Dissociating visual perspective taking and belief reasoning using a novel integrated paradigm: A preregistered online study

Rachel Green[1], Daniel Joel Shaw[1] and Klaus Kessler[12]

[1] Institute of Health and Neurodevelopment, School of Psychology, Aston University, UK

[2] School of Psychology, University College Dublin, Ireland

Running Title: Dissociating Visual Perspective Taking and Belief Reasoning

Corresponding Author:

Prof Klaus Kessler

Email: klaus.kessler@ucd.ie

Phone: +353 (0)1 716 8189

Address: University College Dublin

School of Psychology

G216 Newman Building

University College Dublin,

Belfield, Dublin 4, Ireland

Abstract

There is considerable conceptual overlap between Level-2 Visual Perspective Taking (VPT-2) and Belief Reasoning; both cognitive processes require us to represent another's viewpoint and experience of reality while inhibiting our own egocentric representations. This study investigated if these facets of mentalising are distinct from one another in the general adult population. To do so, we developed a novel "Seeing-Believing Task" with which to compare VPT-2 and true belief (TB) reasoning directly – one in which both judgement types refer to the same state of reality, requiring identical responses, and where self and other perspectives can be dissociated. Across three preregistered online experiments, this task demonstrated consistent differences between these two cognitive processes; specifically, TB judgements were associated with slower response times than VPT-2. This suggests that VPT-2 and TB reasoning are at least partly distinct psychological processes. Further, the greater cognitive effort required for TB reasoning is unlikely to be explained by differences in mnemonic processing. We suggest, therefore, that VPT-2 and TB reasoning differ in the complexity of social processing involved and we discuss theoretical implications based on minimal vs. full-blown Theory of Mind. Future research must attempt to test these conjectures.

Keywords: Theory of Mind, Mentalizing, Belief Reasoning, Level-2 Visual Perspective Taking, Seeing-Believing Task

Statement of Relevance

Social interaction and communication are fundamental human abilities that are supported by various cognitive processes. One crucial process is Perspective Taking (VPT-2), through which we imagine how another person might experience the visual world from their distinct and unique perspective. Another key social process is Belief Reasoning (BR), which refers to the understanding that others' thoughts and beliefs about the world may differ from ours and may even be incorrect. VPT-2 and BR are conceptually similar, both requiring us to represent our own as well as another's viewpoint in mind simultaneously, but their relationship remains unclear. Our findings with a novel "Seeing-Believing Task" reveal that BR, specifically True Beliefs (TB), and VPT-2 are distinct, with TB the more cognitively demanding of the two. Our research provides a novel paradigm and an empirical basis for future research into social abilities and their typical and atypical development.

1. Introduction

Mentalizing refers to the human capacity to represent the mental states of others. Representing another's intentions and beliefs is typically subsumed under the psychological construct Theory of Mind (ToM: Baron-Cohen, Wheelwright, Skinner, Martin, & Clubley, 2001), while inferring their visuospatial experience of the world is referred to commonly as Visuospatial Perspective Taking (VPT; Newcombe, 1989). Of particular importance to the present study is level-2 VPT (Flavell, Everett, Croft, & Flavell, 1981), which refers to our ability to understand *how* the world appears differently from another's viewpoint. Historically, these two aspects of mentalizing have been studied independently (e.g., Schurz, Aichhorn, Martin, & Perner, 2013). There is strong conceptual overlap between them, however; representing others' mental states, such as their beliefs and/or perspectives, requires us to disregard our own egocentric representations. Indeed, an increasing body of research indicates a strong correlational relationship between VPT-2 and ToM (e.g., Hamilton, Brindley, & Frith, 2009; Schurz et al., 2013). Further, both processes seem to develop around 4-5 years of age (Flavell et al., 1981; Wellman, Cross, & Watson, 2001), and various neuroimaging meta-analyses suggest commonalities in their underpinning brain networks (for reviews see Gunia, Moraresku, & Vlček, 2021; Schurz et al., 2013; Van Overwalle, 2009). Here, we investigated whether ToM and VPT-2 involve common or distinct cognitive processes.

1.1. Belief reasoning

False-belief (FB) reasoning paradigms are often employed to measure ToM, the most popular of which is the "Sally-Anne" task (Baron-Cohen, Leslie, & Frith, 1985; originally, Wimmer & Perner, 1983). These tasks require participants, often children, to distinguish between their own privileged representation of reality (they know the true hidden contents of a box) and a protagonist's false representation (Sally is unaware that Anne swapped the contents of the box in Sally's absence). To represent Sally's FB, participants must be able to ignore their own privileged representation. Solving this FB task is therefore taken as a direct reflection of an individual having a developed ToM – i.e., an understanding that others represent reality differently to oneself, sometimes even incorrectly.

While such tasks are valid, albeit conservative measures of ToM ability, they place all emphasis on the representation of FBs – they offer no insight into the representation of true beliefs (TBs) or their potential contribution to ToM (but see Fabricius, Boyer, Weimer, & Carroll, 2010; Huemer et al 2023). TBs are represented faster and more accurately compared with FBs

(Apperly, Back, Samson, & France, 2008; Apperly, Warren, Andrews, Grant, & Todd, 2011). This is due largely to their congruence with reality, but also often due to their congruence with the participant's representation of reality and their visual perspective of the world – a common confound in typical "false location belief reasoning tasks" (e.g., Rahman et al., 2021). Further, belief reasoning tasks often do not establish a clearly defined and unique visual perspective for the protagonist that could be identical or different in content to their belief (e.g. German & Hehman 2006; also Apperly et al., 2008; Apperly et al., 2011). As such, previous paradigms offer little discriminatory value when assessing TB and VPT-2 processing on the basis of behavioural outcomes. Our novel paradigm accounts for these limitations.

1.2. Level-2 visuospatial perspective taking (VPT-2)

Tasks employed to investigate VPT-2 typically require participants to judge whether a target object is located to the left or right of another's perspective (e.g., Michelon & Zacks, 2006), whether the other person sees the front, side or back of a target object (e.g., Hamilton et al., 2009), or if the other perceives a "6" or a "9" from their viewpoint (e.g., Surtees, Apperly, & Samson, 2013). The angular disparity between the participant's egocentric perspective and the other's (target) perspective plays a key role in this process, with response times (RTs) and errors increasing with larger disparities (see Michelon & Zacks, 2006; Kessler & Rutherford, 2010; Kessler & Thomson, 2010; Surtees et al., 2013). This appears to reflect a process of "embodiment"; Kessler and Thomson (2010; also Kessler & Rutherford, 2010; Surtees et al., 2013) demonstrated that typically developed adults perform VPT-2 tasks by mentally rotating themselves into the other's orientation and embodying their viewpoint.

1.3. Levels of complexity

Although embodying another's perspective enables insights into how the world is experienced from that particular viewpoint, it might not necessarily require a complex representation of the other's mind. For example, Kessler and Thomson (2010) demonstrated that the same embodiment process of mental self-rotation was also employed when participants imagined themselves in a different viewpoint that was not defined by another person but indicated by an empty chair (see also Muto et al, 2018); thus, the presence of another's mind was not strictly required for VPT-2. However, the same authors also reported that another person's presence facilitated VPT-2, suggesting that VPT-2 might bridge the gap between spatial cognition and mentalizing (Hamilton, Kessler, & Creem-Regehr, 2014). Some researchers have therefore argued that VPT-2 could be a developmental or even an evolutionary stepping-stone towards more complex forms of mentalizing, such as ToM (Gunia et al., 2021; Kessler, Cao, O'Shea, & Wang, 2014; Michelon & Zacks, 2006; Howlin, Baron-Cohen & Hadwin, 1999).

Such a distinction between perspective taking and belief reasoning vis-à-vis the complexity of representing another's mind is not a new concept. For instance, Howlin, Baron-Cohen and Hadwin, (1999) proposed five developmental levels of increasing complexity in representing another's mind, whereby perspective tracking (level 1) is distinguished from perspective taking (level 2), and representations of other's true (level 4) and false beliefs (level 5) occur at more advanced stages (but see Fabricius, Boyer, Weimer, & Carroll, 2010). In a similar vein, Apperly and Butterfill's (2009; also Butterfill & Apperly, 2013) distinguish between minimal and full-blown ToM. The former is proposed to rely on relational representations comprising objects and their locations relative to agents and what they can or cannot see. In contrast, full-blown ToM involves propositional representations, allowing inferences about the other's mental states and future actions. It has also been proposed that VPT-2 mechanisms develop from basic visuospatial cognitive processes, and then underpin the subsequent development of higher-level ToM abilities (Kessler & Rutherford, 2010; Kessler & Thomson, 2010). Such hierarchies of representational complexity would predict a distinction between VPT-2 and TB reasoning, but this has not yet been investigated directly. Establishing this in the first instance would enable further explorations into *how* the two aspects of mentalising differ in their processing and representational content.

1.4. The current study

In the current pre-registered study (Project registered on OSF: https://osf.io/57wa9/), we devised a novel experimental paradigm that allowed us to compare TB reasoning and VPT-2 directly and in a way that allowed us to dissociate a participant's representation of another's perspective and belief from their own egocentric perspective and representation of reality: the Seeing-Believing Task. In this task, participants are presented with scenarios in which a character's TB was always congruent with their established visual perspective. This is distinct from previous research allowing us to compare TB and VPT-2 judgements directly within the same paradigm, in order to investigate if differences are observed even when the two judgement types are identical in content (both representing the same state of reality) and require identical responses. We also included a manipulation of angular disparity between the viewpoint of the self and the other (0° vs 90°), with which we could distinguish the participant's own egocentric perspective and representation of reality from their representations of another's perspective and their beliefs. At 0° disparity self and other perspectives are identical, while at 90° they are distinct. Therefore, a 90º angular disparity required participants to mentally adopt the other's perspective rather than directly infer their perspective or belief state from their own egocentric representation. To our knowledge, this manipulation is different from any ToM paradigm to date because it varies systematically the disparity between perspectives, resulting in meaningful differences in visuo-spatial representations between self and other perspectives. This allowed us to compare the two other-related judgments of interest, TB and VPT-2, directly and without egocentric confound.

Two alternative hypotheses were contrasted: First, in light of strong correlations between independent performance indices of ToM and VPT-2 (e.g., Hamilton et al., 2009) and a degree of overlap in their associated neural substrates (e.g., Gunia et al., 2021; Schurz et al., 2013), it might be expected that representing another's TB is identical to processing their visual perspective. If so, TB and VPT-2 judgements should take the same time and be equally accurate. Alternatively, should VPT-2 involve less complex social representations than TB reasoning, which is supported by evidence of VPT-2 mechanisms being employed in non-social contexts (for reviews see Hamilton et al., 2014; Kessler & Thomson, 2010; Muto et al., 1018) and sub-served by partially segregated neural substrates (Gunia et al., 2021; Schurz, Radua, Aichhorn, Richlan, & Perner, 2014), VPT-2 judgements should be faster and more accurate than TB judgments.

In a series of three pre-registered experiments, we utilised the novel Seeing-Believing Task to contrast level-2 perspective judgements and (true) belief judgements as our primary research question. In addition, we sought to replicate previous findings regarding the greater processing effort for false (FB) vs true beliefs (TB), and the greater processing effort required for higher compared to lower angular disparities between self and other. Table 1 presents our main pre-registered hypotheses relating to each experiment.

1.4.1. Experiment 1

The first experiment (pre-registration at https://osf.io/efp87) provided an initial evaluation of our basic paradigm and the first test of our main research question. TBs have been shown to be represented faster and with fewer errors compared with FBs due to their congruence with reality (Apperly et al., 2008; Apperly et al., 2011). However, typical "false location belief reasoning tasks" employed to investigate this, often do not distinguish clearly between the other's true belief, their visual perspective, and the participants' own representation of reality and perspective (e.g., Rahman et al., 2021; for exceptions see Edwards & Low, 2019). Although our representations of another's TB and perspective might be identical in content (both representing reality in the current paradigm), they can be distinguished logically (e.g., closing my eyes will affect my visual perspective but not my belief about reality) and are proposed to differ regarding their conceptual complexity (e.g., Howlin et al., 1999). If the latter is true, the two processes will differ in processing time and accuracy.

First, we predicted that TB and VPT-2 judgements would differ in their underlying socio-cognitive processes (Hypothesis 1), leading to differences in response times (RT) and errors (average number of errors). Note that RTs were expected to provide cleaner data due to low overall number of errors. We expected TB to result in higher processing costs compared to VPT-2 (greater RTs and errors, with RTs the main focus of analysis) due to potentially more complex socio-cognitive processes involved in the former. Importantly, the angular disparity manipulation in our paradigm (0° *vs.* 90° disparity between the participant's and the character's perspective) allowed us to dissociate the character's perspective and belief from the participant's own egocentric view. At 0° disparity, the other's perspective and true belief are identical with the participant's current perspective and knowledge and would even require identical responses (as tested in Experiment 2). Thus, other-related processing is likely to be influenced or even dominated by egocentric processing, potentially masking differences between VPT2 and TB judgements. However, at 90° disparity, the responses between VPT-2 and TB judgments are still identical but now differ from the egocentric view; for instance, if the participant sees the front of a rabbit, the character sees its side. This allows for an unhindered comparison between VPT-2 and TB at 90°. We therefore anticipated that the difference between TB and VPT-2 might be revealed more reliably at 90° angular disparity, and thus predicted an interaction between angular disparity and judgement type in addition or instead of a main effect of judgement type (Table 1). We also expected to replicate previously reported effects: we predicted greater RTs and errors for VPT-2 judgements at 90° relative to 0° angular disparity (Hypothesis 2), and faster and more accurate processing for TB compared with FB trials (Hypothesis 3).

1.4.2. *Experiment 2*

It has been suggested that to truly measure ToM processes, paradigms must require the participant to distinguish explicitly between themselves and a protagonist (Quesque & Rossetti, 2020). Such scholars might therefore argue that in Experiment 1, the participant is only required to process the character's perspective, so they might merge the character's (Other) with their own (Self) perspective. While we believe this to be unlikely given our manipulation of angular disparity, whereby the egocentric perspective is dissociated from the character's perspective at 90° disparity, we included explicit egocentric Self judgements in a second experiment (pre-registration at https://osf.io/v87hp) to demonstrate that any effects in Experiment 1 cannot be explained by an absence of Self-related judgements.

Further, it can be argued that TB trials might not only differ from VPT-2 trials in the complexity of social representations. While a question about the other's perspective clearly refers to the current state of reality, a question about the other's expectations includes further variables. These might be social in nature, such as having to represent another's expectation/belief in addition to their visual experience of reality, or they might be unrelated to social processing. For instance, a question about the other's expectations might trigger a check back in memory to the initially presented scene, even if the state of reality has not changed on a particular trial (i.e., TB trials). Following this rationale, higher processing costs (greater RTs and errors) in TB compared to VPT-2 trials would not allow us to disentangle if this increased cognitive effort reflects social or memory-related cognitive processes. To examine any mnemonic aspects of processing in our paradigm (Hypothesis 4), the second experiment included an explicit memory judgement along with "current" Self-judgements; in addition to asking participants about their current perspective ("What will you actually see?"), we also asked them about their memory of reality (the stimulus presented at the start of the trial; "What did you initially see?"). Note that experimental conditions were only distinguishable at the very end of each trial when the required judgement was presented as a question. It is therefore safe to assume that participants encoded events in the same way across all trials up to the point where distinct judgements were required. This allowed us to understand how participants encode the events in any given trial, and if the cognitive costs of a memory check to the initial state of reality (even if unchanged) is a likely explanation for any differences observed between VPT-2 and TB judgements in Experiments 1 and 2 (Hypothesis 5). Adding this second "past" Self-related question also allowed us to balance the overall number of Other- (Perspective *vs.* Belief) and Self-judgements trials (Self-current *vs.* Self-past perspective), avoiding any bias due to asymmetric numbers between judgement types.

