Cognition 278 (2027) 106699



Contents lists available at ScienceDirect

Cognition

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



Full Length Article



Visual perspective-taking and image-like representations: we can see it

Louis Aldrich [a,*], Katrina L. McDonough [b], Bert Timmermans [a], Patric Bach [a]

[a] Psychology Department, King's College, University of Aberdeen, AB24 3FX, UK
[b] Psychology Department, Lawrence Stenhouse, University of East Anglia, NR4 7TQ, UK

ARTICLE INFO

Keywords:
Visual perspective taking
Visuospatial perspective taking
Size constancy
Mental imagery
Theory of mind
Social cognition

ABSTRACT

Visual perspective taking describes the ability to represent how others see the world. It is presumed to be crucial for social reasoning and may underpin the development of theory of mind. Yet, a recent series of experiments by Samuel et al. (2021) purports that people are entirely incapable of representing another person's view. In these experiments, participants were asked to judge which of two parallel and vertical lines would appear larger from another person's perspective. Results indicated a lack of basic awareness that the line closer to the other person would appear larger to them. Due to potential issues with key features of the task however, Samuel et al.'s (2021) conclusion of perspective taking inability may be premature. Specifically, features of how the critical question was phrased, how the stimuli were created, and the way in which participants responded, likely obscured an underlying capacity for visual perspective taking. Over two experiments, we first demonstrate that minor adjustments to the original paradigm enables participants to reliably report another person's perspective. Second, by using a novel multi-trial adaptation of the paradigm, we show a large proportion of participants can make highly accurate judgements that closely track how another person's visual experience changes with location.

1. Introduction

Visual perspective taking (henceforth VPT) describes one's ability to represent the environment from another person's point of view (Flavell et al., 1981). VPT is ubiquitous in our everyday lives, for example, when giving a tourist directions, when warning another driver of a cyclist around a blind corner, or when mirroring another's dance moves. One of the earliest examples is Piaget's seminal 'three mountains' task (Piaget et al., 1967), which tested children's ability to imagine how another's view of three model mountains would be different to their own. It showed that VPT developed in two distinct stages. While most children were able to indicate when a feature was blocked from the other's viewpoint (level-1 VPT, see: Flavell et al., 1981; Lempers et al., 1977; Moll & Tomasello, 2006), those under the age of seven struggled to describe how the scene visually appeared to the other individual (e.g., which feature would appear leftmost to the other, known as 'level-2' VPT, see: Frick et al., 2014; Masangkay et al., 1974), typically making the mistake of reporting their own egocentric viewpoint (see Michelon & Zacks, 2006, for an experimental dissociation between VPT-1 and VPT-2).

Interestingly, a child's ability to appreciate that another's experience, knowledge and/or beliefs might be different to their own also develops around this time, linking VPT-2 to the more sophisticated 'Theory of Mind' (Premack & Woodruff, 1978) abilities that underpin human social interactions. Indeed, both dispositional (Erle & Topolinski, 2015; Kessler & Wang, 2012; Mattan et al., 2016) and neurocognitive (Schurz et al., 2013) evidence links VPT to successful higher-order socio-cognitive reasoning in adulthood. Moreover, there is evidence that VPT generates a shared reference frame (Erle & Topolinski, 2017; Hamilton et al., 2014), promotes a positive opinion of one's interaction partner (Erle et al., 2018) and enables one to ground (Clark & Brennan, 1991), predict (Creem-Regehr et al., 2013) and respond to (Freundlieb et al., 2016) their behaviour.

Several tasks have indicated that VPT does not only emerge when explicitly prompted, but also spontaneously when in the presence of another person. For example, the well-known 'Dot Perspective' task (Samson et al., 2010), showed that participants' judgements of their own perspective become slower and more error-prone, if another person in the scene would make the judgements differently from their perspective (e.g., if participants see 3 markers in a scene, but the person in the scene only sees 2). Taking this 'altercentric interference' (Kampis & Southgate, 2020) as a quantitative index of VPT, the authors argued that representations of others' perspective arise rapidly and effortlessly (for a critical assessment of this task, see Conway et al., 2017; Gardner &

* Corresponding author.
E-mail address: l.aldrich.24@abdn.ac.uk (L. Aldrich).

https://doi.org/10.1016/j.cognition.2026.106699
Received 19 December 2025; Received in revised form 22 August 2026; Accepted 27 August 2026
Available online 11 September 2026
0010-0277/© 2026 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Thorn, 2025). A large body of work now shows that similar interference happens when participants have to indicate whether a numeric character is a '6' or '9' in the presence of another person who would see the number differently (Surtees et al., 2013). Moreover, other studies have indicated that another's viewpoint can also facilitate perceptual reports when this other person's perspective affords a more direct viewpoint than the participant's own, suggesting that VPT makes use of perceptual representations that can stand in for one's own perceptual input (Ward et al., 2019, 2020, 2022).

Together, the available evidence from explicit perspective taking tasks in children and implicit measurement in adults suggests that humans are endowed with robust mechanisms for representing other people's perspectives. Recently, however, surprising findings from a series of studies (Samuel et al., 2021) appeared to challenge this conclusion, suggesting that people possess little to no insight regarding how others perceive the world. In this task, participants simply indicated how long two equally long lines drawn on a wall would visually appear to another obliquely positioned person in the scene (Fig. 1). In stark contrast to prior research, adult participants appeared to lack the simple understanding that the line closer to the other person would appear visually longer to this other person, given the fundamental relationship between object distance and retinal size (i.e., 'perceptual foreshortening'). This surprising finding was supported across multiple experiments, consistently showing that estimates for the size difference between the two lines from the other person's perspective clustered around zero (i.e., that the lines would appear equally long), and not different from chance performance. Given we expect 7–8 year olds to be capable of representing how the environment looks to another person, this is a striking finding. If substantiated, this would suggest that people have little insight into another person's perceptual experience, challenging a broad body of perspective-taking research and raising doubts about whether humans possess visual perspective-taking abilities at all. In Samuel et al.'s (2021) own words: "we found no evidence that participants could make judgements consistent with any visual perspective other than their own".

There are several reasons to believe that people's true VPT ability was likely obscured in Samuel et al.'s (2021) task. First, in several of the experiments, the stimuli lacked crucial three-dimensional depth cues (Fig. 1 top left displays). This absence of depth information likely made it difficult for participants to accurately understand whether the lines were on the same plane as the other person, tiny and close to the participant, or very large and far away. Furthermore, even in the stimuli that did feature depth cues (i.e., the textural convergence of the wooden floorboards, Fig. 1 top right display), the other person's position was not oblique enough to create a substantial distortion in perceived line lengths. In essence, therefore, the experiment required participants to make precise spatial inferences from spatially ambiguous stimuli, likely hindering the expression of VPT ability.

Second, participants used two unmarked sliders to rate the perceived length of the left and right lines as experienced by the other person in the scene, without a clear scale (see Fig. 1, bottom). This made it ambiguous whether the slider's position was meant to represent the ratio of line-lengths, the actual lengths experienced by the other person (scaled to the scene), or simply a binary choice of whether the closer or further line appeared longer. Additionally, the sliders were placed one above the other instead of directly under the to-be-judged lines and simply labelled as "left" and "right". This could have contributed to further confusion, especially given the widespread left-right confusion in even the healthy population (van der Ham et al., 2021).

Lastly and most importantly, asking participants to judge perceptual differences between lines of equal physical length is inherently confusing. Our visual system automatically performs size constancy adjustments, allowing us to perceive objects as keeping their true size regardless of distance. In other words, when we see two lines that are objectively identical, our brain implicitly corrects for distance and continues to 'recognise' them as the same size through preconscious processing (i.e., size constancy; Andrews, 1964; Gregory, 1997; Sperandio et al., 2015). These processes are fundamental components of our visual system and are the basis of several well-known visual illusions (Aglioti et al., 1995; Gregory, 1963; Ponzo, 1910). The question, "How long does each line actually appear from the person's visual perspective?", is therefore highly ambiguous. A socially adept perceiver would know that the other individual's perceptual system would make the same automatic size constancy adjustments as one's own. As a result, the lines would appear identical from that other person's visual perspective, just as they do from one's own. This wording therefore fails to distinguish whether participants should report how the lines are ultimately perceived to be by the other individual (i.e., the same length due to size constancy) or how they appear on the individual's retina before size constancy adjustments (i.e., the closer line would appear larger). Using the term "how long" adds to this confusion as length is a physical property, not a perceptual one. People know a ruler's physical length does not change with distance, but they may be aware its retinal size shrinks. Thus, referring to "length" might have biased participants towards giving a literal rather than a subjective response. A question that clearly establishes a retinal frame of reference would therefore have been needed to elicit any perspective taking skill.

Here, in two experiments, we re-evaluated whether people can represent another person's visual perspective, using the same line length judgement task. The goal of Experiment 1 was to test whether people are capable of making retinal size judgements from another person's perspective. We therefore compared a pseudo-replication of Samuel et al.'s (2021) single-trial line-length task with an "improved" version featuring a new scene stimulus (with cues to depth), a more intuitive response layout (with labelled sliders), and a clearer critical question. We predicted that, in contrast to Samuel et al.'s (2021) null findings, these changes would be sufficient to enable participants to make successful level-2 VPT judgements. In Experiment 2, we used a novel multi-trial adaptation of the line length judgement paradigm to investigate the precision with which participants were able to represent how another person's visual perspective changes with their position relative to the lines. By examining the accuracy of participants' VPT judgements we aim to shed light on whether simple heuristics can account for participants' perspective taking ability, or whether a more complex mechanism (such as perceptual simulation or geometric reasoning) might more parsimoniously account for current findings.

2. Experiment 1

Experiment 1 used a between-subjects design to test whether small changes to the original line-length judgement paradigm would allow participants to demonstrate successful VPT. One group of participants took part in a pseudo-replication of the original line-length judgement task (1a), following the same single-trial procedure as Samuel et al.'s (2021) original experiment, with the exception of a less ambiguous stimulus, with added depth cues (hence, "pseudo"; see Fig. 2A, top). A second group of participants was given a further improved version of the original task (1b), in which we adjusted the sliders' layout/labelling to make them more ergonomic and rephrased the question to more directly target retinal size (see, Fig. 2B). To avoid task ambiguity, we asked participants to imagine they were in the person's location and looking through their eyes, instead of just referring to their perspective. Also, to reduce participants' tendency to report the objective lengths (i.e., that the lines are the same physical length) and instead prompt them to consider how they subjectively appear to the other person in the scene, we asked them to indicate over how much of the other person's view each line would extend.

We predicted that these changes would allow participants to make more accurate VPT judgements. We pre-registered our methodology and analytical strategy on the AsPredicted platform (see, https://aspredicted.org/we3m7.pdf; created in May 2024).

