Cognition 271 (2026) 106478

Contents lists available at ScienceDirect

Cognition

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

I'll believe it unless it's too absurd: Spontaneous visual perspective-taking as prior-based heuristic inference

Xucong Hu [a], Yitong Zheng [a], Qinyi Hu [a], Hui Chen [a,b,*], Mowei Shen [a,b,*], Jifan Zhou [a,b,*]

[a] Department of Psychology and Behavioral Sciences, Zhejiang University, China
[b] Zhejiang Key Laboratory of Neurocognitive Development and Mental Health, Zhejiang University, China

ARTICLE INFO

Keywords:
Spontaneous perspective-taking
Visual perspective-taking
Perceptual simulation
Heuristic inference
Prior information

ABSTRACT

The underlying mechanism of visual perspective-taking (VPT)—the ability to represent what others see—remains contested. Perceptual simulation theory proposes that VPT involves reconstructing others' visual experiences, whereas heuristic accounts argue that it relies on symbolic inference grounded in naïve optics. Evidence for heuristics largely comes from explicit report tasks, leaving open whether spontaneous (implicit) VPT in an agent-irrelevant task is driven by the same mechanism. A further possibility is that apparent "simulation failures" arise because observers lack prior visual information about what the other sees from their viewpoint. Across two experiments, participants performed an agent-irrelevant line-length judgment task while receiving plausible, absent, or implausible prior visual information from the agent's viewpoint. Experiment 1 showed a robust perspective-consistent bias under plausible priors, no bias without priors, and a weaker bias under implausible priors. A control experiment ruled out priming. Experiment 2 parametrically varied implausibility in a Ponzo-style layout and revealed a boundary condition: priors ranging from plausible to moderately implausible continued to bias judgments, whereas highly implausible priors were discounted. These results support a bounded, resource-rational heuristic account in which others' visual information acts as plausibility-weighted cues integrated with one's own visual input, rather than being reconstructed via perceptual simulation.

1. Introduction

Visual perspective-taking (VPT)—the capacity to adopt others' perspectives to represent what they are seeing—plays a critical role in everyday communication and cooperation (Flavell, 1994; Tomasello, 2009). Yet, the mechanism that enables such representations remains contested.

Two prominent theoretical accounts have been proposed to explain how VPT operates. *Perceptual simulation theory* holds that VPT relies on quasi-perceptual simulation: observers generate mental representations that approximate what another agent would see, as if "looking through others' eyes" (Ward, Ganis, & Bach, 2019; Ward, Ganis, McDonough, & Bach, 2020; Ward, Ganis, McDonough, & Bach, 2022). This view is grounded in Kosslyn, Pinker, Smith, and Shwartz's (1979) depictive theory of mental imagery, which assumes that internal images preserve key spatial properties of visual input and support inferences about resulting perceptual experience[1] (Ward et al., 2019; Ward et al., 2020; Ward et al., 2022; Samson et al., 2010). One piece of supporting evidence comes from Ward et al. (2019), who reported that adding an agent to a mental-rotation task facilitated object recognition when the agent's spatial viewpoint aligned with the optimal orientation—an effect interpreted as consistent with observers transforming their own mental representation toward the agent's viewpoint.

In contrast, *heuristic accounts* propose that VPT does not reconstruct another's visual experience, but instead relies on symbolic inference grounded in naïve optics (Samuel et al., 2021; Cole, Millett, Samuel, & Eacott, 2020; Cole & Millett, 2019). This view aligns with Pylyshyn's (2002) propositional theory, which characterizes mental

* Corresponding authors at: Department of Psychology and Behavioral Sciences, Zhejiang University, 866 Yuhangtang Road, Zhejiang, Hangzhou 310058, China.
E-mail addresses: yitongzheng@zju.edu.cn (Y. Zheng), chenhui@zju.edu.cn (H. Chen), mwshen@zju.edu.cn (M. Shen), jifanzhou@zju.edu.cn (J. Zhou).

[1] Throughout this paper, we distinguish between several related but non-identical uses of the term *perspective*.

(1) *Spatial viewpoint* (or viewpoint/view) refers to the agent's physical location and orientation in space.
(2) *Visual input* (or visual information) refers to the optical projection available at the eyes before any perceptual interpretation.
(3) *Visual experience* (or perceptual experience) refers to the subjective perceptual outcome constructed from the visual input.

https://doi.org/10.1016/j.cognition.2026.106478
Received 25 December 2025; Received in revised form 1 February 2026; Accepted 3 February 2026

representations as symbolic "entity-relation" structures rather than depictive images. Under this framework, observers rely on propositional rules or folk beliefs about vision (e.g., "6 and 9 are reversible," "objects at greater depth look larger after compensation"). Growing evidence supports this perspective (Cole et al., 2020; Cole & Millett, 2019; Cole, Samuel, & Eacott, 2022; Samuel et al., 2021; Samuel, Cole, & Eacott, 2023; Samuel, Eacott, & Cole, 2022). For example, Samuel et al. (2021) presented participants with two physically equal-length lines where, due to foreshortening effects, the closer line should look longer from the agent's spatial viewpoint (see Fig. 1). If simulation were operating, participants asked to "imagine the agent's perspective" should consistently choose that line as longer. Instead, judgments were split, converging only when the image from the agent's viewpoint was shown. Samuel et al. argued this reflects reliance on naïve optical rules rather than reconstruction of another's perceptual state.

Despite growing evidence for the *heuristic account*, important gaps remain. First, there is a lack of evidence from implicit (spontaneous)[2] VPT, which is arguably the more ecologically relevant form. Most supporting findings for the *heuristic account* come from explicit VPT tasks in which participants directly report the agent's perspective (Cole et al., 2020; Samuel et al., 2021). In contrast, much of the evidence commonly cited in support of *perceptual simulation theory* originates from implicit VPT paradigms, where participants perform an unrelated task yet exhibit systematic biases consistent with the agent's perspective (Ward et al., 2019; Ward et al., 2020). However, some researchers have argued that effects observed in such paradigms—most notably the dot-perspective task (Samson et al., 2010)—may reflect domain-general processes such as attentional cueing or task-driven strategies rather than genuine perspective-taking, and have therefore favored explicit-report tasks (Cole et al., 2020; Santiesteban, Catmur, Hopkins, Bird, & Heyes, 2014). Importantly, these critiques may reflect limitations of particular paradigms rather than the absence of implicit VPT itself, underscoring the need for implicit designs with more stringent experimental control.

Importantly, implicit VPT is not only theoretically significant but may also differ qualitatively from explicit VPT. Everyday social interaction often involves rapid, automatic sensitivity to others' perspective rather than deliberate perspective reports (Wegner & Bargh, 1998; Samson et al., 2010; Apperly & Butterfill, 2009). Moreover, these two forms of perspective-taking may diverge in important ways. In explicit tasks, strategic, top-down correction can readily mask underlying perceptual processes. For example, in visual illusion research, explicit judgments often incorporate knowledge-based corrections (e.g., "I know these lines should be the same length"), thereby overriding immediate perceptual experience (Firestone & Scholl, 2016). Similarly, eye-tracking studies demonstrate that implicit attentional orienting frequently diverges from explicit verbal reports: individuals may explicitly deny noticing a stimulus even though their gaze has already been drawn toward it (Hayhoe & Ballard, 2005). Consequently, explicit VPT tasks cannot be assumed to transparently reveal the mechanisms that drive spontaneous VPT, underscoring the importance of directly investigating the processes underlying implicit VPT.

Second, the strongest criticism of *perceptual simulation theory* is the claim that observers simply cannot access another person's visual experience directly; any attempted simulation must rely on the observer's own prior similar perceptual experiences (Cole et al., 2020; Samuel et al., 2021). Yet this limitation may reflect a missing ingredient—namely, prior information about what the agent sees from their viewpoint. For instance, in Samuel et al. (2021), participants may have been unable to "simulate" the agent's perspective because they did not know what the agent actually saw (see Fig. 1). However, providing this information in an explicit task would directly reveal the correct answer, making it inappropriate to test whether this prior information is necessary for VPT. In contrast, an implicit task allows this information to be supplied without compromising the task's purpose.