1.4.3. *Experiment 3*

In a third experiment (pre-registration at https://osf.io/cbjfa/), we assessed how the results from Experiments 1 and 2 compared to those of previous studies that have used a different FB manipulation. Frequently, the belief of a protagonist is manipulated by having them witness (TB) or be absent (FB) for a change in an object's location (e.g. "Sally-Anne" task). This differs from Experiments 1 and 2, in which the target object was either swapped for another object (FB) or not (TB; see Methods) but always in the protagonist's absence (similar to Back & Apperly, 2010). This procedural difference was chosen on purpose to avoid a late change in visual experience – an object swap requires a cognitive update, thereby increasing cognitive demands/processing time that could potentially mask differences between TB and VPT-2 judgements. Hence, there was no swap in VPT-2 and TB trials in Experiments 1 and 2. In contrast, belief states (TB *vs.* FB) were induced in Experiment 3 by manipulating whether the character was present or not to witness the object swap. A swap occurred in all trials, equalizing cognitive updating demands across conditions, while keeping the overall number of trials manageable for an online study. Hypotheses for Experiment 3 paralleled those for Experiment 1 (Table 1).

Table 1. Summary of pre-registered hypotheses and associated statistical tests across Experiments 1-3.

No. (Experiment)	Predictions	Statistical test
Hypothesis 1 (Experiments 1-3)	TB ≠ VPT-2 especially at 90º angular disparity: If TB reasoning and VPT-2 are at least partially distinct processes there will be a difference in RTs and/or errors, especially at 90º, where self and other perspectives are distinct.	ANOVA for No-swap trials only: Judgement Type main effect and/or Angular Disparity x Judgement Type interaction. *t*-test TB_90º vs VPT-2_90º.
Hypothesis 2 (Experiments 1-3)	90º > 0º Judgements about another made at 90º angular disparity will be associated with greater RTs and/or errors than judgements made at 0º (replication of previous findings).	H2A: Angular Disparity main effect ANOVA. H2B: Angular Disparity x Judgement Type interaction.
Hypothesis 3 (Experiments 1-3)	FB > TB: FB judgements will be associated with greater RTs and/or errors than TB judgements (replication of previous findings)	Main effect Belief State (FB, TB) ANOVA.
Hypothesis 4 (Experiment 2)	Self_PastPerspective > Self_CurrentPerspective: If the task requires a check back in memory to the originally presented object, past self-perspective judgements should be slower/more error-prone than current self-perspective judgements.	*t*-test Self_Past vs. Self_Current
Hypothesis 5 (if H1 for Hypotheses 1 and 4 are correct) (Experiment 2)	(TB_90º – VPT-2_90º) > (Self_PastPerspective - Self_CurrentPerspective): If there is a "memory effect" in self-perspective judgments (Hypothesis 4), does it explain (is it the same size as) the difference between TB and VPT-2 at 90º (Hypothesis 1)?	*t*-test (TB_90º - VPT-2_90º) vs. (Self_Past - Self_Current) for no-swap trials.

Note. TB = true belief; FB = false belief; VPT-2 = Level-2 Visuospatial Perspective Taking; ToM = Theory of Mind; RT = response times; H_1 and H_0 refer to the alternative and null hypotheses, respectively. For any rejected H_1 Bayesian tests were used to assess support for H_0. Not all Hypothesis numbers in this table map onto those used in the preregistration. Additional Hypotheses 6 and 7 are reported in Supplemental Materials.

2. Methods

2.1. Design

The current study utilised the Seeing-Believing Task in three pre-registered experiments, designed to assess whether TB reasoning can be dissociated from VPT-2. Hypotheses, study designs and planned analyses for all experiments are available on the Open Science Framework (OSF; https://osf.io/57wa9/), and the computerised tasks are publicly available on Pavlovia (https://gitlab.pavlovia.org/RachelGreen). Each experiment employed a three-factor design (see Table 2); factors included Belief State, Angular Disparity, and Judgement Type.

Table 2. Summary of experimental designs and factor levels.

Experiment	Object-Swap (Belief State)	Angular disparity	Judgement type
1 (2x2x2)	No-swap (TB), Swap (FB)	0º, ± 90º	Belief, Perspective
2 (2x2x4)	No-swap (TB), Swap (FB)	0º, ± 90º	Belief, Perspective, Self-Current, Self-Past
3 (2x2x2)	Witnessed Swap (TB), Unwitnessed Swap (FB)	0º, ± 90º	Belief, Perspective

2.2. Participants

Details of the final sample for each experiment are presented in Table 3. Sample size calculations were performed using G*Power (Faul, Erdfelder, Lang, & Buchner, 2007); calculations were computed to achieve a small effect size (f = 0.15) with 95% power and resulted in sample sizes of 48 (planned recruitment of 50) for Experiments 1 and 3, and 64 (planned recruitment of 70) for Experiment 2. Experiments 1 and 3 had a 2x2x2 design (8 variables across 3 factors), whereas Experiment 2 used a 2x2x4 design (16 variables across 3 factors). A correlation of r = .81 for repeated measures was used, based on unpublished pilot data. Participants were aged between 18 and 40 years, had normal or corrected-to-normal vision, were fluent in written and spoken English, had access to a desktop computer, and could sit comfortably at a computer for ~1 hour. Individual participant data was included in our final sample if they completed a minimum of 3 out of 12 trials per condition correctly (see Table 3). This threshold, which is greater than chance level (1/8), was chosen to produce more reliable condition means. Note that the final sample sizes for Experiments 1 and 2 were reduced to 42 and 62, respectively, because additional participants had to be rejected due to issues with the exit questionnaires, which was only noticed after data collection was complete. We decided to continue with the somewhat reduced samples instead of collecting further participants after pre-registration.

Table 3. Participant demographics

Experiment	Age (years)	N *(No. females)*	BAPQ	Participants excluded	
				Performance criterion	Exit questionnaires
1	23.71 *(3.95)*	42 *(21)*	114.55 *(22.98)*	12	8
2	24.15 *(5.69)*	62 *(28)*	113.58 *(20.80)*	38	8
3	26.57 *(6.27)*	51 *(29)*	114.52 *(22.46)*	14	0

Note. Values presents mean *(SD).* BAPQ = Broad Autism Phenotype Questionnaire.

2.3. Procedure

Due to the current global pandemic, which suspended participant-facing research in the UK, all three experiments were designed to be conducted online. All paradigms were programmed in PsychoPy Builder (PsychoPy v2020.1.2), using realistic stimuli created with Unreal Engine Version 4.17.2 (Epic Games, Carry, NC). The table and chairs were selected from 'Startpack' defaults, and the character, hare and locomotive from Adobe's Mixamo (Adobe, San Jose, CA). Participants were recruited via Prolific (https://www.prolific.co/), the experimental task was administered through Pavlovia (Peirce, 2007), and information sheets, consent forms and questionnaires were completed on Qualtrics (https://www.qualtrics.com/uk/). Participation in each experiment was compensated with £6.50, £10 and £6.50, respectively. All experimental procedures were approved by Aston University's Research Ethics Committee, and all participants provided informed consent before taking part. For procedural details and complete experiments see https://gitlab.pavlovia.org/RachelGreen/understanding-how-we-represent-the-beliefs-and-perspectives-of-others.

2.4. Materials

2.4.1. Experiment 1

As shown in Figure 1, each trial began with an image of a character sitting at a table with their back to the participant, such that the participant and the character had identical visual perspectives of the table. Placed on the table was either a hare or locomotive at one of four orientations (front, rear, left side, right side; see Figure 2 for all eight options). In a subsequent image, the hare or locomotive had been hidden by an overturned bucket and the character had left the room; thus, it was suggested that the character was unaware of whatever happened in their absence. At this point, the bucket was removed, and one of two events occurred: on 50% of trials, the object was swapped with another (e.g., the hare swapped with the locomotive) before the bucket was replaced; in the other 50%, the object remained and the bucket replaced over it. The character then returned to sit either in the same chair as before, in front of the participant (0^{o} angular disparity), or another chair to the left or right of their original position ($\pm 90°$ angular disparity). This 90^{o} angular disparity between the participant's egocentric (self) view and the character's new (other) orientation required participants to mentally adopt the character's perspective rather than directly infer their perspective or belief state from their own egocentric representation.

Finally, the participant was asked one of two questions in order to make a belief or a VPT-2 judgement: they were either asked "What will she expect to see?", requiring an inference of the character's belief, or "What will she actually see?", probing the character's perspective. The range of stimuli used resulted in an eight-choice response (2 objects x 4 orientations; see Figure 2). Since the large number of options may have affected participants' response times (RTs), they were asked to press the space bar as quickly as possible once they had visualised their answer before subsequently making a choice between the eight options. RTs were measured in relation to pressing the space bar. Note that due to the complexity of the task (Judgement Type, Angular Disparity, Object Swap) it was advantageous for participants to imagine their answer while the stimulus configuration remained on screen, rather than pressing the space bar as quickly as possible and then determining their answer. Immediately after pressing the space bar, the 8-choice response display was shown and participants could take their time to select which of the eight options reflected their envisioned answer. Their choice determined the accuracy of their response. The experiment consisted of 16 practice trials and 96 experimental trials split into 3 equal blocks of 32 trials, resulting in 12 trials per condition overall (2 x 2 x 2 design; Table 2). Each trial lasted approximately 12 seconds, depending on the duration of the participant's response. RT and error data were collected.

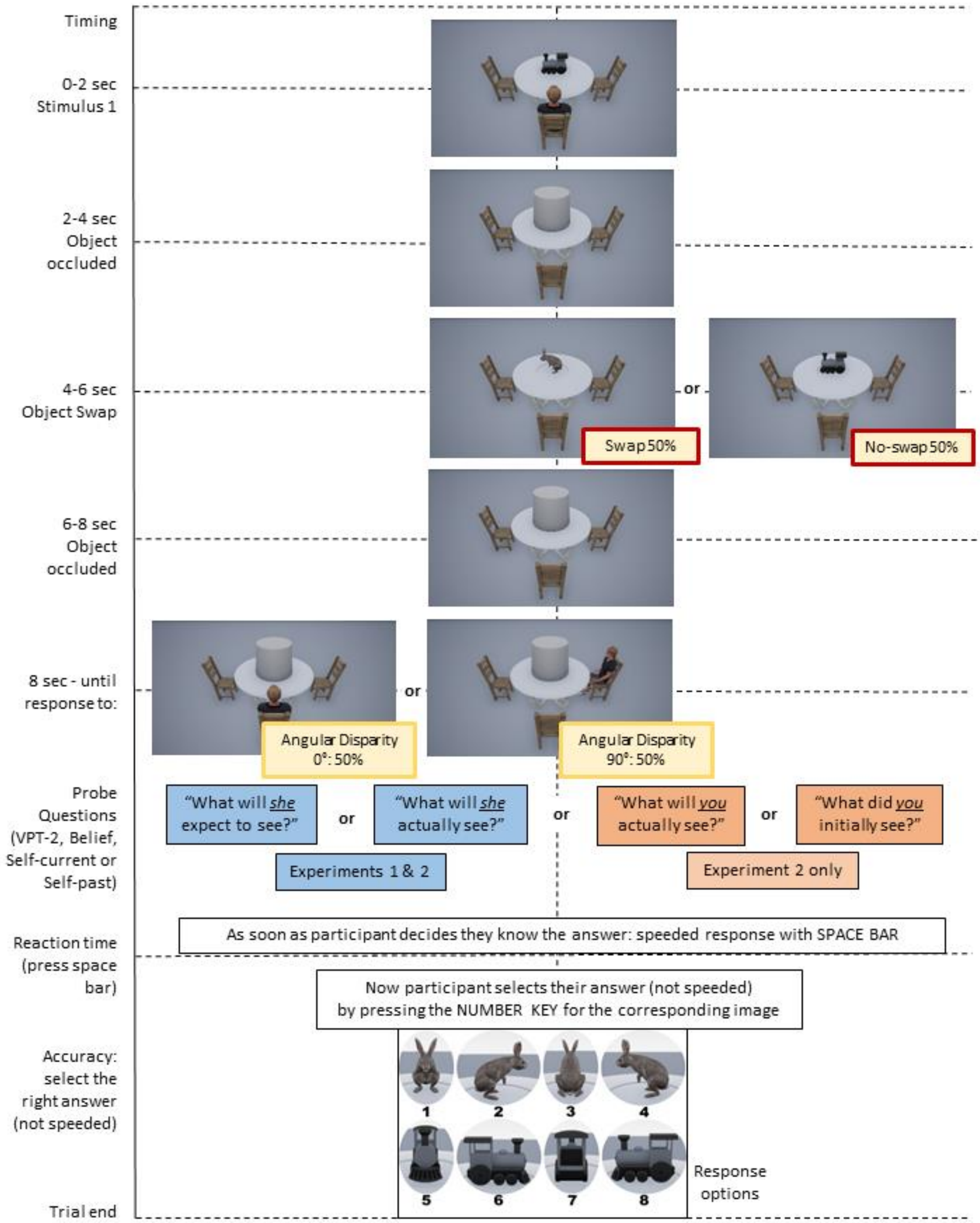

Figure 1. Trial schematic for Experiments 1 and 2.

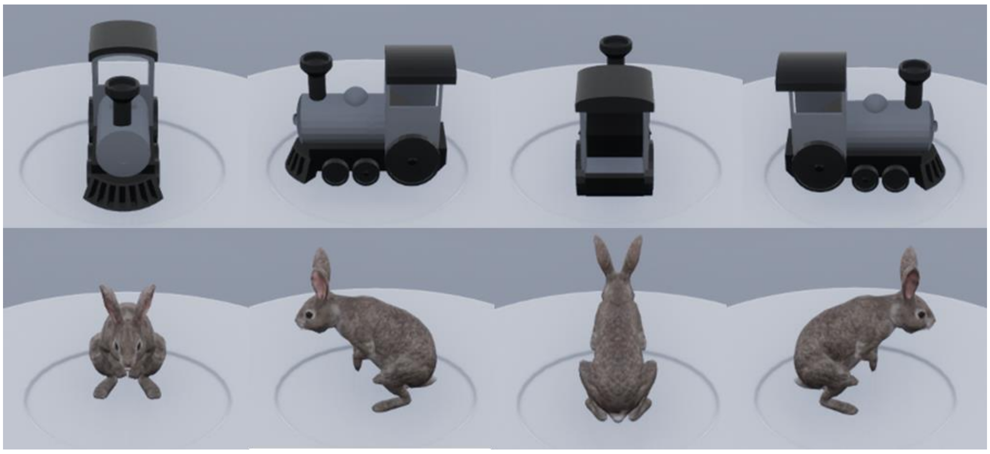

Figure 2. Stimulus options. The target object on the Table in Figure 1 could either be a toy locomotive (top row) or a hare (bottom row), presented in one four different orientations: front, left side, rear, and right side.

2.4.2. Experiment 2

The second experiment included two additional Self-related questions alongside the two Other-related questions from Experiment 1: “What will *you* actually see?” and “What did *you* initially see?” (see Probe Questions in Figure 1). The first question concerned the participant’s current egocentric perspective and representation of reality, whereas the second was designed to mirror the potential memory demands of the belief question, “What will she expect to see?”, whilst not requiring the participant to make a mental state judgement about the character (see Hypotheses 4 and 5). Note that experimental conditions were only distinguishable at the very end of each trial when a question indicated the required judgement.

