Effect of body direction on taking the perspective of a humanoid avatar in a virtual environment

Sachiyo Ueda[1*], Kazuya Nagamachi[1], Maki Sugimoto[2], Masahiko Inami[3], and Michiteru Kitazaki[1]

[1]Department of Computer Science and Engineering, Toyohashi University of Technology, Toyohashi, Aichi, 441-8580, Japan.

[2]Department of Information and Computer Science, Keio University, Yokohama, Kanagawa, 223-8522, Japan.

[3]Research Center for Advanced Science and Technology, The University of Tokyo, Bunkyo-ku, Tokyo, 113-0033, Japan.

*Correspondence: ueda.sachiyo.ms@tut.jp

Abstract

Understanding another's viewpoint is the ability to infer others' minds, which is important for successful social communication. 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. This suggests that the facilitation effect is based not only on attention capture but also the embodied perspective-taking to the avatar.

Introduction

The ability to reasonably understand other people's mental states is important for smooth social communication. This requires considering their perspectives in the visual, conceptual, and emotional domains. Visual perspective-taking, in particular, is a core process of social cognition because understanding the visual content of another person's mind is essential for grasping their thoughts and feelings and predicting their behavior.

Visual perspective-taking is not a unitary ability but can be divided into different levels [1, 2, 3]. 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. Since Level 2 perspective-taking develops later [2, 4, 5], this suggests that it involves cognitively demanding computational processing. Level 2 perspective-taking is underlain by a mechanism that mentally rotates oneself to another person's position [3, 6, 7, 8]. It is presumed that this embodied mental rotation involves a different cognitive process from the ability to mentally rotate objects [9]. The mental rotation of one's body is faster and more accurate than the mental rotation of an object [10, 11, 12]. In addition, while the reaction time (RT) for object mental rotation increases in proportion to the rotation angle, mental self-rotation shows a constant RT up to 60° and a sudden increase in the RT thereafter (e.g., [3, 10, 13]).

However, it is not very clear how mental rotation of one's body is performed when taking another's perspective. An imagined shift of perspective to an arbitrary position is possible

even when no physical body or avatar is present and, regardless of whether there is another person or inanimate object in that imaginary position, the RT becomes longer according to the angular disparity between one's own and imagined position [3]. This confirms that the imagined shift of perspective commonly includes the cognitive process of the mental rotation of one's body. However, the presence of a humanoid avatar can make the transformation of the perspective more efficient [3, 14]. This suggests that Level 2 perspective-taking of another person may involve a unique process in addition to the mere perspective transformation. In this study, we investigated the mechanism underlying efficient Level 2 perspective-taking, and especially examined whether it requires the embodiment of others. There are two possible reasons why the presence of another person facilitates the imagined shift of perspective. First, it may be that humans infer the mental state of another person very quickly and automatically. Ward, Ganis and Bach (2019) used a task that involved the classic mental rotation task, and showed that the facilitation effect of Level 2 perspective-taking was due to the "social" feature of others [4]. They demonstrated that the 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. They concluded that by incorporating another person's visual perspective, the rotation angle of the letter required for the judgment became smaller, and the RT became shorter. This facilitation effect disappeared when a lamp (inanimate object) was

presented instead of the humanoid avatar. Second, it is possible that general attention function explains the facilitation effect. Quesque et al.(2018) investigated to what extent humans spontaneously adopt others' perspectives when interpreting a visual scene [15]. They asked participants to identify the number at the center of a collection of 15 cards that depicted various symbols. The cards were displayed on a table and there was another person looking at the cards across the table. The central number could be interpreted as "6" or "9" depending on the point of view (i.e., the participant's perspective or the other person's perspective). The results showed that, contrary to the authors' expectation, regardless of whether the person could see the visual stimuli (in some conditions, a blindfolded person was presented), an overwhelming number of participants pro-actively and spontaneously took the other person's perspective when interpreting the ambiguous visuo-spatial stimuli. This finding demonstrates that the existence of other people facilitates the perspective transformation without the process of inferring others' minds. The results of Ward, Ganis and Bach (2019) might also be interpreted as showing that humanoid avatars draw more attention than inanimate objects, which facilitates the judgment from the imagined viewpoint [14].

