The effects of body direction and posture on taking the perspective of a humanoid avatar in a virtual environment

Sachiyo Ueda (ueda.sachiyo.ms@tut.jp)
Toyohashi University of Technology
Kazuya Nagamachi
Toyohashi University of Technology
Maki Sugimoto
Keio University
Masahiko Inami
University of Tokyo
Michiteru Kitazaki
Toyohashi University of Technology

Research Article

Keywords: Avatar, perspective taking, embodiment, impossible posture

Posted Date: April 19th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-133156/v2

Version of Record: A version of this preprint was published at PLOS ONE on December 21st, 2021. See the published version at https://doi.org/10.1371/journal.pone.0261063.

Abstract

Visual perspective taking is inferring how the world looks to another person. To clarify this process, we investigated whether employing a humanoid avatar as the viewpoint would facilitate an imagined perspective shift in a virtual environment. We used a task that involved reporting how an object looks by a simple direction judgment either from the avatar's position or an empty chair's position. We found that the humanoid avatar's presence improved task performance. Furthermore, the avatar's facilitation effect was observed only when the avatar was facing the visual stimulus to be judged; performance was worse when it faced backwards than when there was only an empty chair faced forwards. When the directions of the head and the torso were opposite (i.e., an impossible posture), the avatar's facilitation effect disappeared. The performance was better in the order of the condition that both the head and the torso were facing forward, the condition that both the head and the torso were facing backward, the condition that the torso was facing toward the stimulus while the head was facing away, and the condition that the head was facing toward the stimulus while the torso was facing away. Thus, visual perspective taking might not be facilitated by the avatar when its posture is biomechanically impossible. These results suggest that the facilitation effect is based not only on attention capture but also on visual perspective taking of humanoid avatar.

Introduction

Visual perspective taking, or when humans move their virtual viewpoint to another person's perspective [1], may be a basis for more sophisticated social abilities such as empathy [2–4], shared intentionality [5], and theory of mind [6]. Visual perspective taking is not a unitary ability but can be divided into different levels [1,7,8]. Level 1 is the ability to understand whether an object is visible to others, and Level 2 is the ability to understand others' perspectives of the object. Level 2 perspective taking develops later [1,9,10], which suggests that it involves cognitively demanding computational processing. Moreover, visual perspective taking can be divided in two different systems: implicit perspective taking and explicit perspective taking [8,11]. Implicit perspective taking occurs when participants are asked to answer only from the egocentric perspective or the first-person perspective; their responses are affected by existence of another person in the scene. Explicit perspective taking requires participants to take another's perspective, and respond from that other person's viewpoint.

Currently, it is not very clear how visual perspective taking works. There is a debate over whether implicit visual perspective taking is automatic and spontaneous [12– 16] or if it includes a conscious process [17–20]. Samson et al. [12] showed that incongruent information from another person's perspective interferes with performance in the Level 1 visual perspective taking task, even when the participant is ignoring that other person (altercentric intrusions). Surtees et al. [14] showed that performance in the Level 2 visual perspective taking task is spontaneously affected by another person in an interactive task. Ward et al. [15] used a task that involved the classic mental rotation task, and showed that the Level 2 perspective taking occurred spontaneously by the existence of agents even though the agents were irrespective of the task. They demonstrated that the reaction time (RT) to judge letters presented on a

table at various angles was shorter not only when the rotation angle as viewed by the observer was small, but also when the rotation angle as viewed by another person (humanoid avatar) was small. The facilitation effect was enhanced when participants were asked to explicitly take perspective of the avatar. This facilitation effect disappeared when a lamp (inanimate object) was presented instead of the humanoid avatar. They called it “perceptual simulation” of others’ viewing perspective [15]. In a following study, Ward et al. [16] reported that the facilitation effect of another person does not require that person to be looking at an object. However, Cole et al. [17] showed a perspective taking effect even when the agent could not see the same stimuli as the participant due to a barrier, and concluded that humans do not spontaneously take the perspective of others. They concluded that visual perspective taking is a general spatial-navigational ability that is influenced by the spatial location of another person, but not their gaze. Kuhn et al [18] showed that when participants are given ample time to explore the visual environment, gaze following is modulated by another person’s visual perspective; participants fixate a target object faster when the agent can see the object than when it is occluded for the agent. On the other hand, when participants are asked to rapidly discriminate a target, the performance is not modulated by another person’s visibility. Thus, they concluded that the visual perspective taking is not automatic [18]. Samuel et al. [20] showed that visual perspective taking is not perceptual simulation, but a conscious process to solve a problem by using naïve ideas about how vision works.

In explicit perspective taking, participants mentally rotate or transform their own perspective to another person’s position in the Level 2 perspective taking task but not in the Level 1 perspective taking task [7,21]. An imagined shift of perspective to an arbitrary position is possible even when no physical body of another person is present [22–26]. However, the performance of the Level 2 perspective taking task was better when a doll was presented compared to an asterisk [7]. The embodiment to the agent has a crucial role in explicit Level 2 perspective taking [21,27]. Kessler and Thomson [21] showed that congruence between participants’ body posture and direction of mental self-rotation improves visual perspective taking, and the posture of the avatar could not be fully ignored, although it was irrelevant to the task. Vestibular stimulation to participants impairs performance of visual perspective taking [28]. Thus, there may be two critical factors for explicit visual perspective taking: the endogenous (i.e., the participant’s) motoric or vestibular embodiment and the exogenous (i.e., the avatar’s) embodiment [21,28].

In the current study, we aimed to investigate what features of a humanoid avatar facilitate explicit Level 2 perspective taking to know the effect of exogenous embodiment. We developed a virtual reality task involving a simple direction judgment to examine how a humanoid avatar facilitates visual perspective taking in a virtual environment. In Experiment 1, two conditions were created: one with an avatar sitting in a chair and looking at a visual stimulus and another with only the empty chair. Participants were asked to judge the direction of the gap in a circle (i.e., a Landolt ring) from the avatar’s perspective or the empty chair’s position. Thus, participants were explicitly asked to take visual perspective of the location of the avatar or chair. We then examined whether the presentation of the avatar facilitated the perspective transformation. We also manipulated the interval between the presentation of the avatar or the chair and the visual stimulus to examine the time scale of perspective taking.

In Experiment 2, we manipulated the orientation of the avatar with respect to the target stimulus and examined whether the avatar's gaze on the target was necessary for Level 2 perspective taking. If the RT was shortened only when the avatar's body was directed at the target (i.e., in the forward-facing condition), it would indicate that the humanoid avatar is not simply facilitating the perspective judgment from an arbitrary position by strongly attracting attention, but that it is necessary to infer the other's mental state of how the target stimulus is seen by the avatar.

In Experiment 3, we introduced an impossible-posture avatar whose head was oriented in the opposite direction to the torso to test whether the direction of the head or that of the torso was critical for the facilitation effect on visual perspective taking.