Accordingly, the present study adapts Samuel et al.'s (2021) paradigm into an implicit VPT task with experimentally controlled priors. This paradigm has several advantages over widely used alternatives such as the 6/9 task or the dot-perspective task: it avoids reliance on well-known prior knowledge (e.g., that 6 and 9 are reversible) and minimizes general attentional cueing from directional signals, as the agent is depicted with their back facing the participant. Moreover, in our adapted version, participants make only self-viewpoint judgments of line length rather than alternating between self- and agent-viewpoint reports, thereby reducing task-driven strategies. Within this framework, the spontaneous VPT effect is indicated when participants' self-judgments show a systematic bias toward the agent's perspective—for example, judging the close line as longer despite the two lines being physically identical.

Before the task, participants were shown one of three types of prior visual information from the agent's viewpoint: a plausible prior depicting the agent's actual visual input; no prior, in which no visual information about the agent's visual input was provided; or an implausible prior consisting of a deliberately reversed, physically impossible visual input.

If *perceptual simulation theory* is correct, spontaneous VPT should appear only under plausible priors, given that simulation aims to reconstruct the agent's perceptual experience by faithfully mapping to their actual visual input. Thus, neither the No Prior nor Implausible Prior condition should elicit spontaneous VPT effects. By contrast, if the *heuristic account* is correct, prior information should shape judgments regardless of plausibility: the process relies on symbolic, rule-based inference, so even impossible priors may influence responses and potentially produce reversed effects. Our experiments were designed to adjudicate between these competing mechanisms in an implicit VPT context.

2. Experiment 1a

Experiment 1a adapted the paradigm of Samuel et al. (2021) into an implicit VPT task, incorporating experimentally manipulated priors regarding the agent's visual input. The aim was to adjudicate between *perceptual simulation theory* and the *heuristic account* in this implicit VPT context.

2.1. Method

2.1.1. Participants

Since the study employed a 3 × 2 mixed design, we conducted an a priori power analysis using G*Power 3.1 (Faul, Erdfelder, Lang, & Buchner, 2007). Assuming an α level of 0.05, a desired power of 0.95, and a effect size of f = 0.25, which reflects a conservative estimate consistent with prior implicit VPT studies (e.g., Blech, Lembcke, Bouquet, & Liepelt, 2025; He et al., 2024; Ward et al., 2022), with a correlation among repeated measures of 0.50 and a nonsphericity correction of 1, the required total sample size was estimated to be 66. To allow for full counterbalancing and potential exclusions, we recruited 69 participants (M_{age} = 23.1 ± 5.0, 38 males) through campus online forums. Participants received monetary compensation for their time. The study was approved by the Institutional Review Board of the Department of Psychology at the authors' university.

2.1.2. Design

The experiment adopted a 3 (Prior: Plausible Prior, No Prior, Implausible Prior) × 2 (Line Position: Close, Far) mixed design. Prior

[2] Terminology clarification. Throughout this paper, "implicit VPT" and "spontaneous VPT" are used interchangeably, referring to the phenomenon where participants spontaneously integrate an agent's perspective into their own judgments implicitly, without explicit task instructions to do so (O'Grady, Scott-Phillips, Lavelle, & Smith, 2020; Zhou, Peng, Li, Deng, & Chen, 2022).

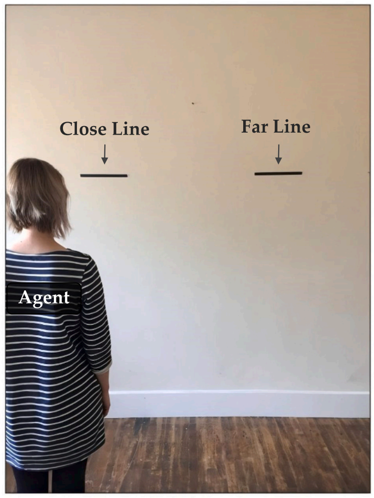

Image from the ***Participant's Viewpoint***

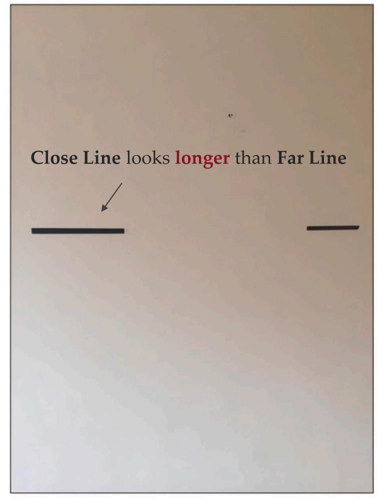

Image from the ***Agent's Viewpoint***

Fig. 1. Stimulus images from the participant's and agent's spatial viewpoints in Samuel, Hagspiel, Eacott, and Cole (2021).
Note. The left panel shows the image photographed from the participant's spatial viewpoint (where the two horizontal lines appear equal in length), whereas the right panel shows the image photographed from the agent's spatial viewpoint (where the closer line projects a longer retinal extent than the farther line because of foreshortening).

served as a between-subjects factor, whereas Line Position served as a within-subjects factor. Line Position was defined relative to the agent's standing location: when the agent stood on the left side of the two lines, the left line was coded as close and the right line as far; when the agent stood on the right, this mapping was reversed. Participants were required to reproduce the perceived length of both line positions, as described in detail below.

2.1.3. Materials

The paradigm was adapted from Samuel et al. (2021). Two identical black paper strips (20 cm each) were affixed to a wall at a height of 159 cm and separated by 47 cm. An agent stood either to the left or right of the strips and fixated the midpoint between them. To reproduce the agent's visual input, we photographed the scene from the agent's actual viewing geometry: the camera was positioned at the agent's eye height (153 cm), placed 75 cm from the wall, and horizontally offset by 50.5 cm from the midpoint between the two strips (see Fig. 2; detailed geometric parameters in the Supplementary Materials). This yielded an image that matched what the agent would see (plausible prior).

To obtain the participant-viewpoint image, a second camera was mounted on a tripod at the midpoint between the two strips, positioned 256 cm from the wall and oriented orthogonally to the wall. This corresponded to the view participants would have in the task. For the Implausible Prior condition, the agent-viewpoint image was horizontally reversed, flipping the line arrangement while keeping the background unchanged, to produce a physically impossible visual input (see Fig. 2; detailed geometric parameters in the Supplementary Materials).

The three Prior conditions differed only in the images and instructions presented at the beginning of the experiment. In the Plausible Prior condition, participants were shown the photograph taken from the agent's actual viewpoint and received the instruction: "The picture shows the agent's perspective[3] in the next scene. But please ignore it and focus only on remembering the line lengths in the upcoming image." In the No Prior condition, no image of the agent's viewpoint was presented; participants saw only the instruction: "Please focus on remembering the line lengths in the upcoming image." In the Implausible Prior condition, participants were shown a deliberately reversed, physically impossible version of the agent-viewpoint image, while the instructions were identical to those in the Plausible Prior condition.

2.1.4. Procedure

We employed a method-of-adjustment procedure rather than the direct length judgment used in Samuel et al. (2021). In this paradigm, participants first viewed a target line and subsequently reproduced its length by adjusting a probe line whose initial value was randomly longer or shorter than the target—a method widely used in visual perception research (Prinzmetal, Shimamura, & Mikolinski, 2001; Shen et al., 2015). This approach ensured that participants performed a primary task unrelated to agent's perspective, a standard requirement for implicit VPT paradigms (Samson et al., 2010; Ward et al., 2019). Moreover, it provided multiple repeated measurements, minimizing the possibility that effects reflected only short-lived priming from the prior.

The experiment was administered online using jsPsych (De Leeuw, 2015). Participants were instructed to judge line length. The instructions, along with the prior image shown at the beginning of the experiment, varied across the three prior conditions. Each trial began with a 2000-ms presentation of the participant-view image, depicting an agent facing two lines on a wall; each of the two lines subtended a constant on-screen length of 2 cm across all displays. After a 500-ms blank screen, a probe line appeared. Its initial length was randomly sampled to fall between 25 and 75% longer or shorter than the true target line length (Shen et al., 2015) (see Fig. 3). Participants adjusted the probe line to match the length of either the left or right target line they had just seen, with no time limit. A 500-ms blank screen followed each response (see Fig. 4).

