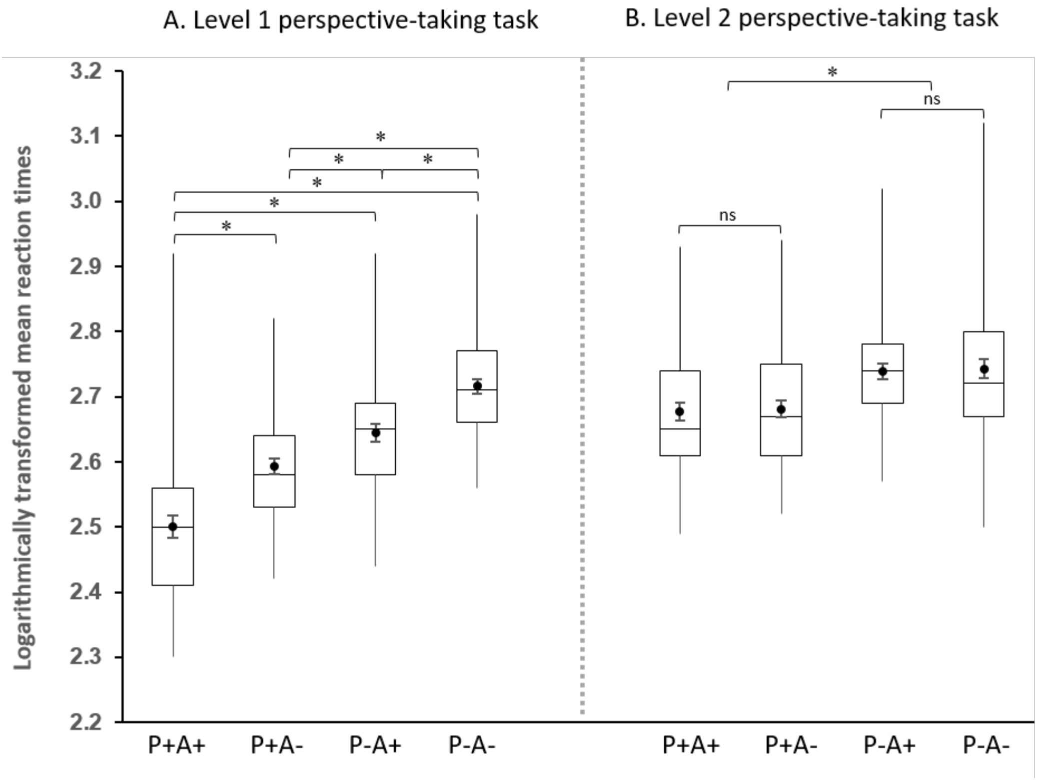

Fig 4. Logarithmically Transformed Mean Response Times for Experiment 1 (N=53).
Panel A shows box plots and logarithmically transformed mean response times for the four conditions in the L1PT task. Panel B shows box plots and logarithmically transformed response times in the L2PT task. Means are represented by dot markers; associated error bars represent the standard error of the mean. Note: * p < .01, two-tailed tests.

An orthogonal analysis was also undertaken to explore the influence of the participant's belief and agent's belief. We performed a 2 (Task: L1PT, L2PT) x 2 (Belief holder: P, A) x 2 (Belief: +, -) repeated measures ANOVA. To do this we first organised the data to create four scenarios, P+, P-, A+ and A-. In P+ scenarios ([P+A+] + [P+A-])/2), participants were led to expect the outcome, whereas in P- scenarios ([P-A+] + [P-A-])/2), events were designed so that the participant did not expect the outcome. In A+ scenarios ([P+A+] + [P-A+])/2) the agent is led to expect the outcome, whereas in A- scenarios ([P+A-] + [P-A-])/2), the outcome is unexpected by the agent. Main effects of Task, F (1, 52) = 151.49, p < .001, ηp^2 = .74, Belief holder, F (1, 52) = 23.15, p < .001, ηp^2 = .31, and Belief, F (1, 52) = 366.29, p < .001, ηp^2 = .88 were revealed. There was no three-way interaction (p = .634) but we a discovered a Task x Belief-holder interaction, F (1, 52) = 6.41, p = .014, ηp^2 = .11, and a Task x Belief interaction, F (1, 52) = 125.34, p < .001, ηp^2 = .71, which were further investigated by task.

L1PT task: A 2 (Belief-holder: P, A) x 2 (Belief: +, -) repeated measures ANOVA revealed a main effect of Belief holder, F (1, 52) = 19.67, $p < .001$, $\eta p^2 = .28$, and a main effect of Belief, F (1, 52) = 477.35, $p < .001$, $\eta p^2 = .90$. However, these main effects were qualified by an interaction, F (1, 52) = 22.44, $p < .001$, $\eta p^2 = .30$. Overall, individuals were quicker to respond when outcomes were expected, compared to when they were not, but the effect of belief depended on the Belief holder. As depicted in Fig 5A, participants were faster to respond when the agent expected the outcome (A+; m = 2.58, sd = .11) compared to when the agent did not expect the outcome (A-; m = 2.66, sd = .08), and they were faster when they themselves expected the outcome (P+; m = 2.55, sd = .10) compared to when they did not (P-; m = 2.68, sd = .09), but the difference between expecting outcomes and not expecting outcomes was greater for the participant-held beliefs.

L2PT task: A 2 (Belief: P, A) x 2 (Belief holder: +, -) repeated measures ANOVA also found main effects of Belief holder, F (1, 52) = 9.60, $p = .003$, $\eta p^2 = .16$, and Belief, F (1, 52) = 47.91, $p < .001$, $\eta p^2 = .48$. Again, these main effects were qualified by an interaction, F (1, 52) = 53.18, $p < .001$, ηp^2 = .501 (see Fig 5B). In this case, whilst individuals were faster to respond when they expected the outcome (P+; m = 2.68, sd = .09), compared to when they did not (P-; m = 2.74, sd = .09) scenarios), there was no significant difference in responding between scenarios in which the agent expected the outcome (A+; m = 2.71, sd = .09)) and those in which agent did not (A-; m = 2.72, sd = .09).