This experiment consisted of 16 practice trials and 192 experimental trials split into 4 equal blocks of 48 trials, resulting in 12 trials per condition overall (2 x 2 x 2 design; see Table 1). Each trial lasted approximately 12 seconds, depending on the duration of the participant’s response. RT and error data were collected.

2.4.3. Experiment 3

In the third experiment, belief state (TB *vs.* FB) was manipulated by whether or not the character was present to witness the object swap, which occurred in all trials (see Figure 3). The experiment consisted of 16 practice trials and 96 experimental trials split into 3 equal blocks of 32 trials, resulting in 12 trials per condition overall (2 x 2 x 2 design; see Table 2).

Each trial lasted approximately 12 seconds, depending on the duration of the participant's response. RT and error data were collected.

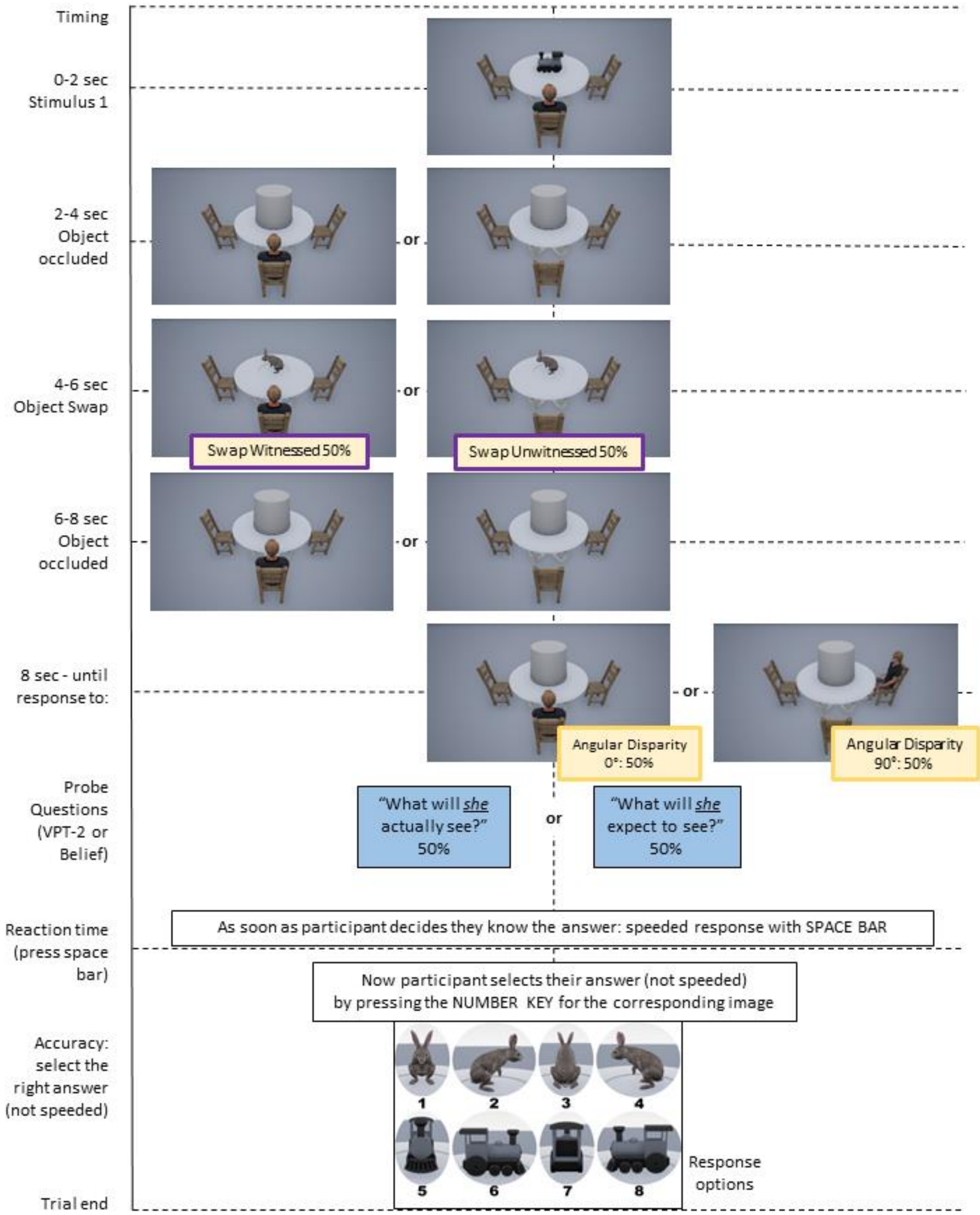

Figure 3. Trial schematic for Experiment 3.

2.4.2 Broad Autism Phenotype Questionnaire

The Broad Autism Phenotype Questionnaire (BAPQ; Hurley, Losh, Parlier, Reznick, & Piven, 2007) was used to assess participants’ self-reported autistic traits. This instrument consists of 36 items and is derived from gold-standard clinical interview methods (Landa et al., 1992; Piven, Palmer, Jacobi, Childress, & Arndt, 1997). Inclusion alongside our main design factors was deemed to be exploratory yet potentially informative for future research. However, when including BAPQ score as a continuous variable into full design ANOVAs for each experiment, we observed no statistically significant model terms that involved this variable (neither main effects nor interactions), and therefore omit this variable from reporting in Results.

2.5 Statistical Analysis

Custom MATLAB (r2019a) code was used to calculate condition averages and remove outliers, and JASP (Version 0.14.1) was used to perform frequentist and Bayesian statistics.

Table 4. Experiment 1.

Object Swap (Belief State)	No-swap (TB)				Swap (FB)			
Angular Disparity	0°		90°		0°		90°	
Judgement Type	VPT-2	ToM	VPT-2	ToM	VPT-2	ToM	VPT-2	ToM
RT	1.264(*.461*)	1.264(*.412*)	1.378(*.462*)	1.482(*.563*)	1.632(*.643*)	1.605(*.590*)	1.793(*.745*)	1.914(*.697*)
Errors	.190(*.552*)	.476(*.804*)	.452(*.772*)	.667(*1.074*)	.524(*1.042*)	1.429(*2.050*)	.833(*1.228*)	1.929(*2.224*)

Table 5. Experiment 2.

Object Swap (Belief State)	No-swap (TB)							
Angular Disparity	0°				90°			
Judgement Type	VPT-2	ToM	Self-current	Self-past	VPT-2	ToM	Self-current	Self-past
RT	1.407(*.428*)	1.408(*.431*)	1.472(*.447*)	1.328(*.393*)	1.586(*.500*)	1.703(*.563*)	1.572(*.452*)	1.411(*.444*)
Errors	.113(*.367*)	.194(*.474*)	.194(*.398*)	.177(*.426*)	.306(*.616*)	.403(*.778*)	.597(*1.166*)	.274(*.657*)

Object Swap (Belief State)	Swap (FB)							
Angular Disparity	0°				90°			
Judgement Type	VPT-2	ToM	Self-current	Self-past	VPT-2	ToM	Self-current	Self-past
RT	1.753(*.519*)	1.736(*.553*)	1.714(*.498*)	1.528(*.435*)	1.991(*.682*)	2.059(*.705*)	1.993(*.669*)	1.634(*.468*)
Errors	.468(.783)	.758(*1.112*)	.500(*.825*)	.645(*.943*)	.645(*.943*)	.952(*1.220*)	1.194(*1.377*)	.806(*1.403*)

Table 6. Experiment 3.

Object Swap (Belief State)	Witnessed Swap (TB)				Unwitnessed Swap (FB)			
Angular Disparity	0°		90°		0°		90°	
Judgement Type	VPT-2	ToM	VPT-2	ToM	VPT-2	ToM	VPT-2	ToM
RT	1.355(*.440*)	1.543(*.528*)	1.875(*.847*)	2.002(*.837*)	1.777(*.589*)	1.870(*.657*)	2.037(*.789*)	2.215(*.908*)
Errors	.569(*1.171*)	.216(*.541*)	.549(*1.064*)	.765(*1.258*)	.824(*1.292*)	.804(*1.249*)	1.000(*1.311*)	1.196(*1.649*)

Note. Values present means (*SD*). RTs = response times (seconds), Errors = Average number of errors.

3. Results

In the sections that follow, we present the results of analyses that evaluate each experiment in relation to the main pre-registered hypotheses specified in Table 1. For the sake of completeness, we report a full-design ANOVA for each experiment as an initial step before a detailed analysis that evaluates each of our main hypotheses. As per our pre-registration, in specific cases where the alternative hypothesis was abandoned in favour of the null, we performed Bayesian analysis to evaluate the strength of support for the null hypothesis. We present the results of statistics evaluating more minor preregistered hypotheses in Supplementary Material. Tables 4-6 report all means and standard deviations for each of the three experiments.

3.1. Experiment 1

We conducted two separate 2x2x2 ANOVAs for response times (RTs) and errors with the full experimental designs: Object Swap (Swap [FB], No-swap [TB]), Angular Disparity (0^0, 90^0), and Judgement Type (Belief, VPT-2). For RTs we observed main effects of Object Swap ($F(1,41) = 43.975$, $p < .001$, n_p^2 =.518), Angular Disparity ($F(1,41) = 40.455$, $p < .001$, n_p^2 =.497), and Judgement Type ($F(1,41) = 4.27$, $p = .045$, n_p^2 =.094). In addition, we observed a significant interaction between Angular Disparity and Judgement Type ($F(1,41) = 7.584$, p = .009, n_p^2 =.156). For errors we also observed main effects of Object Swap ($F(1,41) = 21.196$, $p < .001$, n_p^2 =.341), Angular Disparity ($F(1,41) = 9.72$, $p = .003$, n_p^2 =.192), and Judgement Type ($F(1,41) = 18.103$, $p < .001$, n_p^2 =.306). In addition we observed a significant interaction between Object Swap and Judgement Type ($F(1,41) = 11.583$, $p = .001$, n_p^2 =.220).

Hypothesis 1: The central aim of this study was to investigate if VPT-2 and TB reasoning can be dissociated. This 2x2 ANOVA, applied to RTs on No-swap trials only, revealed a significant main effect of Judgement Type ($F(1,41) = 4.684$, p =.036, n_p^2 = .103) and a significant Angular Disparity-by-Judgement Type interaction ($F(1,41) = 5.706$, p = .022, n_p^2 =.122). Planned comparisons between VPT-2 and TB judgements for each angular disparity revealed a significant difference at 90° ($t(41) = 3.068$, $p = .004$, $d = .473$) but not at 0° ($p = .985$). At 90° VPT-2 was significantly faster than TB (Figure 4). Analysis of errors in Experiment 1 revealed a significant main effect of Judgement Type ($F(1,41) = 5.484$, $p = .024$, n_p^2 =.118), with TB being more error-prone than VPT-2, but the interaction between Angular Disparity and

Judgement Type did not reach significance (p = .645). Significant main effects of Angular Disparity were also observed for RTs (F(1,41) = 23.257, p < .001, n_p^2 =.362) and errors (F(1,41) = 4.612, p = .038, n_p^2 =.101), but these feed into Hypothesis 2 (see below).

As illustrated in Figure 4 (left), RT results in Experiment 1 support the alternative hypothesis (H_1) for Hypothesis 1 by indicating a difference between TB reasoning and VPT-2, whereby TB is associated with greater RTs than VPT-2. Furthermore, this is evident at 90° angular disparity, as indicated by a significant interaction between judgement type and angular disparity. The error results in Experiment 1 further support the alternative hypothesis via a main effect but, no interaction. This must be interpreted with caution due to the overall low number of errors.

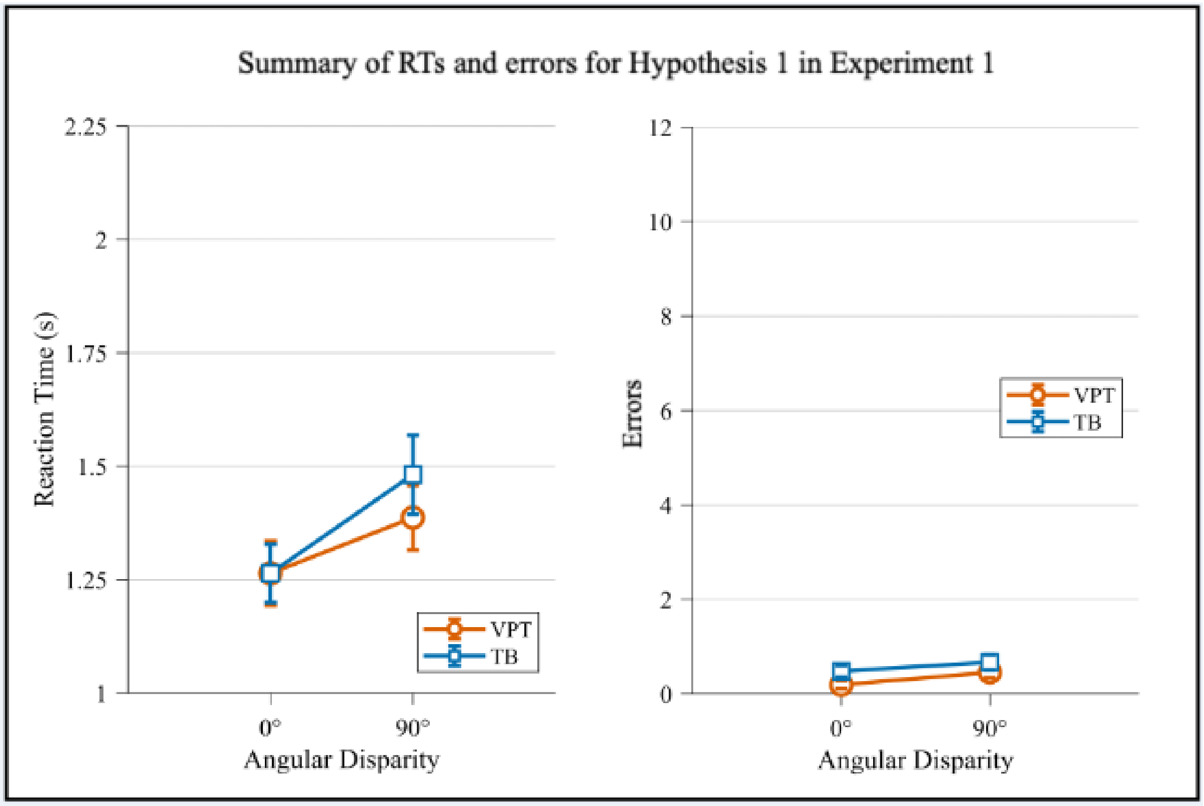

Figure 4. Results for Hypothesis 1 (Table 1) in Experiment 1. Response times (RTs, measured in seconds; correct responses only) are shown on the left and errors (average number of errors per condition and with a maximum of 12) are shown on the right. Error bars are standard error of mean.

Hypothesis 2: As a replication of previous findings, it was hypothesised that judgements about another's mental state (belief or perspective) made at 90º angular disparity would be associated with greater RTs and errors than judgements made at 0º angular disparity. Hypothesis 2A was the more stringent test of a main effect of Angular Disparity, while Hypothesis 2B included the possibility of an *ordinal* interaction between Angular Disparity and Judgement Type. As reported above, in the full-design ANOVA of RT data, this was confirmed by a significant main effect of Angular Disparity ($F(1,41) = 40.455$, $p < .001$, n_p^2 =.497) and a significant interaction between Angular Disparity and Judgement Type ($F(1,41) = 7.584$, $p = .009$, n_p^2 =.156). RTs are slower with increasing angular disparity (90° > 0°) for both judgement types, yet with a more pronounced increase for belief compared with VPT-2 judgements. The full-design ANOVA of error data also revealed a significant main effect of Angular Disparity ($F(1,41) = 9.72$, $p = .003$, n_p^2 =.192) but no significant interaction between Angular Disparity and Judgement Type ($p = .763$). Together, these results support Hypothesis 2 (Figure 5).