The experiment consisted of 60 randomized trials (30 per line position), typically completed within 10 min. Importantly, the prior information was presented only once at the beginning of the experiment to minimize transient priming effects. The agent's standing position (left/right to the lines) was counterbalanced across participants, and close/far coding was defined accordingly for analysis.

2.1.5. Analyses

Data were analyzed using a 3 (Prior: Plausible Prior, No Prior, Implausible Prior) × 2 (Line Position: Close, Far) mixed ANOVA. Normality assumptions were assessed prior to analysis. Our primary focus was the Prior × Line Position interaction, followed by simple-effects analyses examining whether differences between close and far

[3] This colloquial use of "perspective" here is for participant comprehension only, not reflecting the strict terminological clarification defined in Footnote 1.

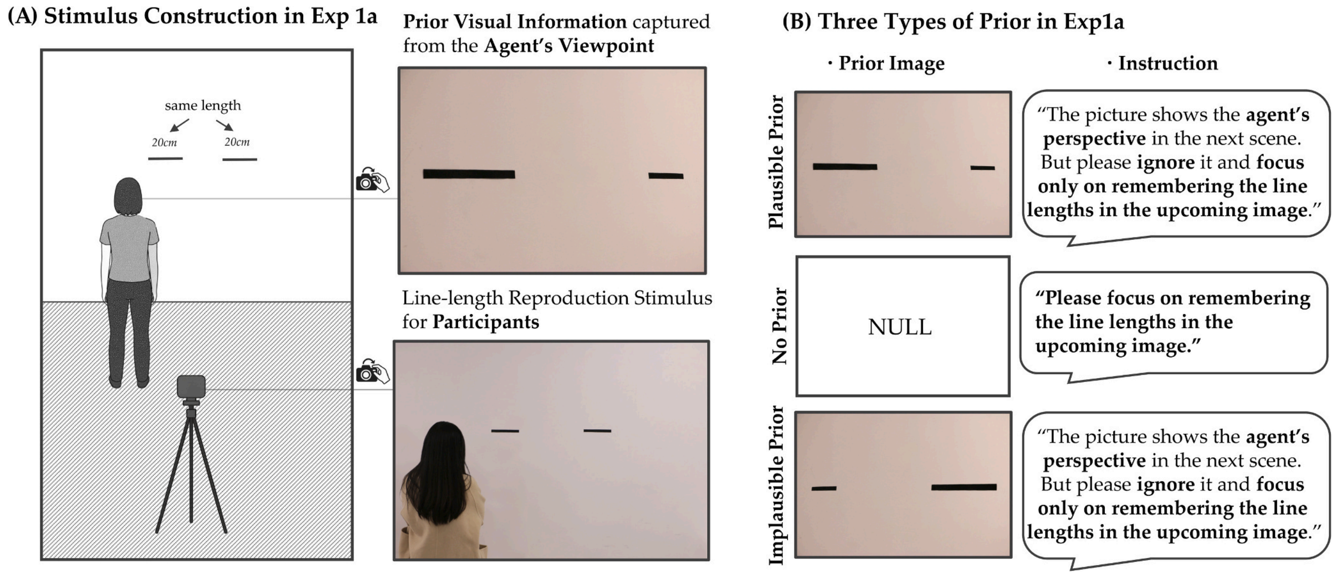

Fig. 2. Schematic illustration of the stimulus setup and prior information in Experiment 1a.
Note. (A) Schematic illustration of the stimulus setup used for photographing materials, including the spatial arrangement of the agent, wall, lines, and cameras in Experiment 1a. The diagram indicates how the images from the agent-viewpoint and participant-viewpoint were captured based on their respective viewing geometries. A female agent standing to the left of the lines is shown here as an illustrative example. (B) Prior images and corresponding instructions shown at the beginning of the experiment under the three prior conditions: Plausible Prior, No Prior, and Implausible Prior.

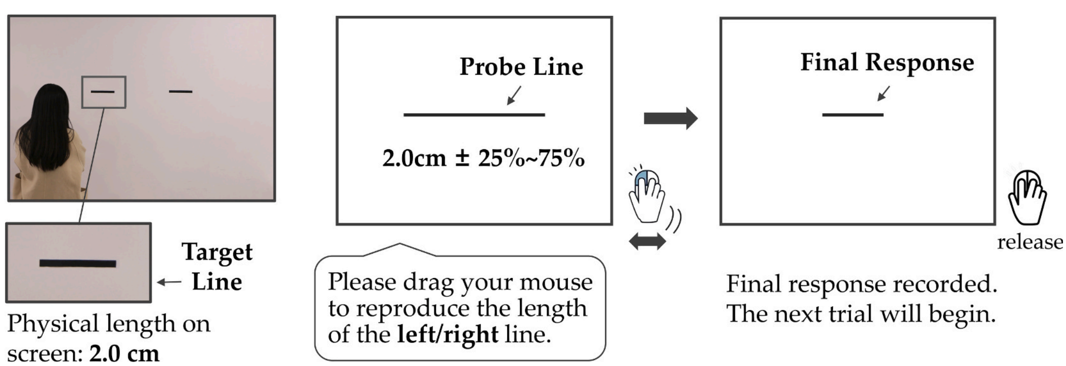

Fig. 3. Schematic illustration of the method-of-adjustment task in Experiment 1a.
Note. The figure illustrates the method-of-adjustment procedure used to measure perceived line length. Participants were instructed to reproduce the length of a target line by adjusting a probe line, whose initial length was randomly set to be 25–75% longer or shorter than the target.

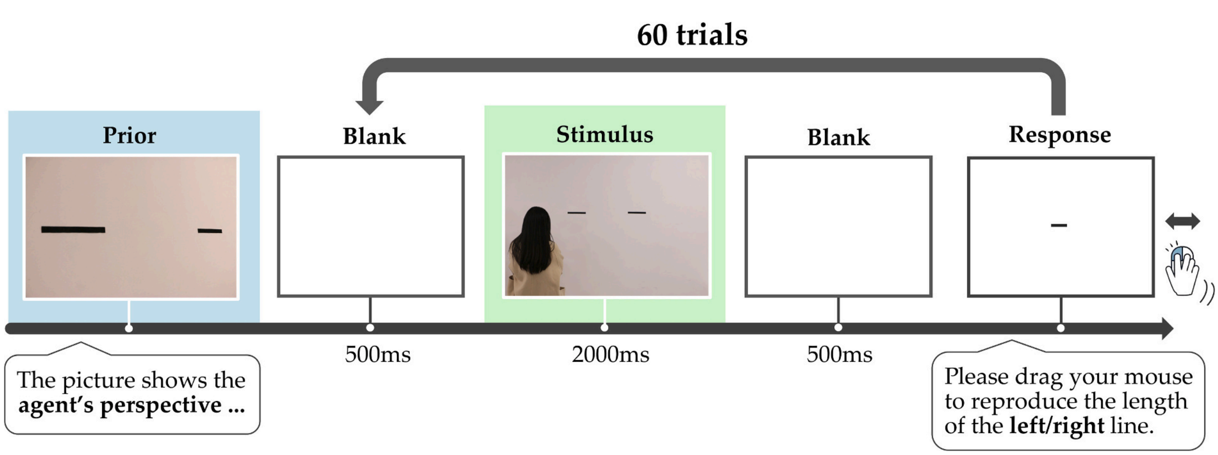

Fig. 4. Experimental procedure of Experiment 1a.
Note. Experimental procedure of Experiment 1a. Blue frames indicate prior information of the agent's viewpoint (including both the image and instructions), whereas green frames indicate the image from the participant's viewpoint. The illustration shows an example from the Plausible Prior condition.

lines emerged within each prior condition as evidence of a systematic bias consistent with spontaneous VPT.

To characterize this bias, participants' reproduction responses were quantified in two complementary ways: (1) Ratio-to-Actual Length. Because the presented lines always subtended 2 cm on screen, each reproduced length was divided by 2 cm to yield a reproduction ratio (e.g., a 2.2-cm reproduction yields 1.10), indexing overestimation (> 1) or underestimation (< 1). These ratios were entered into the mixed ANOVA. (2) Close/Far Ratio. To reduce individual differences in absolute response scale and isolate relative bias, we also computed a ratio by dividing the reproduced length of the close line by that of the far line for each participant. Values greater than 1 indicate judging the close line as longer, values less than 1 indicate the opposite, and a value of 1 reflects no bias. Analyses examined whether the mean ratio within each prior condition deviated from the neutral reference point of 1.0.