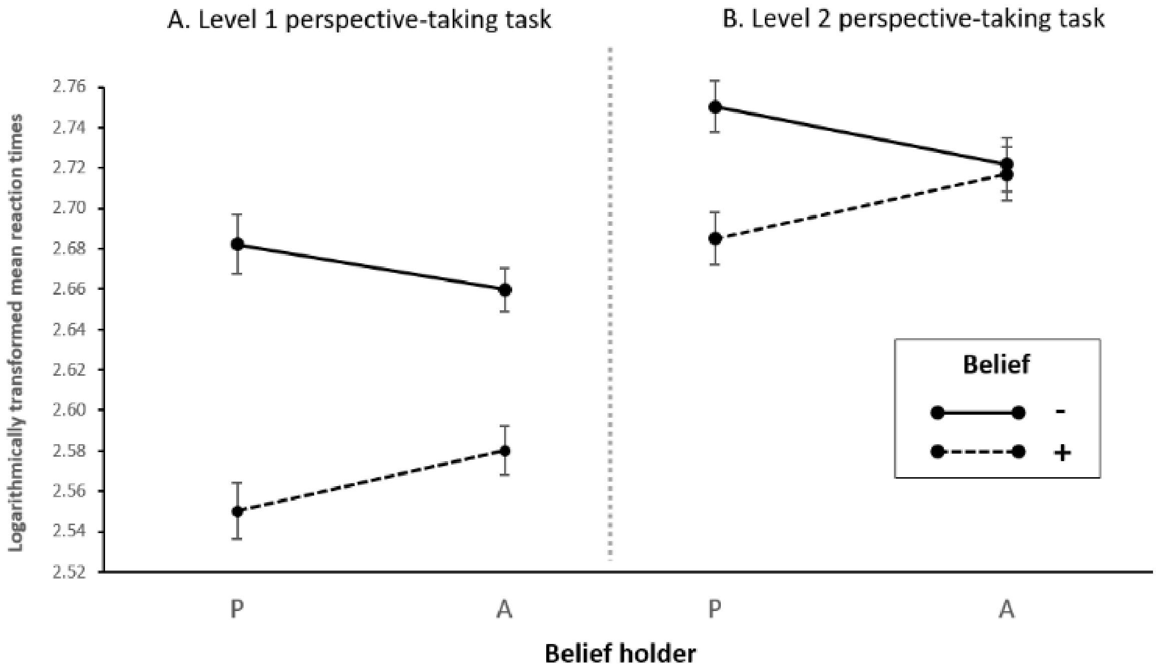

Fig 5. Orthogonal Analyses for Experiment 1 (N=53).
Panels A and B show the interactions between Belief-holder and Belief for the L1PT and L2PT tasks, respectively. Means are represented by dot markers; associated error bars represent the standard error of the mean. Note: 'P' = Participant; 'A' = Agent; '+' = Expected outcome; '-' = Unexpected outcome.

Our finding that participants were faster in P+ compared to P- scenarios, often referred to as the reality bias (e.g., Bardi et al., 2018; Bardi, Six, & Brass, 2017; Deschrijver et al., 2016), suggests that participants were attending to each trial's events and using them to predict outcomes, rather than just waiting for the screens to drop to make their color selection. Moreover, the reality bias was observed in both tasks. Comparing performances in the A+ and A- scenarios, it appears that there was only a facilitating influence of the agent's belief-like state in the L1PT task.

2.2.2 Errors

Overall, participants displayed high accuracy levels; the median error proportion was zero for each of the 16 trial types. The mean error proportions in the L1PT and L2PT tasks were .05 and .04 respectively (see S3 Table for mean error proportions and standard deviations for each condition and trial type). We analyzed mean error rates using non-parametric tests as tests for normality revealed a large positive skew. After collapsing the color of the outcome variable, a Friedman test revealed no difference in error proportions across the 8 conditions (the 4 conditions in each task), $\chi^2(7) = 7.87$, $p = .344$.

To summarise, in keeping with Kovács et al.'s (2010) original study, not only were participants in the L1PT task faster to detect the outcome when *they* expected the outcome, they were also faster to detect the outcome when only the agent expected the outcome (P-A+, compared to P-A- condition). By contrast, in the L2PT task there was no facilitating influence of the agent's belief, indicating that his belief relativized to his visuospatial perspective about the outcome was not automatically processed. A post hoc power analysis using G*Power determined that we had 99.99% power to calculate the critical effect with the current sample. These findings support our hypotheses and elaborate upon the dual-process account of human mindreading by suggesting that registration of perspective differences is likely to be eschewed by the efficient mindreading process. However, to be confident that our findings (that adults automatically track an agent's belief about which of two objects he is expecting to see) are a conceptual extension of classical findings from the original object-detection paradigm (and not a completely different phenomenon), Experiment 2 explored whether the current findings could be replicated when participants had to detect the presence or absence of a single object.

3 Experiment 2

3.1 Method

3.1.1 Participants

Participants in Experiment 2 were 60 right-handed adults, 39 of which were students who participated in partial fulfilment of course requirements, and 21 who were adult volunteers who responded to an advert placed in a community playcentre. There were 38 females and 22 males, with an age mean of 21.88 years (Range 18 to 36). The study was approved by Victoria University of Wellington's Human Ethics Committee. The sample size of 60 participants was greater than the minimum number of participants required to detect Kovács, Téglás and Endress' (2010) critical effect, providing safeguards against potential procedural errors and/or absenteeism.

3.1.2 Materials

Stimuli and instructions were presented using E-Prime 2.0 using the same display parameters as Experiment 1. Each individual watched 40 short videos as part of an object-detection task. Each video was 10 seconds in length (after speeding the original footage by 120% in Adobe Premiere Pro). Sample videos are available in supporting information (S5 and S6 Movies). As in Experiment 1, the videos began with an agent seated at a table (on which were two screens) facing the participant. In contrast to the videos shown in Experiment 1, the to-be-detected object was now a single black ball. In the first movement, the ball moved between the two screens so that it could not be seen by either the participant or agent. Following this movement, the events in the videos varied to create four belief-induction conditions. These conditions differed according to whether the participant expected the ball to be present (P+) or absent (P-) in the outcome phase, and whether the agent expected the ball to be present (A+) or absent (A-) in the outcome phase.

Expectations were induced by manipulating the movements of the ball and by varying the time that the agent left the scene (before or after critical events). The agent's return to the scene signalled the onset of the final phase. There were two possible outcomes in the final phase: the ball was either present or absent when the screens rapidly fell away. As such, participants experienced 8 trial types, comprised of four belief-induction conditions paired with one of two possible outcomes.

Events in the P+A+ condition led both the participant and the agent to expect the presence of the ball in the outcome phase. In the P-A- condition, both participant and agent were led to believe that the ball had left the scene. The P+A- and P-A+ conditions induced inconsistent expectations. In the P+A- condition, the participant and agent saw the ball leave the scene. However, the agent was absent when the ball returned to rest between the screens. In this case, the participant was led to expect the presence of the ball but agent was led to expect its absence. Finally, in the P-A+ condition both participant and agent witnessed the ball moving between the screens but only the participant saw the ball leave the scene. In the outcome phase, the agent's and participant's expectations were inconsistent; the participant expected the ball to be absent while the agent expected it to be present (see S2 Fig for a schematic showing the main belief-inducing events of the four conditions).

3.1.3 Procedure

Participants were tested in the same room (with an identical arrangement) as in Experiment 1. Guidance regarding test format and response requirements was provided via on-screen prompts. Participants were instructed to detect the presence or absence of a single black ball. The initial screen stated, “This is an object-detection task. Your job is to press a key as quickly as you can when you see something appear behind a wall”. Further instructions explained, “You will see 40 videos lasting a total of about 10 minutes. They will look like this (relevant frame of video provided). In each video, the person will leave the scene, then return. Press the ‘Q’ key with your left hand as soon as the person has completely left the scene. When the walls disappear do one of the following with your right hand: Press the ‘N’ key if the ball is present; Press the ‘M’ key if the ball is absent”. The outcome response buttons were not counterbalanced.

As in Experiment 1 the timings of each trial’s events differed by condition (see S3 Fig for timings of critical events). The 40 test trials were presented in a pseudorandom order in two blocks. The first block contained 24 trials comprising three cycles of 8 trials (4 conditions x 2 outcomes) and the second block contained 16 trials (two cycles of 8 trials). A training phase exposed participants to 4 practice trials with feedback. These were undertaken before the experimental trials. No performance feedback was given during the test phase to minimize trial time and distraction. The entire experiment took approximately 15 minutes in total. On completion of the experiment participants were asked to fill out a survey asking them about their experience taking part in the University’s research programme. As in Experiment 1, all survey answers pertaining to the nature of the current task referred to attention and speed of object detection. Finally, participants were debriefed and their data collected.

