Spontaneity of Level-1 Visual Perspective Taking in Multi-Object Contexts

Authors: Dou Yan, Li Jing, Li Jing

Date: 2023-12-06T00:00:00+00:00

Abstract

The present study designed three experiments to investigate the spontaneous manifestation of first-level visual perspective-taking in multi-object scenarios. Experiment 1 adapted the classic perspective-taking task paradigm and verified that the process of first-level visual perspective-taking can occur spontaneously. Experiment 2 introduced an additional virtual agent, and the results revealed that in the presence of multiple agents, particularly when judging from the self-perspective, consistency in the number of objects seen by the virtual agent would produce a group perspective effect, influencing task performance. Experiment 3 set the number of objects seen by the virtual agent to be inconsistent, demonstrating that consistency in gaze direction between agents also captures participants' attention, generating differential effects under self-perspective and other-perspective conditions. In summary, in multi-object scenarios, regardless of whether that perspective is the target perspective, humans highly flexibly and spontaneously adopt others' perspectives; this process is influenced by the tripartite relationship among 'self-target agent-irrelevant agent,' exhibiting context-specific manifestations.

Full Text

Preamble

The Spontaneity of Level-1 Visual Perspective Taking in Multi-Avatar Contexts*

DOU Yan[1,2], **LI Jing**[1]

[1]School of Psychology, Nanjing Normal University, Nanjing 210097, China
[2]Office of Teaching Quality Management, Jiangsu Open University, Nanjing 210036, China

Abstract

This study designed three experiments to investigate the spontaneous manifestation of Level-1 visual perspective taking in multi-avatar contexts. Experiment 1 adapted the classic perspective-taking task paradigm to verify that Level-1 visual perspective taking can occur spontaneously. Experiment 2 added an additional avatar and found that in the presence of multiple avatars, particularly when judging from the self-perspective, consistency in the number of objects seen by the avatars produced a group-perspective effect that influenced task performance. Experiment 3 fixed the number of objects seen by the avatars as inconsistent and demonstrated that consistency in the line of sight between avatars also captured participants' attention, exerting differential effects under self-perspective and other-perspective conditions. In summary, in multi-avatar contexts, individuals spontaneously and flexibly compute others' perspectives regardless of whether those perspectives are task-relevant. This process is influenced by the relationship among the "self-target avatar-irrelevant avatar" triad, exhibiting context-specific patterns.

Keywords: Level-1 visual perspective taking, spontaneity, multiple avatars, egocentric intrusion, altercentric intrusion

Classification Code: B842

Introduction

In human social interaction, reducing ambiguity in communication often requires us to understand others by putting ourselves in their shoes, inferring and predicting their mental states (such as beliefs, attitudes, intentions, and desires) based on their actions and expressions (Samson et al., 2010). Perspective taking serves as a cornerstone of social and spatial interaction, playing an increasingly vital role in our communication with others, such as in empathizing with them (Shao Yuting et al., 2020). Zhang Wenxin (1998) synthesized various definitions, conceptualizing perspective taking as the ability to represent what others see and accurately infer their perspectives or viewpoints based on distinguishing one's own perspective from that of others (Surtees et al., 2013a, 2013b). Its essential characteristic lies in decentering at the cognitive level—being aware of potential differences between self and other perspectives and viewing issues from others' standpoints to achieve accurate understanding and prediction. Researchers have categorized perspective taking into social perspective taking, spatial perspective taking, and visual perspective taking according to the nature of the objects and targets involved (Surtees et al., 2013a).

Visual perspective taking constitutes a fundamental element of human social cognition, reflecting the capacity to perceive others' mental states and thus considered a low-level mechanism of the theory-of-mind system (Apperly & Butterfill, 2009; Schurz et al., 2015). Numerous studies on children and non-human animals indicate that visual perspective taking is not a unitary ability but can be divided into Level-1 and Level-2 visual perspective taking (Flavell et al., 1981). Level-1 visual perspective taking involves understanding what others see—that is, judging what objects others can see (What)—and comprehending that due to occlusion, oneself and others may not see the same objects. Level-2 visual perspective taking entails a more complex judgment of how others see the world (How), understanding that objects visible to both self and other can produce different visual impressions and experiences under different viewing conditions (Zuo Tingting & Hu Qingfen, 2019; Flavell et al., 1981; Kessler & Rutherford, 2010; Surtees et al., 2013b).

Traditional theory-of-mind perspectives posit that perspective taking is a relatively slow, cognitively demanding complex process that is non-spontaneous (Apperly, 2010; Apperly & Robinson, 2003; Gopnik & Meltzoff, 1998). In contrast, recent researchers tend to view the process of adopting others' perspectives as spontaneous and rapid, though these operations can be influenced by intention, attention, or other factors (O'Grady et al., 2020; Westra, 2017).

The dot-perspective task, developed by Samson et al. (2010), is commonly employed to investigate the spontaneity of Level-1 visual perspective taking. In this paradigm, stimuli consist of scene images featuring an avatar standing in a virtual room, facing either the left or right wall, with 0–3 red dots distributed across both walls. Participants must judge whether the number of dots matches a previously presented numerical cue from a prompted perspective (self or other). Results revealed longer reaction times and higher error rates when the number of dots seen by the participant and the avatar was inconsistent compared to consistent trials, demonstrating a consistency effect. Samson et al. argued that perspectives irrelevant to the judgment task interfered with the decision process, producing differential outcomes between consistent and inconsistent conditions. When the participant's own perspective conflicted with the avatar's, judgments of the avatar's perspective suffered from egocentric intrusion—the spontaneous activation of one's own perspective when judging others'. Conversely, judgments of one's own perspective experienced altercentric intrusion—the spontaneous activation of the avatar's perspective when judging the self (Wu Menghui et al., 2022). This paradigm has since been widely adopted in research on the spontaneity of visual perspective taking, with its conclusions receiving increasing support from scholars (Qureshi et al., 2010; Santiesteban et al., 2014; Surtees & Apperly, 2012).

Researchers in the field of perspective-taking spontaneity mechanisms frequently employ two theoretical frameworks for explanation: implicit mentalizing and submentalizing (Pan Wei et al., 2017; Santiesteban et al., 2014). Implicit mentalizing, also known as implicit theory of mind, posits that infants, children, and adults possess an innate capacity to think about others' mental states, with mental representations of individuals being unconscious and spontaneously generated rather than controlled processes (Cole et al., 2017; Heyes, 2014). Submentalizing, in contrast, represents a facilitation of general cognitive processes (such as attention, spatial cognition, and memory) that does not involve thinking about mental states but stimulates mentalizing in social contexts. Both positions have accumulated substantial evidence yet remain unable to refute each other. Consequently, Capozzi and Ristic (2020) proposed an integrated mechanism: submentalizing (e.g., gaze direction, orientation) and implicit mentalizing (e.g., intention, interest) can operate simultaneously and independently, with domain-general cognitive activities having limited explanatory power for social cognitive processes when considered alone, and implicit mentalizing serving as a supplement and modulation of their social significance.