Together, these two analytic approaches provide convergent and robust evidence regarding whether and how prior information shapes spontaneous VPT biases.

2.2. Results

Three participants were excluded for providing at least one zero length judgment. This left a final sample of 66, evenly distributed across the three experimental conditions: Plausible Prior, No Prior, and Implausible Prior.

Given that the data in all six cells met the assumption of normality, parametric tests were conducted. We first performed a 3 (Prior: Plausible Prior, No Prior, Implausible Prior) × 2 (Line Position: Close, Far) mixed ANOVA on the Ratio-to-Actual Length. The main effect of Prior was not significant, $F(2, 63) = 0.064$, $p = .938$, $\eta_p^2 = 0.002$, nor was the main effect of Line Position, $F(1, 63) = 1.385$, $p = .224$, $\eta_p^2 = 0.022$. Crucially, the Prior × Line Position interaction was significant, $F(2, 63) = 8.226$, $p < .001$, $\eta_p^2 = 0.207$. Follow-up simple-effects analyses revealed a clear pattern across prior conditions. In the Plausible Prior condition, close lines ($M = 1.032$) were judged significantly longer than far lines ($M = 1.007$), $p < .001$. In the No Prior condition, there was no significant difference between close ($M = 1.016$) and far lines ($M = 1.011$), $p = .417$. In contrast, in the Implausible Prior condition, close lines ($M = 1.004$) were judged significantly shorter than far lines ($M = 1.021$), $p = .028$ (see Fig. 5).

We further analyzed the Close/Far Ratio with 1.0 as the theoretical baseline indicating no bias. A 3 (Prior: Plausible Prior, No Prior, Implausible Prior) × 2 (Ratio: Close/Far Ratio, 1) mixed ANOVA revealed a significant interaction, $F(2, 63) = 6.791$, $p = .002$, $\eta_p^2 = 0.177$ (additional details are reported in the Supplementary Materials). Follow-up analyses showed that, in the Plausible Prior condition, the ratio was significantly greater than 1 ($p = .001$), indicating that participants reliably judged the close line as longer. Descriptively, 72.7% of participants judged the close line as longer, 13.6% judged the two lines as approximately equal, and 13.6% judged the close line as shorter.[4] In the No Prior condition, the Close/Far Ratio did not significantly differ from 1 ($p = .389$). Here, judgments were broadly distributed across response categories: 50.0% judged the close line as longer, 22.7% as approximately equal, and 27.3% as shorter. In the Implausible Prior condition, the Close/Far Ratio did not significantly differ from 1 ($p = .073$), although responses showed a tendency toward judging the far line as longer. Consistent with this pattern, 22.7% of participants judged the close line as longer, 27.3% as approximately equal, and 50.0% as shorter (see Fig. 5).

[4] For descriptive purposes, responses were classified as “approximately equal” when the Close/Far Ratio fell within ±5% of 1.0, reflecting negligible differences in perceived line length; values greater than 1 indicated that the close line was judged as longer, whereas values less than 1 indicated that the far line was judged as longer.

The results supported the *heuristic account*. Under the plausible prior, participants showed a clear bias consistent with the agent's perspective—judging the close line as longer. Under the No Prior condition, no such bias emerged. Critically, under the implausible prior, participants also showed a bias consistent with the prior—judging the far line as longer. However, this effect was less stable: it was significant in the Ratio-to-Actual-Length measure but did not reach significance in the Close/Far Ratio analysis. This pattern suggests that participants were engaging in plausibility-based integration, weighting the prior against their own visual input rather than blindly following it. To further characterize this integration process, we designed additional experiments manipulating a wider range of prior plausibility levels.

3. Experiment 1b

Before proceeding to further investigations, we conducted a control experiment to rule out the possibility that the effects observed in Experiment 1 were driven by a simple priming effect rather than spontaneous perspective-taking. Although several features of the original design reduced this likelihood (e.g., multiple trials, the prior image presented only once), it remained possible that the prior merely acted as a priming cue that influenced subsequent line reproductions without being interpreted as the visual input from the agent's viewpoint.

In this control experiment, participants were again shown a prior image at the outset, but the accompanying instructions explicitly stated that the image was irrelevant to the upcoming task. If the prior image exerted its influence solely through priming, its effect should persist even under these instructions. Conversely, if the prior shapes judgments only when construed as the agent's perspective, no systematic bias should emerge in this condition. Because only the plausible and implausible prior conditions involved prior images, the control experiment focused exclusively on these two conditions.

3.1. Method

As this experiment employed a 2 × 2 mixed design, we conducted an a priori power analysis using the same parameters as in Experiment 1a. The required total sample size was estimated to be 54. To allow for full counterbalancing and potential exclusions, we recruited 58 participants ($M_{age} = 22.1 \pm 2.3$, 22 males).

This experiment used a 2 (Prior: Plausible Prior, Implausible Prior) × 2 (Line Position: Close, Far) mixed design. All materials and procedures were identical to those in Experiment 1a except for the instructions associated with the prior image. In both prior conditions, the instruction was replaced with: “This picture is unrelated to the task. Please ignore it and focus only on remembering the line lengths in the upcoming image” (See Fig. 6).

3.2. Results

Two participants were excluded for providing at least one zero length judgment. This left a final sample of 56, evenly distributed across the two experimental conditions: Plausible Prior and Implausible Prior.

Given that the data in all four cells met the assumption of normality, parametric tests were conducted. We first performed a 2 (Prior: Plausible Prior, Implausible Prior) × 2 (Line Position: Close, Far) mixed ANOVA on the Ratio-to-Actual Length. The main effect of Prior was not significant, $F(1, 54) = 0.002$, $p = .962$, $\eta_p^2 < 0.001$, nor was the main effect of Line Position, $F(1, 54) = 0.095$, $p = .759$, $\eta_p^2 = 0.002$. and the Prior × Line Position interaction was also not significant, $F(1, 54) = 0.433$, $p = .513$, $\eta_p^2 = 0.008$. Follow-up simple-effects analyses revealed a clear pattern across prior conditions. In the Plausible Prior condition, there was no significant difference between close ($M = 1.008$) and far lines ($M = 1.004$), $p = .620$. Similarly, there was also no significant difference between close ($M = 1.012$) and far lines ($M = 1.016$) in the Implausible Prior condition, $p = .668$ (See Fig. 6).