3.2 Results and Discussion

All statistical analyses were conducted with IBM SPSS Statistics 23 (SPSS Inc., Chicago, IL, USA). Analysis was undertaken on correct responses, defined as those in which the participant accurately detected the presence or absence of the ball. All statistical tests were two-tailed. Error rates are reported separately below. We excluded reaction times for trials in which participants failed to respond to the attention check (4.42% of trials). Following an outlier analysis, we removed all data points greater than 3 standard deviations above or below the participant's overall mean in each task. As a result, 11 individual reaction times were omitted (0.45% of individual responses). Tests for normality revealed a positive skew in reaction times and error rates. We performed a logarithmic transformation of reaction time data to fit the assumptions of an ANOVA before proceeding with further statistical analysis. Transformed and untransformed means for response times are presented in S6 and S7 Tables, respectively. Due to the nature of the error data, analysis was conducted via non-parametric tests (see S8 Table for mean error proportions across conditions). Greenhouse Geisser corrections were used whenever the assumption of sphericity was violated.

3.2.1 Response times and errors

We performed a 2 (Outcome: ball-present, ball-absent) x 4 (Condition: P+A+, P+A-, P-A+, P-A-) repeated measures ANOVA. Main effects of Outcome (F (1, 59) = 35.62, $p < .001$, $\eta p^2 = .63$) and Condition, F (2.72, 160.03) = 27.58, $p < .001$, $\eta p^2 = .32$, were revealed, and a significant Outcome x Condition interaction was confirmed, F (1.54, 91.21) = 35.62, $p < .001$, $\eta p^2 = .38$. To interpret the interaction, a repeated measures ANOVA was performed for each outcome. For the ball-present conditions the repeated measures ANOVA revealed a main effect of Condition, F (1.56, 91.94) = 47.32, $p < .001$, $\eta p^2 = .45$. Post hoc tests showed that the critical prediction was supported: response times were significantly faster when just the agent expected the ball to be present (P-A+), compared to when neither agent nor participant expected it to be present (P-A-), t (59) = 7.83, $p < .001$. We conducted a post hoc power analysis and determined that we had 99.99% power to calculate the critical effect with the current sample size.

A statistical overview of the pairwise comparisons for all conditions in the ball-present and ball-absent trials is presented in S9 Table. Participants were fastest to detect the presence of the ball when both the participant and agent expected it to be present (P+A+ condition), and slowest to detect the ball when neither the participant nor agent expected it to be present (P-A-). Lastly, participants were quicker to detect the ball when they, but not the agent believed it was present compared to when the agent, but not the participant expected it to be present (see S4 Fig). These findings support the hypothesis that participants' reaction times are automatically influenced by the mere presence of others.

A repeated measures ANOVA also revealed a main effect of Condition for ball-absent trials, F (2.44, 144.01) = 9.13, $p < .001$, $\eta p^2 = .13$. Pairwise comparisons revealed no difference between the baseline condition (P-A-), in which neither the participant nor agent was expecting the absence of the ball, and the condition in which only the agent expected there to be no ball present (P+A-). P+A+ and P+A- responding was significantly slower than P-A- and P-A+ responding, though the P+A- versus P-A- comparison did not survive the Bonferroni correction. There was also no difference between response times in the P-A- and P-A+ conditions (see S4 Fig).

Finally, we undertook an orthogonal analysis of the ball-present data with a 2 (Belief holder: P, A) x 2 (Belief; +, -) repeated measures ANOVA. We found a main effect for Belief holder, F (1, 59) = 22.35, $p < .001$, $\eta p^2 = .28$, and a main effect of Belief, F (1, 59) = 71.35, $p < .001$, $\eta p^2 = .55$. However, the main effects were qualified by a Belief holder x Belief interaction, F (1, 59) = 22.97, $p < .001$, $\eta p^2 = .28$. This was explained by an observation that the effect of belief was stronger for P scenarios compared to A scenarios, that is, the difference between P+ (m = 2.62, sd = *.16*) and P- (m = 2.77, sd = *.08*) responding was larger than that between A+ (m = 2.67, sd = *.11*) and A- (m = 2.74, sd = *.10*).

Participants showed a high level of accuracy, revealed by low mean error proportions in both the ball-present and ball-absent conditions (.06 and .05 respectively). Tests for normality revealed that the error data was positively skewed. A Friedman test revealed no statistically significant differences in mean error proportions across the 8 trial types, $\chi^2(7) = 1.86$, $p = .967$.

To conclude, when expecting the ball to be present, responding is fastest when both the participants' and agents' beliefs match the outcome, and slowest when neither are induced to expect the outcome. In keeping with the theoretical basis for the study, not only are participants faster than the baseline condition (P-A-) to detect the ball when they, but not the agent, expect the outcome, they are also speeded when only the agent expects the ball to be present.

Could automatic belief-tracking merely reflect timing variations in the attention check (Phillips et al., 2015)? One possibility is that, in the L1PT task, adults are significantly slower to detect the correct color in the P-A- than in the P-A+ condition because there is a shorter duration between the attention check (which requires the participant to press a button when the agent leaves the scene) in the P-A-condition than in the P-A+ condition. In other words, a shorter stimulus onset asynchrony (SOA) in the P-A- condition than in the P-A+ condition leads to more protracted response times in the former. We do not believe that this is the likely explanation of our findings for several reasons. First, the attention-check hypothesis has been contested (e.g., Nijhof et al., 2016, 2017) on the grounds that the influence of a short SOA on the reaction time to a second stimulus (known as psychological refractory period) is a short-term effect, and only observable at SOAs up to several hundred milliseconds. The shortest SOAs found in the typical object-detection paradigm tend to be over 2,000 milliseconds, and the shortest time between the attention check and detection response in the current paper (> 4000ms) is substantially longer than refractory periods discussed in past literature. Second, in Experiment 1's L1PT task we consistently found faster responding in the P+A+ condition than in the P+A- condition even though the former condition had a shorter SOA. Third, in Experiment 1 adults were not faster to respond in the P-A+ condition than in the P-A- condition of the L2PT task, which would *not* be predicted if the key difference between those conditions was merely the result of a shorter SOA. Nonetheless, it may be argued that some factor associated with tracking a rotating object may have interfered with a potential attention-check effect. To fully mitigate concerns over differences in refractory periods across trial types, we ran a second replication of Experiment 1, removing the attention checks from each condition.

4 Experiment 3

4.1 Method

4.1.1 Participants

A total of 108 right-handed psychology students volunteered to participate in partial fulfilment of course requirements. There were 82 females and 26 males, with an age mean of 18.92 years (Range 17 to 34 years). The study was approved by Victoria University of Wellington's Human Ethics Committee. We recruited a greater number of individuals in this study due to an increase in the availability of students in Victoria University of Wellington's IPRP and because there was concern that the removal of the attention check could result in a greater number of participants failing to meet our accuracy threshold of 75%.

4.1.2 *Materials and procedure*

The materials and procedure were identical to Experiment 1, except that there was no requirement for the participants to respond (by pressing the Q key) when the agent left the scene.