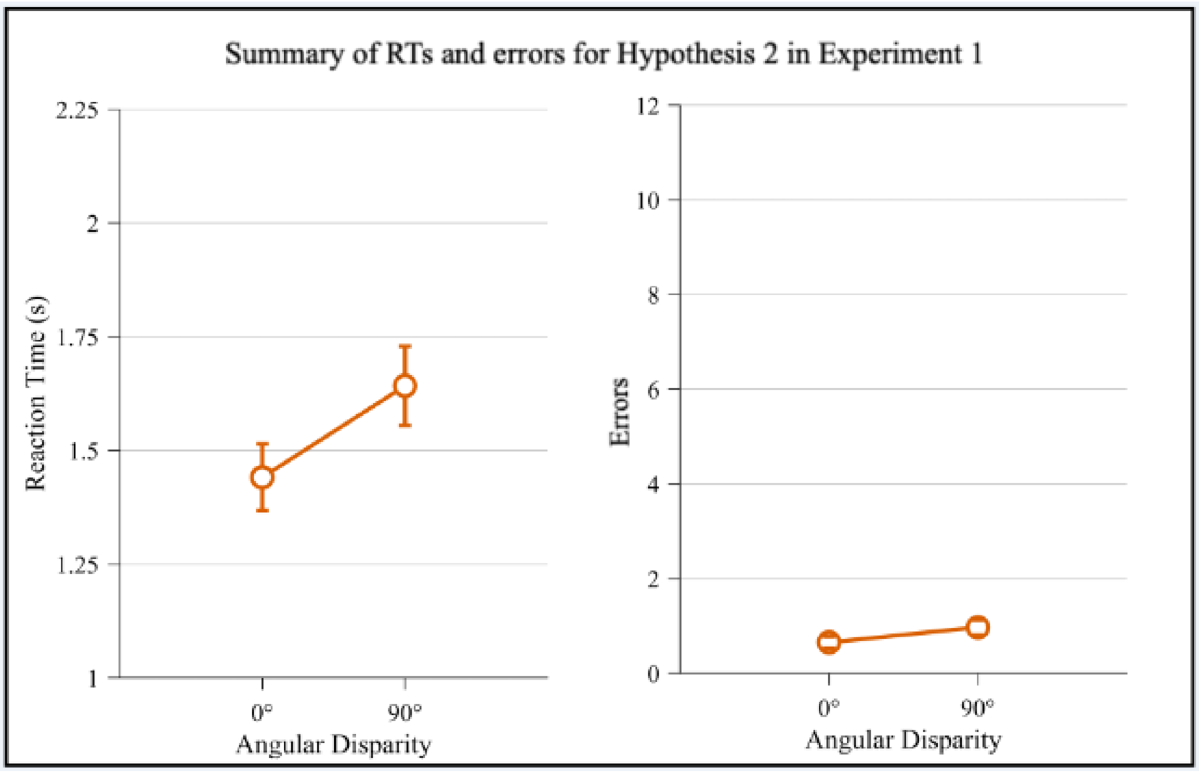

Figure 5. Results for Hypothesis 2 in Experiments 1, with RTs on the left and errors on the right. Error bars denote standard error of mean.

Hypothesis 3: It was hypothesized that FB judgements would be associated with greater RTs and errors than TB judgements. This was evaluated by comparing the average of all FB trials

(Swap) against the average of all TB trials (No-swap), collapsing across angular disparities. Note that perspective judgments were excluded from this particular analysis. Our novel paradigm also allowed us to assess whether any additional cost for FB is greater at 90° relative to 0° angular disparity between self- and other-perspective. It was therefore informative to analyse the interaction between Object Swap (Swap [FB], No-swap [TB]) and Angular Disparity (90°, 0°). Importantly, however, we did not pre-register Hypothesis 3 in this form. Two 2x2 ANOVAs revealed the expected main effects of Object Swap (RT: ($F(1,41) = 37.176$, $p < .001$, $n_p^2 = .476$); error: ($F(1,41) = 19.957$, $p < .001$, $n_p^2 = .327$)) and Angular Disparity (RT: ($F(1,41) = 36.193$, $p < .001$, $n_p^2 = .469$); error: ($F(1,41) = 5.663$, $p = .022$, $n_p^2 = .121$), but no significant interactions in RT ($p = .121$) or in error ($p = .317$). As hypothesized, FB judgements were made with significantly greater RTs and errors than TB judgements, replicating previous results (Figure 6). Extending prior findings, Angular Disparity did not significantly interact with FB costs in our novel paradigm.

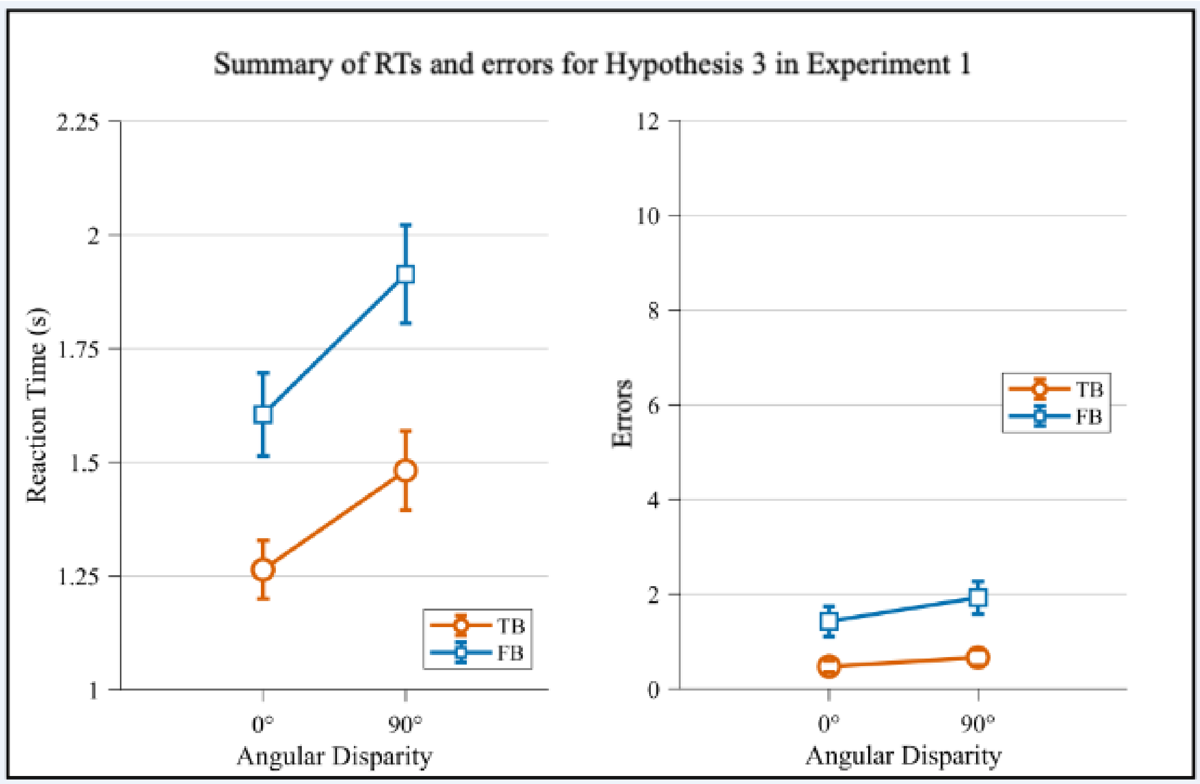

Figure 6. Results for Hypothesis 3 in Experiments 1, with RTs on the left and errors on the right. Error bars denote standard error of mean.

3.2.Experiment 2

The aim of the second experiment was to replicate and extend the results of Experiment 1. First, we included Self-judgements to demonstrate that the effects in the initial experiment cannot be explained by an absence of Self-judgements (Quesque & Rossetti, 2020). Second, in addition to asking participants about their current perspective, we also asked them about their memory of the object presented at the start of the trial (Hypotheses 4 and 5). This allowed us to understand whether the increased processing demands observed in Experiment 1 were likely due to mnemonic processing (checking back to the initial object might be more costly for TB) or more likely due to increased complexity of social processing (representing the other's TB might require a more complex representation of their mind). Since experimental conditions were only distinguishable at the very end of each trial it is safe to assume that participants encoded the events in a trial in the same way across conditions, until the question requiring a specific judgement was presented. Thus, any memory effects observed for self-judgements (current vs. past) would be a good indicator for memory effects that may impact other-judgements.

We conducted two separate 2x2x4 ANOVAs for RTs and errors with the full experimental design (Table 2): Object Swap (Swap [FB], No-swap [TB]), Angular Disparity (0^0, 90^0), and Judgement Type (Other-Belief, Other-Perspective, Self-Current, Self-Past). For RTs we observed main effects of Object Swap (F(1,61) = 110.918, p < .001, n_p^2 =.645), Angular Disparity (F(1,61) = 50.511, p < .001, n_p^2 =.453), and Judgement Type (F(1,61) = 4.539, p = .037, n_p^2 =.069). In addition, we observed a significant interaction between Angular Disparity and Judgement Type (F(1,61) = 8.878, p = .004, n_p^2 =.127). For errors we also observed main effects of Object Swap (F(1,61) = 27.943, p < .001, n_p^2 =.314), Angular Disparity (F(1,61) = 11.634, p = .001, n_p^2 =.160), and Judgement Type (F(1,61) = 6.770, p = .012, n_p^2 =.100). No significant interactions were observed. In order to understand this pattern of results, we evaluated each of our five pre-registered hypotheses (Table 1; additional minor Hypotheses 6 and 7 are reported in Supplementary Material).

Hypothesis 1: As with Experiment 1, we employed a 2x2 within-subjects ANOVA with the factors Angular Disparity (0^0, 90^0) and Judgement Type (VPT-2, TB) for Other-judgements and No-swap trials only. The ANOVA applied to RTs in Experiment 2 revealed a significant main effect of Judgement Type F(1,61) = 6.168, p =.016, n_p^2 = .092) and a significant interaction between Angular Disparity and Judgement Type (F(1,61) = 6.871, p = .011, n_p^2 =.101). Planned comparisons between VPT-2 and TB judgements for each angular disparity revealed a significant difference at 90° (t(61) = 3.324, p = .002, d = .422) but not at 0°(p =

.976). The ANOVA applied to errors in Experiment 2 revealed that neither the main effect of Judgement Type (p = .109) nor the Angular Disparity-by-Judgement Type interaction reached significance (p = .904). A Bayesian analysis seeking to ascertain the robustness of support for the null hypothesis revealed moderate to weak evidence in favour (BF_{10}= 0.374). Significant main effects of Angular Disparity were also observed for RTs (F(1,61) = 65.388, p < .001, n_p^2 =.517) and errors (F(1,61) = 6.618, p = .013, n_p^2 =.098) but feed into Hypothesis 2 (see below).

As illustrated in Figure 7, RTs further supported the alternative hypothesis by revealing significantly increased RTs for TB compared to VPT-2 judgements, especially at 90° angular disparity. The error results were not significant in the frequentist analysis, and the Bayesian analysis provided only weak evidence for the null hypothesis. These error results are therefore inconclusive, possibly reflecting the low number of overall errors.

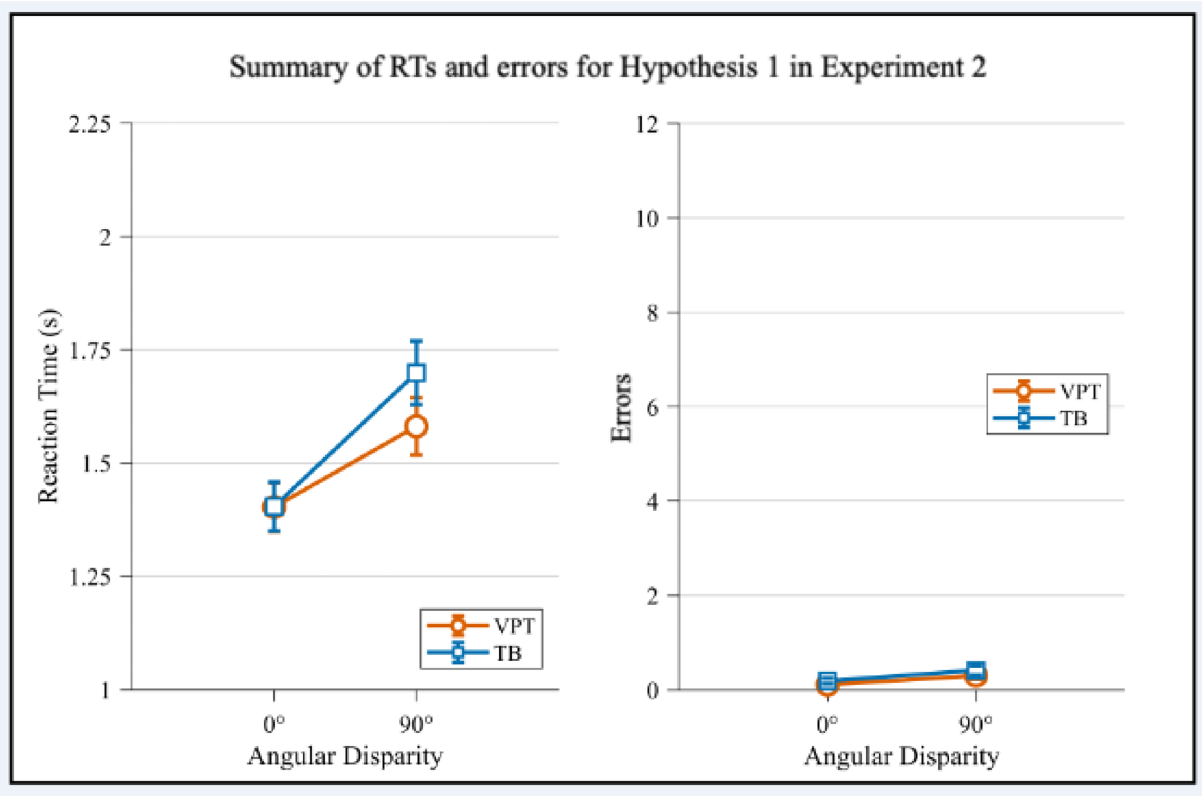

Figure 7. Results for Hypothesis 1 in Experiment 2, with RTs on the left and errors on the right. Error bars are standard error of mean.

Hypothesis 2: In the two full design ANOVAs for Experiment 2 reported above, Hypothesis 2 was confirmed through the significant main effect of Angular Disparity in RTs and errors, and the significant interaction between Angular Disparity and Judgement Type in RTs. However, a more precise test of Hypothesis 2 and a direct replication of Experiment 1 requires a further analysis with Other-judgements only, since Self-judgements were not expected to change with angular disparity.