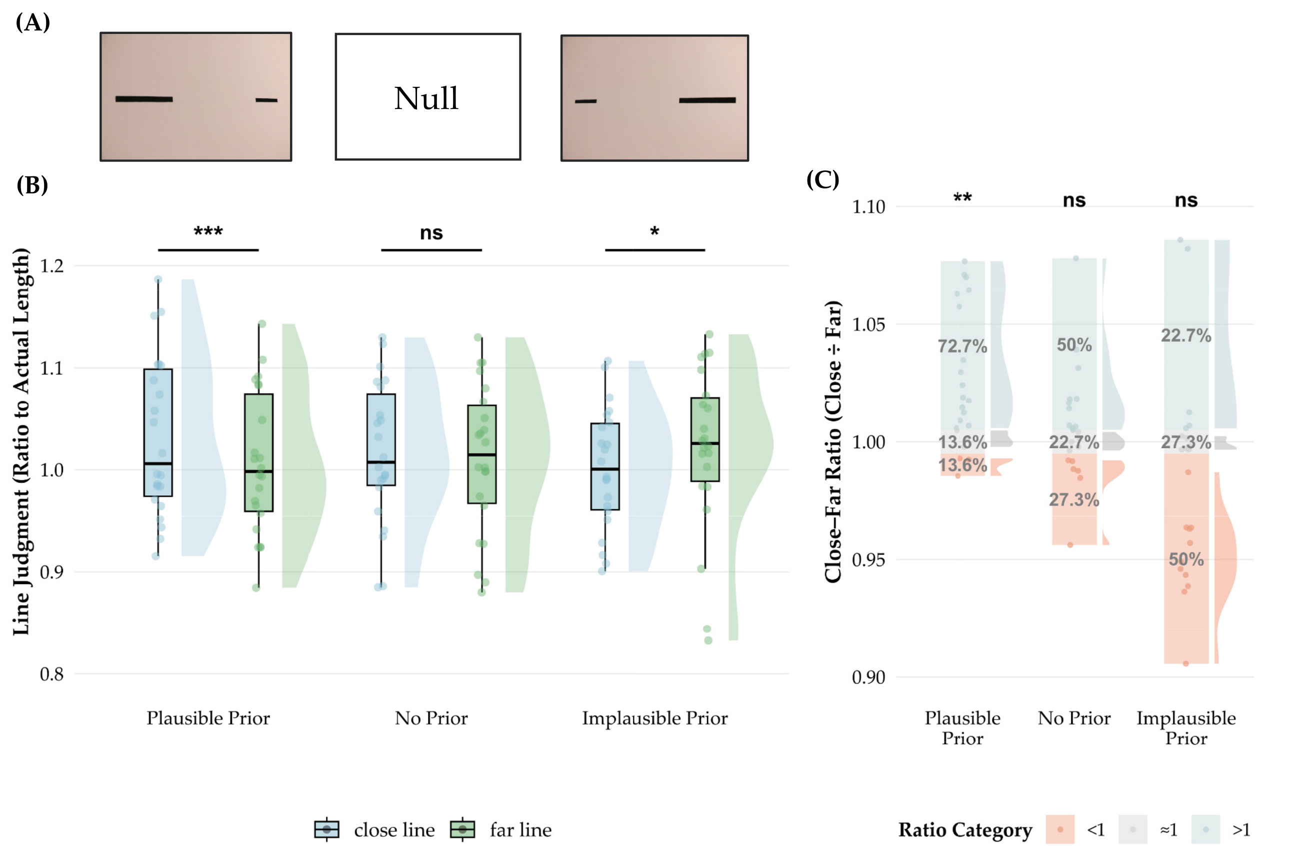

Fig. 5. Stimulus and results of Experiment 1.
Note. (A) Examples of prior images shown at the beginning of the experiment: Plausible Prior (left), No Prior (middle), and Implausible Prior (right). (B) Participants' line-judgment ratios to actual length for close and far lines across prior conditions. Blue indicates close-line judgments and green indicates far-line judgments. Boxplots show the median and interquartile range; dots represent individual participants. (C) Distribution of Close/Far Ratios across prior conditions. Values greater than 1 indicate judgments that the close line was longer (red), values less than 1 indicate judgments that the far line was longer, and values approximately equal to 1 indicate little or no bias (gray). Percentages denote the proportion of participants in each category (shown in black). Dots represent individual participants. Significance markers indicate results of planned comparisons (*$p < .05$, **$p < .01$, ***$p < .001$, ns = not significant).

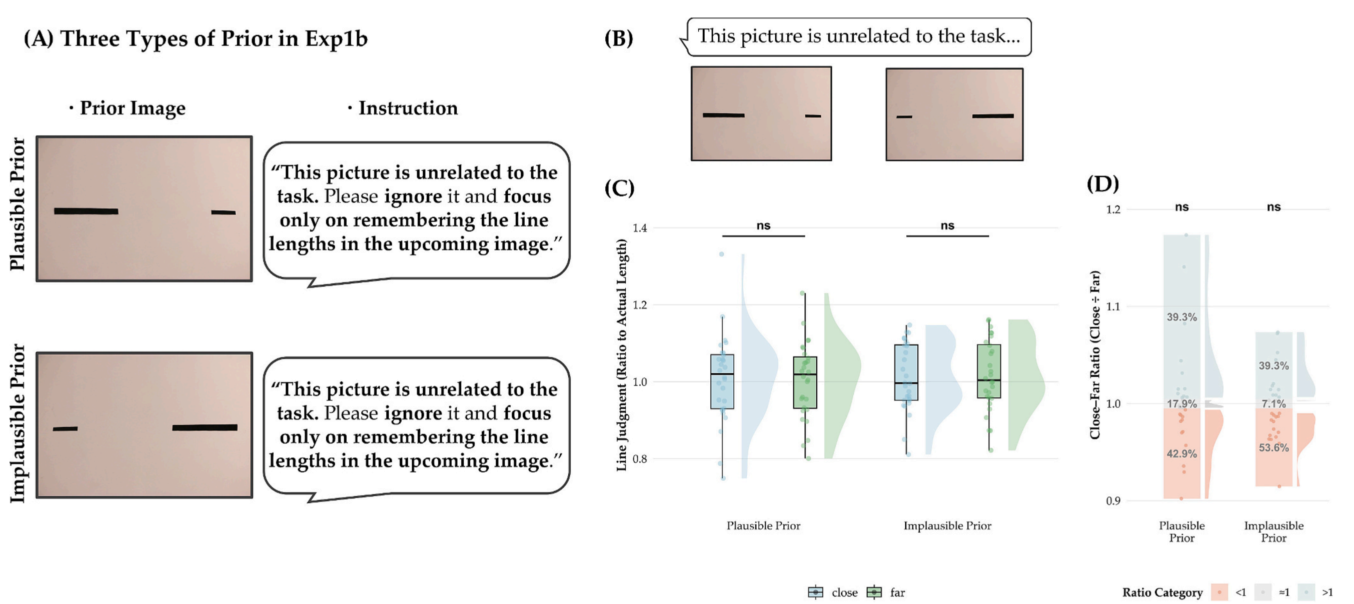

Fig. 6. Stimulus and results of Experiment 1b.
Note. (A) Prior images and corresponding instructions presented at the beginning of Experiment 1b under the Plausible Prior and Implausible Prior conditions. (B) Prior images and instructions used in Experiment 1b. In both conditions, participants were explicitly informed that the image was unrelated to the task. (C) Line-judgment ratios to actual length for close and far lines across prior conditions. (D) Distribution of Close/Far Ratios across conditions. Plotting conventions and significance markers follow those used in Fig. 5.

For the Close/Far Ratio, a 2 (Prior: Plausible Prior, Implausible Prior) × 2 (Ratio: Close/Far Ratio, 1.0) mixed ANOVA revealed a non-significant interaction, $F(1, 54) = 0.295$, $p = .589$, $\eta_p^2 = 0.005$ (additional details are reported in the Supplementary Materials). Follow-up analyses showed that, in the Plausible Prior condition, the Close/Far Ratio did not significantly differ from 1 ($p = .647$). Here, judgments were broadly distributed across response categories: 39.3% judged the close line as longer, 17.9% as approximately equal, and 42.9% as shorter. In the Implausible Prior condition, the Close/Far Ratio did not significantly differ from 1 ($p = .760$). Here, judgments were broadly distributed across response categories: 39.3% judged the close line as longer, 7.1% as approximately equal, and 53.6% as shorter (See Fig. 6).

The results showed that, once participants were told that the prior image was unrelated to the task, the effect disappeared in both the Plausible and Implausible Prior conditions. This indicates that the bias observed in Experiment 1 was not driven by simple priming; rather, participants had treated the prior image as reflecting the agent's viewpoint and integrated it accordingly.

4. Experiment 2

In Experiment 2, we further introduced a wider range of prior conditions to examine how participants integrate prior visual information with their own visual input. Specifically, five prior types were implemented: a plausible prior, a no-prior baseline, and three levels of implausible priors (weak, moderate, and high).

In addition, we developed another set of stimulus materials that served as a conceptual replication. This version preserved the core structure of the task but replaced the vertically aligned lines with a pair of parallel lines arranged in a Ponzo-style configuration. This complementary setting enabled us to assess whether the patterns observed in Experiment 1 would generalize to a different visual layout, and whether the effects were robust across experiments.

4.1. Method

4.1.1. Participants

Since the study employed a 5 × 2 mixed design, we conducted an a priori power analysis using the same parameters as in Experiment 1. The required total sample size was estimated to be 80. To allow for full counterbalancing and potential exclusions, we recruited 110 participants ($M_{age} = 23.2 \pm 4.8$, 47 males).

4.1.2. Design, materials and procedure

The experiment adopted a 5 (Prior: Plausible Prior, No Prior, Weakly Implausible Prior, Moderately Implausible Prior, Strongly Implausible Prior) × 2 (Line Position: Close, Far) mixed design.