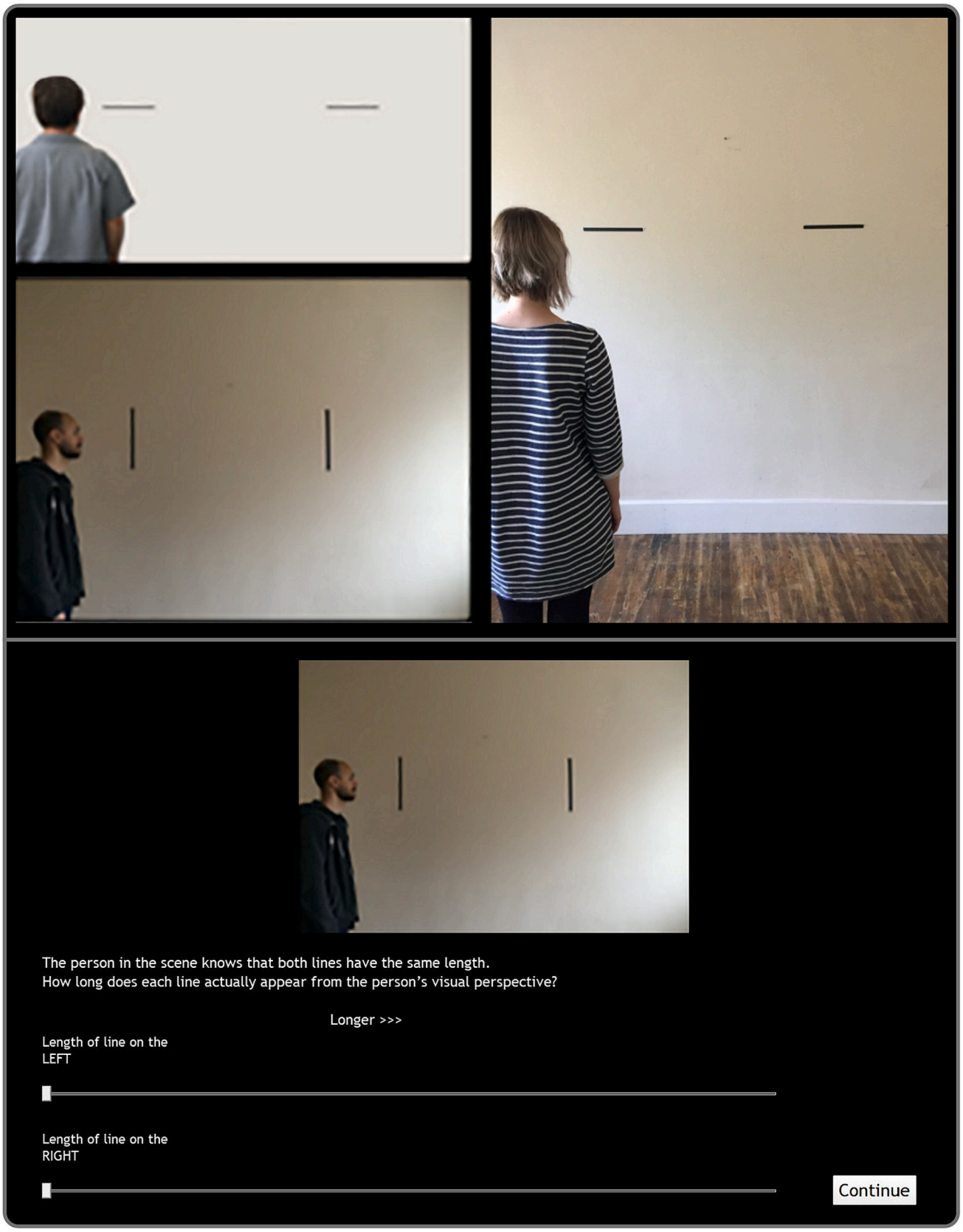


Fig. 1. Stimuli used in Samuel et al. (2021) and example trial. Top: Example scenes judged by participants in Samuel et al.'s (2021) line-length judgement paradigm. Bottom: Example trial in Samuel et al.'s (2021) line-length judgement paradigm.

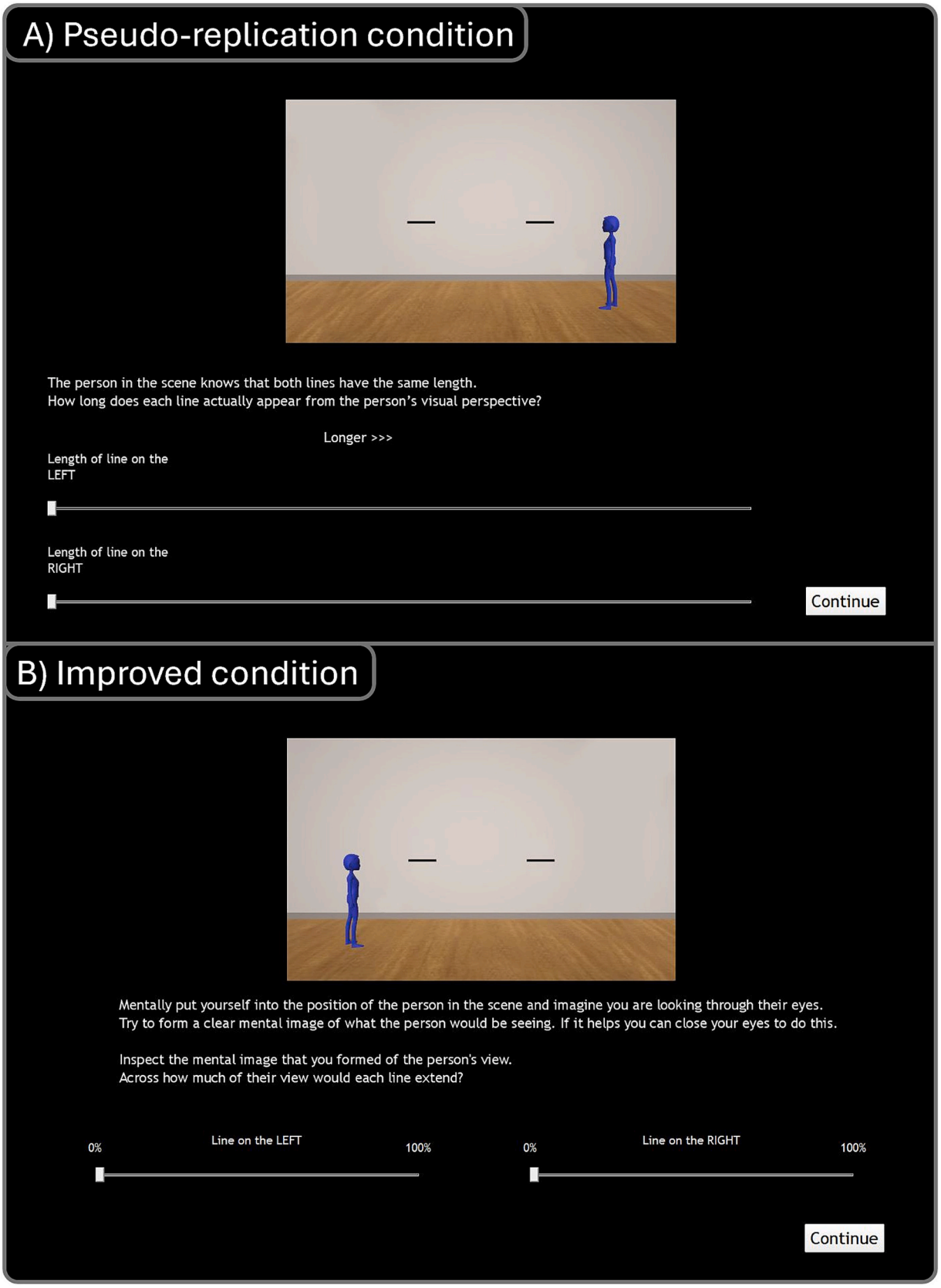


Fig. 2. Example Trials. Example trials in A) our pseudo-replication condition and B) our improved condition.

3. Method (Exp. 1)

3.1. Participants

Following Samuel et al. (2021), a power analysis indicated that 54 participants per experiment would provide 95% power to detect a medium effect ($d = 0.5$) using a two-tailed one-sample *t*-test ($\alpha = 0.05$). The equivalent sample size for a Wilcoxon signed-rank test was 57. Given Samuel et al.'s repeated null findings, 75 participants per group was deemed appropriate. Thus, for the pseudo-replication group, 76 adult (age range 19–50 years; $\bar{X} = 33.4$; $sd = 8$) native English speakers (33 women and 43 men) with self-reported normal (or corrected-to-normal) vision were recruited, and 75 adult (age range 19–50 years; $\bar{X} = 33.7$; $sd = 8.4$) native English speakers (36 women and 39 men) with self-reported normal (or corrected-to-normal) vision were recruited in improved version group. All participants were recruited via the online participant platform 'Testable' (www.testable.org; accessed June 2025) and reimbursed £1.50 for their time. Ethical approval for the experiments was obtained from the University School of Psychology Ethics Committee. Participants gave written consent prior to experimentation and were debriefed upon its completion.

3.2. Apparatus

Both versions of Experiment 1 were developed in Inquisit and were accessed via a web-link hosted on the Millisecond platform (www.millisecond.com; version 6; accessed June 2025). Both experiments were administered via participants' own hardware. Participation over smartphone was not permitted.

3.3. Stimuli

The same stimulus was used for the Pseudo-replication (Fig. 2A) and the Improved version of the experiment (Fig. 2B), closely mirroring Samuel et al.'s stimulus (Samuel et al., 2021), but with the addition of depth cues. The scene was created using a snapshot of a 3D model, depicting a person looking at a wall on which two lines were drawn (created in "Paint 3D"; www.microsoft.com; Windows 10; accessed September 2024). The target person (blue) was positioned so that the line drawn from their head to the midpoint between the two lines formed a 34° angle with the plane of the wall. This person was positioned 35% closer to the midpoint between the lines than the participant's vantage point within the scene (which was perpendicular to the wall, their view centred on the midpoint between the lines), to create an oblique viewing angle that would moderately distort appearance of the lines' relative lengths. The scene also featured a wooden floor, which provided textural cues to depth. Shadows were excluded from the stimuli to mirror Samuel et al.'s (2021) scenes and to reduce potential biases stemming from size constancy effects (Aglioti et al., 1995), given the task involved estimating relative lengths.

Four versions of this stimulus were created, so that the target person appeared either to the left or the right of the participant, and the lines were either horizontal (see Samuel et al., 2021, experiments 1–9 & 11) or vertical (see Samuel et al., 2021, experiment 10). The 'person left' and 'person right' counterbalanced conditions were created by mirroring the scene stimuli and the different line orientation was added to the scene using Adobe Photoshop (www.photoshop.adobe.com; Version 25.0; accessed September 2024), taking care to preserve an equivalent line length.

3.4. Design and procedure

Participants were randomly allocated to either the pseudo-replication group or improved version group, in a single-trial between-subjects design. As in Samuel et al.'s (2021) original paradigm, our dependent variable was the difference between the judged length of the lines closer to and further from the agent in the scene. Within each group, participants were randomly shown one of the four possible scenes, to control for potential effects of agent location (either standing on the left or right) and line orientation (either vertical or horizontal).

Both versions of the experiment began with a brief description of what to expect before informed consent was obtained. Participants then completed a short demographic survey (obtaining age, sex, and handedness data), after which they saw an instruction slide, featuring the scene stimulus (depicting an agent stood looking at two lines of equal length drawn on a wall) and the following written information:

> *"In this experiment you will see the scene above showing a person standing next to a wall. Two lines of the same length have been drawn onto the wall. In this experiment we will ask you a question about how the person sees the lines on the wall. When thinking about your answer, imagine you are stood where the person is stood, and you are seeing the lines through their eyes"*.