As mentioned above, there is still some controversy about the efficiency of Level 2 perspective-taking using a humanoid avatar. In the current study, we hypothesized that embodiment of the avatar facilitates Level 2 perspective-taking. We developed a task to investigate how a target stimulus looks, that involved a simple direction judgment and

83 examined in detail how the humanoid avatar facilitates visual perspective-taking in a virtual
84 environment. Participants observed the three-dimensional scene in the same size as reality on
85 the head-mounted display. In Experiment 1, two conditions were created: one in which an
86 avatar looking at a visual stimulus was presented and another in which only an empty chair
87 was shown (Fig. 1A, 1B). Participants were asked to judge the direction of the crack in a ring
88 (like a Landolt ring) from the avatar's perspective or the empty chair's position, and to
89 respond using the joystick as accurately and quickly as possible. We then examined whether
90 the presentation of the avatar facilitated the perspective transformation. We also manipulated
91 the interval between the presentation of the avatar and/or the chair and the visual stimulus to
92 examine the time scale of perspective-taking. In Experiment 2, we manipulated the
93 orientation of the avatar (forward or backward) with respect to the target stimulus and
94 examined whether the avatar's gaze on the target was necessary for Level 2 perspective-
95 taking (see Fig. 1). If the RT was shortened only when the avatar's gaze was directed at the
96 target (i.e., in the forward-facing condition), it would indicate that the humanoid avatar is not
97 simply facilitating the perspective judgment from an arbitrary position by strongly attracting
98 attention, but that it is necessary to infer the other's mental state of how the target stimulus is
99 seen by the avatar.
100
101 **Figure 1.** here

Results

Experiment 1: Reaction time and accuracy were improved with the humanoid avatar

Individual mean RTs and error rates were calculated for each of the twelve conditions that consisted of the combination of avatar existence (with / without), interval (short / long), and position of the avatar and chair (left, front, and right). The different conditions of the direction of the ring's crack (0°, 45°, 90°, 135°, 180°, 225°, 270°, and 315°) were merged. 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 ring to the time when the joystick reached the end position (i.e., a circumference with a radius of 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 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 [16]. When performing the multiple comparisons after the ANOVAs, we reported

121 the *p*-values that were corrected using Shaffer's modified sequentially rejective Bonferroni
122 procedure [17].

123 The ANOVA of RTs showed significant main effects of the existence of the avatar,
124 $F(1, 19) = 19.570$, $p < 0.001$,$\eta_p^2 = 0.507$, the length of the interval, $F(1, 19) = 73.376$, $p <$
125 0.001, $\eta_p^2 = 0.794$, and the position of the avatar and chair, $F(1.170, 22.222) = 7.864$, $p =$
126 0.008, $\eta_p^2 = 0.293$. There was also a significant interaction between the avatar existence and
127 interval, $F(2, 38) = 7.355$, $p = 0.014$, $\eta_p^2 = 0.279$, and between position and interval, $F(2, 38)$
128 $= 12.068$, $p < 0.001$, $\eta_p^2 = 0.388$. The avatar existence × position interaction, $F(2, 38) =$
129 3.138, $p = 0.055$, $\eta_p^2 = 0.142$, and avatar existence × interval × position interaction, $F(1.351,$
130 $25.670) = 0.383$, $p = 0.604$, $\eta_p^2 = 0.020$, were not significant.
131 Participants' RTs were significantly faster for the "with avatar" condition than the "without
132 avatar" condition, only in the short interval condition ($p<0.001$) (Fig. 2A). In the short
133 interval condition, the RTs were slower in the front position than in the other two positions
134 ($ps < 0.01$). In the long interval condition, the RTs were faster in the left position than in the
135 other two positions ($ps < 0.05$). In all conditions, the long interval condition had faster RTs
136 than the short interval condition.
137
138 **Figure 2.** here
139 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. 2B). 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.