This analysis of RT data revealed a main effect of Angular Disparity (F(1,61) = 50.511, p < .001, n_p^2 = .453) and a significant interaction between Angular Disparity and Other-Judgement Type in RTs (F(1,61) = 8.878, p = .007, n_p^2 = .127). RTs are slower with increasing angular disparity (90° > 0°) for both types of Other-judgement, yet with a more pronounced increase for belief compared with VPT-2 judgements. error data revealed a significant main effect of Angular Disparity (F(1,61) = 11.634, p = .001, n_p^2 = .160), but no significant interaction (p = .882). These findings further support Hypothesis 2 that increasing angular disparity between self and other (90° vs 0°) increases cognitive demand for Other-judgements (Figure 8).

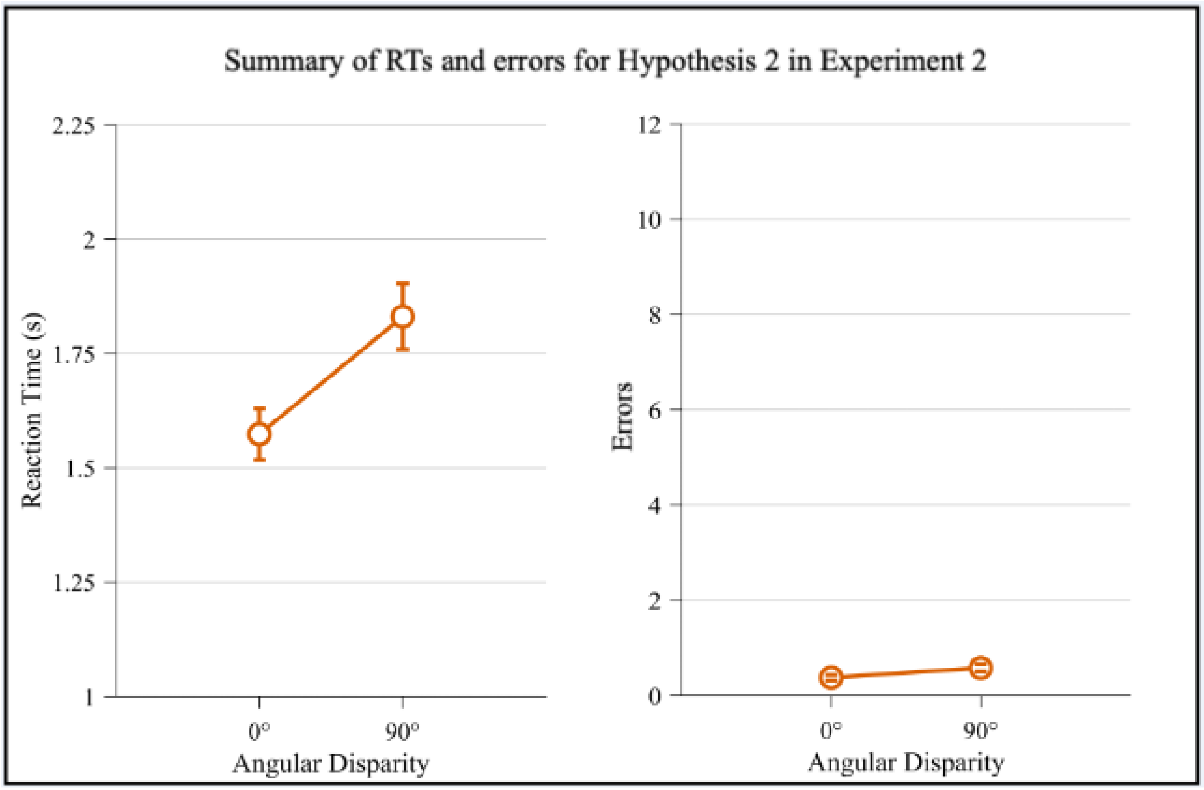

Figure 8. Results for Hypothesis 2 (Table 1) in Experiments 2, with RTson the left and errors on the right. Error bars denote standard error of mean.

Hypothesis 3: As with Experiment 1, we conducted 2x2 ANOVAs with Object Swap (Swap [FB], No-swap [TB]) and Angular Disparity (0°, 90°) as factors for Other-judgement trials only. This analyses revealed the expected main effects of Object Swap (RT: $F(1,41) = 79.416$, $p < .001$, $n_p^2 = .566$; error: $F(1,41) = 19.198$, $p < .001$, $n_p^2 = .241$) and Angular Disparity (RT: $F(1,41) = 47.991$, $p < .001$, $n_p^2 = .440$; error: $F(1,41) = 6.915$, $p = .011$, $n_p^2 = .102$), but no significant interactions in either RT ($p = .641$) nor error ($p = .928$). As shown in Figure 9, FB judgements were made with significantly greater RTs and errors than TB judgements. Replicating Experiment 1, angular disparity did not significantly interact with FB cost.

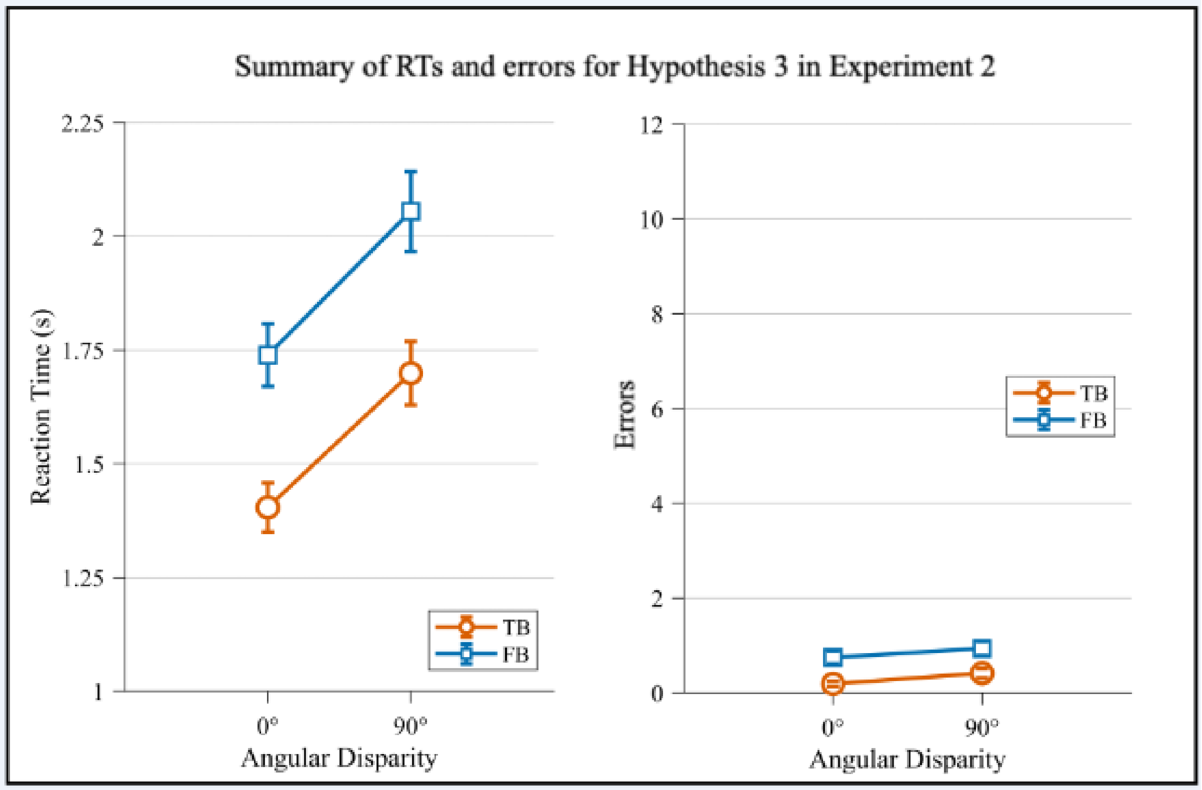

Figure 9. Results for Hypothesis 3 in Experiments 2, with RTs on the left and errors on the right. Error bars denote standard error of mean.

Hypothesis 4: Our results so far indicate consistent differences between TB and VPT-2 processing; TB judgements are associated with greater RTs than VPT-2 judgements, especially at 90ºangular disparity (Hypothesis 1). It could be argued that TB trials required an additional memory component, since participants also had to remember the first stimulus, whereas participants only needed to remember the final stimulus and its appearance to respond accurately in VPT-2 trials. It is therefore unclear whether TB judgements were slower because of an additional memory check or because of additional social processing. To address this question, two additional conditions were added in Experiment 2. A self-past question concerning the first stimulus (“What did *you* initially see?”) and a self-current question related to the second stimulus ("What will *you* actually see?"). It is important to stress that these two questions concern the participant's own perspective – they are not social in nature. Comparing these two non-social questions allowed us to ascertain the relative contributions of a memory check without confounding social processes. Two-tailed t-tests were performed between judgements made about the first stimuli (Self-past) and second stimuli (Self-current).

Contrary to our prediction, Self-past judgements were not associated with increased RTs or errors compared with Self-current judgements; in fact, the former were associated with significantly *lower* RTs and errors than the latter ($t(61) = 7.943$, $p < .001$, $d = 1.009$, and $t(61) = 2.018$, $p = .048$, $d = .256$, respectively; see Figure 10). This indicates that participants may actually anchor their internal representation of the task events to this first stimulus, so they recall the first stimulus (self-past judgements) faster and more accurately than they can recall the second stimulus (self-current judgements).

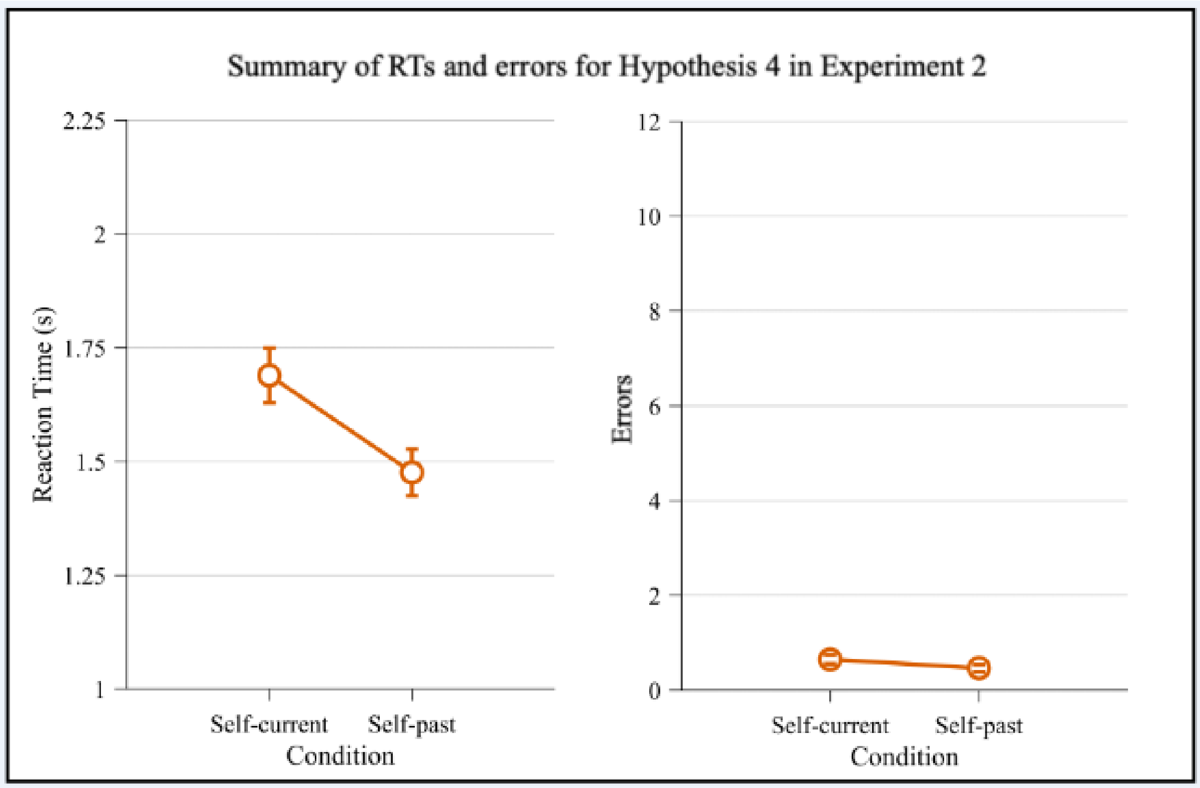

Figure 10. Results for Hypothesis 4 in Experiment 2, with RTs on the left and errors on the right. Error bars denote standard error of mean.

Hypothesis 5: As described in Table 1, this hypothesis was conditional on Hypothesis 4 being confirmed, which was not the case; in fact, the pattern was reversed in comparison to the hypothesized effect direction. It is therefore unlikely that a memory check may explain the slower responses for TB than VPT-2 judgements in Experiments 1 and 2. Nevertheless, we calculated and compared (Self_Past – Self_Current) vs (Other_TB – Other_VPT-2), conducting one-tailed t-tests that resulted in significantly stronger differences for Other- compared to Self-judgements in RTs and errors ($t(61) = 7.267$, $p < .001$, $d = .923$) and ($t(61) = 2.141$, $p = .018$, $d = .272$, respectively; see Figure 11). As previously indicated, these findings suggest that the greater RTs associated with TB calculation relative to VPT-2 processing are more likely to be associated with additional social processes, rather than mnemonic processing (i.e., a check back in memory to the object presented initially).

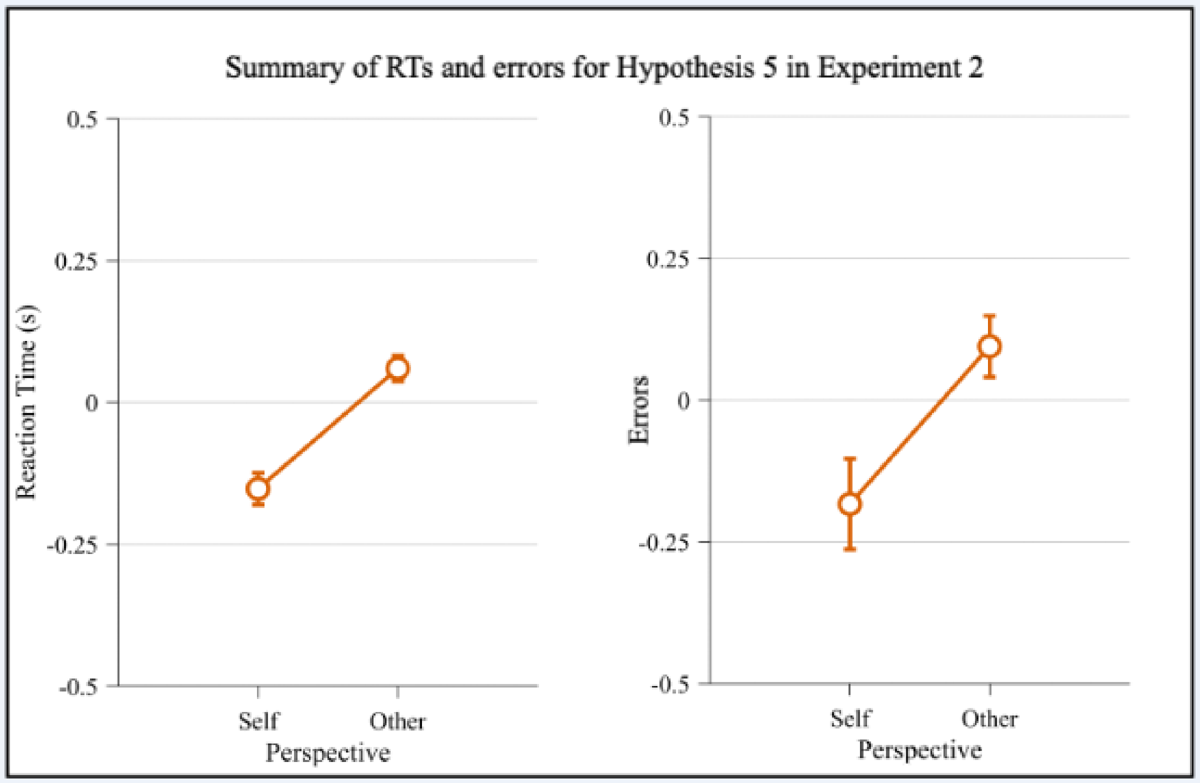

Figure 11. Results for Hypothesis 5 in Experiment 2, with RTs on the left and errors on the right. Error bars denote standard error of mean.