This description was an addition to Samuel et al.'s (2021) original paradigm, who only specified that participants should "pay attention to a picture that would be displayed for a few seconds as they would be asked two questions about it afterwards". In both versions of the experiment, and in line with Samuel et al., participants appraised only a single scene (one of the four counterbalanced versions) in a single trial, making a single VPT judgement. Responses were not time-constrained, and the scene remained visible while participants made their response. Following the instruction slide, the two experiments then deviated in the following manner:

3.4.1. Pseudo-replication of Samuel et al. (2021)

Participants in the pseudo-replication group were presented with the same question phrasing and response method as in Samuel et al.'s the original experiment, but with our new scene stimulus (Fig. 2A). The instruction matched the original study: "The person in the scene knows that both lines have the same length," followed by the question, "How long does each line appear from the person's visual perspective?" Responses were given in participants' own time, using two unmarked sliders (each 65% of the screen in length, ranging between 0 and 100) to indicate the length of each line as they believed it would appear from the agent's visual perspective. These sliders were positioned one above the other, with the top slider labelled 'Length of line on LEFT' and the bottom labelled 'Length of line on RIGHT'. As in Samuel et al.'s (2021) original experiment the sliders were not anchored at their ends.

3.4.2. Improved question and response sliders task

In the improved version of the task (Fig. 2B), we altered Samuel et al.'s (2021) original question to avoid referencing the objective line lengths in the scene. Hence, participants were first told *"Mentally put yourself into the position of the person in the scene and imagine you are looking through their eyes. Try to form a clear mental image of what the person would be seeing. If it helps you can close your eyes to do this."* They were then instructed to *"Inspect the mental image that you formed of the person's view. Across how much of their view would each line extend?"*. We reasoned that this question would make it clear that we were interested in the other person's perception prior to size constancy adjustments. Like in the pseudo-replication condition, participants gave their responses using two feedback sliders. However, in this version the sliders were positioned side by side, under the lines to which they referred (each 30% of the screen in length), rather than being placed one above the other. Percentage labels were also added to each slider's end (0% and 100%).

3.4.3. Participant feedback

Following completion of the task, participants in both groups filled out a short feedback questionnaire. They were asked three questions: *"How clearly could you visualise the person's perspective?", "How well did you understand what was being asked of you in this experiment? Was the question clear?",* and *"In general, how easy do you find it to visualise*

imagined things, be it scenes, people, or objects?". For each question, participants could use a slider (each labelled 0% - 100%) to indicate their answer. We also encouraged participants to elaborate on their reasoning via use of a free response box.

3.5. Data diagnostics

As in Samuel et al.'s (2021) analysis, participants who did not move one of the sliders from the starting zero position were excluded from analysis (13 in the pseudo-replication group and 14 in the improved task group).

3.6. Analytic strategy

All analyses were conducted in RStudio (www.posit.co; version 4.4.0, accessed May 2025). Raw data and R scripts are available on the OSF.

For analysis, we computed a normalised difference between each participant's line length estimates from the other person's perspective as follows:

$$\frac{(\text{LengthEstimate_CloserLine} - \text{LengthEstimate_FurtherLine})}{(\text{LengthEstimate_CloserLine} + \text{LengthEstimate_FurtherLine})/2},$$

This produced a final score for each participant scaled by the overall magnitude of their initial estimates. In line with Samuel et al.'s (2021) original criteria, a positive score would be indicative of awareness that the line closer to the other person will appear longer to them (i.e., successful VPT).

We tested whether scores differed from zero using a one-sample *t*-test to determine whether participants could report the perspective of the person in the scene above chance level. Group differences between the replication of Samuel et al.'s (2021) original paradigm and the improved version were assessed using a two-sample *t*-test, or a Wilcoxon rank-sum test when data were non-normal. We also replicated the analyses using the measure of Samuel et al., which produce identical results (Reported in the Supplementary Material).

Finally, to assess whether participants possessed insight into their own VPT ability, Pearson correlations were calculated by group (i.e., for participants in the pseudo-replication task and for participants in our improved manipulation task) between each participant's normalised difference scores and their responses to the three self-report feedback questions. If meta-awareness of VPT ability was present, we would expect participants with lower difference scores (inaccurate VPT) to report worse comprehension of the task and/or ability to visualise than those with higher difference scores (accurate VPT).

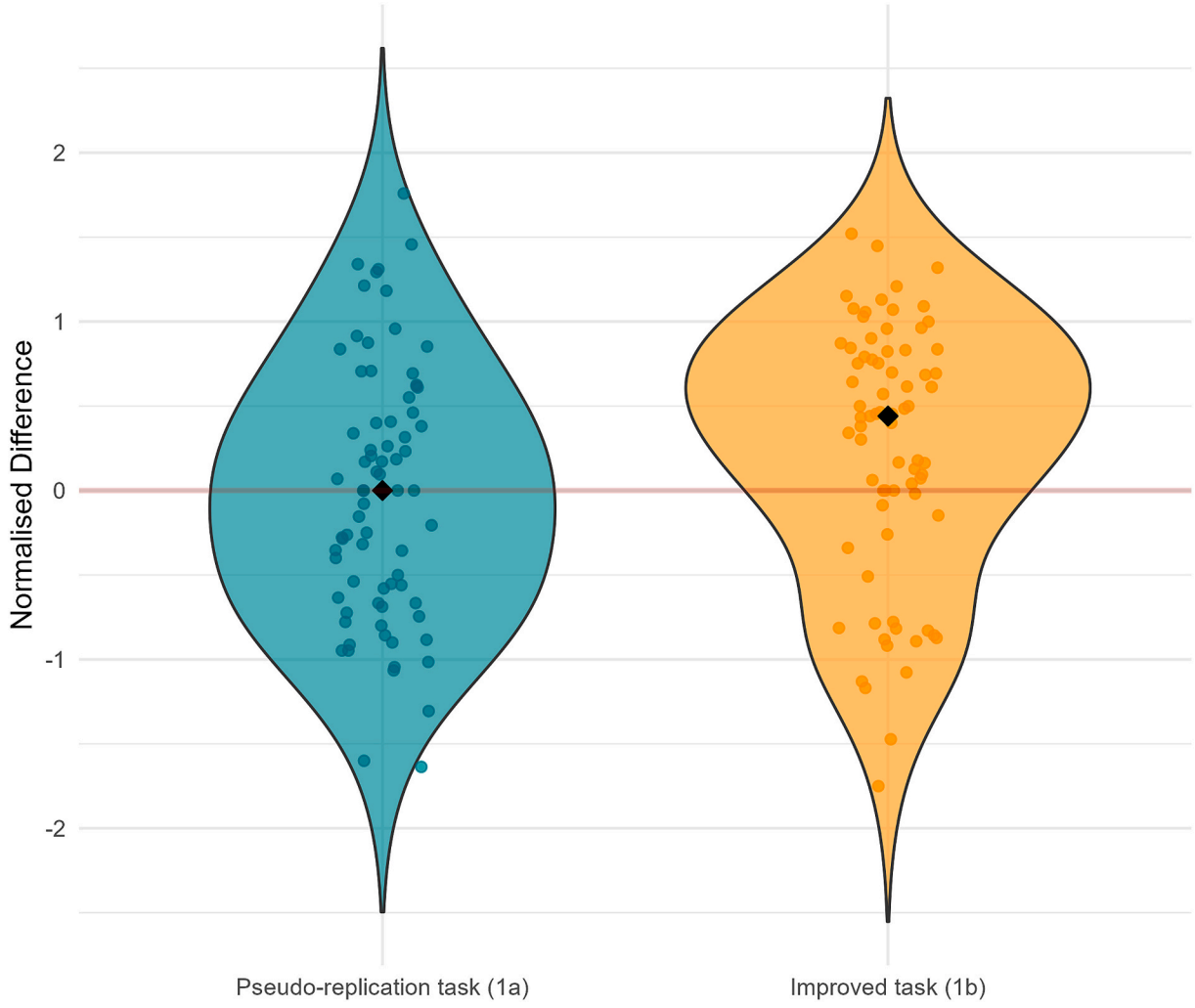


Fig. 3. Experiment 1 Violin plots. Violin plots depicting the distribution of normalised difference scores (i.e., the difference between estimates of the line closer to, resp. further from the avatar, divided by the mean of those estimates) in both the pseudo-replication (1a) and improved task (1b) groups (red line indicative of 'no perceived difference' between the line lengths). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

4. Results (Exp. 1)

4.1. Line judgement task

Means and distributions of normalised difference scores are depicted in Fig. 3. In the Pseudo-replication group, results replicated Samuel et al.'s (2021) original findings. Participants' estimates ($\bar{X} = -0.03$, $SEM = 0.09$, 95% CI [−0.20, 0.15]) were normally distributed ($W = 0.987$, $p = .61$) and did not significantly deviate from zero, $t(75) = -0.293$, $p = .771$, $d = -0.03$. Participants were thus equally likely to report that, from the perspective of the person in the scene, the further line and closer line appeared larger.

In contrast and as hypothesised, participants in the improved task group showed VPT performance above chance. Scores were positively skewed ($W = 0.941$, $p < .01$), indicating that participants tended to judge the line closer to the target person as extending further across their visual field ($Mdn = 0.44$, $IQR = 0.94$, $V = 1781$, $p = .009$, $r = 0.30$). Despite the appearance of multimodality, the distribution is best characterised as unimodal (Hartigan's dip test $D = 0.051$, $p = .153$) with substantial variability and a negatively skewed tail (See Supplementary Materials A: 'Experiment 1b bimodality investigation'). Comparison between the two groups indicated that scores in the improved design group (1b) were significantly higher than scores in the pseudo replication group (1a), $V = 2213$, $p = .018$, $r = 0.19$. Thus, while responses from participants in our pseudo-replication of Samuel et al.'s (2021) original experiment did not differ statistically from zero, the improved question and more ergonomic slider placement in our improved task enabled participants to accurately report the perspective of the other person in the scene, such that a majority (~71%) were aware the line closer to the person would appear larger to them.

Analysis using Samuel et al.'s (2021) original strategy fully replicated this pattern of findings (see Supplemental Material B: 'Experiment 1 Replication of Samuel et al.'s (2021) Analysis Method').

4.2. Feedback questions

We observed significant between-group differences in the distribution of responses to two of our three feedback questions. We conducted three Wilcoxon rank-sum tests due to non-normal distributions for all feedback questions, with effect sizes reported as rank-biserial correlations (r) after correcting for multiple comparisons (Bonferroni). When asked "How clearly could you visualise the person's perspective?", participants in the pseudo-replication group ($Mdn = 92$, $IQR = 24.3$) ranked themselves significantly higher than in our improved design group ($Mdn = 75$ $IQR = 34$), $V = 3604$, $p < .01$, $r = 0.23$. Similarly, when asked "How well did you understand what was being asked of you in this experiment? Was the question clear?", participants in the pseudo-replication group ($Mdn = 78.5$, $IQR = 34.5$) ranked themselves significantly higher than in our manipulated task ($Mdn = 66$, $IQR = 30.5$), $V = 3689$, $p < .01$, $r = 0.25$. Responses to the final question, "In general, how easy do you find it to visualise imagined things, be it scenes, people, or objects?" followed the same pattern between pseudo-replication task ($Mdn = 76.5$, $IQR = 36.25$) and our manipulated task ($Mdn = 71$, $IQR = 29.5$), but was not significant, $V = 3169$, $p = .235$.