In Experiment 2, two identical 60-cm lines were placed on the floor, separated by 160 cm and positioned 30 cm from the wall. As in Experiment 1, we photographed both agent-viewpoint and participant-viewpoint images to serve as experimental stimuli. The agent was seated either to the left or right of the line configuration, positioned approximately 80 cm from the close line and 125 cm from the wall. The agent's eye height was 125 cm. From this position, the agent was aligned with the midpoint between the two lines and instructed to fixate the center of the configuration. Participant-viewpoint images were captured using a second camera mounted at the midpoint between the two lines, positioned 425 cm from the wall at a height of 140 cm. This camera reproduced the stimulus presented to participants during the task (see Fig. 7; detailed geometric parameters are provided in the Supplementary Materials).

The procedure and counterbalancing followed those of Experiment 1a exactly, except that the stimuli were replaced with this new Ponzo-style configuration.

4.2. Results

Ten participants were excluded for providing at least one zero length judgment. This left a final sample of 100, evenly distributed across the five experimental conditions: Plausible Prior, No Prior, Weakly Implausible Prior, Moderately Implausible Prior, Strongly Implausible

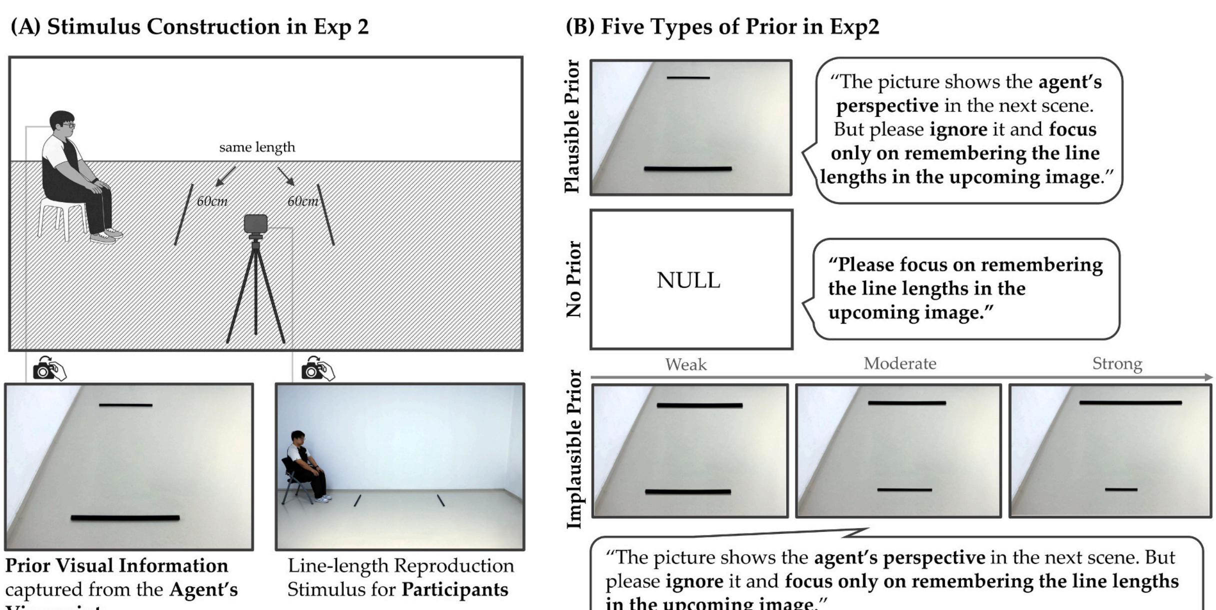

Fig. 7. Schematic illustration of the stimulus setup and prior information in Experiment 2.
Note. (A) Schematic illustration of the stimulus setup used for photographing materials, including the spatial arrangement of the agent, wall, lines, and cameras in Experiment 2. The diagram indicates how the agent-viewpoint and participant-viewpoint images were captured based on their respective viewing geometries. A male agent sitting to the left of the lines is shown here as an illustrative example. (B) Prior images and corresponding instructions shown at the beginning of the experiment under the five prior conditions: Plausible Prior, No Prior, Weakly Implausible Prior, Moderately Implausible Prior, Strongly Implausible Prior.

Prior.

Given that the data in all ten cells met the assumption of normality, parametric tests were conducted. We first performed a 5 (Prior: Plausible Prior, No Prior, Weakly Implausible Prior, Moderately Implausible Prior, Strongly Implausible Prior) × 2 (Line Position: Close, Far) mixed ANOVA on the Ratio-to-Actual Length. The main effect of Prior was not significant, $F(4, 95) = 1.434$, $p = .229$, $\eta_p^2 = 0.057$, nor was the main effect of Line Position, $F(1, 95) < 0.001$, $p = .991$, $\eta_p^2 < 0.001$. Crucially, the Prior × Line Position interaction was significant, $F(4, 95) = 6.357$, $p < .001$, $\eta_p^2 = 0.211$. Follow-up simple-effects analyses revealed a clear pattern across prior conditions. In the Plausible Prior condition, close lines ($M = 1.040$) were judged significantly longer than far lines ($M = 0.980$), $p < .001$. In the No Prior condition, there was no significant difference between close ($M = 1.013$) and far lines ($M = 1.032$), $p = .216$. Similarly, in the Weakly Implausible Prior condition, there was no significant difference between close ($M = 0.991$) and far lines ($M = 1.002$), $p = .471$. In the Moderately Implausible Prior condition, close lines ($M = 1.059$) were judged significantly shorter than far lines ($M = 1.098$), $p = .010$. However, in the Strongly Implausible Prior condition, there was no significant difference between close ($M = 0.979$) and far lines ($M = 0.969$), $p = .532$ (see Fig. 8).

For the Close/Far Ratio, we observed a largely similar pattern. A 5 (Prior: Plausible Prior, No Prior, Weakly Implausible Prior, Moderately Implausible Prior, Strongly Implausible Prior) × 2 (Ratio: Close/Far Ratio, 1.0) mixed ANOVA revealed a significant interaction, $F(4, 95) = 4.212$, $p = .003$, $\eta_p^2 = 0.151$ (additional details are reported in the Supplementary Materials). Follow-up analyses showed that, in the Plausible Prior condition, the ratio was significantly greater than 1 ($p = .003$), indicating that participants reliably judged the close line as longer. Descriptively, 80% of participants judged the close line as longer, and 20% judged the close line as shorter. In the No Prior condition, the Close/Far Ratio did not significantly differ from 1 ($p = .417$). Here, judgments were broadly distributed across response categories: 30% judged the close line as longer, 5% as approximately equal, and 65% as shorter. A similar pattern was observed in the Weakly Implausible Prior condition, where the ratio also did not differ from 1 ($p = .452$; 20.0% close longer, 25% approximately equal, 55% close shorter). In the Moderately Implausible Prior condition, the Close/Far Ratio was