3.3.Experiment 3

The central aim of this third pre-registered experiment was to replicate the relationship between VPT-2 and TB reasoning observed in Experiments 1 and 2, using a different manipulation of TB and FB. Frequently, TBs and FBs have been elicited by manipulating whether a change in an object's location is witnessed (TB) or not (FB; e.g., "Sally-Anne" task). Therefore, in Experiment 3 we manipulated belief state by having the character either be present to witness the object swap (TB) or not (FB; see Figure 3). The target object was always swapped in Experiment 3 to equalize cognitive updating demands across conditions.

The 2x2x2 ANOVA for RTs with the full experimental design included the factors Object Swap (Witnessed [TB], Unwitnessed [FB]), Angular Disparity (0^0, 90^0), and Judgement Type (Belief, Perspective). This revealed main effects of Object Swap ($F(1,41) = 46.289$, $p < .001$, $n_p^2 = .481$), Angular Disparity ($F(1,41) = 42.433$, $p < .001$, $n_p^2 = .449$), and Judgement Type ($F(1,41) = 23.329$, $p < .001$, $n_p^2 = .318$). In addition, we observed a significant interaction between Object Swap and Angular Disparity ($F(1,41) = 9.712$, $p = .003$, $n_p^2 = .163$). For errors, we also observed main effects of Object Swap ($F(1,41) = 17.017$, $p < .001$, $n_p^2 = .254$) and Angular Disparity ($F(1,41) = 6.046$, $p = .017$, $n_p^2 = .108$), but not of Judgement Type ($p = .915$). In addition we observed a significant interaction between Angular Disparity and Judgement Type ($F(1,41) = 7.046$, $p = .011$, $n_p^2 = .124$).

Hypothesis 1: As with Experiments 1 and 2, in Experiment 3 we evaluated this hypothesis with a 2x2 within-subjects ANOVA comprising the factors Angular Disparity (0^0, 90^0) and Judgement Type (VPT-2, TB; see Figure 4), calculated for witnessed trials only. Analysis of RTs revealed that there was a significant main effect of Judgement Type ($F(1,50) = 39.632$, $p < .001$, $n_p^2 = .442$), but the Angular Disparity-by-Judgement Type interaction was not significant ($p = .287$). Analysis of errors revealed no significant main effect of Judgement Type ($p = .458$), but there was a significant interaction between Angular Disparity and Judgement Type ($F(1,50) = 7.741$, $p = .008$, $n_p^2 = .134$). Follow-up analysis of the error results revealed an effect at 0° rather than 90° angular disparity, and in the unpredicted direction (0°: VPT-2 > TB, $t(50) = 2.270$, $p = .028$, $d = 0.318$; 90°: $t(50) = 1.851$, $p = .070$, $d = .259$). Significant main effects of Angular Disparity were also observed for RTs ($F(1,50) = 51.072$, $p < .001$, $n_p^2 = .505$) and errors ($F(1,50) = 4.339$, $p = .042$, $n_p^2 = .080$), but feed into Hypothesis 2 (below).

As illustrated in Figure 12, RT results in Experiment 3 further supported Hypothesis 1 by revealing significantly increased RTs for TBs compared with VPT-2. In this particular paradigm, however, the effect was not significantly more pronounced at 90° compared to 0° angular disparity. Errors revealed an interaction only, showing a small but significant effect in the unpredicted direction at 0°. The latter could be indicative of a speed accuracy trade-off (at

0°) but given the overall low number of errors in this experiment, error findings and possible speed-accuracy trade-offs must be interpreted with care.

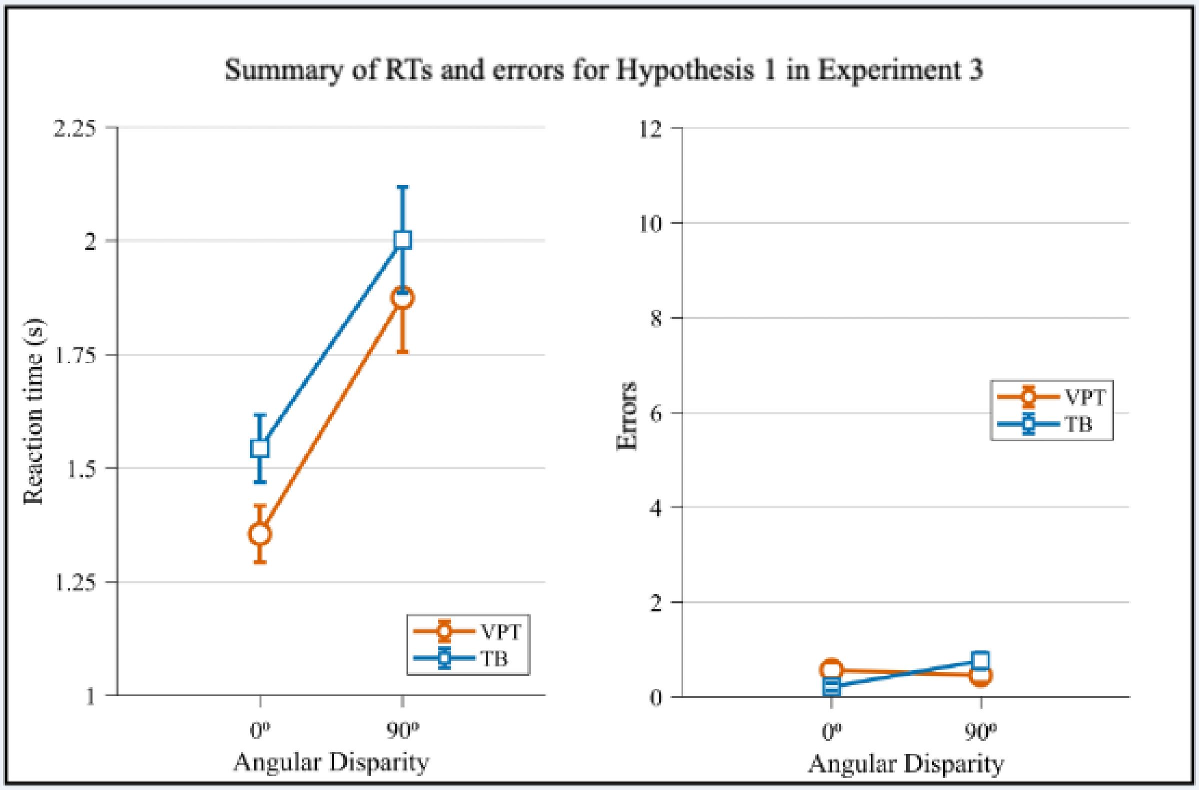

Figure 12. Results for Hypothesis 1 in Experiment 2, with RTs on the left and errors on the right. Error bars denote standard error of mean.

Hypothesis 2: Converging with the results of Experiment 1 and 2, full design analyses of Experiment 3 RT data revealed a significant main effect of Angular Disparity (F(1,50) = 42.433, p < .001, n_p^2 =.449) but no significant interaction between Angular Disparity and Judgement Type (p = .807). error data revealed a significant main effect of Angular Disparity (F(1,50) = 6.046, p = .017, n_p^2 =.108) and a significant interaction between Angular Disparity and Judgement Type (F(1,50) = 7.046, p = .011, n_p^2 =.124). As with Experiments 1 and 2, we urge caution when interpreting these results from errors given the overall low number of errors in this experiment. Together, these results further corroborate the evidence for Hypothesis 2 (Figure 13).

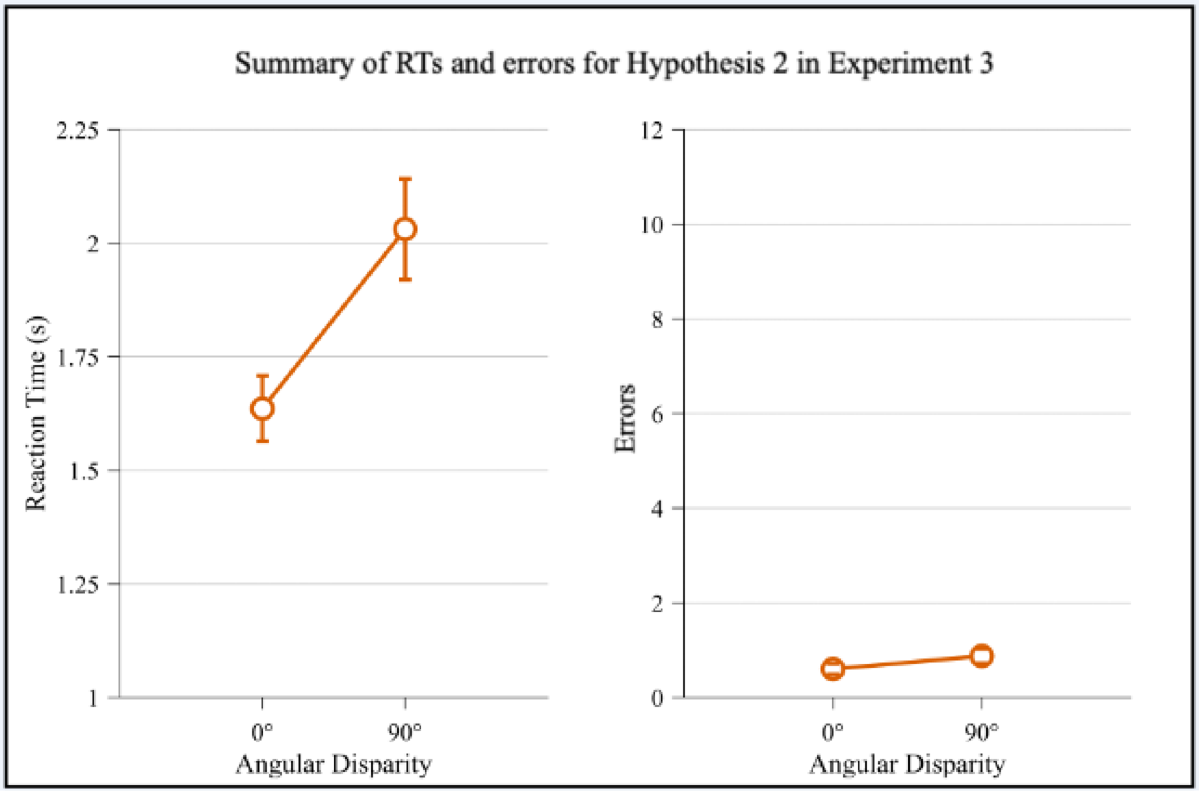

Figure 13. Results for Hypothesis 2 in Experiment 3, with RTs on the left and errors on the right. Error bars denote standard error of mean.

Hypothesis 3: To evaluate this hypothesis, we conducted 2x2 ANOVAs with the factors Object Swap (Swap [FB], No-swap [TB]) and Angular Disparity (0°, 90°) and excluded perspective judgements. This analysis revealed the expected main effects of Object Swap (RT: (F(1,50) = 33.414, p < .001, n_p^2 =.401); error: (F(1,50) = 12.106, p = .001, n_p^2 =.195)) and Angular Disparity (RT: F(1,50) = 38.735, p < .001, n_p^2 =.437; error: F(1,50) = 11.137, p = .002, n_p^2 =.182), but no significant interactions in neither RT (p = .091) nor error (p = .410). Again, FB judgements were made with significantly greater RTs and errors than TB judgements (Figure 14), replicating previous findings. Replicating Experiments 1 and 2, angular disparity did not significantly interact with FB costs.

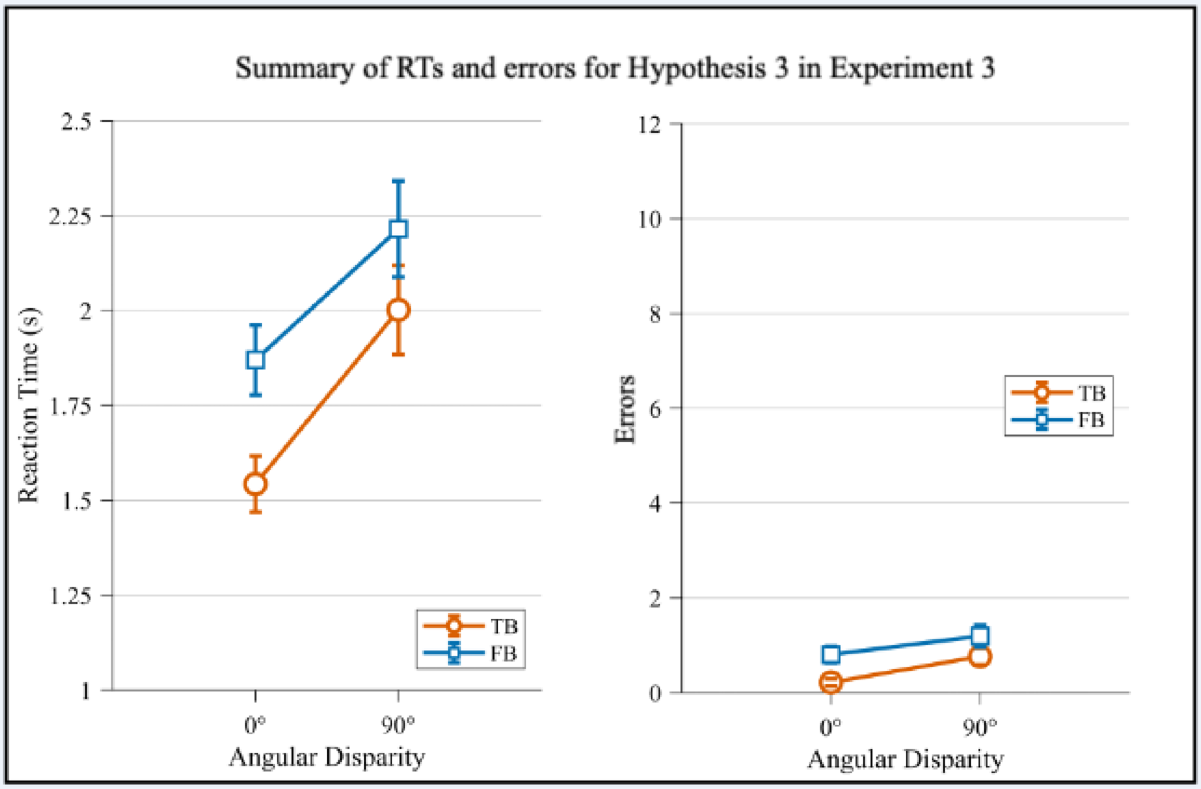

Figure 14. Results for Hypothesis 3 in Experiments 3, with RTs on the left and errors on the right. Error bars denote standard error of mean.