In order to evaluate whether the higher scores for the group who objectively had worse VPT reflected general metacognitive failure specific to the pseudo-replication condition, we correlated the normalised difference scores in the line judgement task and the three self-report measures. However, neither group showed correlations between normalised difference scores and the three self-report measures (for all $p > .11$), indicating that neither self-reported vividness with which participant's introspectively experienced the other person's perspective, nor their reported comprehension of the question, predicted accuracy in either group.

Together, participants' subjective task data show that, compared to the improved version, the original task yielded higher overall confidence in their visualisation and comprehension abilities, despite objectively reduced task success. Within each group, none of the three feedback questions indexed diagnostic information on successful task performance in terms of metacognitive acuity. Note however that linking individual differences measures to behavioural task performance in this manner is likely underpowered, especially given the lack of precision of single-trial measures and the known issue of large sample sizes needed to link robust behavioural measures to individual differences (i.e., "the reliability paradox"; see, Hedge et al., 2018).

4.3. Exploratory analyses

We conducted an exploratory analysis testing for potential effects of avatar location (avatar left vs. avatar right), line orientation (vertical vs. horizontal), handedness (left-handed vs. right-handed), and sex (male vs. female) on VPT performance, using two mixed-effects models (one for each group) with random intercepts. Neither a linear model nor a logistic model (with accurate VPT defined as difference scores >0) identified any significant predictors (all $p > .10$). Thus, we could not detect an influence of age, sex, handedness, avatar location, on VPT performance in either task.

5. Discussion (Exp. 1)

Experiment 1 re-evaluated Samuel et al.'s (2021) line length judgement paradigm. The pseudo-replication group (1a) reproduced the original null finding, with participants unable to identify above chance which line would appear larger to the other person. In contrast, our improved task (1b) featuring a revised question and better-positioned labels, elicited robust VPT, with participants performing above chance.

The current findings challenge Samuel et al.'s (2021) conclusion that people cannot represent another's visual perspective. They show that inability to successfully report another's view of the two lines did not reflect inherent problems with perspective taking, but issues of the experimental design, specifically the ambiguity in the original question phrasing and/or slider layout. In particular, participants may have defaulted to size-constancy assumptions that will apply to others as they apply to themselves (i.e., "the lines will appear as the same length also to the other person"), producing responses centred on zero. Feedback data support this, as the pseudo-replication group reported higher confidence in their ability than our improved task group, despite objectively worse performance. This suggests they adopted a simple but incorrect strategy, whereas reduced confidence in the improved task likely reflects the greater difficulty of reporting retinal (rather than objective) size.

Please note that Experiment 1 was not designed to isolate which specific modification of the original task drove the improved performance. Instead, it was designed as a demonstration that people are generally able to derive how the two lines appear from the other person's perspective, discounting size constancy adjustments. Our suspicion is the disambiguation of reference frame via the question phrasing – understanding the task is about retinal rather than size-constancy adjusted line lengths – is the critical factor. Identification of the necessary and sufficient conditions to ensure this understanding of the task is an important target for future designs that orthogonally manipulate the different sources of ambiguity in the original task.

A minority of participants showed negative scores, reporting the further line as larger, even in our improved task. Although there was no evidence of bimodality in the improved task group (see Supplementary Materials A), it is possible that the observed negative difference scores in some participants may reflect systematic, not random, error. One possibility is that these instances reflect participants overapplying their own egocentric experience of the stimuli (i.e., that the lines are of equal length) when reasoning about the other person's perspective. As a consequence, they may attempt to "correct" the display so that it realises this experience also for the other person, by extending the line that appears further away, thereby producing negative difference scores.

Indeed, this is what was reported anecdotally by some participants. If so, they would therefore fail at the task, not because of representational deficit, but due to incorrectly anchoring on their own egocentric reference frame. As noted however, scores of these participants did not fall statistically outside the expected distribution around the mean and should therefore be interpreted with caution.

6. Experiment 2

Experiment 1 demonstrated that participants can report which line should appear longer to the other person provided they are questioned clearly. However, it remains an open question as to how accurately the other person's perspective is represented. Is it that people just know as a rule (Heyes, 2014) that closer objects will be larger in retinotopic space to another person? Or is the representation more precise, containing information as to the degree of perceived difference between the two lines? We know that perceptual foreshortening is exacerbated by increased obliqueness of the viewing angle towards the other person. Thus, instead of just knowing categorically that objects closer to the person will appear larger (as measured in Experiment 1), people could possess a more accurate representation that captures the degree to which the lines differ depending on the other person's location (Fig. 4).

Experiment 2 investigated whether people possessed this ability for precisely estimating *how* different the two lines would appear the other person. Participants were again tasked with indicating which potential viewpoint the person in the scene would experience given their location relative to the lines. They viewed similar stimuli as in Experiment 1, but now the location of the other agent in the scenes differed across multiple trials. Moreover, instead of sliders, participants were able to select the other person's viewpoint from one of 23 possible visual representations of the lines from different viewing perspectives, each representing a different angle to the wall, resulting in a variety of relative perceived lengths. If participants can accurately represent how retinal sizes change across different viewing locations, this would provide key evidence for sophisticated visual perspective taking abilities.

7. Method (Exp. 2)

7.1. Participants

Given the shift from a single trial to a multi-trial experimental procedure, we estimated a similar effect size to that of experiment one, and planned testing 50 participants pre-exclusion. After excluding 5 participants for not selecting anything but the starting potential viewpoint, 45 adult (age range 19–50; $\bar{X} = 28.9$; $sd = 10$) native English speakers (13 women, 31 men and 1 non-binary person) with self-reported normal (or corrected-to-normal) vision were recruited, providing 90% power to detect effect sizes of $d = 0.49$). Participants were recruited via the online participant platform 'Testable' (www.testable.org; accessed June 2025) and reimbursed £2. Ethical approval was obtained from the University School of Psychology Ethics Committee. Participants gave written consent prior to experimentation and were debriefed upon its completion.

7.2. Apparatus

Experiment 2 was delivered via Inquisit, accessed via a web-link hosted on the Millisecond platform (www.millisecond.com; version 6; accessed June 2025), administered via participants' own hardware.

7.3. Stimuli

Experiment 2 asked participants to make multiple judgements across

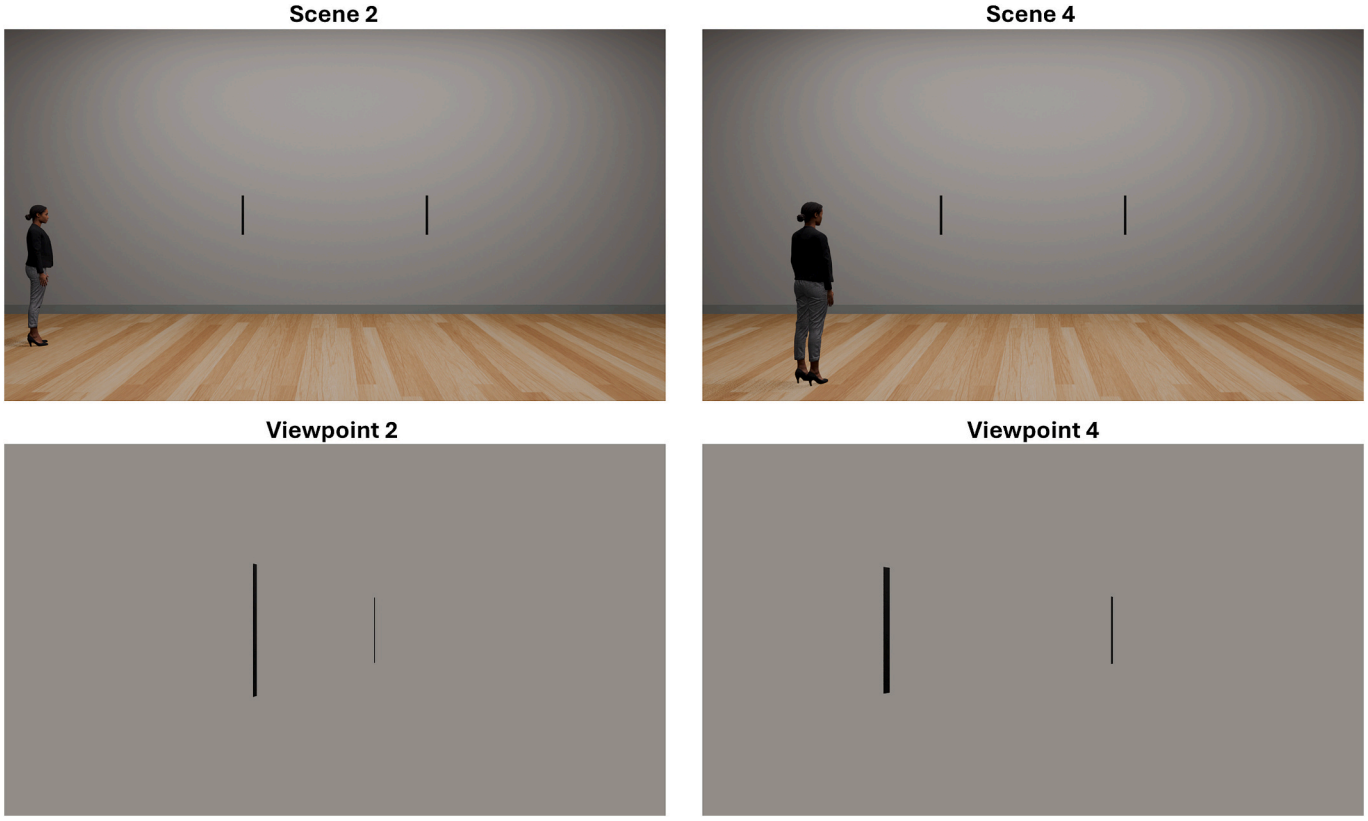


Fig. 4. Example of line length distortion with location. Example of how the relative length of the lines appears distorted at different avatar locations, illustrating how a simplistic heuristic (i.e., 'the line on the left will appear longer if the person is on the left') cannot capture differences that a more accurate representation might (i.e., that the relative difference between the length of the perceived lines is smaller in scene 8 than scene 4).

multiple scenes. To create these scenes, we generated 10 scenes and 23 possible viewpoints using a 3D environment built in Blender (www.blender.org; version 4.2, accessed November 2024). Each scene showed a female avatar standing on a wooden floor, facing two horizontal lines on a wall, which was shown straight on, at a 90° angle to the participant's viewpoint. The avatar could appear in one of 10 positions along an arc in front of the wall. In every position, the avatar stood the same distance from the midpoint of the two lines (approximately 4 m assuming the woman to be 165 cm tall). These 10 positions thus formed a semicircular layout around the lines (see Fig. 5, top panel) and were labelled from left to right as positions 1–10. The avatar's viewing angles towards the centre of the two lines were 15°, 30°, 45°, 60°, 75°, 105°, 120°, 135°, 150°, and 165°, respectively.

Participants could cycle between 23 potential viewpoints using the UP and DOWN arrow keys to indicate the other person's viewpoint in each scene. These viewpoints were taken from the same 3D model such that one viewpoint always perfectly matched the avatar's viewpoint in the scene, the others showing viewpoints between those, one of them being the participant's own perspective. All viewpoints were placed at the same fixed distance from the lines and thus the same arc, starting at 7.5° and increasing in 7.5° increments up to 172.5°.