To this end, this study designed three experiments to investigate the spontaneous manifestation of Level-1 visual perspective taking when two perspective-taking targets are present, compared with previous classic research, and to provide tentative explanations for the underlying mechanisms. The classic dot-perspective task paradigm (Samson et al., 2010) is better suited for scenarios with fewer avatars or relatively uniform dot counts across avatars (e.g., Capozzi et al., 2014). Given the numerous condition variations in the present research, we adapted the partitioned-room scene developed by Mattan et al. (2015). In their study, a room divided by several walls contained two avatars facing away from participants, wearing different colored clothes to represent self and other perspectives while oriented toward different walls. Results showed that when the target perspective was the self-representing avatar, participants tended to prioritize the self-related perspective over the other (irrelevant) perspective, supporting the self-prioritization effect hypothesis. Although this experiment differed from our research focus on spontaneity in multi-avatar visual perspective taking and its procedure diverged from the classic task (Samson et al., 2010), its scene configuration aligned perfectly with our intended presentation. Moreover, the number of avatars could be adjusted to one or two according to research purposes, providing a suitable paradigm reference. Additionally, the avatars in this paradigm always maintained a perspective equivalent to participants, gazing at distant walls, thereby eliminating potential confounds from visuospatial transformation or mental rotation—making it appropriate for our investigation.

Therefore, Experiment 1 adapted this paradigm for validation purposes, exploring whether individuals spontaneously adopt others' perspectives in the presence of a single avatar within the new scene configuration, with results serving as a baseline for subsequent studies. Building upon a viable paradigm from Experiment 1, Experiment 2 introduced an additional avatar to examine how different avatars' perspectives influence perspective-taking spontaneity in a two-avatar context. Experiment 3 further investigated whether multiple avatars' gaze directions affect the spontaneous generation of perspective taking based on Experiment 2. We hypothesized that: (1) in the adapted paradigm with a single avatar present, Level-1 visual perspective taking would occur spontaneously, consistent with previous classic research; (2) with multiple avatars present, Level-1 visual perspective taking would remain spontaneous, and consistency in the number of objects seen by avatars would produce cognitive facilitation, thereby promoting this spontaneous generation; and (3) even when avatars saw inconsistent numbers of objects, Level-1 visual perspective taking would still occur spontaneously, with line-of-sight consistency serving as an important attentional cue that would exert influence to some extent.

Experiment 1

Method

Participants We used G*Power 3.1 to estimate the required sample size (α = 0.05, 1–β = 0.90, effect size f = 0.25), which yielded 30 participants. The experiment ultimately recruited 32 university students (16 male, 16 female) with normal or corrected-to-normal vision. All three experiments in this study were approved by the university's biomedical research ethics committee (approval number NNU202106018), and all participants voluntarily signed informed consent forms prior to participation.

Design This experiment employed a 2 (judgment perspective: self, other) × 2 (self-other object consistency: consistent, inconsistent) within-subjects design. Judgment perspective referred to whether participants judged from their own perspective or the avatar's perspective. Self-other object consistency indicated whether the number of black dots seen by the participant and the avatar was consistent. The dependent variables were accuracy and correct response time.

Apparatus and Stimuli The experimental apparatus consisted of a 20-inch Lenovo computer monitor with a resolution of 1920$×$1080 pixels and a refresh rate of 60 Hz. The experimental program was written in E-Prime 2.0. The stimulus images are illustrated in Figure 1 [Figure 1: see original paper]. The virtual room in the images was modeled using 3D Max software, following the design of Mattan et al. (2015, Experiment 1). The room comprised three exterior walls (left wall, right wall, back wall) and two interior partition walls, dividing the space into three sections. Black dots were positioned at the middle horizontal level of the back wall. To ensure rapid and effortless counting by participants, the number of dots presented never exceeded three per trial (Trick & Pylyshyn, 1994). The avatar, dressed in blue, appeared randomly at either the left or right interior partition wall. Due to the partition walls and head orientation constraints, the avatar could only see dots on the wall it faced. Participants viewed the scene from a 45° overhead angle facing the screen, enabling them to observe from both their own perspective and the avatar's perspective simultaneously.

Procedure Figure 1 shows example stimuli from Experiment 1 (a: self-other object consistent; b: self-other object inconsistent). The experimental procedure is illustrated in Figure 2 [Figure 2: see original paper], with details following the protocol of Samson et al. (2010, Experiment 1). Each trial began with a red fixation cross "+" presented at the center of the screen for 750 ms. Following a 500 ms blank screen, the Chinese character "你" (you) or "他" (he) appeared for 750 ms, instructing participants to judge from their own perspective or the avatar's perspective. After another 500 ms blank screen, a numerical cue (0–3) indicating the number of black dots was presented for 750 ms. Following a final 500 ms blank screen, the virtual room image appeared. Participants then had to judge whether the number of dots seen from the specified perspective matched the previously presented numerical cue, pressing the "F" key for matches and the "J" key for non-matches (e.g., as shown in Figure 2, three dots were visible from the "you" perspective—i.e., the participant's perspective—while two dots were visible from the "he" perspective). Participants were required to respond within 2000 ms; otherwise, the trial automatically advanced to the next one.

After reading the instructions, participants entered a practice phase consisting of 8 trials with correctness feedback provided after each trial. Participants proceeded to the formal experiment only after fully understanding the task rules; otherwise, they repeated the practice phase. Participants were instructed to respond as quickly as possible while maintaining accuracy. The formal experiment comprised 200 test trials, including 96 match trials and 96 non-match trials (48 trials each from self/avatar perspective), plus 8 filler trials. The experiment was divided into four blocks of 50 trials each, with each block containing 48 test trials and 2 filler trials. The trial order within each block was arranged to prevent three consecutive trials of the same type. In half of the trials, the number of dots seen by the avatar was consistent with that seen by the participant (self-other object consistent condition), while in the other half, these numbers were inconsistent (self-other object inconsistent condition). Additionally, the avatar's orientation, position, presentation frequency of numerical cues (0–3), and key-response mapping were all counterbalanced across the experiment.

Data Analysis We conducted repeated-measures ANOVA on the experimental data using SPSS 25.0. Only response time data from correct responses to stimulus images were included in the statistical analysis; response times from incorrect responses and filler trials were excluded.

Results and Analysis

Participants' mean accuracy across all stimulus images was 93.85%, with a mean correct response time of 768.49 ms. A correlation analysis between accuracy and response time across all participants yielded a Pearson correlation coefficient of 0.09 (p = 0.63 > 0.05), indicating no correlation between speed and accuracy during the experiment and thus no speed-accuracy tradeoff. The mean accuracy and correct response times for the 32 participants across different conditions are presented in Table 1 .

A 2$\times 2 repeated-measures ANOVA on accuracy yielded the following results: (1) The main effect of judgment perspective was not significant, $F(1, 31) = 0.01, p = 0.93$. (2) The main effect of self-other object consistency was significant, $F(1, 31) = 32.31, p < 0.001, \eta p^{2}$ = 0.51, 95% CI = [0.03, 0.07], with higher accuracy in the self-other object consistent condition (M = 0.97, SE = 0.01) than in the inconsistent condition (M = 0.92, SE = 0.01). (3) The interaction between judgment perspective and self-other object consistency was not significant, $F(1, 31) = 0.12$, $p = 0.93$.