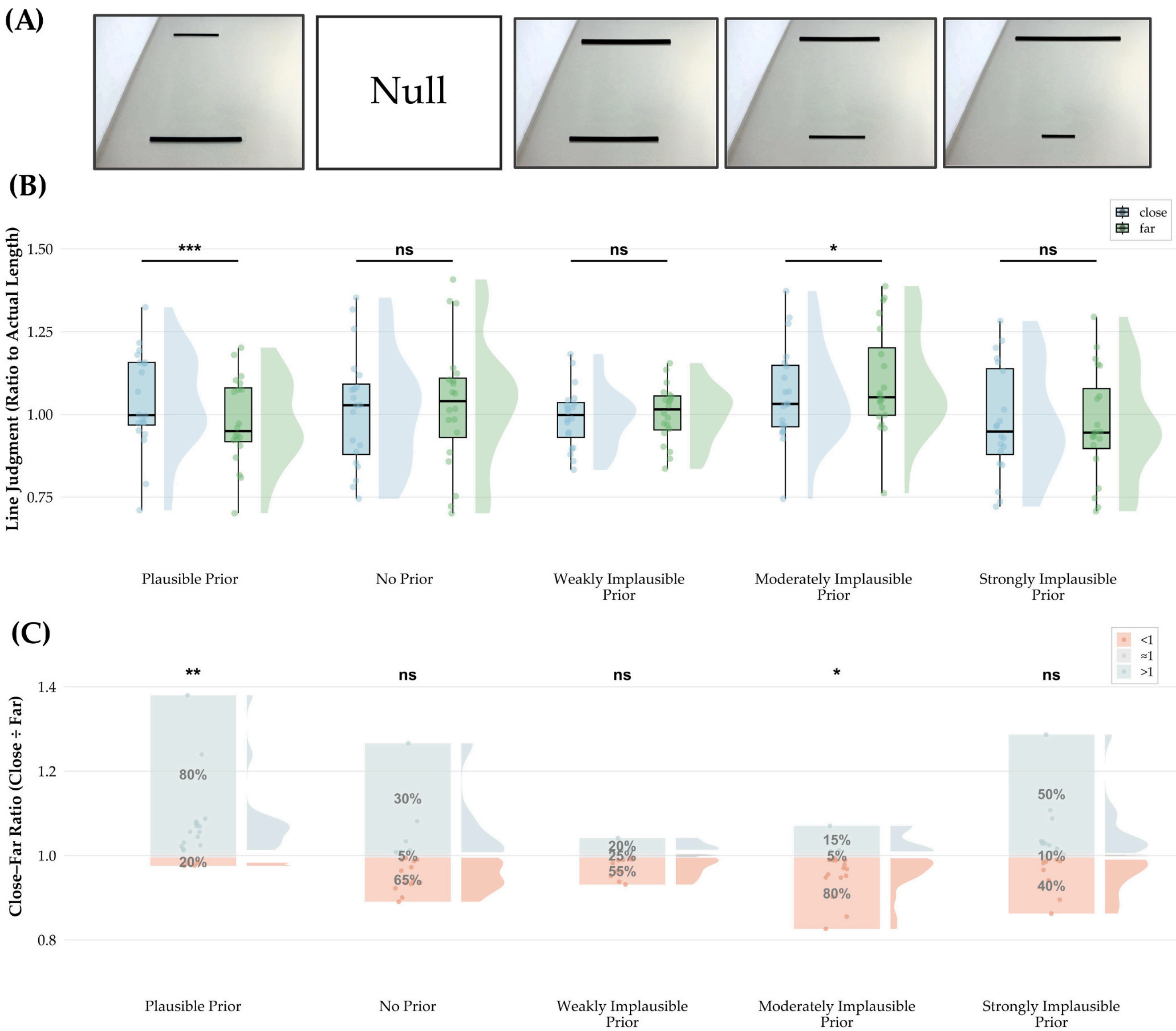

Fig. 8. Stimulus and results of Experiment 2.
Note. (A) Examples of prior images used in Experiment 2, including Plausible Prior, No Prior, Weakly Implausible Prior, Moderately Implausible Prior, and Strongly Implausible Prior. (B) Line-judgment ratios to actual length for close and far lines across prior conditions. (C) Distribution of Close/Far Ratios across prior conditions. Plotting conventions and significance markers follow those used in Fig. 5.

significantly smaller than 1 ($p = .024$), suggesting a tendency to judge the far line as longer (15% close longer, 5% approximately equal, 80% close shorter). In contrast, in the Strongly Implausible Prior condition, the ratio again did not differ from 1 ($p = .363$), with judgments broadly distributed across response categories (50% close longer, 10% approximately equal, 40% close shorter) (see Fig. 8).

Importantly, although neither the No Prior nor the Weakly Implausible Prior condition showed a mean difference between close and far line judgments in either the Ratio-to-Actual-Length or Close/Far Ratio analyses, response variability differed reliably between the two conditions. Levene's tests revealed that responses were consistently more stable in the Weakly Implausible Prior condition than in the No Prior condition—both for Ratio-to-Actual-Length in separate close or far line judgments ($p = .019$ and $p = .015$, respectively) and for the Close/Far Ratio itself ($p = .047$; see details in Supplementary Materials). This pattern suggests that weakly implausible prior information reduced response variability even in the absence of a systematic mean bias.

The results revealed an informative pattern that both replicated and extended the findings of Experiment 1. Under the Plausible Prior, participants again showed a clear spontaneous VPT bias consistent with the agent's perspective, judging the close line as longer. No such bias emerged in the No Prior condition. Critically, under the Weakly and Moderately Implausible Prior conditions, participants' judgments continued to align with the prior information: responses either shifted toward equality (with reduced variance relative to the No Prior condition) or showed a bias favoring the farther line. However, when the prior became strongly implausible, this bias disappeared entirely. This graded pattern supports a *heuristic account*: observers integrate prior information with their own visual input only when the prior remains within a plausibility range. Once the prior exceeds this threshold, it is no longer incorporated, and judgments revert to reliance on their own visual input. We return to the implications of this plausibility-based integration process in the General Discussion.

5. General discussion

The present research aimed to fill a critical gap in current VPT research by directly adjudicating between *perceptual simulation theory* and *heuristic accounts* in implicit VPT. Building on Samuel et al. (2021), we adapted their paradigm into an implicit task and systematically manipulated prior visual information from the agent's viewpoint to test competing predictions of the two accounts across two experiments.

Across both experiments, the results consistently favored the *heuristic account* over *perceptual simulation theory*. In Experiment 1, participants' judgments were influenced by prior information not only when the prior was plausible, but also—albeit less stably—when it was implausible. Crucially, the effect under implausible priors was weaker and less robust (i.e., not significant in the Close/Far Ratio analysis), suggesting that participants were not blindly following the prior but instead integrating it with their own visual input in a plausibility-based manner. Experiment 2 further clarified this pattern by introducing a graded manipulation of prior plausibility. Participants integrated prior information when it was plausible, weakly implausible, or moderately implausible, but ceased to do so when the prior became strongly implausible.

Importantly, these findings rule out a critical alternative interpretation of *perceptual simulation theory*: namely, that prior failures to observe simulation effects merely reflected observers' lack of prior knowledge regarding what an agent sees from its viewpoint. However, our results indicate that even when participants were explicitly provided with this prior visual information, they still failed to selectively reconstruct the agent's veridical visual experience. Instead, even implausible priors were incorporated into participants' judgments, suggesting that agent-related information acted as a potent cue that interfered with their own perceptual judgments. In other words, participants spontaneously integrated others' viewpoints with their own, rather than simulating another's visual experience in a manner that faithfully maps to veridical visual input.

A methodological consideration worth noting is that the present design does not involve 60 independent stimuli per participant, but rather 60 repeated judgments of the same physical stimulus configuration. Although this approach increases measurement precision at the individual level, it also raises the possibility that effects observed across trials may originate from judgments formed on the initial trials, with subsequent responses reflecting repetition around an initial judgment rather than integration processes per se. To address this possibility, we conducted complementary analyses focusing exclusively on first-trial responses. Across the 10 conditions of 3 experiments, the qualitative pattern observed in the first-trial analyses closely mirrored that of the main analyses based on all trials. One notable exception occurred in Experiment 2 for the Strongly Implausible Prior condition: whereas no reliable bias was observed in the main analysis, first-trial judgments showed a significant deviation, with the closer line judged as shorter than the farther line. Importantly, trial-by-trial inspection revealed that this deviation attenuated rapidly over subsequent trials, with judgments converging toward the neutral reference value (see Supplementary Materials). This temporal profile suggests that initial exposure to strongly implausible prior information may transiently bias perceptual judgments, but that such influences are not sustained once participants accumulate task-relevant perceptual evidence. Crucially, this pattern underscores the value of repeated measurements in the present paradigm. While first-trial judgments may be disproportionately influenced by short-lived effects following prior exposure, repeated trials provide a more reliable estimate of how prior information is integrated—or discounted—over time. Future work may further examine the temporal dynamics by which initial biases give way to more stable integrative judgments.

A particularly informative aspect of Experiment 2 was the emergence of a nonlinear, plausibility-based integration pattern: prior information influenced judgments only within a limited range of plausibility. This pattern closely resembles principles from cue integration research (Ernst & Banks, 2002), which show that humans combine multiple sensory information sources by weighting them according to their plausibility, down-weighting or discarding cues when conflicts become too large. Although the present study does not involve multiple sensory modalities, a similar computational logic may apply here: participants appear to evaluate the plausibility of prior visual information about the agent and integrate it with their own visual input only when it falls within an acceptable range.