4. Discussion

We present a novel paradigm, the "Seeing-Believing Task", designed specifically to investigate differences between cognitive processes that allow us to infer another's belief states vs. their visual perspective of the world. This is a crucial question in ToM research (Gunia et al., 2021; Hamilton et al., 2009; Schurz et al., 2013), insights into which would afford a more precise neurocognitive categorisation of the processing mechanisms involved. Previously, research has focused on false belief (FB) rather than true belief (TB) reasoning (Baron-Cohen et al., 1985) because a more complex understanding of another's mental representation is required when their belief is known to the observer to be false (Apperly et al., 2008). Indeed, representing another's FB has been shown reliably to be slower and more error prone than representing a TB that is consistent with the observer's own representation of reality (Apperly et al., 2008; Apperly et al., 2011). Here, we replicate and extend this finding consistently across three pre-registered online experiments.

Given these considerations about added costs for processing FBs, it would be unsurprising if FB judgements also lead to slower and more error-prone processing compared to VPT-2 judgements. For our primary hypothesis (Hypothesis 1), we therefore compared judgements about another's perspective ("What will she actually see?") with judgements about their TB – and not their FB – ("What will she [*correctly*] expect to see?") since in our paradigm, both, VPT-2 and TB, concern representations of the actual state of reality, as it is known to the participant. We considered two alternative theoretical notions. The first predicted identical or largely overlapping cognitive processing for VPT-2 and TB judgments in light of strong correlations between independent performance indices of VPT-2 and ToM (e.g., Hamilton et al., 2009; Quesque & Rossetti, 2020; Złotogórska-Suwińska & Putko, 2019) and overlap in associated neural substrates (e.g., Gunia et al., 2021 for review; Schurz et al., 2013). Furthermore, both reflect the same (true) state of reality and required exactly the same responses in our novel paradigm.

In contrast, our alternative hypothesis predicted that there would be a significant performance difference when making VPT-2 and TB judgements, because the two types of mentalising might rely – at least in part – upon psychologically distinct mechanisms. Firstly, another's visuospatial perspective can be dissociated from their belief if they close their eyes, thus, a simple perceptual manipulation affecting their perspective but not their belief. It was further argued earlier that representing another's (true) belief might require additional or more sophisticated processing of their mental states than merely representing their visuospatial perspective (e.g., Howlin et al., 1999). For instance, the latter could be resolved by imagining the self occupying another virtual viewpoint, rather than fully representing the other's mental state in relation to their viewpoint. Indeed, Kessler and Thomson (2010) reported that the same embodied mental self-rotation was also employed when participants imagined themselves in a different viewpoint that was defined by an empty chair rather than another person. A wealth of research conducted in the context of "spatial updating" further suggests that VPT-2 mechanisms are widely employed in non-social, spatial contexts (for review see Hamilton et al., 2014; Kozhevnikov, Motes, Rasch, & Blajenkova, 2006; Wang et al., 2006; Zacks & Michelon, 2005).

Our primary hypothesis (Hypothesis 1) also predicted that if TB imposed greater cognitive load than VPT-2, this difference could be more pronounced when the other's perspective differed from the participant's egocentric perspective (90° angular disparity); that is, when the other's perspective could not be merely assimilated into the participant's own perspective. More precisely, Hypothesis 1 assumed that a mental transformation into the other's viewpoint was required for all other-related judgements at 90° angular disparity, but additional processing would then be needed for TB compared with VPT-2 judgements. Importantly, the required mental transformation for all other-related judgements at 90° angular disparity (Hypothesis 2) was also evaluated. Both hypotheses received consistent support in RT and error analyses across all three experiments, replicating previous findings in relation to Hypothesis 2 (e.g., Kessler & Thomson, 2010; Michelon & Zacks, 2006).

Regarding Hypothesis 1, in all three experiments, TB judgements were associated with greater cognitive effort than VPT-2 judgements (larger RTs and, more inconsistently, number of errors). Further, Experiments 1 and 2 also revealed that the difference in RTs was greater when self- and other-perspectives differed (90° angular disparity). Experiment 3, however, revealed equally higher RT costs for TB compared to VPT-2 at 0° and 90° angular disparity. This is likely due to the more costly manipulation in this paradigm, where the target object was always swapped but was either witnessed by the character or not. Response times were generally longer in Experiment 3, confirming these added costs (Tables 4-6). The required object update and the need for keeping track of the character's perspective in the scene (what they had witnessed or not) may have prevented participants to assimilate the other's perspective and belief into the egocentric perception of reality at 0° angular disparity as effectively as in Experiment 1. This would explain the additional cost of TB over VPT-2 judgements at 0° angular disparity in Experiment 3, which was nullified in Experiments 1 and 2. Future research should address these considerations.

Error data presented a somewhat inconsistent pattern, where Experiment 1 revealed a significant main effect of Judgement Type but no interaction with Angular Disparity, Experiment 2 showed no effects, while Experiment 3 revealed a significant interaction between Judgement Type and Angular Disparity but no main effect of Judgement Type. In a Bayesian follow-up analysis of Experiment 2 weak to medium support for the H_0 (no difference between VPT-2 and TB) was found, which is the only piece of evidence in the entire study that supports the notion that TB and VPT-2 might not be distinct processes. Error findings have to be interpreted with care due to the overall low number of errors, but more importantly, error data did not indicate speed-accuracy trade-offs, which could have jeopardised the interpretation of the RT findings. Therefore, in the light of overwhelming evidence in RT data we conclude that inferences concerning another's true belief require additional resources as compared to judgements about their visual perspective.

What exactly is the difference between VPT-2 and TB? Given that perspective judgements always referred to the final object in a trial, while FB judgments related to the initial object in a trial, other non-social variables may have influenced processing throughout belief trials, for FB as well as TB. For example, as FB judgements depended on a larger number of events in a trial, participants may have also been more uncertain about making their decision and taken more time on TB trials. Such uncertainty when making a belief judgement might have resulted in a check back in memory to the beginning of a trial (to determine if the object really was swapped or not). This was investigated directly in Experiment 2 by including past and current judgements from the participant's egocentric perspective (self-judgements).

Crucially, including self-judgements did not change the basic pattern of difference between other-related VPT-2 and TB judgements observed in Experiment 1, indicating that the initial results did not simply reflect an absence of self-judgements (Quesque & Rossetti, 2020). Surprisingly, however, self-related judgements did not reveal the hypothesised processing advantage for current- relative to past-judgements ("What *will* you actually see" compared with "What *did* you initially see?"; Hypothesis 4). Instead, past-self judgements were significantly faster than current-self judgements, ruling out the possibility of a memory check for the former. While surprising and worth further pursuit in future research, this effect suggests that participants anchor their representation of trial events to the initially presented scene rather than updating their representation as the trial unfolds. We contend that an extrapolation of this conclusion from self- to other-judgements is legitimate because experimental conditions were only distinguishable at the very end of each trial when the different judgements were required. Thus, given the lack of any early distinguishing features, it is valid to conclude that all trials were encoded in the same way until the final question.

As discussed, we considered the possibility that, on TB trials, participants might alleviate their uncertainty by initiating an unnecessary check back in memory to confirm whether or not the target object had been swapped (Hypothesis 5). In the light of the surprising result for self-related judgements, this now appears unlikely. However, while we can rule out a memory check, other sources of uncertainty should be considered in future studies as possible explanations for the difference between VPT-2 and TB. Uncertainty is often the reason for slower decisions, and TB might involve more uncertainty than VPT-2. However, it is important to understand what the source of uncertainty might be. Considering “what the other ought to know” during belief judgments might increase “representational” uncertainty (we know less about the other’s belief and desires) compared with mere judgements of reality from another’s perspective. In other words, belief reasoning is likely to involve more complex and/or uncertain representations in the sense that more sources of information are available to the other to inform their beliefs as compared to their visual perspective of reality. There is more uncertainty about what exactly may inform the other’s belief, but we (as participants) may settle for the assumption that what they have seen will be their primary source for informing their belief. The process of settling for the latter could then result in longer processing times.

In addition to “representational” uncertainty that might be perceived as inherent to the other’s true beliefs (“She is uncertain about her belief”), participants might also be more uncertain about their correct processing of all events on trials that require belief representation (“I am uncertain about my judgement of her belief”). Future research could address these sources of uncertainty through distinct confidence scales after each trial (e.g., “How certain is She?” vs “How certain are You?”).

Importantly, the notion of uncertainty that may emerge from various sources of information being available to the other to inform their belief ties in with a more mechanistic interpretation of our main result. We conjecture that the added effort for TB compared to VPT-2 judgements indexes a transition from minimal to full-blown ToM processing, as described by Apperly and Butterfill (2009; also Butterfill and Apperly, 2013). Minimal ToM is defined as relational, whereby objects and their locations are registered with another agent (e.g., what the other can or cannot see at any given time, and whether their last registration of an object is correct or false with respect to the object’s current location), thus enabling goal-directed actions that allow participants to complete ToM tasks quickly and efficiently without the need for slower, more effortful, full-blown ToM. Apperly and Butterfill conceived of the latter as a full representation in propositional format, allowing inferences about the other’s mental states and future actions (Butterfill and Apperly, 2013). This is further underpinned by philosophical distinctions between perceptual and inferential mentalizing, where only the latter requires propositional processing (e.g., Michael & De Bruin, 2015; McNeill, 2012).

While the embodied transformation engaged for VPT-2 requires substantial processing effort when angular disparity is high (in contrast to line-of-sight mechanisms involved in level-1 perspective tracking; e.g., Michelon & Zacks, 2006; Kessler & Rutherford, 2010), due to its analogue embodied nature (Kessler & Thomson, 2010) this transformation would not automatically generate a propositional representation of the other’s visual experience (for a connectionist model see Kessler, 1999). The outcome of the transformation could still be a relational representation of the target perspective – conforming to minimal ToM – that simply registers the correct perceptual orientation of the object from the mentally embodied viewpoint (also McNeill, 2012). It is therefore not strictly required to generate a propositional mental representation of the other’s experience for resolving a perspective judgement correctly (e.g., Kessler & Thomson, 2010, Experiment 3; same embodied transformation without an avatar).

However, when explicitly asked to take another’s belief into account, participants might feel compelled to generate a propositional representation of the other’s belief at the expense of additional, inferential processing resources (i.e., full-blown ToM). This seems plausible when considering the details of our paradigm. The question “What will she expect to see?” might not only depend on what she last saw. While she was away, she may have obtained further information about what to expect, or she may have become aware that the object has been swapped in the past while she was away, or she may have forgotten what object she initially saw. These considerations, akin to considering counterfactual possibilities (see Rafetseder, O’Brien, Leahy, & Perner, 2021 for a developmental relationship between belief- and counterfactual reasoning), tie in with considerations of uncertainty as described above. As indicated by their responses, participants settle for what she initially saw as the best source of information for her belief, but such a decision process would come at extra costs.

The crucial point here is that considering various sources of information that may inform the other’s (true) belief, akin to considering counterfactual possibilities, requires inferential processing (full-blown ToM), while a factual judgement about the object’s identity and orientation from the other’s viewpoint (VPT-2) can be resolved through relational processing (minimal ToM). Of course, further findings are required to corroborate this conjecture and future studies should aim at testing this interpretation of our effects (see Section 4.2. Limitations and Outlook).

Note that Apperly and Butterfill's framework (2009; Butterfill & Apperly, 2013) includes the possibility that belief judgements could also be resolved using minimal ToM. However, a relational representation that simply registers objects and their locations with another observer's past and current viewpoints would not differ between VPT-2 and TB judgements in our paradigm. Therefore, our findings speak against this possibility.

Finally, the angular disparity manipulation in our novel paradigm allowed us to determine possible effects of overlapping vs disparate self and other perspectives on the processing costs of FB vs TB (Hypothesis 3). Recently, overlapping self and other perspectives have been identified as a common confound in typical "false location belief reasoning tasks" (e.g., Rahman et al., 2021). However, when we included Angular Disparity as a factor in the analysis of the difference between FB and TB, we did not observe any significant interactions between Belief State (TB, FB) and Angular Disparity (0°, 90°). Thus, an overlapping perspective between self and other does neither increase nor decrease the added costs for FBs compared to TBs. We therefore conclude that the observed differences are likely to reflect the additional demands of inhibiting one's own *perspective-unspecific* privileged representation of reality (i.e., object identity under the bucket) in FB tasks, unlike TB tasks where the other's belief is congruent with the participants' privileged representation (but not necessarily identical in perspective).

4.1. Developmental considerations

VPT-2 judgements were consistently quicker than TB judgments. One could therefore argue that perspective taking might be a developmental stepping-stone towards more sophisticated forms of mentalizing such as full-blown belief reasoning (Gunia et al., 2021; Kessler et al., 2014; Kessler & Thomson, 2010; Michelon & Zacks, 2006). However, several objections will have to be considered. In our task, TB judgements had to be made based on the other's viewpoint. Thus, a viewpoint transformation was required at 90º in any case and while longer processing times for TBs may indicate additional/different processing compared to VPT-2, it does not necessarily indicate that TB processing depends on fully developed perspective taking abilities. In other words, TBs that do not rely in their content on a VPT-2 transformation, e.g., TBs about the correct location of an object (rather than its viewpoint-dependent orientation), could actually develop before VPT-2.

Indications that children may pass simple forms of TB (around the age of 3+) before FB (e.g., Fabricius, Boyer, Weimer, & Carroll, 2010) and VPT-2 (around the age of 4-5; e.g., Hamilton et al., 2009) have indeed been reported. However, a debate has ensued whether TB processing is the same in all age groups. That is, TB processing at an early age, before FB tests are passed, could be different to processing at a later age, when FB tests are also passed (e.g., Fabricius, Boyer, Weimer, & Carroll, 2010; but see Huemer et al 2023). Young children may pass TB tests in a simpler way, potentially using minimal ToM, while older children may process them in a similar fashion to FBs and counterfactual thinking, eventually engaging full-blown ToM. VPT-2 could therefore be a stepping-stone for FB more specifically, which in turn may affect the way children process beliefs altogether, i.e., using inferential reasoning. An interesting implication would be that children, who can complete VPT-2 and TB tasks, but have not yet mastered FB reasoning, would not show a difference between VPT-2 and TB judgements in our Seeing-Believing Task, as they might process VPT-2 and TB judgements in the same relational fashion, using minimal ToM. We hope that our novel paradigm may benefit developmental research as well as further research with adults.

4.2. Limitations and Outlook

This study is the first to employ the newly developed Seeing-Believing Task and must therefore be treated as preliminary. This is reflected in our suboptimal pre-registration of Hypothesis 3, for instance, where we did not maximise the possibilities of our novel paradigm. We pre-registered three experiments to address some of the most obvious experimental questions, such as the influence of self vs other judgements, memory (Experiment 2) and different ways of manipulating true and false beliefs (Experiment 3). We further chose to pre-register our study given that experimentation and data collection was conducted online. While we believe that the presented results are consistent and powerful, a replication in more controlled laboratory settings might reveal further subtleties of this novel paradigm.

A surprising observation across all three experiments was that average number of errors were so low that variability was limited, affecting analysis and interpretation of this dependent measure. This is surprising, given that a substantial number of participants in each experiment failed to achieve our minimal performance criterion in at least one condition. It appears that these participants may not have understood the instructions fully. This is perhaps understandable given the lack of interactive explanation and instruction in this online version of a complex paradigm. However, the majority of participants who understood the instructions committed low numbers of mistakes. As this may reflect a selection bias in terms of levels of motivation and attention in these successful participants, a replication under laboratory conditions might help achieve more realistic estimates of errors.