7.4. Design and procedure

Each participant completed 40 trials over two blocks, with a short break in between. In each block, participants were shown each of the 10 scenes twice, in a randomised order. The dependent variable was the visual angle between the participant's selected viewpoint and the avatar's actual location in the scene.

Experiment 2 began with a brief description of the task and what would be expected of participants, before a short demographic survey (obtaining age, sex, and handedness data) before being shown the initial

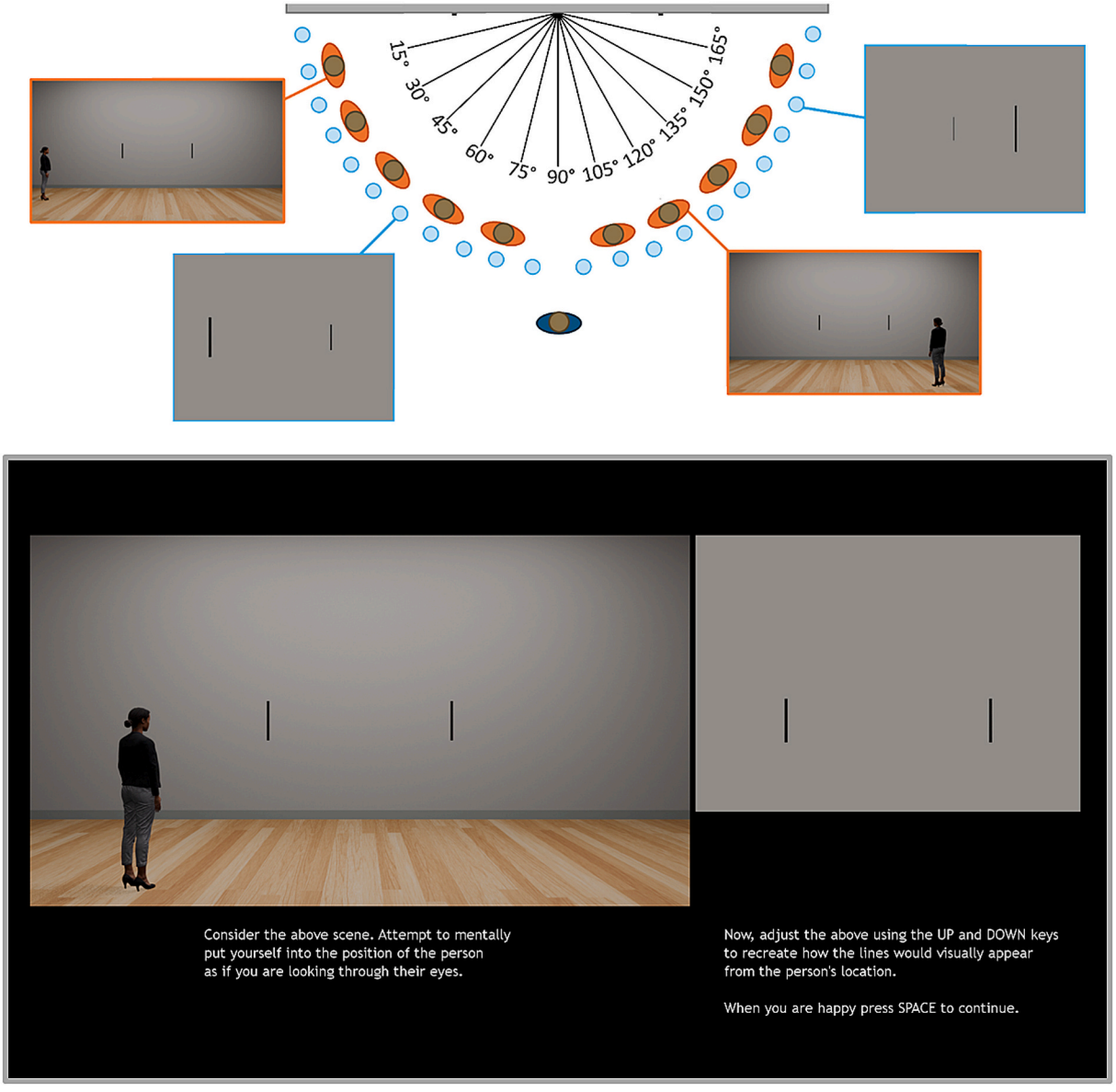


Fig. 5. Experiment 2 Stimuli and trial setup. Top panel: a plan view diagram depicting the 10 potential locations of the avatar (orange figure) and the 23 potential viewpoints participants could select from (light blue circles). The blue figure represents the location from which the participant observes the scene, with the wall and lines (grey rectangle and black squares) depicted at the top of the image. Bottom panel: Example trial in our multi-trial novel adaptation of Samuel et al.'s (2021) line-length judgement paradigm, with participants responding by rotating the right image using the up and down arrows. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

instruction slide featuring scene 1 and the following written information:

> *"In this experiment you will see multiple scenes like the one above. Each scene will feature a wall upon which two lines of the same length have been drawn. Each scene will also feature a person, shown standing next to the wall, who is looking at the lines. The person's location will vary over the course of the experiment.*
> *Your task is to estimate how the lines visually appear to the person in each scene. When thinking about your answer, imagine you are stood where the person is stood and you're seeing the lines through their eyes. Try and be as precise as possible in each of your estimations."*

After pressing continue, participants were then shown the first experimental trial (Fig. 5, bottom panel). This first scene was always the same as the scene displayed on the instruction slide; all 39 scenes that followed were randomised. Each experimental trial slide featured the to be judged scene (left) alongside a potential viewpoint (right). Initially, it showed viewpoint 12 (depicting the lines as the same length and matching the participant's own view). Participants could cycle between possible views using the UP and DOWN keys. Under the scene were displayed the following instructions: *"Consider the above scene. Attempt to mentally put yourself into the position of the person as if you are looking through their eyes."*

Participants were not timed when cycling through the 23 potential viewpoints and could go back and forwards between viewpoints as many times as they wished. When they were happy with their selection, pressing space confirmed their estimation of the avatar's view and advanced to the next trial.

Upon completion of the 40 trials, participants completed a short questionnaire. They were asked the three following questions, to which they were able to give an open typed response: *"How did you estimate the other person's view? Did you visualise what they would see, or use a different technique?"*, *"Did you experience any issues during the experiment, or feel something was unclear?"*, and *"What do you think we're trying to find out by conducting this experiment?"*

They were also asked *"Over the course of the experiment, did you visualise the other persons view more at the start or end?"* To which they could respond using a slider (labelled left to right: *'More at the beginning', 'equally across the experiment', and 'more at the end'*).

7.5. Data diagnostics

Similar to Samuel et al.'s (2021) analysis, participants who did not adjust the potential viewpoint from the starting position over the course of the experiment were considered to have not understood the task instructions and were removed from analysis (5 individuals).

7.6. Analytic strategy

All analyses were conducted in R (version 4.4.0, 2025). All data and analysis scripts are available on the OSF, including additional analyses of participant's responses to the feedback questions (see Supplementary Materials C: 'Experiment 2 analysis of feedback questions').

First, to quantify whether and how accurately participants represented the avatar's visual perspective, we ran a simple linear regression for each participant separately. In each, the selected viewpoint was regressed onto the correct viewpoint for each scene (i.e., the true avatar viewing angle; see Fig. 5 top panel), after centring both variables around zero (perpendicular to the wall, where the lines appear the same length). In this way, the β coefficient for each participant would depict any systematic tendency to over- or underestimate how the difference between the line lengths would vary with changes to the other's location (i.e., $B = 1$ indicates perfect tracking of the other's perception, whereas $B = 5$ represents a tendency to over-emphasise the view of the avatar by 5°, and a B of 0.2 (1/5) represents a tendency to underestimate the view of the avatar, meaning the avatar has to shift 5° before the participant indicates a 1° shift).

Second, in a similar fashion to the Wilcoxon signed-rank and one-sample t-tests against 0 in Experiment 1, we tested participants' individual B coefficients against 0 to evaluate whether overall participants track the avatar's perspective.

Third, to explore potential subgroups of VPT strategies, we applied k-means clustering to participants' viewpoint selections. Participant responses were first averaged across the four trials for each scene, and the data were then standardised (z-scored across scenes) before being submitted to a k-means cluster analysis. The optimal number of clusters was determined using the elbow method. Results (Exp. 2).

8. Results (Exp. 2)

8.1. Per-participant linear regressions

Overall, Wilcoxon signed-rank test indicated that participant-level regression slopes were significantly greater than zero, with a large effect size, $V = 976$, $p < .001$, $r = 0.89$. Thus, replicating Experiment 1b, participants were generally able to indicate which line was visually larger from the target person's viewpoint.

When investigating individual performance, 32 of 45 participants (71%) showed significant positive regression slopes, indicating they were able to track the relationship between retinal line length and distance and would have been considered capable of VPT under Samuel et al.'s (2021) original criteria. More specifically, using Cohen's (1988) cut-offs for regression R^2 *(0.02 = small, 0.13 = medium, 0.26 = large)*, 30 participants (67%) showed a large effect, with the model explaining more than 26% of the variance. An additional 7 participants (16%) demonstrated medium or small effects. Only 2 participants showed no effect, and 5 participants exhibited effects in the opposite direction, ranging from small to large.

8.2. Cluster analysis

To further analyse and qualify individual response patterns, we conducted a cluster analysis. Using the elbow method, a four-group solution of sizes 12, 8, 21 and 4 was deemed most appropriate. The four clusters are represented in different colours in Fig. 6, and their characteristics described below.

8.2.1. Cluster 1 – strong visual perspective takers

The largest of the four clusters constituted 21 participants (~47% of our sample, Fig. 6, red). Within this cluster, average regression slopes were positive ($\bar{B} = 0.86$, $sd = 0.15$), with an average model fit of $X(R^2) = 0.88$ ($sd = 0.1$), and a residual standard error $\bar{RSE} = 1.97$ ($sd = 0.76$). This cluster of participants showed close-to-perfect match between avatar locations and potential viewpoints, characterised by slopes approaching 1 and low residual error. These participants therefore did not only appreciate that the line closer to the other person will appear longer on their retinal image but also accurately indicated the precise viewpoint that corresponded to the person's location, revealing a near-perfect perspective taking ability as measured by this task.