4.2 *Results and Discussion*

All statistical analyses were conducted with IBM SPSS Statistics 23 (SPSS Inc., Chicago, IL, USA). Analysis was undertaken on correct responses, defined as those in which the participant detected a color that matched the revealed object. Five participants were excluded from analysis as their performances were below the 75% accuracy threshold across all trials. Of the 103 remaining participants there were 79 females and 24 males, with a mean age of 18.8 years (range 17 to 34). We removed all individual data points greater than 3 standard deviations above or below the participant's overall mean in each task. As a result, 103 individual reaction times were omitted (0.6% of individual responses in the L1PT task and 0.6% of individual responses in the L2PT task). All statistical tests were two-tailed. Tests for normality revealed a positive skew in reaction times and error rates. We performed a logarithmic transformation of the reaction time data to fit the assumptions of an ANOVA before proceeding with further statistical analysis. Mean response times are presented in S10 Table (transformed) and S11 Table (untransformed). Error rates were compared across conditions using non-parametric tests (see S12 Table for mean error proportions). Greenhouse Geisser corrections were used whenever the assumption of sphericity was violated.

4.2.1 *Response times*

Informed by previous research, we performed a 2 (Task: L1PT, L2PT) x 2 (Order: L1PT first, L2PT first) x 4 (Condition: P+A+, P+A-, P-A+, P-A-) mixed model ANOVA. There was no three-way interaction (p = .597), Task x Order interaction (p = .311), Condition x Order interaction (p = .876), or main effect of Order (p = .556). However, there was main effect of Task, F (1, 101) = 30.68, p < .001, ηp^2 = .23; the mean reaction time in the L1PT task (m = 2.60, sd = .07) was smaller than that of the L2PT task (m = 2.62, sd = .07). There was also a main effect of Condition, F (2.67, 269.75) = 144.45, p < .001, ηp^2 = .57, and a two-way Task x Condition interaction, F (2.71, 273.58) = 10.78, p < .001, ηp^2 = .10, which we explored further after separating the data by task.

L1PT task: A repeated measures ANOVA showed that performance significantly differed across conditions, F (2.91, 296.78) = 85.26, $p < .001$, $\eta p^2 = .45$. Supporting our critical prediction, we determined that response times were significantly faster in the P-A+ condition than in the P-A- condition, t (102) = 8.05, $p < .001$. Bonferroni-corrected pairwise comparisons between the other L1PT conditions (see S13 Table for an overview) provided a pattern of findings that is illustrated in Fig 6A. Fastest responding was found in the P+A+ condition and slowest responding in the P-A- condition, but there was no significant difference between the P+A- and P-A+ conditions. These findings indicate that speed of response was modulated by both the participants' and the bystander's beliefs.

L2PT task: Reaction times differed between conditions, as revealed by a repeated measures ANOVA, F (2.43, 247.59) = 79.71, $p < .001$, $\eta p^2 = .44$. Focusing on the critical conditions we found support for our primary L2PT hypothesis: there was no difference between response times in the P-A+ and P-A- conditions ($p = .75$). As depicted in Fig 6B, the pattern of responding diverged from the L1PT task. In the L2PT task there was no difference between the P+A+ and P+A- conditions, and no difference between the P-A- and P-A+ conditions, indicating that participants were not influenced by the bystander's belief. A statistical overview of the pairwise comparisons for each condition is provided in S13 Table.

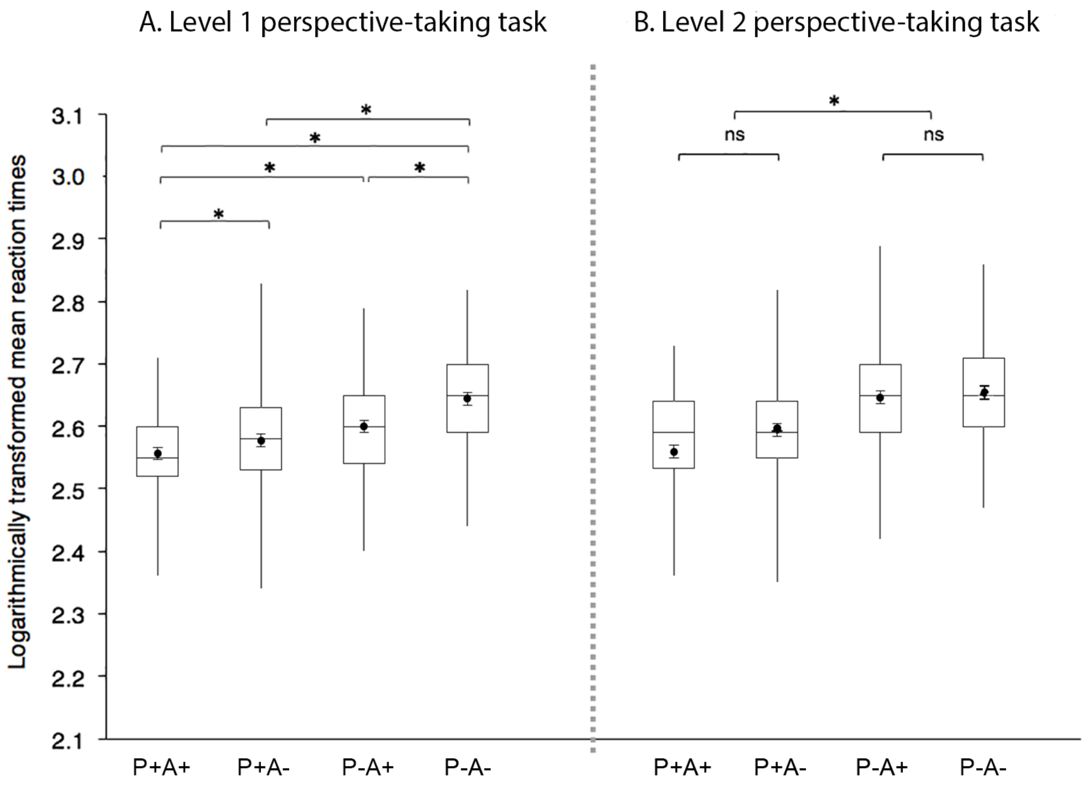

Fig 6. Logarithmically Transformed Mean Response Times for Experiment 3 (N=103). Panel A shows box plots and logarithmically transformed mean response times for the four conditions in the L1PT task. Panel B shows box plots and logarithmically transformed response times in the L2PT task. Means are represented by dot markers; associated error bars represent the standard error of the mean. Note: * p < .01, two-tailed tests.

As in Experiments 1 and 2, we conducted orthogonal analyses to examine the influence of participants' and agent's beliefs. A 2 (Task: L1PT, L2PT) x 2 (Belief holder: P, A) x 2 (Belief: +, -) repeated measures ANOVA revealed main effects of Task, F (1, 102) = 35.66, $p < .001$, $\eta p^2 = .26$, Belief holder, F (1, 102) = 7.09, $p = .009$, $\eta p^2 = .07$, and Belief, F (1, 102) = 259.79, $p < .001$, $\eta p^2 = .72$. There was no Task x Belief interaction ($p = .066$), but we did find a three-way interaction, F (1, 102) = 32.80, $p < .001$, $\eta p^2 = .24$, and two-way interactions between Task and Belief holder, F (1, 102) = 5.52, $p = .021$, $\eta p^2 = .05$ and between Belief holder and Belief, F (1, 102) = 62.70, $p < .001$, $\eta p^2 = .38$. These were investigated further by task.