Experiment 1

Methods

Participants

Twenty paid volunteers participated in the experiment (17 men, 3 women, all aged 19–24 years). Sample size was determined by our previous experiences before conducting experiments. This sample size corresponds to an effect size f of 0.195, alpha = 0.05, power = 0.8 using the G*Power 3.1 [29,30]. Participants were undergraduate and graduate students of Toyohashi University of Technology. All had normal or corrected-to-normal vision and were naïve to the purpose of the study.

In all of the studies, all participants provided written informed consent before the experiment. All of the experiments were approved by the Ethical Committee for Human-Subject Research of Toyohashi University of Technology, and all experiments were performed in accordance with the committee's guidelines and regulations.

Apparatus

The visual stimuli were generated and controlled by a computer using Unity Pro and presented on a head-mounted display (HTC Vive Pro: 1,440 × 1,600 pixels, 90 Hz refresh). The participants responded in the task by moving a joystick.

Stimuli and conditions

In the virtual space, a table was in the center of the room, and either an empty chair or an avatar sitting in a chair was presented in one of three positions, that is, on the left, right, or other side of the table, based on the participant's perspective (Fig. 1A and 1B). Then, a broken circle was presented on the table. The gap in the broken circle was angled in one of eight directions (0°, 45°, 90°, 135°, 180°, 225°, 270°, and 315°), like a Landolt ring. We created two conditions for the interval between the presentation of the chair or avatar and the presentation of the circle (short: 200 ms; long: 1,000 ms). There were 96 combinations

of trials (2 with/without the avatar, 2 short/long intervals, 3 positions of the avatar and chair, and 8 directions of the gap in the circle). The directions of the gap in the circle were merged in the analysis.

Procedure

Each trial began with a blank black screen for 1,000 ms, which was followed by a red fixation dot. Then, 1,000 ms later, the fixation dot disappeared and the room with the table, chair, and/or avatar appeared. After 200 ms (short interval) or 1,000 ms (long interval), the broken circle was presented on the table. The participants were asked to judge the direction of the gap in the circle from the avatar's perspective or the empty chair's position and to respond with the joystick as accurately and quickly as possible (Fig. 2). If the gap direction was 135° counterclockwise from the participants' point of view (when the 6 o'clock position is defined as 0°), as in Figure 2, the participant's task was to adjust the direction from the perspective of the avatar, that is, the joystick should have been moved to the 45° counterclockwise position. Participants received no feedback. The next trial began immediately after the joystick response.

Before the practice trials, a direction judgment task from the participant's perspective was conducted first. Then, 12 practice trials (2 with/without the avatar, 2 short/long intervals, and 3 positions of the avatar and chair) were presented, and the participants judged the direction of the gap in the circle from the avatar's perspective or the empty chair's position.

In each test session, all 96 combinations of the conditions were repeated twice in a random order, for a total of 192 trials. Each participant completed four test sessions, for a total of 768 test trials. It took approximately 90 minutes for each participant to finish this experiment, including the time required to provide the experimental instructions, to conduct the practice trials and the test sessions, and to take breaks between sessions.

Data analysis

Individual mean RTs and error rates were calculated for each of the twelve conditions (i.e., with/without the avatar, short/long, and left/front/right position of the avatar/chair). For the analysis, we treated the trials in which the participant moved the joystick within the range of ± 22.5° from the correct angle, as the correct response. RTs were determined as the time from the onset of the broken circle to the time when the joystick reached the end position (i.e., 2.5 cm from the center of the joy stick). Trials for which the RT was shorter than 150 ms (0 %) and trials for which the RT was longer than three standard deviations from the mean RT of each condition for each participant (1.5 %) were excluded as outliers from the analysis. Trials in which the participants made an error were also excluded from the RT analysis (approximately 6.8% of the trials). The RTs and error rates were submitted to a 2 × 2 × 3 repeated-measures analysis of variance (ANOVA) with the avatar existence, interval, and position of the avatar and chair as the within-subject factors. If there was a lack of sphericity, the reported values were adjusted using the Greenhouse-Geisser correction [31]. When performing the multiple comparisons after the ANOVAs, we reported the *p*-values that were corrected using Shaffer's modified sequentially rejective Bonferroni procedure [32].

Results

The ANOVA of RTs showed significant main effects of the existence of the avatar, $F(1, 19) = 19.570$, $p < 0.001$, $\eta_p^2 = 0.507$, the length of the interval, $F(1, 19) = 73.376$, $p < 0.001$, $\eta_p^2 = 0.794$, and the position of the avatar and chair, $F(1.170, 22.222) = 7.864$, $p = 0.001$, $\eta_p^2 = 0.293$. There was also a significant interaction between the avatar existence and interval, $F(2, 38) = 7.355$, $p = 0.014$, $\eta_p^2 = 0.279$, and between position and interval, $F(2, 38) = 12.068$, $p < 0.001$, $\eta_p^2 = 0.388$. The avatar existence × position interaction, $F(2, 38) = 3.138$, $p = 0.055$, $\eta_p^2 = 0.142$, and avatar existence × interval × position interaction, $F(1.351, 25.670) = 0.383$, $p = 0.604$, $\eta_p^2 = 0.020$, were not significant.

Participants' RTs were significantly faster for the "with avatar" condition than the "without avatar" condition, only in the short interval condition ($p < 0.001$) (Fig. 3A). In the short interval condition, the RTs were slower in the front position than in the other two positions ($ps < 0.01$). In the long interval condition, the RTs were faster in the left position than in the other two positions ($ps < 0.05$). In all conditions, the long interval condition had faster RTs than the short interval condition.

The ANOVA of error rates revealed a significant main effect of the existence of the avatar, $F(1, 19) = 7.335$, $p = 0.014$, $\eta_p^2 = 0.279$ (Fig. 3B). The participants were more accurate when the avatar was presented than when it was not. The main effect of position was also significant, $F(2, 38) = 11.696$, $p < 0.001$, $\eta_p^2 = 0.381$. Participants responded more accurately when the avatar was in the front and left positions than in the right position. No other main effect or interactions were found to be significant.

Discussion

The humanoid avatar was associated with improved performance of identifying the orientation of a visual stimulus from an imagined position only in the short interval condition (200 ms). When the avatar was present, the participant's viewpoint moved quickly to the position of the avatar, but this process either did not occur or took a long time in the chair condition. In the long interval condition, participants may have enough time to transform the viewpoint to an arbitrary position regardless of the presence of the avatar.

This facilitation effect of the humanoid avatar was basically consistent with the findings of Ward et al. [15] and Michelon and Zacks [7]. They also showed that the presence of humanoid avatar makes it faster to determine how the visual stimulus looks from that position than it does with an inanimate object [7, 15]. However, one may argue that the results were obtained because humanoid avatars capture attention more easily than an empty chair. To examine this, in Experiment 2, we added a condition in which the humanoid avatar was sitting backwards. We hypothesized that if there is no facilitation effect on the RT in the backward avatar condition, then the embodiment to the avatar is important for efficient perspective taking.