8.2.2. Cluster 2 – weak visual perspective takers

Cluster two represents what we classify as 'weak' visual perspective takers (Fig. 6, green). This cluster constitutes 12 participants (~26% of our sample), who show some sensitivity to the relationship between retinal size and distance with respect to another's perspective. On average, these participants' regression slopes were lower than those in Cluster 1, but still positive ($\bar{B} = 0.22$, $sd = 0.11$), with an average model fit $X(R^2) = 0.44$ ($sd = 0.35$) and residual standard error $\bar{RSE} = 2.09$ ($sd = 1.55$). These shallower slopes suggest that this cluster underestimated perceptual foreshortening at more extreme viewing angles. Hence, whilst they might appreciate that, for example, a person on the left will perceive the left line as being larger than the right line, they

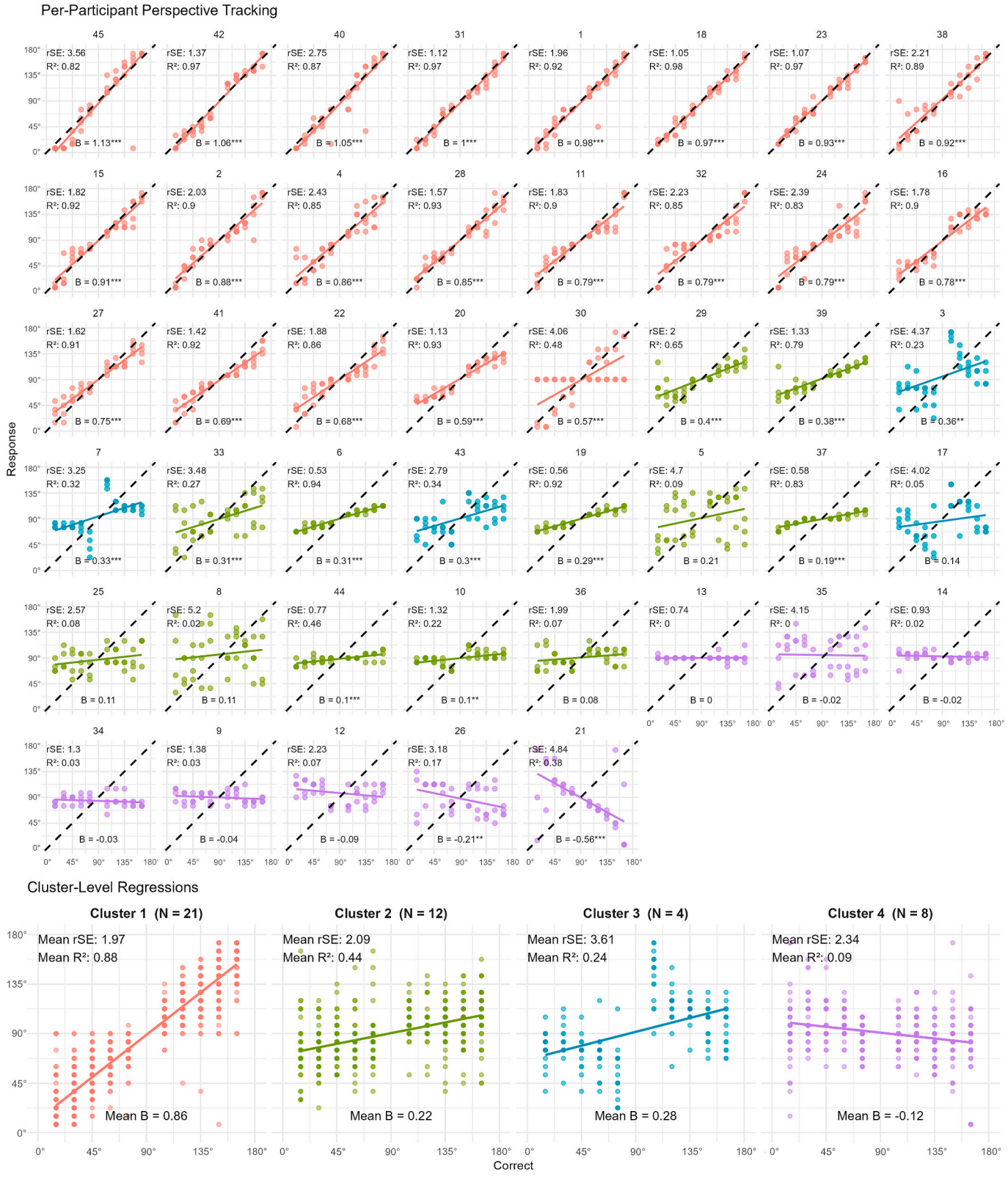


Fig. 6. Experiment 2 Scatterplots. Top: Scatterplots showing linear regressions between participants' average selected viewpoints (y-axis) and the avatar's actual viewpoints (x-axis) across 40 trials, coloured by cluster. Panels are arranged by regression slope. The dashed black line indicates the ideal 1:1 relationship. Bottom: Cluster-level regressions showing the same relationship, coloured by cluster.

underestimated the extent of the visual differences that would be induced. Nevertheless, these individuals would be classed as successful visual perspective takers under Samuel et al.'s (2021) criteria, given they clearly appreciate the relationship between retinal size and distance.

8.2.3. Cluster 3 – categorical visual perspective takers

Cluster three contained only four participants (~9% of our sample; Fig. 6, blue). On average, regression slopes in this cluster were positive ($\bar{B} = 0.28$, $sd = 0.08$), with an average model fit of $X(R^2) = 0.24$ ($sd = 0.12$) and residual standard error $\bar{RSE} = 3.61$ ($sd = 0.63$). While these individuals correctly indicated that the line closer to the avatar should appear longer to them, their estimates of perceptual distortion relative to the person's obliqueness was inverted. At extreme angles (i.e., when the person was depicted very close to the wall) these individuals reported lower amounts of perceptual distortion compared to lower angles (i.e., when the avatar was depicted closer to the participant's location). We suspect that these participants may have been working purely from the simple rule 'the line closer to the person will appear longer to them', while disregarding the specific amount of foreshortening the changing viewing location would induce. Their responses (though unorthodox) would be classified as successful perspective taking under Samuel et al.'s (2021) criteria.

8.2.4. Cluster 4 – non-perspective takers

Finally, 8 individuals in cluster four (~18% of our sample) do not show any understanding about the avatar's visual experience of the lines (i.e., the purple themed plots in Fig. 6). On average, participants' regression slopes were negative ($\bar{B} = -0.12$, $sd = 0.18$), with an average model fit of $X(R^2) = 0.09$ ($sd = 0.12$) and residual standard error $\bar{RSE} = 2.34$ ($sd = 1.45$). We observe a combination of negative and non-correlations here, suggesting participants were either responding randomly or with the incorrect intuition that objects further away from another individual should appear larger on their retinal image than closer objects.

8.3. Response behaviour

We also analysed participants' behaviour across trials when cycling through the potential viewpoints. The following measures were extracted using participant's 'step counts' (shifts from one potential view to another in the response) in each trial:

- Error: The number of steps away from the correct viewpoint the participant's final selection was, on a given trial.
- Inefficiency: The number of steps made minus the minimum required to reach the correct viewpoint on a given trial (hence, positive values show additional steps and negative values show a lack of required steps).

To gain insight into changes in cluster level behaviour across trials of the experiment, cluster-level mean trajectories were computed for each measure across trials (see Fig. 7). These measures show participants in the Strong VPT cluster were able to accurately identify the other person's visual perspective from the very first trial. At the start of the experiment, those in the Strong VPT cluster also explored more potential viewpoints than members of the other cluster, before becoming the most efficient group after trial 5. The negative efficiency scores for the other three clusters suggests that, in contrast to the Strong VPT cluster, these participants were not making enough button presses (and exploring enough potential viewpoints) to reach the correct answer, in line with the well-known performance problems of online participants (Uittenhove et al., 2023). As also seen in Fig. 7 (left), these participants therefore consistently underestimated the true difference in line lengths.

9. Discussion (Exp. 2)

Experiment 2 provided a novel test of the precision with which people can represent another person's visual experience. The findings challenge Samuel et al.'s (2021) claim that people are unable to simulate how a scene appears from another's viewpoint, instead revealing a highly precise ability to derive the visual experience of another person. The majority of participants (82%) correctly identified that the line nearer to the other person would appear longer, confirming the findings of Experiment 1.

Cluster analyses further specified this pattern. A large cluster of participants (47%) were able to match the avatar's viewpoint and tracked perceptual foreshortening with high precision (within approximately 18° of angle). These results reveal an ability to form a precise

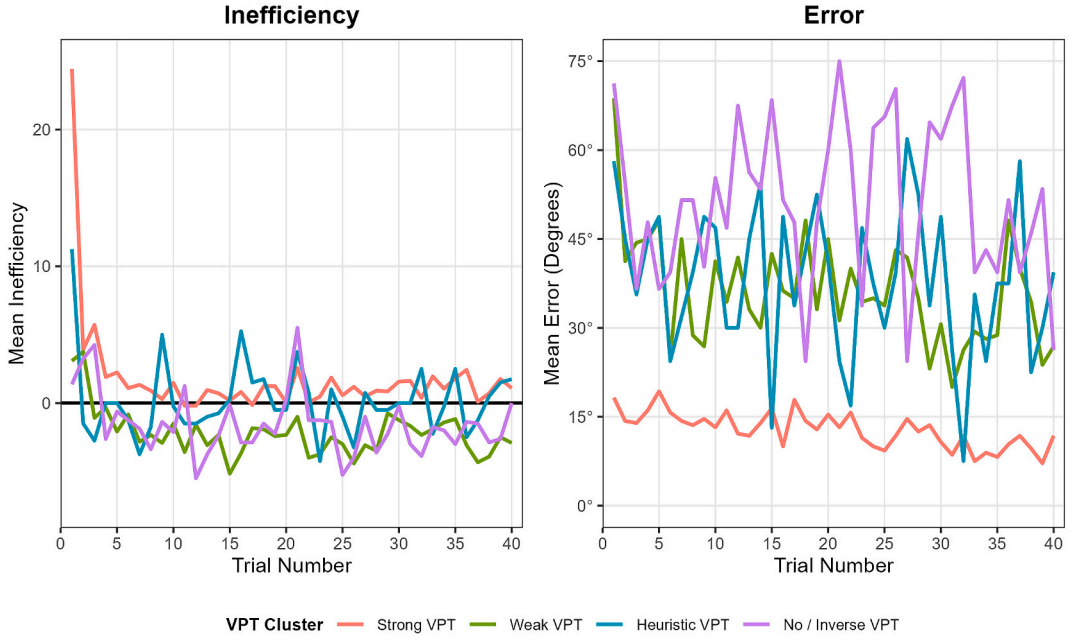


Fig. 7. Response behaviour by VPT cluster across the experiment.
Note. The cluster-level average Inefficiency and Error across the 40 trials.

representation of another person's visual perspective that is difficult to explain with simple categorical rules (e.g., "closer means larger"). In contrast, the "weak VPT" group (~26%) showed consistent underestimation of line-length disparities, suggesting egocentric anchoring, and a small "categorical VPT" group (~9%) appeared to rely on simplified rules ("closer means larger"), which precluded accurate matching of viewpoints to person locations. Inverse VPT was rare, and slopes rarely exceeded 1, consistent with a persistent egocentric influence on participant's estimates, wherein participants' own view biases their estimates of the other's experience (Keysar et al., 2000).

Strikingly, qualitative inspection of performance across trials suggests that strong perspective taker's ability to track foreshortening was present from the first trial of the experiment, arguing against a prominent learning effect when participants learn to associate different viewpoints to scenes across the experiment. Moreover, after the first trial in which they explored a wide range of response options available in the task, these strong perspective takers were highly efficient: the number of steps they took towards the correct response closely mapped the minimum number needed, suggesting little decisional uncertainty or "going back and forth" between options.