L1PT task: A 2 (Belief holder: P, A) x 2 (Belief: +, -) repeated measures ANOVA uncovered a main effect of Belief holder, F (1, 102) = 5.77, $p = .020$, $\eta p^2 = .05$, and a main effect of Belief, F (1, 102) = 261.84, $p < .001$, $\eta p^2 = .72$. However, these findings were qualified by an interaction, F (1, 102) = 5.69,

$p = .019$, $\eta p^2 = .05$. Replicating Experiment 1's findings, we found that, overall, individuals were quicker to respond when beliefs contained an expectation of the outcome (+), compared to when they did not (-), however the effect of Belief depended on the Belief holder. As illustrated in Fig 7A, the response differential between P+ scenarios ($m = 2.57$, $sd = .06$) and P- scenarios ($m = 2.63$, $sd = .06$), was greater than the response differential between A+ ($m = 2.58$, $sd = .05$) and A- scenarios ($m = 2.62$, $sd = .06$).

L2PT task: A 2 (Belief holder: P, A) x 2 (Belief: +, -) repeated measures ANOVA revealed a main effect of Belief holder, F (1, 102) = 6.37, $p < .013$, $\eta p^2 = .06$, and a main effect of Belief, F (1, 102) = 107.12, $p < .001$, $\eta p^2 = .05$. Again, these main effects were qualified by an interaction, F (1, 102) = 106.10, $p < .001$, $\eta p^2 = .05$, which is depicted in Fig 7B. It was observed that individuals were faster to respond in P+ scenarios ($m = 2.59$, $sd = .06$) compared to P- scenarios ($m = 2.65$, $sd = .06$), but there was no significant difference in responding for A+ ($m = 2.62$, $sd = .05$) versus A- ($m = 2.63$, $sd = .06$) scenarios.

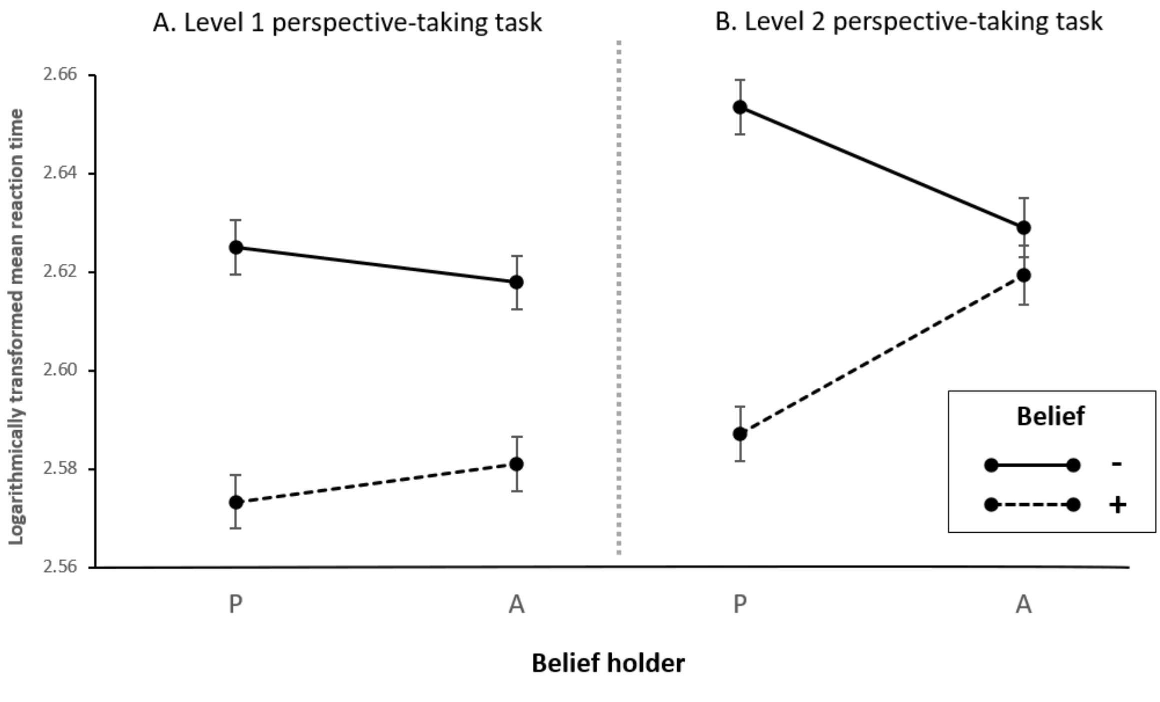

Fig 7. Orthogonal Analyses for Experiment 3 (N=103).
Panels A and B show the interactions between Belief-holder and Belief for the L1PT and L2PT tasks, respectively. Means are represented by dot markers; associated error bars represent the standard error of the mean. Note: 'P' = Participant; 'A' = Agent; '+' = Expected outcome; '-' = Unexpected outcome.

Replicating Experiments 1 and 2, these findings suggest the presence of a reality bias (Bardi et al., 2018, 2017; Deschrijver et al., 2016) in both tasks, inferring that participants do use their own beliefs about the position and/or orientation of the object/s when detecting the color outcome. However it seems that the agent's beliefs are only taken into account in the L1PT task which does not involve contrasting perspectives.

4.2.2 Errors

Overall, participants displayed high accuracy levels; the median error proportion was zero for each of the 16 trial types. The mean error proportions in the L1PT and L2PT tasks were .05 and .04 respectively (see S8 Table for mean error proportions and standard deviations for each trial type). We analyzed mean error proportions using non-parametric tests as tests for normality revealed a large positive skew. A Friedman test revealed no significant difference in mean error proportions across the 8 conditions (4 conditions in each task), $\chi^2(7) = 13.32$, $p = .065$.

In sum, the pattern of reaction-times emulates that of Experiment 1, even in the absence of the attention check. In the L1PT task not only are participants faster to detect the outcome when they *expect* the outcome, they are also faster to detect the outcome when only the agent expects the outcome. By contrast, in the L2PT task there is no facilitating influence of the agent, indicating that his belief about the outcome is not automatically processed in this instance. A post hoc power analysis determined that we had 99.99% power to calculate the critical effect with the current sample size. Removing the attention check did not have any impact on participants' accuracy compared to Experiment 1, implying that this procedural change did not adversely affect engagement with either task.

5 General discussion

Experiment 1 tested the extent to which participants automatically tracked the beliefs of a passive bystander in two closely-matched but conceptually distinct tasks. In the L1PT object-detection task involving homogenous objects, adults' reaction times were involuntarily influenced by the presence of a passive bystander. Participants were faster to detect the color of an object when the agent, but not the participant (P-A+), expected the outcome, compared to a baseline condition in which neither expected the outcome (P-A-). By contrast, in our L2PT task, the presence of the agent did not influence adults' response times when the to-be-detected object could be differently perceived depending on where the agent was located in relation to that object. In this scenario, reaction times for the pairs of conditions in which a participant expected a certain color to be revealed (P+A+, P+A-) were significantly faster than the pair of conditions (P-A-, P-A+) in which the participant did not expect a certain color to be revealed. The pattern of responding in the L2PT task indicated that reaction times were contingent on participants' expectations only. In Experiment 2, we replicated the critical effect of automatic belief-tracking when only one homogenous ball was used, and the agent's perspective was not relevantly different. Experiment 3 sought to rule out the possibility that response times in the object-detection paradigm may be influenced by differences in the timings of the attention checks across conditions. Using the same procedure and materials as in Experiment 1, we found that the overall pattern of responding was not affected when we removed the requirement to respond when the agent left the scene. These findings were also supported by an orthogonal analysis investigating the influence of participants' own beliefs (P+, P-) and the belief of the agent (A+, A-). Overall, we conclude that adults automatically track others' beliefs concerning where an object is located but not their beliefs of how an object is perceived from a certain perspective.