Experiment 2

Methods

Participants

Twenty paid volunteers participated in the experiment (15 men, 5 women, all

aged 20–24 years). Eight of them had participated in Experiment 1. Sample size was determined by our previous experiences before conducting experiments. This sample size corresponds to an effect size f of 0.21, alpha = 0.05, power = 0.8 using the G*Power 3.1 [29,30]. They all had normal or corrected-to-normal vision and were naïve to the purpose of the study.

Apparatus

The same apparatus used in Experiment 1 was used in Experiment 2.

Stimuli and conditions

This experiment differed from Experiment 1 in that we added the backward avatar condition (Fig. 1C). Since the avatar's facilitation effect was obtained only in the short interval condition in Experiment 1, only the short interval was employed in Experiment 2. There were 72 combinations of trials (3 types of avatar: no avatar, forward avatar, and backward avatar; 3 positions of the avatar and chair; and 8 directions of the gap in the circle). The directions of the gap in the circle were combined in the analysis.

Procedure

The procedure was the same as in Experiment 1, but the experimental conditions were changed. In the backward avatar condition, participants were instructed to imagine the perspective when looking from the avatar's position toward the center of the table. Before the practice trials, the direction judgment task from the participant's perspective was conducted first. Then, a block of nine practice trials (3 types of avatar and 3 positions of the avatar and chair) was presented.

In each test session, all 72 combinations of the conditions (3 types of avatar, 3 positions of the avatar and chair, and 8 directions of the gap in the circle) were repeated twice in a random order, for a total of 144 trials. Each participant completed four test sessions, for a total of 576 test trials. It took approximately 60 minutes for each participant to finish this experiment.

Data analysis

Individual mean RTs and error rates were calculated for each of the nine conditions (i.e., without/forward/backward avatar and left/front/right position) as in Experiment 1. Trials for which RT was shorter than 150 ms (0 %), trials for which RT was longer than three standard deviations from the mean RT (1.6 %), and error trials (6.2 %) were excluded as outliers from the RT analysis. The RTs and error

rates were submitted to a 3 × 3 repeated-measures ANOVA with the avatar type and position of the avatar and chair as the within-subject factors. When conducting the multiple comparisons after the ANOVAs, we reported the *p*-values that were corrected using Shaffer's modified sequentially rejective Bonferroni procedure [32].

Results

The ANOVA of RTs showed significant main effects of the type of avatar, $F(2, 38) = 25.459$, $p < 0.001$, $\eta_p^2 = 0.573$, and position of the avatar and chair, $F(2, 38) = 14.952$, $p < 0.001$, $\eta_p^2 = 0.440$. The avatar existence × position interaction was not significant, $F(4, 76) = 1.113$, $p = 0.357$, $\eta_p^2 = 0.055$. The RTs were significantly faster in the order of the forward avatar condition, without avatar condition, and backward avatar condition (*ps* < 0.01). The RTs were significantly slower in the front position condition than in the right position and left position conditions (*ps* < 0.001) (Fig. 4A). The ANOVA of error rates revealed that there were no significant main effects or interactions for the accuracy (Fig. 4B).

Discussion

We found that the facilitation of the perspective transformation occurred only when the avatar's body and head were facing the visual stimulus. Thus, the direction of a humanoid avatar's head and body is an effective cue for the facilitation effect of spatial cognition in perspective taking. This suggests that the facilitation effect of the presence of humanoid avatars on Level 2 perspective taking is not based on attention capture due to the saliency of the humanoid avatars, but on the efficiency of the cognitive process of perspective taking.

In contrast, Ward et al. [16] showed that the facilitation effect of perspective taking exists even when the avatar does not look at an object. This difference in the findings may be due to the fact that our avatar was placed facing away from the object with a full body including the head, but the avatar in the study by Ward et al. [16] averted its gaze while its torso faced the object. Thus, the direction of the torso may be more important than the gaze direction. To investigate this, in Experiment 3, we tested whether the direction of the head or that of the torso was critical for the facilitation effect on visual perspective taking by employing an avatar in an impossible posture.

Experiment 3

Methods

Participants

Twenty paid volunteers participated in the experiment (17 men, 3 women, all

aged 20–23 years). None of them had participated in either Experiment 1 or 2. Sample size was determined by our previous experiences before conducting experiments. This sample size corresponds to

an effect size f of 0.22, alpha = 0.05, power = 0.8 using the G*Power 3.1 [29,30]. They all had normal or corrected-to-normal vision and were naïve to the purpose of the study.

Apparatus

The same apparatus used in Experiments 1 and 2 were used in Experiment 3.

Stimuli and conditions

In this experiment, we added the avatar whose head was oriented in the opposite direction to the torso (Fig. 5). There were 4 conditions: (1) both the head and the torso were facing toward the stimulus, (2) both the head and the torso were facing away from the stimulus, (3) the torso was facing toward the stimulus while the head was facing away, (4) the head was facing toward the stimulus while the torso was facing away. The direction of the chair was always same as that of the torso. The short interval (200 ms) was employed. The gap in the broken circle was angled in one of four directions (45°, 135°, 225°, and 315°). The avatar sitting in the chair was presented either on the left or right side of the table. There were 32 combinations of trials (4 types of avatar; 2 positions of the avatar and chair; and 4 directions of the gap in the circle). The directions of the gap in the circle were combined in the analysis.

Procedure

The procedure was the same as in Experiments 1 and 2, but the experimental conditions were changed. Before the practice trials, the direction judgment task from the participant's perspective was conducted first. Then, a block of eight practice trials (4 types of avatar and 2 positions of the avatar and chair) was presented.

In each test session, all 32 combinations of the conditions (4 types of avatar, 2 positions of the avatar and chair, and 4 directions of the gap in the circle) were repeated twice in a random order, for a total of 64 trials. Each participant completed four test sessions, for a total of 256 test trials. It took approximately 40 minutes for each participant to finish this experiment, including the time required to provide the experimental instructions, take a break, and conduct the practice trials.

Data analysis

Individual mean RTs and error rates were calculated for each of the eight conditions by the same procedure as in Experiment 1 and 2. Trials for which RT was shorter than 150 ms (0 %), trials for which RT was longer than three standard deviations from the mean RT (1.5 %), and error trials (13.6 %) were excluded as outliers from the RT analysis. The RTs and error rates were submitted to a 4 × 2 repeated-measures ANOVA with the avatar type and position of the avatar and chair as the within-subject factors. If there was a lack of sphericity, the reported values were adjusted using the Greenhouse-Geisser correction [31]. When conducting the multiple comparisons after the ANOVAs, we reported the *p*-values that were corrected using Shaffer's modified sequentially rejective Bonferroni procedure [32].