Experiment 2: The forward avatar facilitated the imaged shift of perspective over the empty chair, and the backward avatar was worse than the empty chair

Individual mean RTs and error rates were calculated for each of the nine conditions that consisted of the combination of the avatar type (without, forward, or backward avatar) and position of the avatar and chair (left, front, or right) by the same procedure 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 (6.2s%) were excluded 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. If there was a lack of sphericity, the reported values were adjusted using the Greenhouse-Geisser correction [16]. When conducting the multiple comparisons after the ANOVAs, we reported the *p*-values that were corrected using Shaffer's modified sequentially rejective Bonferroni procedure [17].

159 The ANOVA of RTs showed significant main effects of the type of avatar, $F(2, 38) = 25.459$,
160 $p < 0.001$, $\eta_p^2 = 0.573$, and position of the avatar and chair, $F(2, 38) = 14.952$, $p < 0.001$, η_p^2
161 $= 0.440$. The avatar existence × position interaction was not significant, $F(4, 76) = 1.113$, $p =$
162 0.357, $\eta_p^2 = 0.055$. The RTs were significantly faster in the order of the forward avatar
163 condition, without avatar condition, and backward avatar condition (*ps* < 0.01). The RTs were
164 significantly slower in the front position condition than in the right position and left position
165 conditions (*ps* < 0.001) (Fig. 3A).
166 The ANOVA of error rates revealed that there were no significant main effects or interactions
167 for the accuracy (Fig. 3B).
168
169 **Figure 3.** here
170
171 **Discussion**
172 We can imagine observing a scene from a viewpoint that is different from ours. Above all,
173 knowing what or how the environment appears from another person's point of view is an
174 automatic and very efficient cognitive process [e.g., 14, 18, 19]. The present study was
175 conducted to examine whether the presence of a humanoid avatar enhanced the imagined
176 shift of perspective; we found that it improved the performance of identifying the orientation
177 of a visual stimulus from an imagined position only in the short interval condition (200 ms)

178 in Experiment 1. This indicates that, with or without avatars, the perspective transformation
179 to an arbitrary position can be performed sufficiently for 1000 ms (long interval), but the
180 perspective-taking is a faster and more efficient cognitive process than imagining the
181 perspective from the object position. The results of Experiment 2 showed that the facilitation
182 of the perspective transformation occurred only when the avatar's body and head were facing
183 the visual stimulus. Thus, the direction of a humanoid avatar's head or body is an effective
184 cue for the facilitation effect of spatial cognition in perspective-taking. This suggests that the
185 facilitation effect of the presence of humanoid avatars on Level 2 perspective-taking is not
186 based on attention capture due to the saliency of the humanoid avatars, but on the efficiency
187 of the cognitive process of perspective-taking.
188 To the best of our knowledge, this is the first study to provide direct evidence that calculating
189 an avatar's perspective of a scene is faster than imagining the perspective of an inanimate
190 object that is in the same position, in Level 2 perspective-taking. Previous studies have
191 reported that Level 1 perspective-taking occurs automatically and implicitly, but Level 2
192 perspective-taking does not (e.g., [18, 19, 20, 21]) (note: except when collaborating with
193 others). In these studies, the participant and avatar were in different spatial positions, and the
194 participants were asked to judge what or how a scene looked from their own perspective. In
195 this task, although the participants did not need to consider the avatar's viewpoint, the RT
196 was longer when the two viewpoints did not match each other than when they matched. That

197 is, the participants were automatically influenced by the viewpoint of the avatar. Since the
198 interference effect occurs only in Level 1 perspective-taking, this implies that Level 1
199 perspective-taking is implicit and spontaneous, whereas Level 2 perspective-taking is a
200 voluntary and more effortful cognitive process (but, see also [14]). The interference effect is a
201 good way to study the automaticity of perspective-taking; however, it is not sufficient for
202 studying the speed and efficiency of the process.
203
204 Recently, using a classic mental rotation task, Ward, Ganis and Back (2019) reported that a
205 humanoid avatar could lead to the participant making a judgment faster if the scene that
206 appeared from the avatar's position helped when responding to the task, while an inanimate
207 object (a lamp) did not show this effect [14]. However, in their study, it was not clear whether
208 this facilitation effect was due to the saliency of the humanoid avatar or the efficiency of
209 calculating the content of the avatar's perspective. In addition, Michelon and Zacks (2006)
210 showed that a Level 2 perspective-taking task was performed faster when a doll was
211 presented compared to an asterisk; however, they discussed the cost of the avatar's absence
212 by comparing two experiments indirectly, and the doll was larger than the asterisk [3]. The
213 saliency due to the size was not controlled. In the current study, we investigated the
214 mechanism that underlies the avatar's facilitation effect on perspective-taking by
215 manipulating only the bodily orientation of the humanoid avatar. As a result, the RT was

faster only when the humanoid avatar was facing forward. This suggests that not only spatial attention but also the embodiment of the avatar facilitates Level 2 perspective-taking.