The present findings raise a fundamental point that proponents of the dual-process account of human mindreading have not addressed in the literature. The P-A+<P-A- critical effect, as detected in responses to the L1PT task, is readily explained in terms of a minimal model of the mind: humans efficiently model other people's minds in terms of registrations (relationships to objects), even when the encoding of others' belief-like states is completely irrelevant to the task being performed. However, the obliteration of the critical effect in the L2PT task cannot be explained by a breakdown in the ability to efficiently process object identity per se. Explorations of signature limits on efficient processing often rely on belief-reasoning tasks that are designed to exploit the subtle understanding that attributions of identity can generate mistakes in the numerical sense. To clarify, there are two kinds of numerical identity mistakes: compression, in which there are in fact two entities but someone falsely believes there is one, and expansion, in which there is in fact one entity but someone falsely believes there are two. The rotation of the dog-robot toy was revealed to the agent so there is nothing to suggest that the agent is

necessarily going to make mistakes about identity in the numerical sense, that is, to think that there are two robots when there is really one.

One conjecture is that that representations underlying automatic belief-tracking either do not specify agents' locations or do not specify objects' orientation properties, or perhaps neither. This conjecture generates the prediction that automatic belief-tracking alone will not yield expectations about agents' perspectives, which would explain the elimination of the critical effect in the L2PT task. If the participant has not encoded where the agent was when she last encountered the object (the agent could have been on either side of the table), she cannot make a prediction about what the agent expects to see. If the participant has encoded the agent's location but only encoded the object as a bare object (that is, its orientation is not part of the registration), then the participant has the object, the registration, and the agent's current location, but he or she cannot go back and work out what the agent is expecting to see.

Before our current findings, it was an open question as to whether registration, being a relationship to an object and its location, might include detailed information about the agent. Our findings in the L1PT task suggest that the P-A+<P-A- effect can be explained by registration alone (where the object was at time of registration) without the need to assume that the registered location amounts to a belief state. The elimination of the P-A+<P-A- effect in our L2PT task suggests that registration as a belief-like state is further impoverished in not taking into account the agent's position in space in relation to the object. In belief-tracking, representing the agent's location and orientation would be relevant to understanding how someone perceives and expects the world to be, but perhaps there is a distinction between representing the agent merely as an individual when assigning the representation, and representing the agent's position in space as *part* of the registration. Thus, one possibility is that the registration comprises the spatial location of the agent and all entities in the agent's field. Another possibility is that the agent's presence may trigger the generation of a registration containing only [Objects seen by agent] (see Surtees, Samson, & Apperly, 2016). In other words, the agent's visual as well as spatial perspective can be important for what the agent registers, but the efficient mindreading process may not necessarily encode and/or store those parameters within the registration itself. If we are asking a question, as applied to our L2PT task, about what the agent expects to see or happen when the screens drop, we can answer that question using a flexible mindreading process based both on what the

agent believed he last perceived from that spatial position and imagining ourselves in the agent's current position. Our findings suggest that efficient mindreading is not set to handle different beliefs in combination with perspectives, as it seems that tracking registration encodes where the dog-robot-object's is placed in the scene but perhaps not how the agent is located with respect to the dog-robot, or how the dog-robot is represented from that location. Our findings showing adults' resistance to the influence of an agent's perspective and belief in the L2PT task reveals important information about the specific parameters of the signature limit that constrains the efficient and relatively automatic mindreading process. If the encoding of someone's belief, vis-à-vis how the person's location in space restricts the aspects of the object in focus, is naturally eschewed by an efficient mindreading process, it would explain why studies show that adults are immune to altercentric interference over how others experience the meaning of rotationally asymmetrical digits (e.g., a number that looks like a 6 to the participant and a 9 to the agent) (Surtees et al., 2012).

Adults were slower to react in the L2PT task than in the L1PT task in Experiments 1 and 3 (error rates were very low in both tasks). A potential concern might be that the critical P-A+<P-A- effect was present in the L2PT task but hidden by the longer detection responses. For example, participants may have acknowledged the difference in perspective between self and agent and slowed down accordingly, masking the effect of the automatic processing. However, for this claim to be substantiated we would have seen greater reaction times in the L2PT task than in the L1PT task only when there was a difference of belief between the participant and agent (i.e., the inconsistent conditions: P+A- and P-A+). On comparing reaction times in each condition, we found that this was not the case. One explanation of the condition-wide slowing down of L2PT reaction times may be that the participant, made aware of the perspective-relevant nature of the object for the self and other, is motivated to engage in flexible off-line mindreading by using an embodied representation of the self that is then rotated to the current bodily position of the agent's position in space (Surtees et al., 2013).

A different explanation which still preserves a dual-process account of human mindreading is that the content of the agent's registration that is efficiently tracked differs between tasks. For example, in the L1PT task, the participant tracks the agent's registration that a blue ball left the scene in the P-A- condition. When the blue ball is revealed, the encoded registration interferes with the color detection response, prolonging the reaction time in comparison with the P-A+ condition in which there is no such interference. By contrast, in the P-A- condition of the L2PT task, the participant may simply compute the agent's registration that the 'dog-robot' moved behind the occluders, so when the object is revealed in the outcome phase there is no such interference in comparison with the P-A+ condition. The nature of the task provokes the idea that neither participant nor agent tracked the dog-robot's color as it moved through the scene: participants may have paid no attention to the movements of the heterogeneous object during the trial and relied only upon the final revelation to make a color selection. Experiment 1's reality bias (P+<P-) was reduced for the L2PT task compared to the L1PT task, which in some part supports this conjecture. However, there was no replication of this finding with Experiment 3's larger sample, with the P+ versus P- differential being greater in the L2PT task then the L1PT task.

Another consideration is that reaction times are potentially influenced by three factors: the accuracy of participants' own beliefs, the accuracy of the agent's belief, and the content of the agent's belief (which may or may not be accurate). We should consider the possibility that there is a confound between the latter two factors, so that when we refer to the tracking of an agent's beliefs we are not clear whether it is the accuracy of the belief that is influencing the participant's behavior, or the content of the belief, or both. That said, the current paradigm is designed to de-confound the first two factors, as is standard in false-belief testing; the four experimental conditions exist precisely to separate the participant's own beliefs and expectations from the agent's beliefs or expectations. The distinction between belief content and belief accuracy is an important one but an experiment to de-confound them would need to be the subject of a future project.