Results

The ANOVA of RTs showed significant main effects of the type of avatar, $F(1.820, 34.574) = 15.450$, $p < 0.001$, $\eta_p^2 = 0.448$. The main effects of position of the avatar and the interaction were not significant. The RTs were fastest in condition 1, followed by conditions 2, 3, and 4 ($ps < 0.05$, Fig. 6A). The ANOVA of error rates revealed that there were no significant main effects or interactions for the accuracy (Fig. 6B).

Discussion

We found that the facilitation of the perspective transformation disappeared when the avatar's head was oriented in the opposite direction to the torso. The RTs for the impossible-posture avatar conditions were slower than the full-body backward avatar. This suggests that visual perspective taking might not be facilitated by the avatar when its posture is biomechanically impossible. As for the impossible-posture avatars, the torso-only forward condition was responded to faster than the head-only forward condition. Thus, the direction of the torso may be more effective for visual perspective taking than the direction of the head or gaze.

General Discussion

Our results showed that visual perspective taking was not facilitated by the avatar when its posture was biomechanically impossible. Participants' performance in the impossible-position avatar condition was worse than the backward avatar condition. We speculate that it was difficult for participants to imagine the perspective of the impossible-posture avatar, so the facilitation effect disappeared. However, we did not measure sense of embodiment to the avatar. Thus, we need further investigation on this issue. The RT for the torso-only forward avatar was faster than the head-only forward avatar. This is partly consistent with the study by Ward et al. [16], in which the visual perspective taking occurred even when the avatar diverted its gaze. Furthermore, Longo et al. [33] showed that both the head and torso contribute to perspective taking, but the torso more so than the head.

There was a difference in RT depending on the position of the avatar. In the short interval condition in Experiments 1 and 2, the RTs were longer in the front position than in the right and left positions, and in the long interval condition in Experiment 1, the RTs were longer in the front and right positions than in the left position. The longest RT in the front position obtained in short interval condition is consistent with previous studies which showed that Level 2 perspective taking involves the process of mental self-rotation to a different viewpoint from one's own (e.g., [7,21,35]). That is, the large angular disparity between the front position and the individual's own body position required a longer time for the mental self-rotation, resulting in a longer RT. In Experiment 1, the main effect of the position was significant (no interaction with other factors) for the error rate, and the performance was significantly accurate in the left and front positions, so that the longest RT in the front position can be partly explained by the trade-off with accuracy. However, since Experiment 2 did not show significant main effects or interactions for the accuracy (error rate), the trade-off alone is not a sufficient explanation. The gradient of RT weakened in

the long interval condition and there was no difference between the RT in the right position and in the front position. This might be due to the time-gap between the presentation of the avatar or chair and the presentation of the target stimulus. In that case, the observer can finish perspective transformation to any position before presentation of the target stimulus, and therefore, the RT would not differ depending on the avatar's position. Of course, if this account is reasonable, the result of the left position having a shorter RT than the other two positions appears to be strange, but this may have been due to the sofa presented in the upper left corner of the display. The sofa may have attracted attention and the RT in the left position may have been shortened. The attention capture effect of the sofa may not work under the short interval condition. However, this is just speculation.

Declarations

Acknowledgements

We would like to thank Naho Isogai for collecting some of the data. We would like to thank Editage (www.editage.jp) for English language editing.

References

1. Flavell JH, Everett BA, Croft K, Flavell ER. Young children's knowledge about visual perception: Further evidence for the level 1–level 2 distinction. Dev Psychol. 1981;17: 99–103.
2. Batson CD, Early S, Salvarani G. Perspective taking: Imagining how another feels versus imaging how you would feel. Pers Soc Psychol Bull. 1997;23(7): 751–758.
3. Erle TM, Topolinski S. Spatial and empathic perspective taking correlate on a dispositional level. Soc Cogn. 2015;33(3): 187–210.
4. Mattan BD, Rotshtein P, Quinn KA. Empathy and visual perspective taking performance. Cogn Neurosci. 2016;7(1–4): 170–181. doi: 10.1080/17588928.2015.1085372.
5. Tomasello M, Carpenter M, Call J, Behne T, Moll H. Understanding and sharing intentions: The origins of cultural cognition. Behav Brain Sci. 2005;28(5): 675–691.
6. Hamilton AFDC, Brindley R, Frith U. Visual perspective taking impairment in children with autistic spectrum disorder. Cognition. 2009;113(1): 37–44.
7. Michelon P, Zacks JM. Two kinds of visual perspective taking. Percept Psychophys. 2006;68: 327–337.
8. Apperly IA, Butterfill SA. Do humans have two systems to track beliefs and belief-like states? Psychol Rev. 2009;116: 953–970.
9. Masangkay ZS, McClusky KA, McIntyre CW, Sims-Knight J, Vaughn BE, Flavell JH. The early development of inferences about visual percepts of others. Child Dev. 1974;45: 357–366.
10. Gzesh SM, Surber CF. Visual perspective taking skills in children. Child Dev. 1985;56: 1204–1213.

11. Martin AK, Perceval G, Davies I, Su P, Huang J, Meinzer M. Visual perspective taking in young and older adults. J Exp Psychol Gen. 2019;148(11): 2006–2026. doi: 10.1037/xge0000584.
12. Samson D, Apperly IA, Braithwaite JJ, Andrews BJ, Scott SEB. Seeing it their way: Evidence for rapid and involuntary computation of what other people see. J Exp Psychol Hum Percept Perform. 2010;36: 1255–1266.
13. Furlanetto T, Becchio C, Samson D, Apperly I. Altercentric interference in level 1 visual perspective taking reflects the ascription of mental states, not submentalizing. J Exp Psychol Hum Percept Perform. 2016;42: 158–163.
14. Surtees A, Samson D, Apperly I. I've got your number: Spontaneous perspective taking in an interactive task. Cognition. 2016;150: 43–52.
15. Ward E, Ganis G, Bach P. Spontaneous Vicarious Perception of the Content of Another's Visual Perspective. Curr Biol. 2019;5: 874–880.
16. Ward E, Ganis G, McDonough KL, Bach P. Perspective taking as virtual navigation? Perceptual simulation of what others see reflects their location in space but not their gaze. Cognition. 2020;199: 104241.
17. Cole GG, Atkinson M, Le AT, Smith DT. Do humans spontaneously take the perspective of others?. Acta Psychol. 2016;164: 165–168.
18. Kuhn G, Vacaityte I, D'Souza AD, Millett AC, Cole GG. Mental states modulate gaze following, but not automatically. Cognition. 2018;180: 1–9.
19. Cole GG, Millett AC. The closing of the theory of mind: A critique of perspective taking. Psychon Bull Rev. 2019;26(6): 1787–1802.
20. Samuel S, Hagspiel K, Eacott MJ, Cole GG. Visual perspective taking and image-like representations: We don't see it. Cognition. 2021;210: 104607.
21. Kessler K, Thomson LA. The embodied nature of spatial perspective taking: Embodied transformation versus sensorimotor interference. Cognition. 2010;114: 72–88.
22. Amorim MA, Glasauer S, Corpinot K, Berthoz A. Updating an object's orientation and location during nonvisual navigation: A comparison between two processing modes. Percept Psychophys. 1997;59(3): 404–418.
23. Amorim MA, Stucchi N. Viewer- and object-centered mental explorations of an imagined environment are not equivalent. Brain Res Cogn Brain Res. 1997;5(3): 229–239
24. Presson CC, Montello DR. Updating after rotational and translational body movements: Coordinate structure of perspective space. Perception. 1994;23(12): 1447–1455.
25. Wraga M, Shephard JM, Church JA, Inati S, Kosslyn SM. Imagined rotations of self versus objects: An fMRI study. Neuropsychologia. 2005;43: 1351–1361.
26. Zacks JM, Mires J, Tversky B, Hazeltine E. Mental spatial transformations of objects and perspective. Spat Cogn Comput. 2000;2(4): 315–332.