A 2$\times$2 repeated − measures ANOVA on correct response times revealed : (1) A significant main effect of judgment perspective, $F(1, 31) = 53.83, p < 0.001, \eta p^{2}$ = 0.64, 95% CI = [–91.45, –51.66], with shorter correct response times when judging from the self-perspective (M = 732.72 ms, SE = 29.01 ms) than from the other-perspective (M = 804.27 ms, SE = 29.24 ms). (2) A significant main effect of self-other object consistency, $F(1, 31) = 96.09$, $p < 0.001$, $p^2 = 0.76$, 95% CI = [–80.78, –52.96], with faster correct response times in the self-other object consistent condition (M = 735.06 ms, SE = 27.38 ms) than in the inconsistent condition (M = 801.93 ms, SE = 30.38 ms). (3) A significant interaction between judgment perspective and self-other object consistency, $F(1, 31) = 42.86$, $p < 0.001$, $p^2 = 0.58$.

Simple effects analysis indicated that in the self-judgment perspective condition, correct response times were significantly shorter when self-other object consistency was present (M = 715.49 ms, SE = 27.87 ms) than when it was absent (M = 749.94 ms, SE = 30.72 ms), $F(1, 31) = 16.12$, $p < 0.001$. In the other-judgment perspective condition, correct response times were also significantly shorter under self-other object consistency (M = 754.62 ms, SE = 27.72 ms) than under inconsistency (M = 853.92 ms, SE = 31.24 ms), $F(1, 31) = 143.93$, $p < 0.001$. However, the regression slope of self-other object consistency affecting correct response time was steeper in the other-judgment perspective condition, as illustrated in Figure 3 [Figure 3: see original paper].

Discussion

Figure 3 shows correct response times under self-other object consistency conditions. The results of this experiment demonstrate significant main effects of self-other object consistency on both accuracy and correct response time, replicating the "consistency effect" found in Samson et al.'s (2010) study. Whether judging from self- or other-perspective, participants showed lower accuracy and slower response times when the number of black dots they saw was inconsistent with what the avatar saw, compared to consistent trials. This indicates that participants found it difficult to ignore irrelevant perspectives, and inconsistency with the avatar's visual objects interfered with their judgments. Moreover, both perspectives were spontaneously activated at an implicit level: when participants judged from their own perspective, the avatar's perspective was spontaneously activated, producing altercentric intrusion; when judging from the avatar's perspective, participants' own perspective was spontaneously activated, producing egocentric intrusion. The new paradigm's results indicate that Level-1 visual perspective taking can occur spontaneously with a single avatar present, validating the paradigm's effectiveness. This new paradigm enables subsequent experiments to explore the spontaneity of Level-1 visual perspective taking with multiple avatars. Experiment 2 will build upon this by adding another avatar for further investigation.

Experiment 2

Method

Participants Using G*Power 3.1 to estimate sample size (α = 0.05, 1–β = 0.90, effect size f = 0.25) yielded 20 participants. The experiment ultimately recruited 34 undergraduate and master's students from the Nanjing area (17 male, 17 female) with normal or corrected-to-normal vision.

Design This experiment used a 2 (judgment perspective: self, Avatar A) × 2 (self-other object consistency: consistent, inconsistent) × 2 (other-other object consistency: consistent, inconsistent) within-subjects design. Other-other object consistency referred to whether the number of black dots seen from the perspectives of the two avatars—target avatar (A) and irrelevant avatar (B)—was consistent. The dependent variables were accuracy and correct response time.

Apparatus and Procedure The apparatus and procedure were identical to Experiment 1. The stimulus images for Experiment 2 are shown in Figure 4 [Figure 4: see original paper]. The room scene and avatar configurations remained consistent with Experiment 1. Building upon Experiment 1's scene, Experiment 2 added an additional avatar, resulting in stimulus images containing two avatars. To correspond with and distinguish from the judgment perspective cue "you," we assigned Chinese character labels "甲" (A) and "乙" (B) to the two avatars. To prevent participants from distinguishing A and B through other means, we controlled for these extraneous variables: Avatars A and B appeared at the left and right interior partition walls in random positions, both wore identical blue clothing, ensuring participants could only differentiate them by their character labels.

Figure 4 [Figure 4: see original paper] shows example stimuli from Experiment 2 (the top row depicts self-other object consistent conditions; the bottom row depicts self-other object inconsistent conditions; the left column shows other-other object consistent conditions; the right column shows other-other object inconsistent conditions).

The experimental procedure and detailed settings were essentially the same as Experiment 1, except that the Chinese characters changed from "你" or "他" to "你" or "甲," instructing participants to judge from either their own perspective or Avatar A's perspective. The data analysis method was identical to Experiment 1.

Results and Analysis

Participants' mean accuracy across all stimulus images was 93.29%, with a mean correct response time of 875.31 ms. A correlation analysis between accuracy and response time across all participants revealed a Pearson correlation coefficient of –0.15 ($p = 0.46 > 0.05$), indicating no speed-accuracy tradeoff in the experiment.

The accuracy and correct response times for the 34 participants across different conditions are presented in Table 2 .

A 2$×2×2 repeated−measures ANOVA on accuracy revealed: (1) A significant main effect of judgment perspec
13.18, $p < 0.01$, ηp^{2}$ = 0.29, 95% CI = [0.02, 0.06], with higher accuracy when judging from the self-perspective (M = 0.95, SE = 0.01) than from the other-perspective (M = 0.91, SE = 0.01). (2) A significant main effect of self-other object consistency, F(1, 33) = 25.10, p < 0.001, p^2 = 0.43, 95% CI = [0.03, 0.07], with higher accuracy in the self-other object consistent condition (M = 0.96, SE < 0.01) than in the inconsistent condition (M = 0.91, SE = 0.01). (3) The main effect of other-other object consistency was not significant, F(1, 33) = 0.53, p = 0.47. (4) A significant interaction between judgment perspective and self-other object consistency, F(1, 33) = 10.80, p < 0.01, p^2 = 0.25. When judging from the other-perspective, accuracy was significantly higher under self-other object consistency (M = 0.96, SE = 0.01) than under inconsistency (M = 0.87, SE = 0.02), F(1, 33) = 21.34, p < 0.001; however, no significant difference emerged when judging from the self-perspective, F(1, 33) = 2.60, p = 0.12. (5) The interaction between judgment perspective and other-other object consistency was not significant, F(1, 33) = 3.24, p = 0.08. (6) The interaction between self-other object consistency and other-other object consistency was not significant, F(1, 33) = 0.77, p = 0.39. (7) The three-way interaction among judgment perspective, self-other object consistency, and other-other object consistency was significant, F(1, 33) = 6.90, p < 0.05, p^2 = 0.17. Specifically, when judging from the self-perspective under self-other object inconsistency, accuracy was significantly higher when other-other object consistency was present (M = 0.96, SE = 0.01) than when it was absent (M = 0.93, SE = 0.01), F(1, 33) = 5.41, p = 0.03. No significant effects were found in other directions, Fs < 2.45, ps > 0.13, as shown in Figure 5 [Figure 5: see original paper].