In contrast, performance in the other clusters (weak, categorical, inverse clusters) was characterised by considerably less exploration in the first trial, and consistently higher errors across subsequent trials, with participants consistently making of fewer-than-needed button presses to arrive at the correct viewpoints. This suggests that, rather than egocentric anchoring, weaker performance might be due simply to task disengagement, particularly given the online format of the task. Because each trial began from a perpendicular viewpoint, more oblique avatar perspectives required substantially more keypresses to reach the correct position. Thus, some apparent egocentric responding may partly reflect an unwillingness to exert the required effort, rather than a limitation in VPT ability per se. Indeed, as noted, all apart from the Strong VPT cluster possessed on average negative Inefficiency scores (i.e., they were not exploring enough potential viewpoints to reach the correct answer). Thus, it could be that our task, and its online format, underestimates high perspective taking performance in the population, masked by participant disengagement due to the experimental set-up. Future studies could resolve this issue by randomising the starting viewpoint or presenting a small set of candidate viewpoints for participants to select between.

Regarding the range of VPT ability apparent in our task, one possibility is that some participants might experience cycling through potential avatar viewpoints as a rotation of self around the lines whereas others a rotation of the lines themselves. Given the evidence that VPT might be supported by imagined self-motion or embodied self-rotation processes (Elk & Blanke, 2013; Kessler & Thomson, 2010; Surtees et al., 2013; Ward et al., 2022), it is possible that those participants who experienced a self-rotation effect also displayed enhanced VPT ability in matching possible perspective to person locations. Future work should gather feedback on participants' experiences to help clarify this potential link.

Together, these findings show, like in Experiment 1, that people can in fact represent another's viewpoint and track the resulting perceptual foreshortening across viewing perspectives with a surprising degree of accuracy. A minority of atypical responses likely reflects a combination of factors, including egocentric anchoring, variability in cognitive strategy (e.g., geometric vs. embodied reasoning), and, given the online context, misunderstanding or task disengagement.

10. General discussion

In two experiments, we evaluated Samuel et al.'s (2021) claim that people have no insight into another's visual experience, using two modified versions of their original line-length judgement task. In contrast to their findings, both experiments showed (a) that the majority of people were sensitive to the fundamental relationship between retinal size and distance, with regard to another person's visual perspective, and that (b) almost half of participants were highly precise in their understanding of perceptual foreshortening from another's perspective.

Experiment 1 tested whether participants could identify which of two identical lines on a wall would appear longer to a person with a viewing location that is closer to one line than the other. In a pseudo-replication of Samuel et al.'s (2021) original experiment, we replicated their original finding, that participants perform at chance when asked which line should appear longer to the person in the scene. In an improved version of this task, we phrased the question so that it communicated more explicitly that participants were to represent the other person's retinal experience of the lines without size-constancy adjustment (i.e., that the lines are of the same length), in addition to ergonomically situating the sliders. Given these minor improvements, the majority of participants correctly indicated that the line closer to the other person would appear longer on their retina. These results show that, even in Samuel et al.'s (2021) original task, participants can make correct perspective judgements, provided the representation type (i.e., the retinal rather than consciously experienced size of the lines) is disambiguated, and other methodological sources of confusion are eliminated.

Results from the post-experiment questions are in line with this interpretation. Participants in the pseudo-replication group expressed greater confidence in their understanding and visualisation ability than those in the improved group, despite worse VPT performance. Although this may seem paradoxical, it is consistent with the mistaken belief that the task requires judging actual relative line length (i.e., the physical line lengths, not retinal size for the person). In contrast, the greater subjective difficulty reported in the improved task is consistent with engagement with the more demanding process of representing retinal size from another viewing perspective.

In Experiment 2, participants viewed scenes in which an avatar observed two identical lines from different locations and selected the matching view from 23 possible depictions. A cluster analysis revealed that approximately 82% of participants were able to take the visual perspective of the avatar under Samuel's own criteria (Samuel et al., 2021). Notably, 47% of participants could predict perceptual foreshortening across viewpoints with near-perfect precision. Another 26% showed slightly weaker ability, consistently underestimating the degree of line-length disparity, A smaller 9% of these participants showed a simplified rule-based strategy, being able to indicate which line should appear longer, but were less sensitive to the magnitude of this difference.

These results suggest most people are highly sensitive to the relationship between retinal size and distance and can apply this to estimate another person's visual experience. Furthermore, there is variation in people's aptitude at judging another's view, with a subset of individuals able to form highly precise representations of how the world looks from another's visual perspective (i.e., level-2 VPT). The current results therefore suggest that the original null findings in Samuel et al.'s (2021) work most likely stem from unclear task instructions that leave open, to participants, which level of representation it is targeted. When looking at objects, an observer's own visual system naturally accounts for different viewing distances, so that objects of different retinal sizes will appear perceptually identical (Sperandio et al., 2015), and that objects of identical retinal sizes appear illusorily enlarged when depth cues suggest a greater distance from the observer (see size constancy illusion; Gregory, 1997; Ponzo, 1910). Nevertheless, people can introspect into this otherwise automatic adjustment, and recover the retinal sizes of the objects, but this a highly controlled, skilled process that is difficult for people to undertake (Bednarek et al., 2024; Perdreau & Cavanagh, 2011; Wincza et al., 2025). Samuel's task essentially asks participants to perform this same process on the (inferred) retinal input of another participant. Our studies show that, if properly instructed, participants are highly able to solve this task, pointing to advanced abilities for representing participant's visual input before size constancy processing

takes place.

An important question is which type of mechanism allowed participants to precisely report the other person's retinal experience of the relative line lengths. Although the flexibility and accuracy of some participants' judgements is suggestive of a detailed representation that can be mapped onto a potential target image (i.e., the 'Strong VPT' group in Experiment 2), we cannot rule out the possibility that participants are instead deploying an abstract spatial representation of the scene (e.g., Erle & Topolinski, 2017) or even a set of complex heuristics that might allow one to arrive at the correct response, which does not require visualisation at all (Heyes, 2014). This lack of clarity is mirrored in the current literature as most interference-based tasks (e.g., the dot perspective task, Samson et al., 2010; the 6/9 task, Surtees et al., 2012) leave open the possibility that interference stems from sources other than visual representation. Our present findings offer an important grounding to this debate, by demonstrating that participants can report what the world looks like to another person at a fundamental (retinal input) level, in contrast to Samuel et al.'s (2021) previous conclusions. Moreover, the precision of participants' estimates puts substantial bounds on the potential underlying mechanisms. As previously discussed, a simple "closer is larger" rule cannot account for the performance of the best perspective taking groups, which were able to precisely track the amount of visual foreshortening across different avatar locations in Exp 2. Instead, this result suggests involvement of either a perceptual simulation mechanism that can precisely represent another's perspective (e.g., Ward et al., 2019, 2020, 2022), or a sophisticated heuristic process involving geometric reasoning about perspective distortions. Note however that those participants that reported to have used such a geometric reasoning strategy generally achieved worse performance (see Supplementary Materials C: Experiment 2 analysis of feedback questions) than the best VPT groups, potentially pointing towards a perceptual simulation origin (for other evidence, see Ward et al., 2019, 2020, 2022). Together, the two experiments presented here are thus vital in demonstrating people not only have knowledge of other's perceptual experience (Exp.1), but that this knowledge can be highly precise (Exp.2), indicating that more sophisticated geometric or perceptual simulation mechanisms underpin VPT ability in our task.

As Samuel et al. (2021, 2024) and others (Cole et al., 2020, 2022; Cole & Millett, 2019) point out, the lack of theoretical clarity about what perspective taking could mean is an important problem for the field to address. Partly to blame for this explanatory gap is the field's paradigmatic shift away from explicit methodologies (in which successful performance constitutes a verbal indication of a 'correct' perspective) to methods which utilise interference as a proxy measure of VPT. Through this process, VPT has garnered a variety of different operationalisations. This is true even for the current paradigm. For example, compared to the seminal 'three-mountains' task (Piaget et al., 1967) in which successful level-2 VPT required participants to select a correct depiction of the person's view out of three potentials, participants within the original line length judgement paradigm were required to generate this depiction themselves and then abstractly translate it onto a pair of sliders. In this manner, one could fairly construe the criteria for demonstrating level-2 VPT as more stringent in Samuel et al.'s (2021) original paradigm. Further to this, orthographic depictions were used in the 'three-mountains' task (Piaget et al., 1967), meaning that they did not feature any perceptual foreshortening due to depth. Consequently, one did not need to understand that a mountain further away from the other person would appear smaller to them (or vice versa) in order to identify the correct viewpoint. Instead, one would rely on the organisation of objects relative to the observer, denoted primarily by occlusion (i.e., how the flat projections of the mountains overlapped each other, see, Anderson, 2003) and location of the features on top of each mountain (i.e., their order across the depiction from left to right). Hence, while possessing an understanding of perceptual foreshortening as a result of distance is vital for both Samuel et al.'s (2021) original line length judgement paradigm and our multi trial adaptation, earlier measures of level-2 (i.e., the turtle task, Flavell et al., 1981; Masangkay et al., 1974) would not require this ability. In essence therefore, the experiments presented here pose a novel (and arguably more challenging) measure of level-2 VPT. This raises questions as to what degree these abilities measured are cognitively comparable. That is to say, when scholars postulate VPT ability underlies higher order socio-cognitive reasoning (Clark & Brennan, 1991; Creem-Regehr et al., 2013; Freundlieb et al., 2016), are they referring to one's ability to represent another's preconscious retinal image? Given VPT has strong links to socio-cognitive reasoning, it makes more sense that we should represent how others actually perceptually experience the world, rather than how it projects onto their retina. In short, it could be argued that both the original Samuel et al. (2021) "null" results as well as our current, seemingly contradictory, results, are both indicative of level-2 VPT: Samuel et al.'s (2021) data might show that indeed people take others' perspective but including their cognitive adjustments such as perceptual foreshortening, hence experiencing the lines as equally long, whereas our data show that when properly queried to visualise what the other sees, people can inhibit that type of perspective taking and introspect the actual preconscious retinal image.

11. Conclusion

Our studies reveal that participants are accurate at deriving other's visual experience that results from different viewing perspectives. Experiment 1 illustrated that minor clarifying changes enable successful perspective taking, contrasting Samuel et al.'s (2021) original findings. Experiment 2 demonstrated that these representations are accurate and track how visual input changes across different viewing perspectives. Together these findings offer strong evidence against Samuel et al.'s (2021) claim that people are incapable of representing how the environment visually appears to another person.

CRediT authorship contribution statement

Louis Aldrich: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Katrina L. McDonough:** Conceptualization, Software, Methodology. **Bert Timmermans:** Conceptualization, Methodology, Supervision, Writing – review & editing. **Patric Bach:** Conceptualization, Methodology, Supervision, Writing – review & editing.

Funding

This work was supported by a Postgraduate Research Grant to L.A., funded by the University of Aberdeen's School of Psychology. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of competing interest

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

Acknowledgments

This work was supported by the University of Aberdeen, which was not involved in the design of the study nor preparation of the manuscript. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.

org/10.1016/j.cognition.2026.106699.

Data availability

I have shared a link to my data and code at the attach files step (through an OSF view only link).

References

Aglioti, S., DeSouza, J. F., & Goodale, M. A. (1995). Size-contrast illusions deceive the eye but not the hand. *Current Biology, 5*(6), 679–685.