27. Kessler K, Rutherford H. The two forms of visuo-spatial perspective taking are differently embodied and subserve different spatial prepositions. Front Psychol. 2010;1: 213. doi: 10.3389/fpsyg.2010.00213.
28. Gardner MR, Stent C, Mohr C, Golding JF. Embodied perspective taking indicated by selective disruption from aberrant self-motion. Psychol Res. 2017;81: 480–489.
29. Faul, Erdfelder, Lang, & Buchner, 2007
30. Faul, Erdfelder, Buchner, & Lang, 2009
31. Geisser S, Greenhouse SW. An extension of Box's results on the use of the F distribution in multivariate analysis. The Annals of Mathematical Statistics. 1958;29: 885–891. doi: 10.1214/aoms/1177706545.
32. Shaffer JP. Modified sequentially rejective multiple test procedures. J Am Stat Assoc. 1986;81: 826–831.
33. Morey RD. Confidence intervals from normalized data: A correction to Cousineau (2005). Tutor Quant Methods Psychol. 2008;4: 61–64.
34. Longo MR, Rajapakse SS, Alsmith AJ, Ferrè ER. Shared contributions of the head and torso to spatial reference frames across spatial judgments. Cognition. 2020;204: 104349. doi: 10.1016/j.cognition.2020.104349.
35. Surtees ADR, Apperly IA, Samson D. Similarities and differences in visual and spatial perspective taking processes. Cognition. 2013;129: 426–438.
36. Neely KN, Heath M. Visuomotor mental rotation: Reaction time is determined by the complexity of the sensorimotor transformations mediating the response. Brain Res. 2010;1366: 129–140.
37. Neely KN, Heath M. The Visuomotor Mental Rotation Task: Visuomotor Transformation Times Are Reduced for Small and Perceptually Familiar Angles. J Mot Behav. 2011;43: 393–402.
38. Quesque F, Foncelle A, Chabanat É, Jacquin-Courtois S, Rossetti, Y. Take a Seat and Get Into Its Shoes! When Humans Spontaneously Represent Visual Scenes From the Point of View of Inanimate Objects. Perception. 2020;49(12): 1333–1347.
39. Lenggenhager B, Tadi T, Metzinger T, Blanke O. Video ergo sum: manipulating bodily self-consciousness. Science. 2007;317: 1096–1099.
40. Kondo R, Sugimoto M, Minamizawa K, Hoshi T, Inami M, Kitazaki M. Illusory body ownership of an invisible body interpolated between virtual hands and feet via visual-motor synchronicity. Sci Rep. 2018;8: 1–8.
41. Fribourg R, Argelaguet F, Lécuyer A, Hoyet L. Avatar and sense of embodiment: studying the relative preference between appearance, control and point of view. IEEE Trans Vis Comput Graph. 2020;26: 2062–2072.