The avatar facilitation effect was observed only in the short (200 ms) interval condition. This shows the time scale on which the viewpoint is transformed to the position of another person or inanimate object. That is, when the avatar is present, the viewpoint jumps quickly to the position of avatar, but this process does not occur for the chair. Whereas, in the case where there is sufficient time between the presentation of the avatar or chair and the presentation of the target stimulus (i.e., long interval condition), the viewpoint can be consciously transformed to an arbitrary position regardless of the presence of the avatar. This may lead to no difference in RT for perspective-taking.

There was also a difference in RT depending on the position of the avatar. In the short interval condition (both in Experiment 1 and 2), the RTs were longer in the front position than in the right and left positions, and in the long interval condition, the RTs were longer in the front and right positions than in the left position. The longest reaction time 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., [3, 7, 8]). 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), 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, the trade-off alone is not a sufficient explanation. The gradient of reaction time weakened in the long interval condition and there was no difference between the reaction time 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 (broken ring). In that case, the observer can finish perspective transformation to any position before presentation of the target stimulus, and therefore, the reaction time 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 reaction time 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 reaction time 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.

One may argue that participants learned the pattern between the position of the avatar and the

correct answer, and adopted the strategy of reacting according to that pattern without imagining the perspective from the position of the avatar or the chair. For example, when the avatar is present on the right, the correct response can always be calculated by rotating 90 degrees clockwise the crack direction of the ring from the perspective of the participants. However, this possibility is unlikely. This is because the position of the avatar changes within the block, so it is not easy to switch between different patterns depending on the position of the avatar. Furthermore, if the participants adopted the angular rotation pattern strategy, the 180 degrees pattern was the simplest and the reaction time was expected to be shortest in the front position condition (e.g., [22, 23]). However, the result showed the longest reaction time in the front position. Therefore, it is reasonable to suppose that the participants did not adopt the pattern rotation strategy but judged the perspective from the position of the avatar and the chair.

This study showed that the direction of a humanoid avatar's head or body is crucial for the facilitation effect on the imagined shift of perspective. However, it is unclear whether the orientation of the face or body is important. The embodiment of the avatar could enhance the imagined shift of perspective. If so, the enhancement should be further improved by illusory ownership to the avatar using the full body illusion (e.g., [24, 25, 26]). These points should be examined in future research.

Methods

Experiment 1

Participants: Twenty paid volunteers participated in the experiment (17 men, 3 women, all aged 19–24 years). 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. 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 placed in the center of the room, and either an empty chair or an avatar that was sitting in the chair was presented in one of three positions, that is, on the left, right, or other side of the table (Fig. 1). Then, a broken ring was presented on the table. The crack in the broken ring 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 ring

(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 crack in the ring). The directions of the crack in the ring were merged in the analysis.

Procedure: Each trial began with a black blank 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 (the "without avatar condition" included just the table and chair). Subsequently, the broken ring was presented on the table after 200 ms (short interval) or 1,000 ms (long interval), depending on the condition. The participants were asked to judge the direction of the crack in the ring from the avatar's perspective or the empty chair's position and to respond with the joystick as accurately and quickly as possible (Fig. 4). Participants received no feedback. The next trial began immediately after the joystick response.

Figure 4. here.