Crucially, this integration does not appear to rely on fully rational computation that demands substantial computational resources, as assumed by full Bayesian models. Instead, the observed pattern is more consistent with heuristic, resource-rational inference (Griffiths, Lieder, & Goodman, 2015; Hawkins, Gweon, & Goodman, 2021), in which observers draw on coarse, experience-derived intuitions—namely, folk beliefs about how vision works (i.e., naïve optics)—to integrate prior information with their own perceptual judgments. Importantly, while such intuitions may occasionally be erroneous, they function as fast, resource-efficient judgmental shortcuts; simultaneously, they reflect regularities abstracted from everyday visual experience (e.g., objects tend to appear larger at closer distances, and the same object can appear different from different viewpoints). Furthermore, these heuristics are context-sensitive and selectively engaged: in the absence of prior information, judgments were relatively variable; by contrast, the presence of a moderately implausible prior stabilized responses in a direction consistent with that prior. When prior information violated plausibility constraints too severely, however, it was effectively discounted and no longer incorporated into the judgment process.

Taken together, these findings suggest that implicit VPT may rely on the same heuristic processes as explicit VPT. However, implicit VPT tasks may be particularly well suited for probing underlying mechanisms, for instance, prior information can be manipulated without explicitly instructing participants to reason about another's perspective.

In contrast, explicit tasks are more likely to invite strategic, top–down correction or to directly reveal the target perspective, potentially obscuring the processes of interest. Importantly, implicit paradigms also place stronger demands on experimental control. Because participants are not explicitly guided to adopt a perspective, alternative explanations—such as general attentional orienting or priming—must be carefully ruled out through clean task designs and appropriate control conditions (Cole et al., 2022).

More broadly, our results highlight the need for computational models of VPT. Although the present findings suggest that spontaneous VPT operates via heuristic inference and plausibility-based integration, the underlying computational mechanisms remain underspecified. For instance, future work should formalize how the plausibility of prior information is represented, how such priors are weighted relative to one's own visual input, and how folk beliefs about vision are represented and recruited during judgment. Addressing these questions will likely require computational models that make these components explicit, allowing competing theoretical accounts to be evaluated against observed behavioral patterns. By integrating psychophysical paradigms with modeling, future research can move toward a mechanistic account of when and how humans spontaneously adopt—or selectively disregard—others' visual perspectives.

CRediT authorship contribution statement

Xucong Hu: Writing – original draft, Software, Methodology, Formal analysis, Conceptualization. **Yitong Zheng:** Writing – original draft, Visualization, Data curation. **Qinyi Hu:** Visualization, Methodology. **Hui Chen:** Resources, Investigation. **Mowei Shen:** Supervision, Funding acquisition. **Jifan Zhou:** Writing – review & editing, Supervision, Project administration, Conceptualization.

Compliance with ethical standards

The study was approved by the Institutional Review Board (IRB) of the Department of Psychology and Behavioral Sciences at Zhejiang University (Approval No. 2023–002).

Funding

This research was supported by the National Natural Science Foundation of China (32371088, 62337001, and 32471093), the Space Medical Experiment Project of the China Manned Space Program (HYZHXMR01003), the Science and Technology Innovation 2030—"Brain Science and Brain-like Research" Major Project (2022ZD0210800), and the Fundamental Research Funds for the Central Universities (226-2024-00118).

Declaration of competing interest

The authors have no competing interests to declare.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cognition.2026.106478.

Data availability

All data are openly available on the Open Science Framework (OSF): https://osf.io/vs5c4/.

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.
Blech, C., Lembcke, H., Bouquet, C. A., & Liepelt, R. (2025). Level-1-visual perspective taking for human and robot avatars. *Psychological Research, 89*(6), 176.
Cole, G. G., & Millett, A. C. (2019). The closing of the theory of mind: A critique of perspective-taking. *Psychonomic Bulletin & Review, 26*(6), 1787–1802.
Cole, G. G., Millett, A. C., Samuel, S., & Eacott, M. J. (2020). Perspective-taking: In search of a theory. *Vision, 4*(2), 30.
Cole, G. G., Samuel, S., & Eacott, M. J. (2022). A return of mental imagery: The pictorial theory of visual perspective-taking. *Consciousness and Cognition, 102*, Article 103352.
De Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a web browser. *Behavior Research Methods, 47*, 1–12.
Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature, 415*(6870), 429–433.
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.
Firestone, C., & Scholl, B. J. (2016). Cognition does not affect perception: Evaluating the evidence for "top-down" effects. *Behavioral and Brain Sciences, 39*, Article e229.
Flavell, J. H. (1994). *Cognitive development: Past, present, and future*.
Griffiths, T. L., Lieder, F., & Goodman, N. D. (2015). Rational use of cognitive resources: Levels of analysis between the computational and the algorithmic. *Topics in Cognitive Science, 7*(2), 217–229.
Hawkins, R. D., Gweon, H., & Goodman, N. D. (2021). The division of labor in communication: Speakers help listeners account for asymmetries in visual perspective. *Cognitive Science, 45*(3), Article e12926.
Hayhoe, M., & Ballard, D. (2005). Eye movements in natural behavior. *Trends in Cognitive Sciences, 9*(4), 188–194.
He, G., Song, W., Dong, P., Xu, C., Zhou, Y., Ye, H., & Zhang, Q. (2024). Is working memory necessary for both selection and calculation in level 1 visual perspective taking? Evidence from children and adults. *Behavioral Science, 14*(10), 897.
Kosslyn, S. M., Pinker, S., Smith, G. E., & Shwartz, S. P. (1979). On the demystification of mental imagery. *Behavioral and Brain Sciences, 2*(4), 535–548.
O'Grady, C., Scott-Phillips, T., Lavelle, S., & Smith, K. (2020). Perspective-taking is spontaneous but not automatic. *Quarterly Journal of Experimental Psychology, 73*(10), 1605–1628.
Prinzmetal, W., Shimamura, A. P., & Mikolinski, M. (2001). The Ponzo illusion and the perception of orientation. *Perception & Psychophysics, 63*(1), 99–114.
Pylyshyn, Z. W. (2002). Mental imagery: In search of a theory. *Behavioral and brain sciences, 25*(2), 157–182.
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.
Samuel, S., Cole, G. G., & Eacott, M. J. (2023). It's not you, it's me: A review of individual differences in visuospatial perspective taking. *Perspectives on Psychological Science, 18*(2), 293–308.
Samuel, S., Eacott, M. J., & Cole, G. G. (2022). Visual perspective taking without visual perspective taking. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 48*(7), 959.
Samuel, S., Hagspiel, K., Eacott, M. J., & Cole, G. G. (2021). Visual perspective-taking and image-like representations: We don't see it. *Cognition, 210*, Article 104607.
Santiesteban, I., Catmur, C., Hopkins, S. C., Bird, G., & Heyes, C. (2014). Avatars and arrows: Implicit mentalizing or domain-general processing? *Journal of Experimental Psychology: Human Perception and Performance, 40*(3), 929.
Shen, M., Xu, H., Zhang, H., Shui, R., Zhang, M., & Zhou, J. (2015). The working memory Ponzo illusion: Involuntary integration of visuospatial information stored in visual working memory. *Cognition, 141*, 26–35.
Tomasello, M. (2009). *The cultural origins of human cognition*. Harvard university press.
Ward, E., Ganis, G., & Bach, P. (2019). Spontaneous vicarious perception of the content of another's visual perspective. *Current Biology, 29*(5), 874–880.
Ward, E., Ganis, G., McDonough, K. L., & Bach, P. (2020). Perspective taking as virtual navigation? Perceptual simulation of what others see reflects their location in space but not their gaze. *Cognition, 199*, Article 104241.
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.
Wegner, D. M., & Bargh, J. A. (1998). Control and automaticity in social life. *The Handbook of Social Psychology, 2*(4).
Zhou, J., Peng, Y., Li, Y., Deng, X., & Chen, H. (2022). Spontaneous perspective taking of an invisible person. *Journal of Experimental Psychology: Human Perception and Performance, 48*(11), 1186.