We should also consider the possibility that human beings only have a single mindreading system that is sufficiently sophisticated to also enable speedy calculations of wide ranging mental-contents. The bystander might not have influenced adults' own responses in the L2PT scenario due to extraneous demands associated with that task. That said, in Experiment 1, we compared reactions times between well-matched situations: a task involving two distinct sides and a task involving two distinct objects. Moreover, we found that adults were influenced automatically by the agent's beliefs in the L1PT two-ball task and in the L1PT one-ball task, even when tracking beliefs about the path of two distinct objects is more cognitively demanding than tracking beliefs about the paths of one distinct object (Horowitz &

Cohen, 2010). Despite our efforts to match the cognitive structure of each task, it is tempting to claim that the L2PT task may be challenging because it makes unnecessary demands on rotation skills that mask the expression of sophisticated mindreading that is both flexible and efficient. In response, let us consider an alternative version of our L2PT task. Suppose that the participant sees the dog-robot object zip behind the screens, and then we attenuate demands on rotation by having the agent move around to the participant's position in space when he returns. Now if we are reasoning with a single mindreading system that is context sensitive, it is possible to predict that participants' reaction time will be modulated by what the agent believes he is expecting to see from his new position in space. We are currently testing this possibility, but that prediction involves participants successfully tracking both the nature of the object and the agent's position in space. On the other hand, if we have an efficient mindreading process where the agent's location is just not encoded or stored as part of the registration itself (as our current findings would suggest), then there should no evidence of adults' being automatically influenced by the agent's belief relative to how his expectations change as he moves in relation to the object. If the latter turns out to be the case, it is less about differences in demands between tasks that mask expression of mindreading competency but more about embodied mental rotation being conceptually and mechanistically closer to flexible rather than efficient mindreading processes.

While there is a growing number of studies utilizing the object-detection paradigm for measuring whether and to what extent certain mindreading inferences can be automatic, the conclusions drawn have been contentious given criticisms that the critical effects are just artefacts of the timings in the attention checks used by the researchers to ensure participants' task compliance (Phillips et al., 2015). However, a recent object-detection study found that Kovács and colleagues' (2010) critical P-A+<P-A- effect was maintained despite ensuring that the attention check occurred at exactly the same time across all trials (El Kaddouri et al., 2019). Another study, involving a group of adults with high functioning autism (Deschrijver et al., 2016), found a negative correlation between the size of the critical effect and the severity of autism spectrum disorder symptoms. Assuming that attention check performances were consistent across the group, this finding does not support the idea that attention check timings alone drive the difference between P-A+ and P-A- responding. In addition, Bardi et al. (2018) showed that whilst a critical effect was uncovered in a ball-detection task involving a human-like bystander, it was not

revealed in a ball-detection task involving a dog bystander, despite the attention check timings being the same for the two tasks. Our findings from Experiment 3 suggest that the critical P-A+<P-A- effect is stable and maintained even when attention checks are removed completely from the current task context. However, we cannot rule out the possibility that the P-A+<P-A- effect may be the result of other methodological factors. Furthermore, as per Phillip's et al.'s (2015) findings, it is too soon to make any firm conclusions about the confounding role of attention checks given that our work involves different materials and set up (e.g., forced choice instead of go-no-go response, real-life as opposed to animated agent, two occluders rather than one).

A legitimate question is why a relatively separate, and restricted, mindreading process – which persists beyond infancy and childhood - would have evolved in humans: how adaptive is a mental-state calculator that, under certain circumstances, breaks down? One possibility is that fast, but limited processing in adulthood may be an adaptive reaction to the demands of complex environments (Payne, Bettman, & Johnson, 1993). As social animals we have always needed to quickly predict the motives and actions of others (especially dangerous ones). As fully matured humans we routinely come across hurried instances in which we erroneously infer others' intentions, desires and beliefs, and our experiences also inform us that even the most studious deliberation of others' minds is far from fool proof. Although limited processing may lead to erroneous judgements, it is important to grasp that cognitive limitations are not exclusively linked to negative outcomes (Hertwig & Todd, 2003). Even in a simple visual detection task involving a homogenous object, we have shown that performance is enhanced by the automatic belief ascription of other agents.

6 Conclusions

In conclusion, our findings lean towards a dual-process account of mindreading and our study represents a move away from debating whether a mindreading process uses a minimal-theory-of-mind model, to assuming that it does and then working out what exactly the signature limit of the process might be. The current study's new and innovative version of an object-detection task also provides a promising tool for assessing the competing theories that seek to explain the cognitive architecture underlying humans' automatic and non-automatic mindreading abilities.

References

Apperly, I. A. (2011). *Mindreaders: The Cognitive Basis of "Theory of Mind."* Psychology Press.

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

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

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

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

Bardi, L., Desmet, C., & Brass, M. (2018). Spontaneous Theory of Mind is reduced for nonhuman-like agents as compared to human-like agents. *Psychological Research*, *0*, 1–10. https://doi.org/10.1007/s00426-018-1000-0

Bardi, L., Six, P., & Brass, M. (2017). Repetitive TMS of the temporo-parietal junction disrupts participant's expectations in a spontaneous Theory of Mind task. *Social Cognitive and Affective Neuroscience*, ??(September), ?? https://doi.org/https://doi.org/10.1093/scan/nsx109

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

Bull, R., Phillips, L. H., & Conway, C. A. (2008). The role of control functions in mentalizing: Dual-task studies of Theory of Mind and executive function. *Cognition*, *107*(2), 663–672. https://doi.org/10.1016/j.cognition.2007.07.015

Butterfill, S. A., & Apperly, I. A. (2013). How to construct a minimal theory of mind. *Mind & Language*, *28*(5), 606–637.

Cole, G. G., Atkinson, M., Le, A. T. D., & Smith, D. T. (2016). Do humans spontaneously take the perspective of others? *Acta Psychologica*, *164*, 165–168. https://doi.org/10.1016/j.actpsy.2016.01.007

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

Davidson, D. (1980). Towards a unified theory of meaning and action. *Grazer Philosophische Studien*, *11*, 1–12. https://doi.org/doi.org/10.5840/gps19801120

Davidson, D. (1990). *The Structure and Content of Truth*. *87*(6), 279–328.

Deschrijver, E., Bardi, L., Wiersema, J. R., & Brass, M. (2016). Behavioral measures of implicit theory of mind in adults with high functioning autism. *Cognitive Neuroscience*, *7*(1–4), 192–202. https://doi.org/10.1080/17588928.2015.1085375

Doherty, M., & Perner, J. (1998). Metalinguistic awareness and theory of mind: Just two words for the same thing? *Cognitive Development*, *305*(1998), 279–305.

Edwards, K., & Low, J. (2017). Reaction time profiles of adults' action prediction reveal two mindreading systems. *Cognition*, *160*, 1–16. https://doi.org/http://dx.doi.org/10.1016/j.cognition.2016.12.004

El Kaddouri, R., Bardi, L., De Bremaeker, D., Brass, M., & Wiersema, J. R. (2019). Measuring spontaneous mentalizing with a ball detection task: Putting the attention-check hypothesis by Phillips and colleagues (2015) to the test. *Psychological Research*, 1–9. https://doi.org/10.1007/s00426-019-01181-7

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175–191. https://doi.org/10.3758/BF03193146

Fizke, E., Butterfill, S., van de Loo, L., Reindl, E., & Rakoczy, H. (2017). Are there signature limits in early theory of mind? *Journal of Experimental Child Psychology*, *162*, 209–224. https://doi.org/10.1016/j.jecp.2017.05.005

Flavell, J. H. (1978). The development of knowledge about visualperception. In C.B.Keasey (Ed.), *Nebraska Symposium on Motivation (Vol. 25)*. Lincoln: University of Nebraska Press.

Flavell, J. H. (1992). Perspectives on perspective taking. In H. Beilin & P. B. Pufall (Eds.), *Piaget's theory: Prospects and possibilities.* (pp. 107–139). Hillsdale, NJ: Erlbaum.