In the practice trials, the direction judgment task from the participant's perspective was conducted first. Then, a block of 12 practice trials (2 with/without the avatar, 2 short/long intervals, and 3 positions of the avatar and chair) was presented, and the participants judged

311 the direction of the crack in the ring from the avatar's perspective or the empty chair's
312 position, same as in the test session. In the test session, all combinations of the conditions
313 (two with/without the avatar, two short/long intervals, three positions of the avatar and chair,
314 and eight directions of the crack in the ring) were repeated twice in a random order, and 192
315 trials were conducted within the same block. There were four blocks, and a total of 768 test
316 trials. It took approximately 90 minutes for each participant to finish this experiment,
317 including the time required to provide the experimental instructions, to take a break, and to
318 conduct the practice trials.
319
320 **Experiment 2**
321 **Participants:** Twenty paid volunteers participated in the experiment (15 men, 5 women, all
322 aged 20–24 years). Eight of them had participated in Experiment 1. They all had normal or
323 corrected-to-normal vision and were naïve to the purpose of the study. The procedure for
324 informed consent and the ethical review were the same as in Experiment 1.
325
326 **Apparatus:** Experiment 2 used the same apparatus as in Experiment 1.
327
328 **Stimuli and conditions:** This experiment differed from Experiment 1 in that we added the
329 backward avatar condition (Fig. 1C). Since there was no difference in the avatar's effect

between the two interval conditions in Experiment 1, only the short interval was employed in Experiment 2. There were 72 combinations of trials (three types of avatar: no avatar, forward avatar, and backward avatar; three positions of the avatar and chair; and eight directions of the crack in the ring). The directions of the crack in the ring 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 did not imagine the avatar's perspective (in which case the ring is not visible), but imagined the perspective when looking from the avatar's position at the center. In the practice trials, the direction judgment task from the participant's perspective was conducted first. Then, a block of nine practice trials (three types of avatar and three positions of the avatar and chair) was presented. In the test session, all combinations of the conditions (three types of avatar, three positions of the avatar and chair, and eight directions of the crack in the ring) were repeated twice in a random order, and 144 trials were conducted within the same block. There were four blocks, and a total of 576 test trials. It took approximately 60 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 Availability: Original data for figures in the paper is available at Mendeley Data

DOI: 10.17632/6n953kv7f8.1

References

1. Apperly, I. A. & Butterfill, S. A. Do humans have two systems to track beliefs and belief-like states? *Psychol. Rev*. **116**, 953-970 (2009).
2. Flavell, J. H., Everett, B. A., Croft, K. & Flavell, E. R. Young children's knowledge about visual perception: Further evidence for the level 1–level 2 distinction. *Dev. Psychol.* **17**, 99-103 (1981).
3. Michelon, P. & Zacks, J. M. Two kinds of visual perspective taking. *Percept. Psychophys.* **68**, 327-337 (2006).
4. Gzesh, S. M. & Surber, C. F. Visual perspective-taking skills in children. *Child Dev.* **56**, 1204-1213 (1985).
5. Masangkay, Z. S., McClusky, K. A., McIntyre, C. W., Sims-Knight, J., Vaughn, B. E. & Flavell, J. H. The early development of inferences about visual percepts of others. *Child Dev.* **45**, 357-366 (1974).
6. Gardner, M. R., Stent, C., Mohr, C. & Golding, J. F. Embodied perspective-taking indicated by selective disruption from aberrant self-motion. *Psychol. Res.* **81**, 480-489 (2017).
7. Kessler, K. & Thomson, L. A. The embodied nature of spatial perspective taking: Embodied transformation versus sensorimotor interference. *Cognition*, **114**, 72-88 (2010).
8. Surtees, A. D. R., Apperly, I. A. & Samson, D. Similarities and differences in visual and

370 spatial perspective-taking processes. *Cognition*, **129**, 426-438 (2013).

371 9. Shepard, R. N. & Metzler, J. Mental rotation of three-dimensional objects. *Science*, **171**,
372 701-703 (1971).

373 10. Keehner, S. A. Guerin, M. B. Miller, D. J. Turk. & M. Hegarty, Modulation of neural
374 activity by angle of rotation during imagined spatial transformations. *Neuroimage*, **33**,
375 391-398 (2006).

376 11. Wraga, M., Shephard, J. M., Church, J. A., Inati, S. & Kosslyn, S. M. Imagined rotations
377 of self versus objects: An fMRI study. *Neuropsychologia*, **43**, 1351-1361 (2005).

378 12. Zacks, J. M. & Michelon, P. Transformations of visuospatial images. *Behav. Cogn.*
379 *Neurosci. Rev.* **4**, 96–118 (2005).