Figures

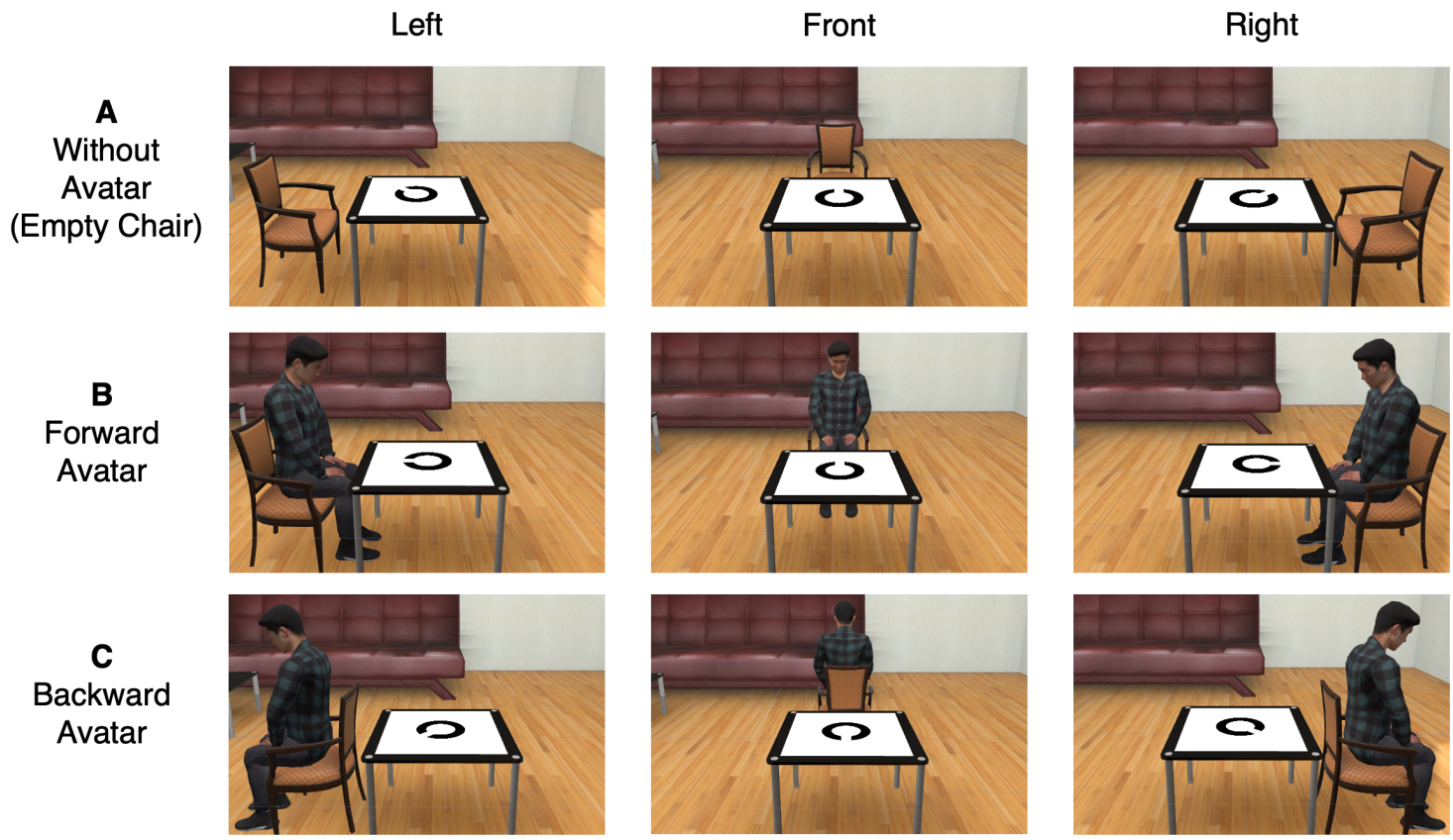

Figure 1

A subset of the stimuli used in the experiments. An empty chair (A) or a forward avatar (B) was presented in either the left, front, or right position from the participant's point of view in Experiment 1. In addition to the avatar presentation conditions in Experiment 1, there was a backward avatar condition in which the avatar was sitting backward against the desk (C) in Experiment 2. Participants were asked to judge the direction of the gap in the circle from the avatar's perspective or the empty chair's position and to respond with the joystick as accurately and quickly as possible.

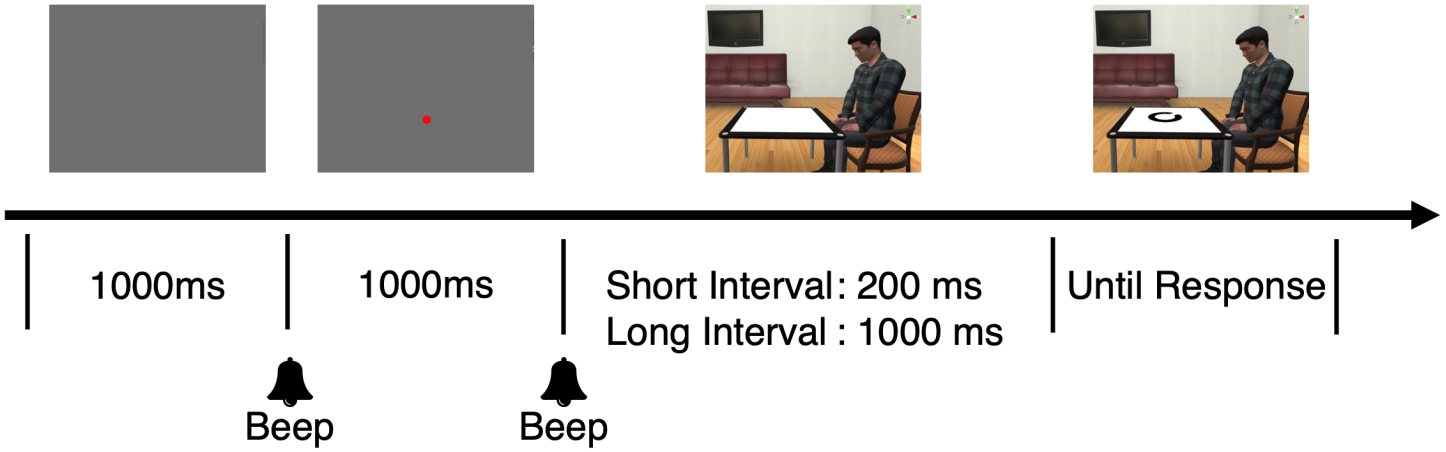

Figure 2

Procedure of Experiment 1.

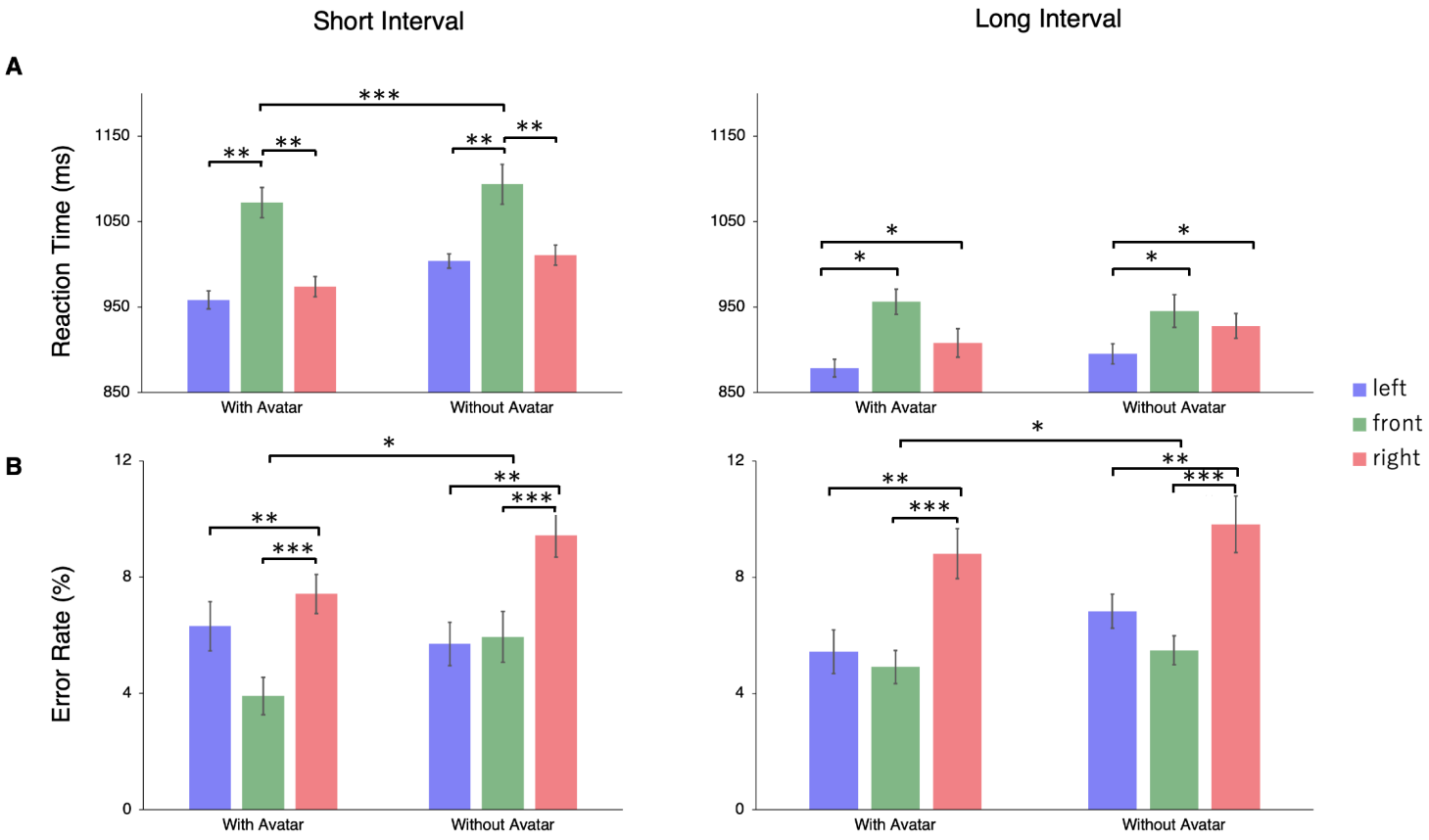

Figure 3

Results of Experiment 1. (A) Mean reaction times (RTs). Three-way repeated measures ANOVA and Shaffer's post hoc tests were conducted (*p < .05, **p < .01, ***p < .001). The RTs were significantly faster for the "with avatar" condition than the "without avatar" condition, only in the short interval condition. In the short interval condition, the RTs were slower in the front position than in the other two positions. In the long interval condition, the RTs were faster in the left position than in the other two positions. (B) Error rates. The participants were more accurate when the avatar was presented than when it was not. The responses were more accurate when the avatar was in the front and left positions than in the right position. Error bars represent 95% within-subjects confidence intervals [33].