A 2$×2×2 repeated − measures ANOVA on correct response times revealed: (1) A significant main effect of judgment perspective, $F(1, 33) = 99.37$, $p < 0.001$, ηp^{2}$ = 0.75, 95% CI = [−189.28, −125.11], with shorter correct response times when judging from the self-perspective (M = 796.72 ms, SE = 36.88 ms) than from the other-perspective (M = 953.91 ms, SE = 40.29 ms). (2) A significant main effect of self-other object consistency, F(1, 33) = 121.44, p < 0.001, p^2 = 0.79, 95% CI = [−79.47, −54.70], with faster correct response times under self-other object consistency (M = 841.77 ms, SE = 37.35 ms) than under inconsistency (M = 908.86 ms, SE = 38.50 ms). (3) The main effect of other-other object consistency was not significant, F(1, 33) = 0.04, p = 0.85. (4) A significant interaction between judgment perspective and self-other object consistency, F(1, 33) = 75.07, p < 0.001, p^2 = 0.70. When judging from the other-perspective, correct response times were significantly shorter under self-other object consistency (M = 888.16 ms, SE = 38.75 ms) than under inconsistency (M = 1019.67 ms, SE = 42.48 ms), F(1, 33) = 146.30, p < 0.001; however, this difference was not significant when judging from the self-perspective, F(1, 33) = 0.11, p = 0.75. (5) A significant interaction between judgment perspective and other-other object consistency, $F(1, 33) = 11.04$, $p < 0.01$, $p^2 = 0.25$. When judging from the self-perspective, correct response times were significantly shorter under other-other object consistency ($M = 787.07$ ms, $SE = 37.09$ ms) than under inconsistency ($M = 806.36$ ms, $SE = 37.21$ ms); this effect was not significant when judging from the other-perspective, $F(1, 33) = 2.29$, $p = 0.14$. (6) A significant interaction between self-other object consistency and other-other object consistency, $F(1, 33) = 21.25$, $p < 0.001$, $p^2 = 0.39$. When self-other object consistency was present, correct response times were significantly shorter under other-other object consistency ($M = 825.45$ ms, $SE = 38.13$ ms) than under inconsistency ($M = 858.10$ ms, $SE = 37.05$ ms), $F(1, 33) = 14.42$, $p < 0.001$. When self-other object consistency was absent, correct response times were significantly shorter under other-other object inconsistency ($M = 894.13$ ms, $SE = 36.78$ ms) than under consistency ($M = 923.58$ ms, $SE = 41.11$ ms), $F(1, 33) = 5.63$, $p < 0.05$. (7) The three-way interaction among judgment perspective, self-other object consistency, and other-other object consistency was not significant, $F(1, 33) = 2.58$, $p = 0.12$, as shown in Figure 6 [Figure 6: see original paper].

Discussion

Figure 6 shows correct response times under Experiment 2 conditions. The results from Experiment 2 reveal several key findings: First, the main effect of judgment perspective on both accuracy and correct response time replicated the pattern observed in Experiment 1. Second, the interaction between judgment perspective and self-other object consistency was significant for both measures. However, unlike Experiment 1, Experiment 2 showed the trend of higher accuracy and faster correct response times under self-other object consistency only in the other-perspective condition. This indicates that the consistency effect existed only when judging from the other-perspective, meaning that when participants judged from an avatar's perspective in a multi-avatar context, they could still spontaneously activate their own perspective, demonstrating egocentric intrusion. When participants judged from the self-perspective, the pattern changed, likely because the increased number of avatars and the introduction of other-other object consistency complicated the composition of interference effects during self-perspective judgments.

Third, when judging from the self-perspective, performance was better under other-other object consistency than under inconsistency. A plausible explanation is that in the self-judgment perspective condition, when A and B saw the same number of dots (other-other object consistency), participants recognized this consistency and integrated them into a single interfering perspective. If the two avatars saw different numbers of dots, however, there would be two separate interfering avatar perspectives, resulting in relatively poorer performance. This effect was particularly pronounced when the participant and Avatar A saw inconsistent numbers of dots. This result further demonstrates, from another angle, that spontaneously adopting others' perspectives influences the adoption and judgment of one's own perspective.

Fourth, the interaction between self-other object consistency and other-other object consistency on correct response time indicates that observation and judgment were most efficient—yielding the fastest response times—when the participant and Avatar A saw consistent numbers of dots and when A and B also saw consistent numbers. If the self and A saw inconsistent numbers, their association was broken, and in this case, consistency between A and B actually created interference.

In summary, consistent with Experiment 1, Experiment 2 demonstrated a self-other object consistency effect. However, Level-1 visual perspective taking with multiple avatars differed from single-avatar perspective taking, as other-other object consistency did influence the perspective-taking process to some extent. When facing multiple avatars, Level-1 visual perspective taking could still occur spontaneously. Simultaneously, the introduction of the other-other object consistency variable exerted both facilitative and interfering effects on the perspective-taking process. Additionally, we observed in the experiment that although Avatars A and B might see different numbers of dots, their gaze directions could be the same (e.g., both looking left). This consistency might capture participants' attention (Li Jing et al., 2017), potentially influencing the results. Therefore, Experiment 3 will control for this condition based on Experiment 2 to further investigate how relationships between virtual objects affect the process.

Experiment 3

Purpose

In social interaction, the attentional direction of others' eyes can induce observers to orient attention toward the indicated direction (Zhang Zhijun et al., 2011), and avatars' line of sight also serves as an attentional cue that triggers spontaneous visual perspective taking (Samson et al., 2010). In our paradigm, when avatars see consistent objects, their lines of sight must be inconsistent. However, when avatars see inconsistent objects, the two avatars' lines of sight can be either consistent or inconsistent. Therefore, Experiment 3 aimed to fix the object consistency condition and explore how avatars' line of sight influences perspective-taking spontaneity.

Method

Participants Using G*Power 3.1 to estimate sample size ($\alpha = 0.05$, $1-\beta = 0.90$, effect size f = 0.25) yielded 20 participants. We ultimately recruited 37 undergraduate and master's students from the Nanjing area (18 male, 19 female) with normal or corrected-to-normal vision.

Design This experiment employed a 2 (judgment perspective: self, Avatar A) × 2 (self-other object consistency: consistent, inconsistent) × 2 (other-other line-of-sight consistency: consistent, inconsistent) within-subjects design. Other-other line-of-sight consistency referred to whether the gaze directions of Avatar A and Avatar B were consistent (e.g., both looking left or right). The dependent variables were accuracy and correct response time.

Apparatus and Procedure The apparatus was identical to Experiment 1. Line-of-sight consistency included two avatars facing the left wall simultaneously or facing the right wall simultaneously. Line-of-sight inconsistency involved one avatar facing the first partition wall and the other facing the third partition wall (see Figure 7 [Figure 7: see original paper]), with avatars seeing different numbers of black dots. The number of formal experiment trials and data analysis methods were the same as in Experiment 1.

Figure 7 shows example stimuli from Experiment 3 (a: other-other line-of-sight consistent; b: other-other line-of-sight inconsistent).

Results and Analysis

Participants' mean accuracy across all stimulus images was 93.20%, with a mean correct response time of 850.80 ms. A correlation analysis between accuracy and response time across all participants revealed a Pearson correlation coefficient of –0.36 ($p < 0.05$), indicating that higher accuracy was associated with shorter response times. This suggests that participants did not sacrifice response time for accuracy, and no speed-accuracy tradeoff existed in the experiment. The accuracy and correct response times for the 37 participants across different conditions are presented in Table 3 .