Flavell, J. H., Everett, B. A., Croft, K., & Flavell, E. R. (1981). Young children's knowledge about visual perception: Further evidence for the Level 1-Level 2 distinction. *Developmental Psychology*, *17*(1), 99–103. https://doi.org/10.1037/0012-1649.17.1.99

Furlanetto, T., Becchio, C., Samson, D., & Apperly, I. A. (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. https://doi.org/http://dx.doi.org/10.1037/xhp0000138

Hamilton, A. F., Brindley, R., & Frith, U. (2009). Visual perspective taking impairment in children with autistic spectrum disorder. *Cognition*, *113*(1), 37–44. https://doi.org/10.1016/j.cognition.2009.07.007

Hamilton, A. F., & Ramsey, R. (2013). How are the actions of triangles and people processed in the human brain? In D. Rutherford, M. & M. Kuhlmeir, V. (Eds.), *Social Perception: Detection and interpretation of animacy, agency, and intention* (pp. 231–257). https://doi.org/10.7551/mitpress/9780262019279.003.0010

Hertwig, R., & Todd, P. M. (2003). More is not always better: The benefits of cognitive limits. In D. Hardman & L. Macchi (Eds.), *Thinking: Psychological Perspectives on Reasoning, Judgment and Decision Making* (pp. 213–231). John Wiley & Sons, Ltd.

Heyes, C. (2014). Submentalizing: I Am Not Really Reading Your Mind. *Perspectives on Psychological Science*, *9*, 131–143. https://doi.org/10.1177/1745691613518076

Horowitz, T. S., & Cohen, M. A. (2010). Direction information in multiple object tracking is limited by a grade resource. *Attention, Perception & Psychophysics*, *72*(7), 1765–17. https://doi.org/10.3758/APP.72.7.1765

Karg, K., Schmelz, M., Call, J., & Tomasello, M. (2016). Differing views: Can chimpanzees do Level 2 perspective-taking? *Animal Cognition*, *19*(3), 555–564. https://doi.org/10.1007/s10071-016-0956-7

Kessler, K., & Rutherford, H. (2010). The two forms of visuo-spatial perspective taking are differently embodied and subserve different spatial prepositions. *Frontiers in Psychology*, *1*, 1–12. https://doi.org/10.3389/fpsyg.2010.00213

Keysar, B., Barr, D. J., Balin, J. A., & Brauner, J. S. (2000). Taking perspective in conversation: the role of mutual knowledge in comprehension. *Psychological Science : A Journal of the American Psychological Society*, *11*(1), 32–38. https://doi.org/10.1111/1467-9280.00211

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

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. https://doi.org/10.1126/science.1190792

Low, J., Drummond, W., Walmsley, A., & Wang, B. (2014). Representing how rabbits quack and competitors act: Limits on preschoolers' efficient ability to track perspective. *Child Development*, *85*(4), 1519–1534. https://doi.org/10.1111/cdev.12224

Low, J., & Watts, J. (2013). Attributing false beliefs about object identity reveals a signature blind spot in humans' efficient mind-reading system. *Psychological Science*, *24*(3), 305–311. https://doi.org/10.1177/0956797612451469

Masangkay, Z. S., McCluskey, K. A., McIntyre, C. W., Sims-Knight, J., Vaughn, B. E., & Flavell, J. H. (1974). The early development of inferences about the visual percepts of others. *Child Development*, *45*(2), 357–366. https://doi.org/10.1111/1467-8624.ep12154629

McKinnon, M. C., & Moscovitch, M. (2007). Domain-general contributions to social reasoning: Theory of mind and deontic reasoning re-explored. *Cognition*, *102*(2), 179–218. https://doi.org/10.1016/j.cognition.2005.12.011

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

Moll, H., & Meltzoff, A. N. (2011). How does it look? Level 2 perspective taking at 36 months of age. *Child Development*, *82*(2), 661–673. https://doi.org/10.1111/j.1467-8624.2010.01571.x

Moll, H., Meltzoff, A. N., Merzsch, K., & Tomasello, M. (2012). Taking Versus Confronting Visual Perspectives in Preschool Children. *Developmental Psychology*, *49*(4), 646–654.

Mozuraitis, M., Chambers, C. G., & Daneman, M. (2015). Privileged versus shared knowledge about object identity in real-time referential processing. *Cognition*, *142*, 148–165. https://doi.org/10.1016/j.cognition.2015.05.001

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*(10), 1–15. https://doi.org/10.1371/journal.pone.0164373

Nijhof, A. D., Brass, M., & Wiersema, J. R. (2017). Spontaneous mentalizing in neurotypicals scoring high versus low on symptomatology of autism spectrum disorder. *Psychiatry Research*, *258*(March), 15–20. https://doi.org/10.1016/j.psychres.2017.09.060

Oktay-Gür, N., Schulz, A., & Rakoczy, H. (2018). Children exhibit different performance patterns in explicit and implicit theory of mind tasks. *Cognition*, *173*(January), 60–74. https://doi.org/10.1016/j.cognition.2018.01.001

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The Adaptive Decision Maker*. New York: Cambridge University Press.

Perner, J., Stummer, S., Sprung, M., & Doherty, M. (2002). Theory of mind finds its Piagetian perspective: why alternative naming comes with understanding belief. *Cognitive Development*, *103*, 1–22. https://doi.org/http://dx.doi.org/10.2307/1129379

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*(9), 1353–1367. https://doi.org/10.1177/0956797614558717

Qureshi, A. W., Apperly, I. A., & Samson, D. (2010). Executive function is necessary for perspective selection, not Level-1 visual perspective calculation: evidence from a dual-task study of adults. *Cognition*, *117*(2), 230–236. https://doi.org/10.1016/j.cognition.2010.08.003

Rowe, A. D., Bullock, P. R., Polkey, C. E., & Morris, R. G. (2001). "Theory of mind" impairments and their relationship to executive functioning following frontal lobe excisions. *Brain*, *124*(3), 600-616. https://doi.org/10.1093/brain/124.3.600

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. *J Exp Psychol Hum Percept Perform*, *36*(5), 1255–1266. https://doi.org/10.1037/a0018729

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), 929–937. https://doi.org/10.1037/a0035175

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

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

Schneider, D., Slaughter, V. P., & Dux, P. E. (2017). Current evidence for automatic Theory of Mind processing in adults. *Cognition*, *162*, 27–31. https://doi.org/10.1016/j.cognition.2017.01.018

Surtees, A. D. R., Apperly, I. A., & Samson, D. (2013). The use of embodied self-rotation for visual and spatial perspective-taking. *Frontiers in Human Neuroscience*, *7*(November), 698. https://doi.org/10.3389/fnhum.2013.00698

Surtees, A. D. R., Butterfill, S. A., & Apperly, I. A. (2012). Direct and indirect measures of Level-2 perspective-taking in children and adults. *British Journal of Developmental Psychology*, *30*, 75–86. https://doi.org/10.1111/j.2044-835X.2011.02063.x

Surtees, A. D. R., Samson, D., & Apperly, I. A. (2016). Unintentional perspective-taking calculates whether something is seen, but not how it is seen. *Cognition*, *148*, 97–105. https://doi.org/10.1016/j.cognition.2015.12.010

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

Zeman, S. (2017). Confronting perspectives: Modeling perspectival complexity in language and cognition. *Glossa: A Journal of General Linguistics*, *2*(1), 6. https://doi.org/10.5334/gjgl.213