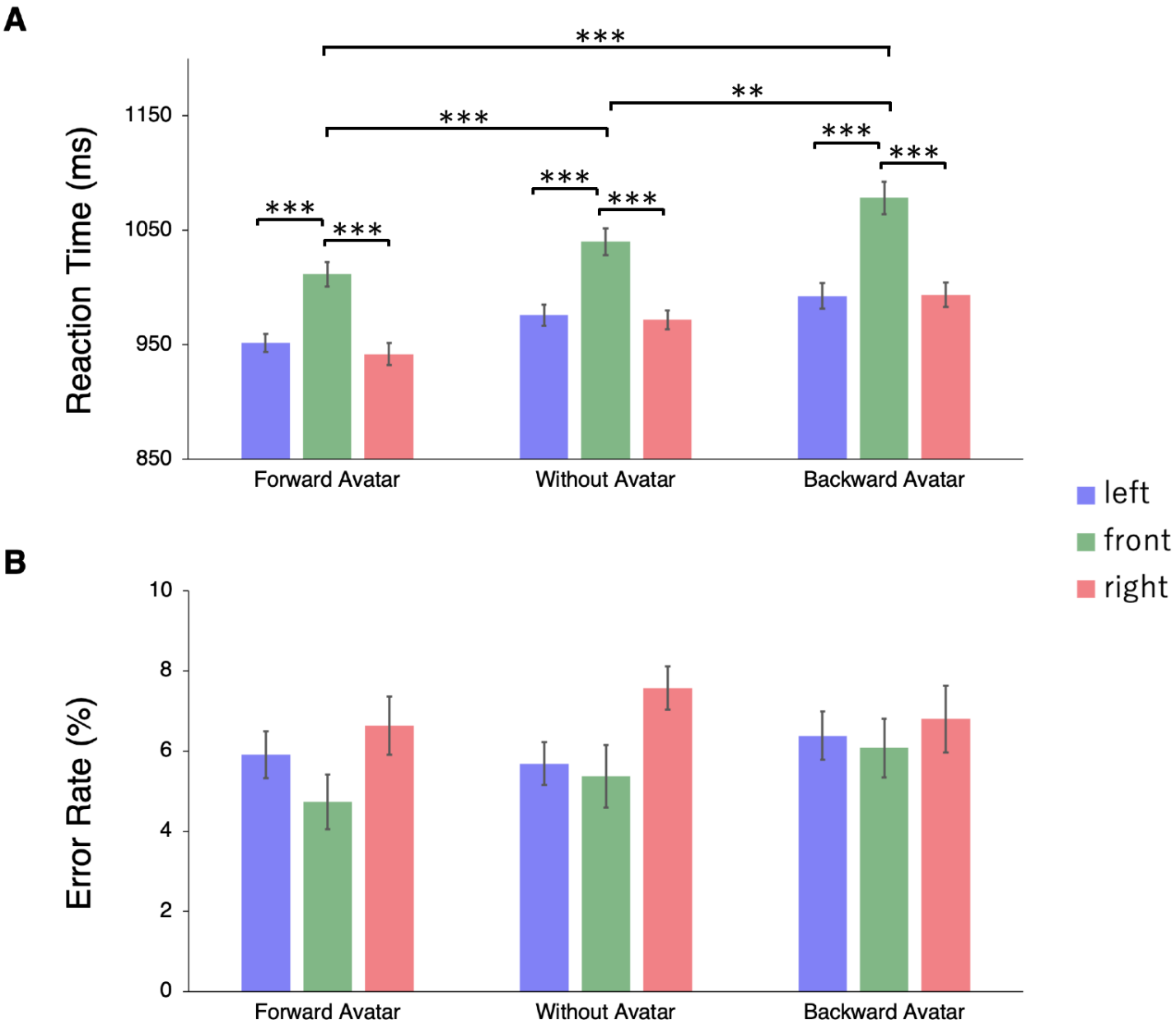

Figure 4

Results of Experiment 2. (A) Mean reaction times (RTs). Two-way repeated measures ANOVA and Shaffer's post hoc tests were conducted (*p < .05, **p < .01, ***p < .001). The RTs were significantly faster in the order of the forward avatar condition, without avatar condition, and backward avatar condition. The RTs were significantly slower in the front position condition than in the right position and left position conditions. (B) Error rates. There were no significant main effects or interactions for the accuracy. Error bars represent 95% within-subjects confidence intervals.