A 2×2×2 repeated-measures ANOVA on accuracy revealed: (1) A significant main effect of judgment perspec 106.09, $p < 0.001$, $\eta p^{2} = 0.75$, 95% CI = [0.04, 0.06], with significantly higher accuracy when judging from the self-perspective (M = 0.96, SE = 0.01) than from the other-perspective (M = 0.91, SE = 0.01). (2) A significant main effect of self-other object consistency, $F(1, 36) = 105.11$, $p < 0.001$, $p^2 = 0.75$, 95% CI = [0.04, 0.06], with significantly higher accuracy under self-other object consistency (M = 0.96, SE = 0.01) than under inconsistency (M = 0.91, SE = 0.01). (3) A significant main effect of other-other line-of-sight consistency, $F(1, 36) = 10.36$, $p < 0.01$, $p^2 = 0.22$, 95% CI = [0.01, 0.03], with significantly higher accuracy when other-other line-of-sight consistency was present (M = 0.94, SE = 0.01) than when it was absent (M = 0.92, SE = 0.01). (4) A significant interaction between judgment perspective and self-other object consistency, $F(1, 36) = 74.76$, $p < 0.001$, $p^2 = 0.68$. When judging from the other-perspective, accuracy was significantly higher under self-other object consistency (M = 0.96, SE = 0.01) than under inconsistency (M = 0.86, SE = 0.01), $F(1, 36) = 129.44$, $p < 0.001$; this difference was not significant when judging from the self-perspective, $F(1, 36) = 0.67$, $p = 0.42$. (5) The interaction between judgment perspective and other-other line-of-sight consistency was not significant, $F(1, 36) < 0.001$, $p = 0.99$. (6) The interaction between self-other object consistency and other-other line-of-sight consistency was not significant, $F(1, 36) = 0.72$, $p = 0.40$. (7) The three-way interaction among judgment perspective, self-other object consistency, and other-other line-of-sight consistency was significant, $F(1, 36) = 12.16$, $p < 0.001$, $p^2 = 0.25$. When judging from the self-perspective under self-other object consistency ($F(1, 36) = 10.60$, $p < 0.01$) or when judging from the other-perspective under self-other object inconsistency ($F(1, 36) = 9.52$, $p < 0.01$), accuracy was significantly higher under other-other line-of-sight consistency (M = 0.97, SE = 0.01 and M = 0.88, SE = 0.01, respectively) than under inconsistency (M = 0.94, SE = 0.01 and M = 0.84, SE = 0.01, respectively). No significant effects were found in other directions, Fs < 1.10, ps > 0.30, as shown in Figure 8 [Figure 8: see original paper].

A 2$\times 2\times 2 repeated - measures ANOVA on correct response times revealed: (1) A significant main effect of judgment perspective, F(1, 36) = 116.77, p < 0.001, \eta p^{2}$ = 0.76, 95% CI = [–152.23, –104.12], with shorter correct response times when judging from the self-perspective (M = 786.71 ms, SE = 28.17 ms) than from the other-perspective (M = 914.89 ms, SE = 27.66 ms). (2) A significant main effect of self-other object consistency, $F(1, 36) = 62.78$, $p < 0.001$, $p^2 = 0.64$, 95% CI = [–66.50, –39.40], with faster correct response times under self-other object consistency (M = 824.32 ms, SE = 27.10 ms) than under inconsistency (M = 877.27 ms, SE = 27.87 ms). (3) A significant main effect of other-other line-of-sight consistency, $F(1, 36) = 11.20$, $p < 0.01$, $p^2 = 0.24$, 95% CI = [6.84, 27.73], with shorter correct response times when line-of-sight was inconsistent (M = 842.16 ms, SE = 27.37 ms) than when it was consistent (M = 858.43 ms, SE = 27.44 ms). (4) A significant interaction between judgment perspective and self-other object consistency, $F(1, 36) = 18.53$, $p < 0.001$, $p^2 = 0.34$. When judging from either the self-perspective ($F(1, 36) = 7.58$, $p = 0.01$) or other-perspective ($F(1, 36) = 70.78$, $p < 0.001$), correct response times were significantly shorter under self-other object consistency (M = 774.27 ms, SE = 28.01 ms and M = 874.38 ms, SE = 28.04 ms, respectively) than under inconsistency (M = 799.15 ms, SE = 29.05 ms and M = 955.40 ms, SE = 28.11 ms, respectively), though the regression slope was steeper in the other-perspective condition. (5) A significant interaction between judgment perspective and other-other line-of-sight consistency, $F(1, 36) = 12.00$, $p < 0.01$, $p^2 = 0.25$. When judging from the other-perspective, correct response times were significantly shorter under line-of-sight inconsistency (M = 896.55 ms, SE = 27.39 ms) than under consistency (M = 933.23 ms, SE = 28.37 ms), $F(1, 36) = 26.79$, $p < 0.001$; this difference was not significant when judging from the self-perspective, $F(1, 36) = 0.07$, $p = 0.79$. (6) A significant interaction between self-other object consistency and other-other line-of-sight consistency, $F(1, 36) = 5.22$, $p < 0.05$, $p^2 = 0.13$. When self-other object consistency was absent, correct response times were significantly shorter under line-of-sight inconsistency (M = 863.01 ms, SE = 27.64 ms) than under consistency (M = 891.54 ms, SE = 28.61 ms), F(1, 36) = 12.88, p < 0.01; this difference was not significant when self-other object consistency was present, F(1, 36) = 0.93, p = 0.34. (7) The three-way interaction among judgment perspective, self-other object consistency, and other-other line-of-sight consistency was not significant, F(1, 36) = 0.96, p = 0.33, as shown in Figure 9 [Figure 9: see original paper].

Because the results for other-other line-of-sight consistency and its related interactions were not entirely consistent between accuracy and correct response time, we calculated efficiency scores (accuracy ÷ response time, in units of 1/ms). The results were as follows: (1) The main effect of other-other line-of-sight consistency was not significant, F(1, 36) = 0.16, p = 0.69. (2) A significant interaction between judgment perspective and other-other line-of-sight consistency, F(1, 36) = 4.34, p = 0.04, p^2 = 0.11. When judging from the self-perspective, efficiency scores were marginally significantly higher under line-of-sight consistency (M = 1.29$×10$^{\text{\{-3\}}, SE = 4.7×10}$\{-5\}$)thanunderinconsistency(M = 1.27×10^{\text{\{-3\}}, SE = 4.9×10}\{-5\}), F(1, 36) = 2.90, p = 0.09; thiseffectwasnotpresentwhenjudgingfromtheother − perspective, F(1, 36) = 1.96, p = 0.17.(3)Theinteractionbetweenself − otherobjectconsistencyandother−otherline−of−sightconsistencywasnotsignificant, F(1, 36) = 1.06, p = 0.31.(4)Thethree−wayinteractionamongjudgmentperspective, self− otherobjectconsistency, andother−otherline−of−sightconsistencywassignificant, F(1, 36) = 12.60, p < 0.01, ηp^{2}$ = 0.26. When judging from the self-perspective under self-other object consistency, efficiency scores were significantly higher under line-of-sight consistency (M = 1.34$×10$^{\text{\{-3\}}, SE = 4.9×10}$\{-5\}$)thanunderinconsistency(M = 1.27×10^{\text{\{-3\}}, SE = 5.0×10}\{-5\}), F(1, 36) = 7.90, p = 0.01.Whenjudgingfromtheother − perspectiveunderself − otherobjectconsistency, efficiencyscoresweresignificantlyhigherunderline− of−sightinconsistency(M = 1.16×10^{\text{\{-3\}}, SE = 3.7×10}\{-5\})thanunderconsistency(M = 1.12×10^{\text{\{-3\}}, SE = 3.5×10}\{-5\}$), F(1, 36) = 9.52, p < 0.01. No significant effects were found in other directions, Fs < 0.64, ps > 0.43.