Anderson, B. L. (2003). The role of occlusion in the perception of depth, lightness, and opacity. *Psychological Review, 110*(4), 785–801. https://doi.org/10.1037/0033-295X.110.4.785

Andrews, D. P. (1964). Error-correcting perceptual mechanisms. *Quarterly Journal of Experimental Psychology, 16*(2), 104–115. https://doi.org/10.1080/17470216408416355

Bednarek, H., Przedniczek, M., Wujcik, R., Olszewska, J. M., & Orzechowski, J. (2024). Cognitive training based on human-computer interaction and susceptibility to visual illusions. Reduction of the Ponzo effect through working memory training. *International Journal of Human-Computer Studies, 184*, Article 103226. https://doi.org/10.1016/j.ijhcs.2024.103226

Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In *Perspectives on socially shared cognition* (pp. 127–149). American Psychological Association. https://doi.org/10.1037/10096-006.

Cole, G. G., & Millett, A. C. (2019). The closing of the theory of mind: A critique of perspective-taking. *Psychonomic Bulletin and Review, 26*(6), 1787–1802. https://doi.org/10.3758/s13423-019-01657-y

Cole, G. G., Millett, A. C., Samuel, S., & Eacott, M. J. (2020). Perspective-taking: In search of a theory. *Vision, 4*(2), Article 30. https://doi.org/10.3390/vision4020030

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

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

Creem-Regehr, S. H., Gagnon, K. T., Geuss, M. N., & Stefanucci, J. K. (2013). Relating spatial perspective taking to the perception of other's affordances: Providing a foundation for predicting the future behavior of others. *Frontiers in Human Neuroscience, 7*. https://doi.org/10.3389/fnhum.2013.00596

Elk, M., & Blanke, O. (2013). Imagined own-body transformations during passive self-motion. *Psychological Research, 78*, 18–27. https://doi.org/10.1007/s00426-013-0486-8

Erle, T. M., Ruessmann, J. K., & Topolinski, S. (2018). The effects of visuo-spatial perspective-taking on trust. *Journal of Experimental Social Psychology, 79*, 34–41. https://doi.org/10.1016/j.jesp.2018.06.006

Erle, T. M., & Topolinski, S. (2015). Spatial and empathic perspective-taking correlate on a dispositional level. *Social Cognition, 33*(3), 187–210. https://doi.org/10.1521/soco.2015.33.3.187

Erle, T. M., & Topolinski, S. (2017). The grounded nature of psychological perspective-taking. *Journal of Personality and Social Psychology, 112*(5), 683–695. https://doi.org/10.1037/pspa0000081

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

Freundlieb, M., Kovács, Á. M., & Sebanz, N. (2016). When do humans spontaneously adopt another's visuospatial perspective? *Journal of Experimental Psychology: Human Perception and Performance, 42*(3), 401–412. https://doi.org/10.1037/xhp0000153

Frick, A., Möhring, W., & Newcombe, N. S. (2014). Picturing perspectives: Development of perspective-taking abilities in 4-to 8-year-olds. *Frontiers in Psychology, 5*, 386.

Gardner, M. R., & Thorn, L. (2025). Submentalizing: Clarifying how domain general processes explain spontaneous perspective-taking. *Journal of Experimental Psychology: Human Perception and Performance, 51*(1), 7.

Gregory, R. L. (1963). Distortion of visual space as inappropriate constancy scaling. *Nature, 199*(4894), 678–680. https://doi.org/10.1038/199678a0

Gregory, R. L. (1997). *Eye and brain: The psychology of seeing*. Princeton University Press.

van der Ham, I. J., Dijkerman, H. C., & van Stralen, H. E. (2021). Distinguishing left from right: A large-scale investigation of left–right confusion in healthy individuals. *Quarterly Journal of Experimental Psychology, 74*(3), 497–509. https://doi.org/10.1177/1747021820968519

Hamilton, A. F. D. C., 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.

Hedge, C., Powell, G., & Sumner, P. (2018). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior Research Methods, 50*(3), 1166–1186.

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

Kampis, D., & Southgate, V. (2020). Altercentric cognition: How others influence our cognitive processing. *Trends in Cognitive Sciences, 24*(11), 945–959.

Kessler, K., & Thomson, L. (2010). The embodied nature of spatial perspective taking: Embodied transformation versus sensorimotor interference. *Cognition, 114*(1), 72–88. https://doi.org/10.1016/j.cognition.2009.08.015

Kessler, K., & Wang, H. (2012). Spatial perspective taking is an embodied process, but not for everyone in the same way: Differences predicted by sex and social skills score. *Spatial Cognition & Computation, 12*(2–3), 133–158. https://doi.org/10.1080/13875868.2011.634533

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

Lempers, J. D., Flavell, E. R., & Flavell, J. H. (1977). The development in very young children of tacit knowledge concerning visual perception. *Genetic Psychology Monographs, 95*(1), 3–53.

Masangkay, Z. S., McCluskey, K. A., McIntyre, C. W., Sims-Knight, J., Vaughn, B. E., & Flavell, J. H. (1974). The early development of inferences about the visual percepts of others. *Child Development, 45*(2), 357. https://doi.org/10.2307/1127956

Mattan, B. D., Rotshtein, P., & Quinn, K. A. (2016). Empathy and visual perspective-taking performance. *Cognitive Neuroscience, 7*(1–4), 170–181. https://doi.org/10.1080/17588928.2015.1085372

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

Moll, H., & Tomasello, M. (2006). Level 1 perspective-taking at 24 months of age. *British Journal of Developmental Psychology, 24*(3), 603–613.

Perdreau, F., & Cavanagh, P. (2011). Do artists see their retinas? *Frontiers in Human Neuroscience, 5*, 171.

Piaget, J., Inhelder, B., Langdon, F. J., & Lunzer, J. L. (1967). *The child's conception of space*. W.W. Norton & Company, Inc.

Ponzo, M. (1910). *Intorno ad alcune illusioni nel campo delle sensazioni tattili, sullillusione di Aristotele e fenomeni analoghi*. Leipzig, Germany: Wilhelm Engelmann.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences, 1*(4), 515–526. https://doi.org/10.1017/S0140525X00076512

Samson, D., Apperly, I. A., Braithwaite, J. J., Andrews, B. J., & Bodley Scott, S. E. (2010). Seeing it their Way: Evidence for rapid and involuntary computation of what other people see. *Journal of Experimental Psychology: Human Perception and Performance, 36*(5), 1255–1266. https://doi.org/10.1037/a0018729

Samuel, S., Erle, T. M., Kirsch, L. P., Surtees, A., Apperly, I., Bukowski, H., Auvray, M., Catmur, C., Kessler, K., & Quesque, F. (2024). Three key questions to move towards a theoretical framework of visuospatial perspective taking. In *, Vol. 247. Cognition*. Elsevier B.V. https://doi.org/10.1016/j.cognition.2024.105787.

Samuel, S., Hagspiel, K., Eacott, M. J., & Cole, G. G. (2021). Visual perspective-taking and image-like representations: We don't see it. *Cognition, 210*. https://doi.org/10.1016/j.cognition.2021.104607

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*. https://doi.org/10.3389/fnhum.2013.00712

Sperandio, I., Sperandio, I., & Chouinard, P. A. (2015). The mechanisms of size constancy. *Multisensory Research, 28*(3–4), 253–283. https://doi.org/10.1163/22134808-00002483

(a) 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.(b) Surtees, A., Apperly, I., & Samson, D. (2013). Similarities and differences in visual and spatial perspective-taking processes. *Cognition, 129*(2), 426–438.

Surtees, A. D. R., Butterfill, S. A., & Apperly, I. A. (2012). Direct and indirect measures of Level-2 perspective-taking in children and adults. *British Journal of Developmental Psychology, 30*(1), 75–86. https://doi.org/10.1111/j.2044-835X.2011.02063

Uittenhove, K., Jeanneret, S., & Vergauwe, E. (2023). From lab-testing to web-testing in cognitive research: Who you test is more important than how you test. *Journal of Cognition, 6*. https://doi.org/10.5334/joc.259

Ward, E., Ganis, G., & Bach, P. (2019). Spontaneous vicarious perception of the content of another's visual perspective. *Current Biology, 29*(5), 874–880.e4. https://doi.org/10.1016/j.cub.2019.01.046

Ward, E., Ganis, G., McDonough, K. L., & Bach, P. (2020). Perspective taking as virtual navigation? Perceptual simulation of what others see reflects their location in space but not their gaze. *Cognition, 199*. https://doi.org/10.1016/j.cognition.2020.104241

Ward, E., Ganis, G., McDonough, K. L., & Bach, P. (2022). Is implicit Level-2 visual perspective-taking embodied? Spontaneous perceptual simulation of others' perspectives is not impaired by motor restriction. *Quarterly Journal of Experimental Psychology, 75*(7), 1244–1258. https://doi.org/10.1177/17470218221077102

Wincza, R., Hartley, C., Donovan, T., Linkenauger, S., Crawford, T., Griffiths, D., & Doherty, M. (2025). Specific visual expertise reduces susceptibility to visual illusions. *Scientific Reports, 15*(1), 5948. https://doi.org/10.1038/s41598-025-88178-y

Further-reading

Blender Foundation. (2024). Blender (Version 4.2) [Computer software] https://www.blender.org Accessed November 2024.

Millisecond Software. (2020). Inquisit (Version 6) [Computer software] https://www.millisecond.com Accessed June 2025.

Open Science Framework. (2025). OSF project repository. https://osf.io/hgnxy/overview?view_only=337ec3c69a1f4f7da1a136a2a889c9cf. Created December 17, 2025. Accessed June 2026.

Rezlescu, C., Danaila, I., Miron, A., & Amariei, C. (2020). More time for science: Using Testable to create and share behavioral experiments faster, recruit better participants, and engage students in hands-on research. *Progress in Brain Research, 253*, 243–262.

Glossary

Altercentric interference: A performance cost (e.g., slower responses or increased errors) observed when another person's perspective conflicts with one's own.

Depth cues: Visual information that allows an observer to infer the three-dimensional structure of a scene, such as texture gradients, perspective convergence, shading, occlusion, and relative size.

Egocentric viewpoint: A perspective anchored to one's own position in space, in which spatial relations are represented relative to the self rather than another individual.

Grounding: The process by which interlocutors establish shared understanding during communication, often by aligning reference frames, assumptions, or perceptual access.

Level-1 visual perspective taking (VPT-1): The ability to understand whether another person can or cannot see an object or feature (i.e., judgements about visibility or line-of-sight).

Level-2 visual perspective taking (VPT-2): The ability to understand how an object or scene visually appears to another person, including left–right orientation, relative position, or appearance from a different viewpoint.

Perceptual foreshortening: The reduction in retinal size or apparent extent of an object as its distance from the observer increases, due to the geometry of projection onto the retina.

Retinal size: The size of an object's image as projected onto the retina, which depends on both the object's physical size and its distance from the observer.

Shared reference frame: A common spatial or conceptual coordinate system adopted by interacting individuals, allowing them to align perspectives and interpret information consistently.

Size constancy: A perceptual mechanism by which the visual system maintains a stable perception of an object's physical size despite changes in viewing distance and corresponding changes in retinal size.

Theory of Mind: The capacity to attribute mental states—such as beliefs, intentions, knowledge, or perceptions—to others, and to understand that these states may differ from one's own.

Visual perspective taking (VPT): The ability to represent what the environment looks like from another person's point of view, encompassing both judgements about what others can see (VPT-1) and how they see it (VPT-2).