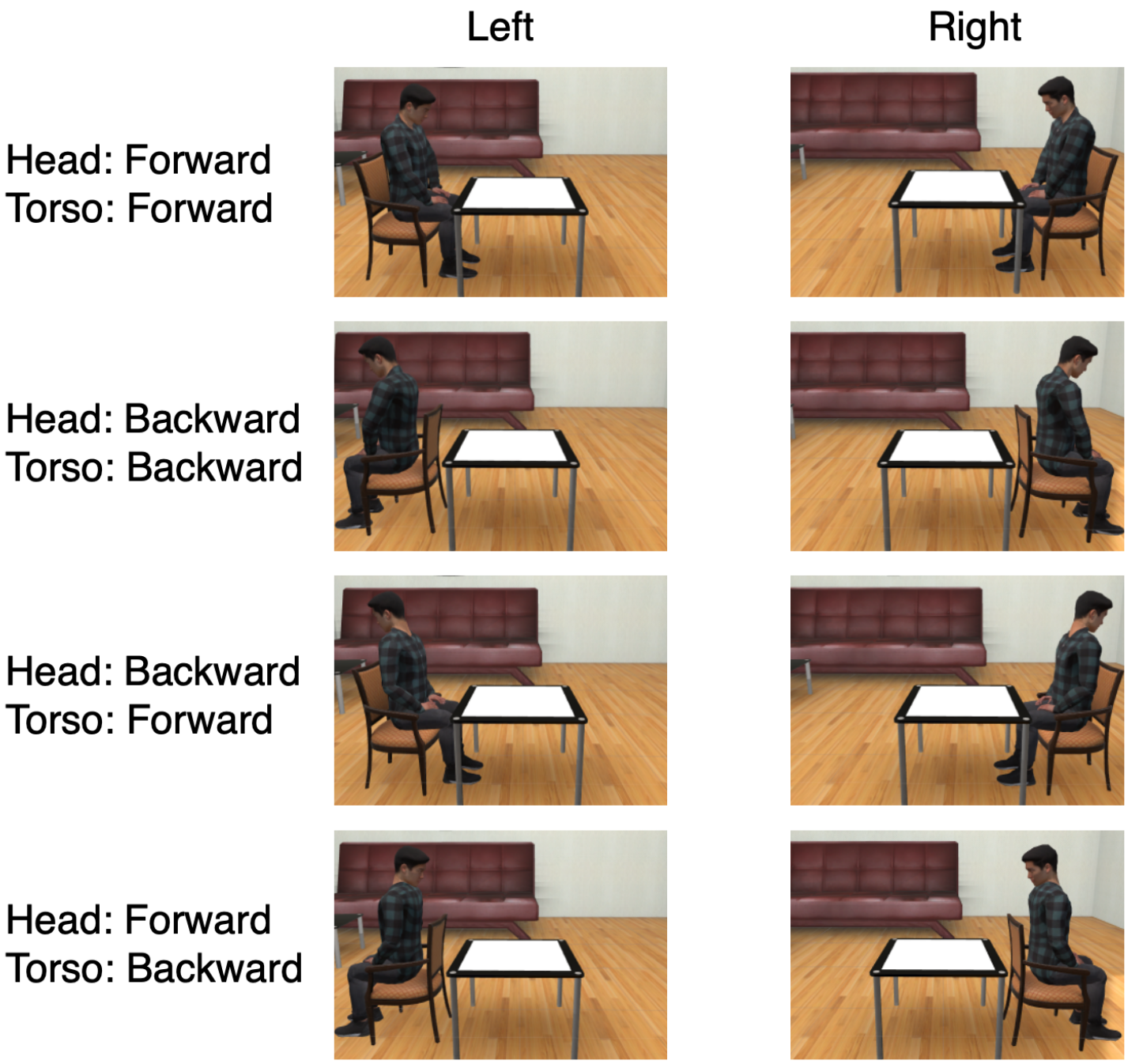

Figure 5

A subset of the stimuli used in Experiment 3. An avatar sitting in a chair was presented in either the left or right position from the participant's point of view. In addition to the forward avatar and the backward avatar, a torso-only forward avatar and a head-only forward avatar were employed.

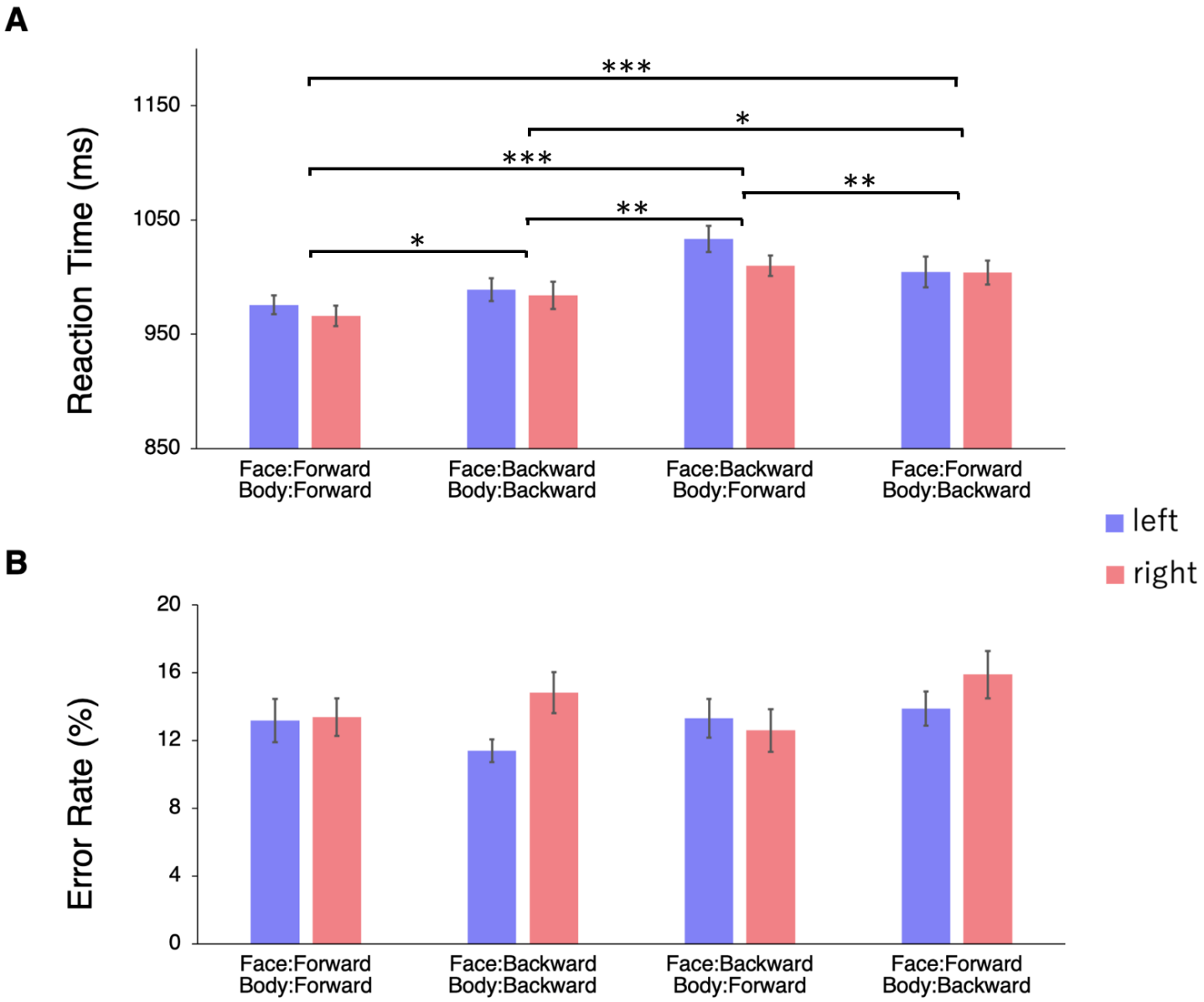

Figure 6

Results of Experiment 3. (A) Mean reaction times (RTs). Two-way repeated measures ANOVA and Shaffer's post hoc tests were conducted (*p < .05). The RTs were significantly faster in the order of the condition (1) both the head and the torso were facing forward, (2) both the head and the torso were facing backward, (3) the torso was facing toward the stimulus while the head was facing away, and (4) the head was facing toward the stimulus while the torso was facing away. (B) Error rates. There were no significant main effects or interactions for the accuracy. Error bars represent 95 % within-subjects confidence intervals.