Experiment 3 replicated the main effects of judgment perspective and self-other object consistency, as well as their interaction, on both accuracy and correct response time, consistent with Experiments 1 and 2. This again demonstrates that Level-1 visual perspective taking remains spontaneously activated even when avatars see inconsistent objects in multi-avatar contexts.

Under the premise of other-other object inconsistency in this experiment, the newly introduced variable "other-other line-of-sight consistency" played an important role in dot-perspective task judgments. Specifically, examining efficiency scores: First, when judging from the self-perspective under self-other object consistency, participants saw black dots on the same wall. The consistency in the two avatars' line of sight focused participants' attention on the total number of black dots, enabling more accurate judgments compared to line-of-sight inconsistency. Second, when judging from the other-perspective, participants needed to focus more on the avatars' perspectives. Consistent line of sight between the two avatars quickly indicated their gaze directions, but since the avatars saw inconsistent objects, participants required more time to differentiate between them, reducing response efficiency. We term this phenomenon—where an irrelevant other (Avatar B) interferes with participants' perspective taking of the target other (Avatar A) when judging from A's perspective—the "irrelevant avatar interference effect." When A and B's lines of sight were consistent, the irrelevant avatar was more easily noticed, creating more pronounced interference. Conversely, when lines of sight were inconsistent, participants could more quickly locate the target avatar's perspective to be adopted and ignore the other irrelevant avatar, resulting in smaller irrelevant avatar interference effects and shorter response times. This result also reveals that participants followed the avatars' line of sight to locate the dots on the walls before making judgment responses (Baker et al., 2016).

The results of Experiment 3 indicate that in multi-avatar contexts, under the uniform premise of other-other object inconsistency, consistency in the two avatars' line of sight also affects participants' attention to objects. Line-of-sight consistency may generate interference from irrelevant avatars, influencing the spontaneous manifestation of Level-1 visual perspective taking to some extent. This influence exists under both self-perspective and other-perspective conditions.

General Discussion

Multi-avatar Perspective Taking Spontaneity

The spontaneity of Level-1 visual perspective taking has long attracted researchers' attention. Building upon this foundation, our study introduced two perspective-taking targets, creating multiple independent non-group perspectives, to investigate how these perspectives influence the spontaneity of Level-1 visual perspective taking. First, Experiment 1 validated the newly adapted paradigm, revealing a consistency effect identical to that found in classic experiments, demonstrating that Level-1 visual perspective taking occurs spontaneously in the new paradigm. Experiment 2 added a new avatar and found that this spontaneity persisted, though its level was influenced by the consistency in the number of objects seen by the two avatars. Experiment 3 further introduced other-other line-of-sight consistency based on Experiment 2, finding that under the premise of inconsistent object numbers seen by the two avatars, their line-of-sight consistency also affected the spontaneous manifestation of perspective taking, with this influence being more complex.

Effects of Other-Other Object Consistency

Beyond conclusions about spontaneity, we primarily investigated how the relationship among the self, target avatar, and irrelevant avatar influences perspective-taking judgment processes. Experiment 2 found that judgment perspective and self-other object consistency each interacted with other-other object consistency. When judging from the self-perspective, if A and B saw the same number of dots, this facilitated self-perspective judgment. Additionally, under self-other object consistency, A and B seeing the same number of dots also facilitated judgment. We propose that in the self-perspective condition, when A and B saw identical dot counts, they could be stored as a group in memory. Furthermore, due to the presence of three overlapping and compatible dimensions—self, target avatar, and irrelevant avatar (Kornblum et al., 1990)—working memory load during judgment was lower compared to when A and B saw different numbers, resulting in better overall performance. However, this facilitative effect disappeared when judging from the other-perspective, presumably because the presence of the self-perspective created a clear egocentric bias (Samuel et al., 2019), making it difficult for consistency to exert a facilitative effect. We borrow the cognitive economy principle common in linguistics (Zipf, 1949, cited in Linders & Louwerse, 2023) to explain this phenomenon: combining two individuals of the same nature into a single entity for judgment processing requires minimal cognitive effort. It is this tendency toward cognitive resource economy that produces the facilitative effect when judging from the self-perspective, also indicating that we cannot suppress the adoption of others' perspectives even when judging from the self-perspective.

In this process, examining correct response time and accuracy, the influence of self-other object consistency appears relatively stable, whereas other-other object consistency only manifests under specific perspective conditions, making its effect less stable than that of self-other object consistency.

When disregarding judgment perspective, the interaction between self-other object consistency and other-other object consistency aligns with previous results. When participants and Avatar A saw different numbers of dots, the association between self and target was broken. If A and B saw the same number of dots in this case, according to the cognitive economy principle, they would be stored as a unified entity, and unified entities attract more attention. Consequently, this unified entity conflicted with the self-perspective, actually prolonging judgment time. This finding is similar to results obtained in group perspective research by He et al. (2021). Different from their study, however, we found that when participants and Avatar A saw the same number of dots, if A and B also saw identical numbers, this facilitated the formation of a self-A-B unified entity, allowing participants to complete all judgments by remembering only one number, resulting in extremely fast judgment speed. If A and B saw different numbers, the number seen by B created interference, slowing judgment speed. This result similarly demonstrates that during task execution, we not only attend to our own perspective and adopt target Avatar A's perspective but also spontaneously attend to the visual objects seen by irrelevant Avatar B.

Effects of Other-Other Line-of-Sight Consistency

Experiment 3 fixed other-other object consistency as an inconsistent condition and found that under this premise, other-other line-of-sight consistency also influenced the specific manifestation of spontaneous perspective taking. When judging from the self-perspective under self-other object consistency, if A and

B's lines of sight were consistent, the special nature of this cue led participants to perceive them again as a unified entity, resulting in better judgment performance compared to line-of-sight inconsistency. Pesimena and Soranzo (2023) reported similar conclusions, noting that the perspective-taking target's line of sight causes participants to spontaneously orient their own gaze toward the location indicated by the target's gaze. When multiple cues share the same direction, this orienting effect is additive; when these directions conflict, the orienting effect is canceled out. However, the same situation created strong interference for other-perspective judgments. Specifically, when judging from the other-perspective, if A and B's lines of sight were identical and participants saw the same number of dots as Avatar A, judging Avatar A's perspective involved both connecting with the participant and being constrained by Avatar B. This constraining effect manifested as follows: In this case, Avatar A as the target object carried two conflicting stimulus attributes—line-of-sight consistency with irrelevant Avatar B and object inconsistency—while responses needed to be object-based. According to Kornblum and Lee's dimensional overlap model (1995), as two relevant attributes, their conflict inevitably leads to changes in response performance, producing the aforementioned results. This indicates that when judging from others' perspectives, we not only spontaneously activate our own perspective but also follow avatars' lines of sight to observe corresponding objects. Consistent line of sight makes this behavior more natural and more likely to cause irrelevant avatars to interfere with target avatars, producing irrelevant perspective interference effects. Therefore, we can infer that the influence of other-other line-of-sight consistency occurs before object consistency comparison. Finally, we note that in this experiment, under other-other line-of-sight consistency conditions, the walls gazed at by A and B were relatively close, both located in the left or right visual field; whereas under inconsistency conditions, the walls gazed at by A and B were relatively far apart. Compared to the former situation, the latter may have experienced less interference. Future experiments could further optimize the design to reduce visual field effects.