A further surprising outcome that would warrant further investigation was the strong effect favouring past-self judgments over current-self judgements in Experiment 2. This indicates that participants anchor their trial representation to the initial scene rather than update their representations continuously as the trial unfolds. A new version of Experiment 2, where Self and Other judgments are presented in separate blocks might shed light on effects induced by the complexity of having to maintain the possibility of four different judgement types until the very end of each trial. Differences between past- and current-self judgements might be diminished in such a paradigm (see Hypothesis 6 in Supplementary Material). It would also be of interest to replicate the observed difference between past- and current-self in the context of Experiment 3 (e.g., adding self-past and self-current judgements), in order to understand if the more difficult manipulation of un/witnessed object swaps changes participants' representational anchoring and updating during trials.

Finally, future research should aim at corroborating the interpretation our main effect. We proposed that VPT-2 judgements are processed relationally, while TB judgments are processed propositionally. One option for investigating this further would be to include target objects that either instil a symbolic (e.g., digits/symbols) or a pictorial (e.g., abstract shapes) processing format. Denying a switch between formats due to the nature of the target objects could reduce the observed difference between VPT-2 and TB. A second option would be to tap into distinct processing properties attributed to each format. For instance, a relational format should be more susceptible to priming than a propositional format. By including trials where the character sits at 90° disparity at the start, the character's disparate perspective could be primed through relational processing (by registering the object on the table with the character's perspective from the outset), resulting in faster VPT-2 judgements at the end of such trials. However, if TB requires a switch to a propositional format, relational priming of the other's perspective should not affect the additional costs for processing their TB. Thus, the difference between TB and VPT-2 should be preserved or even appear increased as VPT-2 becomes faster. Finally, only a propositional format enables inferences regarding potential actions and consecutive mental states. By including additional statements at the very end of trials, such as "The character is surprised", that require a yes/no forced choice response, one could probe this property. The prediction would be that these responses would be given faster after TB/FB judgments as compared to VPT-2 judgements. For the latter a propositional representation would have to be generated first, before inferences could be drawn about ensuing mental states.

4.3. Conclusions

Using a novel Seeing-Believing Task, the experiments comprising this study provide consistent evidence that inferences concerning another's visual perspective (how the world appears visually to them) involves, at least in part, psychologically distinct processes to those supporting inferences about their beliefs (what they believe to be true about the visual world). This is the case even when the other's belief is true, referring to the same true state of reality and requiring identical responses in the paradigm we developed. Perspective judgements were found to be reliably faster and less error prone when compared to those concerning beliefs, indicating that the former involves representations of another's mental states that are easier to generate. We propose that the observed differences could reflect higher demands in social processing (rather than mnemonic processing) for representing another's mind when generating their true belief as compared to merely imagining another's visual viewpoint. The higher processing demands could be indicative of a transition from relational (perspectives) to propositional (beliefs) processing. The presented work supports future research endeavours by providing a novel but replicated paradigm along with open materials and data (https://osf.io/57wa9/).

Acknowledgements. RG was supported by an Aston University PhD Scholarship. We would like to thank Prof Steven Butterfill for his insightful comments on our interpretation of the findings. We wish to thank Nicholas Alexander for his help with generating the experimental stimuli with Unreal Engine Version 4.17.2.

References

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

Apperly, I. A., 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.

Apperly, I. A., Warren, F., Andrews, B. J., Grant, J., & Todd, S. (2011). Developmental continuity in theory of mind: Speed and accuracy of belief–desire reasoning in children and adults. Child Development, 82(5), 1691-1703.

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.

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

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The autism-spectrum quotient (AQ): Evidence from asperger syndrome/high-functioning autism, malesand females, scientists and mathematicians. Journal of Autism Developmental Disorders, 31(1), 5-17.

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

Fabricius, W. V., Boyer, T. W., Weimer, A. A., & Carroll, K. (2010). True or false: Do 5-year-olds understand belief?. Developmental Psychology, 46(6), 1402.

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.

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.

Gunia, A., Moraresku, S., & Vlček, K. (2021). Brain mechanisms of visuospatial perspective-taking in relation to object mental rotation and the theory of mind. Behavioural Brain Research, 113247.

German, T. P., & Hehman, J. A. (2006). Representational and executive selection resources in ‘theory of mind’: Evidence from compromised belief-desire reasoning in old age. Cognition, 101(1), 129-152.

Hamilton, A., Brindley, R., & Frith, U. (2009). Visual perspective taking impairment in children with autistic spectrum disorder. Cognition, 113(1), 37-44.

Hamilton, A., Kessler, K., & Creem-Regehr, S. H. (2014). Perspective taking: building a neurocognitive framework for integrating the “social” and the “spatial”. Frontiers in human neuroscience, 8, 403.

Howlin, P., Baron-Cohen, S., & Hadwin, J. A. (1999). Teaching children with autism to mind-read: A practical guide for teachers and parents. John Wiley & Sons.

Huemer, M., Schröder, L. M., Leikard, S. J., Gruber, S., Mangstl, A., & Perner, J. (2023). The knowledge (“true belief”) error in 4-to 6-year-old children: When are agents aware of what they have in view?. Cognition, 230, 105255.

Hurley, R. S., Losh, M., Parlier, M., Reznick, J. S., & Piven, J. (2007). The broad autism phenotype questionnaire. Journal of Autism Developmental Disorders, 37(9), 1679-1690.

Kessler, K. (2000). Raumkognition und Lokalisationsäußerungen: Ein konnektionistisches Modell des Verstehens von Richtungspräpositionen. Wiesbaden: Deutscher Universitäts-Verlag – Studien zur Kognitionswissenschaft.

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, 213.

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

Kessler, K., Cao, L., O'Shea, K. J., & Wang, H. (2014). A cross-culture, cross-gender comparison of perspective taking mechanisms. Proceedings of the Royal Society B: Biological Sciences, 281(1785), 20140388.

Kozhevnikov, M., Motes, M. A., Rasch, B., & Blajenkova, O. (2006). Perspective-taking vs. mental rotation transformations and how they predict spatial navigation performance. Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory, 20(3), 397-417.

Landa, R., Piven, J., Wzorek, M., Gayle, J., Chase, G., & Folstein, S. (1992). Social language use in parents of autistic individuals. Psychological Medicine, 22(1), 245-254.

McNeill, W. E. S. (2012a). Embodiment and the perceptual hypothesis. The Philosophical Quarterly, 62, 569–591.

Michael, J., & De Bruin, L. (2015). How direct is social perception?. Consciousness and Cognition, 36, 373–3.

Michelon, P., & Zacks, J. M. (2006). Two kinds of visual perspective taking. Perception & psychophysic, 68(2), 327-337.

Muto, H., Matsushita, S., & Morikawa, K. (2018). Spatial perspective taking mediated by whole-body motor simulation. Journal of Experimental Psychology: Human Perception and Performance, 44(3), 337.

Newcombe, N. (1989). The development of spatial perspective taking. Advances in child development behavior, 22, 203-247.

Peirce, J. W. (2007). PsychoPy—psychophysics software in Python. J Journal of neuroscience methods, 162(1-2), 8-13.

Piven, J., Palmer, P., Jacobi, D., Childress, D., & Arndt, S. (1997). Broader autism phenotype: evidence from a family history study of multiple-incidence autism families. American Journal of Psychiatry, 154(2), 185-190.

Quesque, F., & Rossetti, Y. (2020). What do theory-of-mind tasks actually measure? Theory and practice. Perspectives on Psychological Science, 15(2), 384-396.

Rafetseder, E., O'Brien, C., Leahy, B., & Perner, J. (2021). Extended difficulties with counterfactuals persist in reasoning with false beliefs: Evidence for teleology-in-perspective. Journal of Experimental Child Psychology, 204, 105058.

Rahman, F., Kessler, K., Apperly, I. A., Hansen, P. C., Javed, S., Holland, C. A., & Hartwright, C. E. (2021). Sources of Cognitive Conflict and Their Relevance to Theory-of-Mind Proficiency in Healthy Aging: A Preregistered Study. Psychological science, 09567976211017870.

Schurz, M., Aichhorn, M., Martin, A., & Perner, J. (2013). Common brain areas engaged in false belief reasoning and visual perspective taking: a meta-analysis of functional brain imaging studies. Frontiers in human neuroscience, 7, 712.

Schurz, M., Radua, J., Aichhorn, M., Richlan, F., & Perner, J. (2014). Fractionating theory of mind: a meta-analysis of functional brain imaging studies. Neuroscience Biobehavioral Reviews, 42, 9-34.

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

Van Overwalle, F. (2009). Social cognition and the brain: a meta-analysis. Human brain mapping, 30(3), 829-858.

Wang, R. F., Crowell, J. A., Simons, D. J., Irwin, D. E., Kramer, A. F., Ambinder, M. S., . . . Hsieh, B. B. (2006). Spatial updating relies on an egocentric representation of space: Effects of the number of objects. Psychonomic bulletin review, 13(2), 281-286.

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

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

Zacks, J. M., & Michelon, P. (2005). Transformations of visuospatial images. Behavioral cognitive neuroscience reviews, 4(2), 96-118.

Złotogórska-Suwińska, A., & Putko, A. (2019). Visual Perspective-Taking: Are Theory of Mind and Executive Function Involved? Polish Psychological Bulletin, 50(2).

Supplementary Materials
Comparing Visual Perspective Taking and Belief Reasoning

Supplementary Results

Hypothesis 6 (Experiment 2 only): Self- vs Other-Perspective differ, particularly at high angular disparity

In replication of previous perspective taking results (e.g. Surtees, Samson & Apperly, 2016; Surtees, Butterfill & Apperly, 2012), it was expected that self judgements would be associated with shorter RT and/or lower ERs than judgements about another's perspective, especially – or even only - when another's perspective is different to one's own (i.e., ± 90^0 angular disparity). This hypothesis was examined in Experiment 2 using a 2 x 2 ANOVA with Angular Disparity (0°, 90°) and Judgement Type (Self-Perspective, Other-Perspective) as factors. (Note that Self-Past and Self-Current judgements were collapsed into Self-Perspective and only no-swap trials were included in this analysis). Main effects of Angular Disparity ($F(1,61)$ = 49.795, p < .001, n_p^2 =.449) and Judgement Type ($F(1,61)$ = 4.908, p = .03, n_p^2 =.074), as well as the Angular Disparity-by-Judgement Type interaction ($F(1,61)$ = 4.587, p = .036, n_p^2 =.07) reached significance (Figure S-1). With respect to errors, only the main effect of Angular Disparity ($F(1,61)$ = 11.438, p = .001, n_p^2 =.101) was significant.

Two paired samples *t*-tests were employed to follow up on the ANOVA for RTs, one comparing Other-Perspective and Self-Perspective judgements made at 90^0, and another comparing Other-Perspective and Self-Perspective judgements made at 0^0. Other-Perspective judgements made at 90^0 where associated with significantly greater RTs ($t(61)$ = 2.682, p = .009, d = .341) than Self-Perspective judgements, whereas, other-perspective judgements made at 0^0 were not associated with greater RTs (p = .162).

Based on the RT results we conclude that there is little difference in processing speed between Self- and Other-Perspective judgements when there is no angular disparity, i.e., the other's perspective is identical to the egocentric perspective. In contrast, when there is an angular disparity of 90^0 it is more difficult to adopt the other's perspective (RTs), which replicates extant literature.

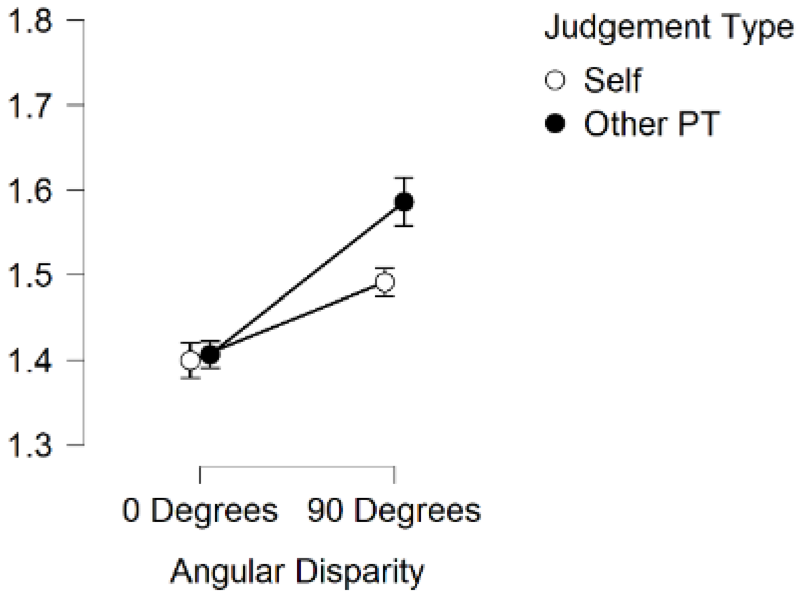

Figure S-1. Results for Hypothesis 7. Angular Disparity-by-Judgement Type interaction for RT data (in seconds). Error bars are standard error of mean.

However, it is important to note that these differences between self and other perspectives in RTs were only observed when Self-Past and Self-Current judgements were averaged together. As shown in Table 5 (main text), only Self-Past but not Self-Current judgements actually appear to differ from Other-Perspective judgements, which is also the case for no-swap trials, where Self-Past and Self-Current judgements required identical responses. As we argue for Hypotheses 4 and 5 in the main text, for the complex paradigm we have employed here, participants appeared to anchor each trial to the scene at the start, resulting in Self-Past judgements being processed the fastest, while Self-Current judgements being significantly slowed down, to the level of Other-Perspective judgements (while Other-Belief judgements being slowest, see Hypothesis 1 in the main text). Further research is therefore required to understand the slow responses for Self-Current judgements. For instance, a blocked design separating Self and Other judgements might reveal diminished differences between Self-Past and Self-Current, when answer options are narrowed down to Self judgements only. In contrast, a block of Other judgements should replicate the findings from Experiment 1, revealing a difference between Other-Perspective (VPT-2) and Other-TrueBelief (TB) judgements (at 90°). With such a design we would also expect to observe differences between Self-Current and Other-Perspective judgements, e.g. at 90° angular disparity.

Hypothesis 7 (Experiment 2 only): Are TBs ascribed by default?

Based on previous research (Back & Apperly, 2010) it was predicted that TB judgements would be associated with greater RTs and/or ERs relative to “reality” (Self-Current) judgements, indicating that TB judgements are *not* ascribed by default (e.g., Back & Apperly, 2010). This was tested in Experiment 2 using one-tailed *t*-tests. Contrary to this prediction, judgements about “reality” (Self-Current) were numerically associated with greater RTs than TB judgements (Table 5), but the difference did not reach significance (p = .0975). There was also no difference in ERs (p = .500). These results might suggest that TBs are in fact ascribed by default in some circumstances. However, it is important to note that the current comparison was conducted for conditions where the other’s perspective and their TB about reality were fully aligned with the participants’ own perspective (0^0 angular disparity) and privileged knowledge of reality. This differs from typical designs in previous research, where the other always occupied a different viewpoint than the self (e.g., Back & Apperly, 2010, on which our hypothesis was based; see pre-registration). Such misalignment of perspectives is likely to trigger additional processing in itself. It could therefore be that processing another’s (true) belief and perspective only require additional resources when their representation cannot be simply “assimilated” into the egocentric perspective and knowledge of reality. This latter statement is further corroborated by the analysis in relation to Hypothesis 7 below, which confirms that processing times (RTs) for other-perspectives are not significantly longer than processing times for self-perspectives at 0^0 angular disparity in Experiment 2.

References

Surtees, A., Samson, D., & Apperly, I. (2016). Unintentional perspective-taking calculates whether something is seen, but not how it is seen. Cognition, 148, 97-105.

Surtees, A. D., 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(1), 75-86.