380 13. Kozhevnikov, M. & Hegarty, M. A dissociation between object manipulation spatial
381 ability and spatial orientation ability. *Mem. Cogn.* **29**, 745-756 (2001).

382 14. Ward, E., Ganis, G. & Bach, P. Spontaneous Vicarious Perception of the Content of
383 Another's Visual Perspective. *Curr. Biol.* **5**, 874-880.e4 (2019).

384 15. Quesque, F., Chabanat, E. & Rossetti, Y. Taking the point of view of the blind:
385 Spontaneous level-2 perspective-taking in irrelevant conditions. *J. Exp. Soc. Psychol.* **79**,
386 356-364 (2018).

387 16. Geisser, S. & Greenhouse, S. W. An extension of Box's results on the use of the F
388 distribution in multivariate analysis. *Ann. Math. Stat.* **29**, 885-891 (1958).

389 doi:10.1214/aoms/1177706545

390 17. Shaffer, J. P. Modified sequentially rejective multiple test procedures. *J. Am. Stat. Assoc.*
391 **81**, 826-831 (1986).

392 18. Furlanetto, T., Becchio, C., Samson, D. & Apperly, I. Altercentric interference in level 1
393 visual perspective taking reflects the ascription of mental states, not submentalizing. *J.*
394 *Exp. Psychol. Hum. Percept. Perform.* **42**, 158-163 (2016).

395 19. Samson, D., Apperly, I. A., Braithwaite, J. J., Andrews, B. J. & Scott, S. E. B. Seeing it
396 their way: Evidence for rapid and involuntary computation of what other people see. *J.*
397 *Exp. Psychol. Hum. Percept. Perform.* **36**, 1255-1266 (2010).

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

400 21. Surtees, A., Samson, D. & Apperly, I. I've got your number: Spontaneous perspective-
401 taking in an interactive task. *Cognition*, **150**, 43-52 (2016b).

402 22. Neely, K. N. & Heath, M. Visuomotor mental rotation: Reaction time is determined by
403 the complexity of the sensorimotor transformations mediating the response. *Brain Res.*
404 **1366**, 129-140 (2010).

405 23. Neely, K. N. & Heath, M. The Visuomotor Mental Rotation Task: Visuomotor
406 Transformation Times Are Reduced for Small and Perceptually Familiar Angles. *J. Mot.*
407 *Behav.* **43**, 393-402 (2011).

24. Lenggenhager, B., Tadi, T., Metzinger, T. & Blanke, O. Video ergo sum: manipulating bodily self-consciousness. *Science,* **317**, 1096-1099 (2007).

25. 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.* **8**, 1-8 (2018).

26. 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 T VIS COMPUT GR*, **26**, 2062-2072 (2020).

27. Morey, R.D. Confidence intervals from normalized data: A correction to Cousineau (2005). *Tutor Quant Methods Psychol.* **4**, 61–64 (2008).

Figure Legends

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). Participants were asked to judge the direction of the crack in the ring 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. Results of Experiment 1. (A) Mean reaction times. Three-way repeated measures ANOVA and Shaffer's post hoc tests were conducted (*p < .05, **p < .01, ***p < .001). The reaction times (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 [27].

Figure 3. Results of Experiment 2. (A) Mean reaction times. Two-way repeated measures ANOVA and Shaffer's post hoc tests were conducted (*p < .05, **p < .01, ***p < .001). The reaction times (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 [27].

Figure 4. Procedure of Experiment 1.

Acknowledgements

We would like to thank Naho Isogai for collecting some of the data. This research was supported by the Japan Science and Technology Agency's ERATO grant (number JPMJER1701) (Inami JIZAI Body Project) and the Japan Society for the Promotion of Science's KAKENHI grant (numbers JP20H04489 and JP20K20147). We would like to thank Editage (www.editage.jp) for English language editing.

Author Contributions

S.U., K.N., M.S., M.I. and M.K. conceived and designed the experiments. S.U. and K.N. collected and analyzed the data. S.U. and M.K. contributed to the preparation of the manuscript. All authors reviewed the manuscript.

Additional Information

The authors declare no competing interests.