Interpretation and Future Directions

Our findings provide some evidence supporting the integrated mechanism of implicit mentalizing and submentalizing. When multiple avatars are present, visual perspective taking is highly flexible and spontaneously generated, influenced by both other-object consistency and line-of-sight consistency. Other-object consistency may be represented as a group, and this social factor of group perspective can only be explained by implicit mentalizing theory—participants recognize that avatars can see, that gaze can be followed, and that avatars seeing consistent objects can be perceived as a unified entity. Therefore, this study suggests that altercentric effects can be partially explained by implicit mentalizing perspectives (Cole et al., 2016; Conway et al., 2017; Heyes, 2014; Santiesteban et al., 2014). However, we do not exclude the influence of submentalizing on spontaneous perspective taking, as avatars' gaze and directional cues are difficult to separate (Baker et al., 2016). Clearly, the direction avatars face also provides low-level cues to guide attention (Cole et al., 2016), further influencing the visual perspective-taking process. That is, domain-general cognitive processing also plays a fundamental role in visual perspective taking (Fan Ya et al., 2021). In summary, we propose that when multiple avatars are present, both implicit mentalizing and submentalizing underlie the spontaneous generation of Level-1 visual perspective taking. The implicit mentalizing cue that objects are "visible" is an important factor in adopting others' perspectives, while submentalizing cues are also needed to influence perspective taking from the bottom up, such as through domain-general cognitive activities. Therefore, this study actually provides supporting evidence for both theories, leaning toward an integrated mechanism of implicit mentalizing and submentalizing.

Visual perspective taking constitutes an important research domain in spatial interaction. How we adopt others' perspectives, or even group perspectives, to better understand viewpoints other than our own and reach consensus is key to comprehending, predicting, and transmitting spatial information, thereby facilitating social interaction among people (Xiao Chengli et al., 2021). Starting from multiple avatars, this study explored whether the spontaneity of visual perspective taking differs from single-avatar contexts, preliminarily revealing that the number of objects seen by others and line-of-sight relationships in multi-avatar contexts influence the spontaneity of Level-1 visual perspective taking. We discovered some patterns not found in single-avatar or group-avatar contexts. Our findings help people understand the processing mechanisms of perspective taking in groups and draw attention to how other-object consistency and line-of-sight consistency affect perspective taking.

Looking at daily life, people often need to cooperate and empathize within groups and collectives. This study provides insights into how to help people in groups avoid interference, better understand others, and communicate more effectively in complex interpersonal relationships. Additionally, future research could start from the social attributes of avatars, altering individuals' own social-cognitive traits or social cues in spatial environments. By optimizing experimental paradigms and enhancing the social-interactive properties of research, we can improve ecological validity, continuously refine research in the domain of perspective taking with multiple avatars, and better apply perspective-taking research findings to social practice.

Conclusion

(1) The adapted experimental paradigm in this study is feasible and successfully replicates previous findings: when a single avatar is present, Level-1 visual perspective taking occurs spontaneously, with mutual influence between self and other perspectives. (2) When multiple avatars are present, the Level-1 visual perspective-taking process remains spontaneous, and consistency in the number of objects seen by avatars produces a group-perspective effect, particularly when judging from the self-perspective. (3) Consistency in others' line of sight captures participants' attention,

exerting differential effects under self-perspective and other-perspective conditions. From these conclusions, we understand that we spontaneously and flexibly compute others' perspectives, whether they are target or irrelevant perspectives. The outcomes of perspective taking can either facilitate or interfere with performance in dot-perspective judgment tasks. In summary, the spontaneity of Level-1 visual perspective taking persists in multi-avatar contexts but is influenced by the relationship among the "self-target avatar-irrelevant avatar" triad, exhibiting context-specific patterns.

References

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

Apperly, I. A. (2010). *Mindreaders: The Cognitive Basis of "Theory of Mind"*. Psychology Press. https://doi.org/10.4324/9780203833926

Apperly, I. A., & Robinson, E. J. (2003). When can children handle referential opacity? Evidence for systematic variation in 5- and 6-year-old children's reasoning about beliefs and belief reports. *Journal of Experimental Child Psychology, 85*(4), 297–311. https://doi.org/10.1016/s0022-0965(03)00099-7

Baker, L. J., Levin, D. T., & Saylor, M. M. (2016). The extent of default visual perspective taking in complex layouts. *Journal of Experimental Psychology: Human Perception and Performance, 42*(4), 508–516. https://doi.org/10.1037/xhp0000164

Capozzi, F., Cavallo, A., Furlanetto, T., & Becchio, C. (2014). Altercentric intrusions from multiple perspectives: Beyond dyads. *Plos One, 9*(12), e114210. https://doi.org/10.1371/journal.pone.0114210

Capozzi, F., & Ristic, J. (2020). Attention AND mentalizing? Reframing a debate on social orienting of attention. *Visual Cognition, 28*(2), 97–105. https://doi.org/10.1080/13506285.2020.1725206

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

Cole, G. G., Atkinson, M. A., D'Souza, A. D. C., & Smith, D. T. (2017). Spontaneous Perspective Taking in Humans? *Vision, 1*(2), 17. https://doi.org/10.3390/vision1020017

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

Cracco, E., & Cooper, R. P. (2019). Automatic imitation of multiple agents: A computational model. *Cognitive Psychology, 113*, 101224. https://doi.org/10.1016/j.cogpsych.2019.101224

Fan, Y., Xiao, C. L., & Zhou, R. L. (2021). Regulating role of attention in spontaneous visual perspective-taking. *Space Medicine & Medical Engineering, 34*(1), 31–38. [范丫, 肖承丽, 周仁来. (2021). 注意对视觉视角采择自发产生的调节作用. 航天医学与医学工程, 34(1), 31–38.]

Ferguson, H. J., Brunsdon, V. E. A., & Bradford, E. E. F. (2018). Age of avatar modulates the altercentric bias in a visual perspective-taking task: ERP and behavioral evidence. *Cognitive, Affective, & Behavioral Neuroscience, 18*, 1298–1319. https://doi.org/10.3758/s13415-018-0641-1

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-

Gopnik, A., & Meltzoff, A. N. (1998). *Words, Thoughts, and Theories.* MIT Press. https://doi.org/10.7551/mitpress/7289.001.0001

He, X. Y., Yang, Y. Q., Wang, L., & Yin, J. (2021). Tracking multiple perspectives: Spontaneous computation of what individuals in high entitative groups see. *Psychonomic Bulletin & Review, 28*, 879–887. https://doi.org/10.3758/s13423-020-01857-x

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

Kessler, K., & Rutherford, H. (2010). The two forms of visuo-spatial perspective taking are differently embodied and subserve different spatial prepositions. *Frontiers in Psychology, 1*, 213. https://doi.org/10.3389/fpsyg.2010.00213

Kornblum, S., Hasbroucq, T., & Osman, A. (1990). Dimensional overlap: Cognitive basis for stimulus–response compatibility-A model and taxonomy. *Psychological Review, 97*(2), 253–270. https://doi.org/10.1037/0033-295X.97.2.253

Kornblum, S., & Lee, J-W. (1995). Stimulus-response compatibility with relevant and irrelevant stimulus dimensions that do and do not overlap with the response. *Journal of Experimental Psychology: Human Perception & Performance, 21*(4), 855–875. https://doi.org/10.1037/0096-1523.21.4.855

Li, J., Su, X. Y., Wang, Y., Liu, C., & Chen, Q. R. (2017). The effect of different object orientation conditions on the construction of intrinsic frame of reference system in Virtual Scene. *Journal of Psychological Science, 40*(6), 1322–1327. [李晶, 苏晓雨, 王垚, 刘昌, 陈庆荣. (2017). 虚拟场景中不同客体朝向条件对内在参照系建立的影响. 心理科学, 40(6), 1322–1327.]

Linders, G. M., & Louwerse, M. M. (2023). Zipf's law revisited: Spoken dialog, linguistic units, parameters, and the principle of least effort. *Psychonomic*

Bulletin & Review, 30(1), 77–101. https://doi.org/10.3758/s13423-022-02142-9

Mattan, B., Quinn, K. A., Apperly, I. A., Sui, J., & Rotshtein, P. (2015). Is it always me first? Effects of self-tagging on third-person perspective-taking. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41*(4), 1100–1117. https://doi.org/10.1037/xlm0000078

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. https://doi.org/10.1177/1747021820942479

Pan, W., Wang, Y., & Chen, W. (2017). Evolvement and Development Trend of Mentalizing Accounts of Social Cognition: Reflections from Submentalizing. *Journal of Psychological Science, 40*(5), 1274–1279. [潘威, 汪寅, 陈巍. (2017). 心智化社会认知观的演变及发展——来自潜心智化的思考. 心理科学, 40(5), 1274–1279.]

Pesimena, G., & Soranzo, A. (2023). Both the domain-general and the mentalising processes affect visual perspective taking. *Quarterly Journal of Experimental Psychology, 76*(3), 469–484. https://doi.org/10.1177/17470218221094310

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

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. http://doi.org/10.1037/a0018729

Samuel, S., Roehr-Brackin, K., Jelbert, S., & Clayton, N. S. (2019). Flexible egocentricity: Asymmetric switch costs on a perspective-taking task. *Journal of Experimental Psychology: Learning Memory and Cognition, 45*(2), 213–218. https://doi.org/10.1037/xlm0000582

Santiesteban, I., Catmur, C., Hopkins, S. C., Bird, G., & Heyes, C. (2014). Avatars and arrows: Implicit mentalizing or domain-general processing? *Journal of Experimental Psychology: Human Perception and Performance, 40*(3), 929–937. https://doi.org/10.1037/a0035175

Schurz, M., Kronbichler, M., Weissengruber, S., Surtees, A., Samson, D., & Perner, J. (2015). Clarifying the role of theory of mind areas during visual perspective taking: Issues of spontaneity and domain-specificity. *NeuroImage, 117*, 386–396. http://doi.org/10.1016/j.neuroimage.2015.04.031

Shao, Y. T., Li, W. J., Sun, B. H., & Zhang, W. H. (2020). The effect of visuospatial perspective-taking on teacher empathy: Different functions of two components. *Journal of Psychological Science, 43*(4), 871–878. [邵雨婷, 李伟健, 孙炳海, 张文海. (2020). 视觉空间观点采择对教师共情的影响: 自我表征抑制和自我视空转换的不同作用. 心理科学, 43(4), 871–878.]

Surtees, A. D. R., & Apperly, I. A. (2012). Egocentrism and Automatic Perspective Taking in Children and Adults. *Child development, 83*(2), 452–460. http://doi.org/10.1111/j.1467-8624.2011.01730.x

Surtees, A., Apperly, I., & Samson, D. (2013a). Similarities and differences in visual and spatial perspective-taking processes. *Cognition, 129*(2), 426–438. http://doi.org/10.1016/j.cognition.2013.06.008

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

Trick, L. M., & Pylyshyn, Z. W. (1994). Why are small and large numbers enumerated differently? A limited-capacity preattentive stage in vision. *Psychological Review, 101*(1), 80–102. https://doi.org/10.1037/0033-295X.101.1.80

Westra, E. (2017). Spontaneous mindreading: A problem for the two-systems account. *Synthese, 194*(11), 4559–4581. https://doi.org/10.1007/s11229-016-1159-0

Wu, M. H., Xie, J. S., & Deng, Z. (2022). The debate between inhibition and attribution of egocentric bias in visual perspective taking. *Advances in Psychological Science, 30*(1),179–187. [吴梦慧, 谢久书, 邓铸. (2022). 视觉观点采择中自我中心性偏差的抑制和归因之争. 心理科学进展, 30(1), 179–187.]

Xiao, C. L., Sui, Y. Q., Xiao, S. H., & Zhou, R. L. (2021). A new perspective on spatial interaction research: The effects of multiple social factors. *Advances in Psychological Science, 29*(5), 796–805. [肖承丽, 隋雨棨, 肖苏衡, 周仁来. (2021). 空间交互研究新视角：多重社会因素的影响. 心理科学进展, 29(5), 796–805.]

Ye, T., Furumi, F., Catarino da Silva, D., Hamilton, A. (2021). Taking the perspectives of many people: Humanization matters. *Psychonomic Bulletin & Review, 28*, 888–897. https://doi.org/10.3758/s13423-020-01850-4

Zhang, W. X. (1998). The perspective taking and the issues existing in these studies (In Chinese). *Journal of East China Normal University (Educational Sciences), 4*(7), 58–63+24. [张文新. (1998). 论观点采择及其研究中存在的若干问题. 华东师范大学学报 (教育科学版), 4(7), 58–63+24.]

Zhang, Z. J., Zhao, Y. J. & Zhan, Q. T. (2011). Effects of gaze direction perception on gaze following behavior. *Acta Psychologica Sinica, 43*(7), 726−738. [张智君, 赵亚军, 占琪涛. (2011). 注视方向的知觉对注视追随行为的影响. 心理学报, 43(7), 726−738.]

Zuo, T. T., & Hu, Q. F. (2019). The Automaticity of Perspective-taking. *Psychological Development and Education, 35*(2), 236–245. [左婷婷, 胡清芬. (2019). 视角采择的自动性. 心理发展与教育, 35(2), 236–245.]

Note: Figure translations are in progress. See original paper for figures.