To appear in
Journal of Experimental Psychology: Human Perception and Performance (2019)

Seeing the World through the Eyes of an Avatar?

Comparing Perspective Taking and Referential Coding

Sophia von Salm-Hoogstraeten, Katharina Bolzius, and Jochen Müsseler

RWTH Aachen University, Germany

Author's note

Correspondence is asked to be addressed to Sophia von Salm-Hoogstraeten, Institute of Psychology, RWTH Aachen University, 52066 Aachen, Germany (salm@psych.rwth-aachen.de). The authors declare that the study was conducted and the manuscript prepared without financial considerations or other conflicts of interest. This study was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation; project number MU 1298/11) and was associated with the DFG Priority Program "The Active Self" (DFG SPP 2134). The data collected in the experiments are archived at the APA repository hosted by the Center for Open Science https://osf.io/bgqwn/. We would like to thank Axel Mayer very much for his statistical advices.

Abstract

Previous studies have shown that users spontaneously take the position of a virtual avatar and solve spatial tasks from avatar's perspective. The common impression is that users develop a spatial representation that allows them to "see" the world through the eyes of the avatar, that is, from its virtual perspective. In the present paper, this perspective taking assumption is compared with a referential coding assumption that allows the users to act on the basis of changed reference points. Using a spatial compatibility task, Experiment 1 demonstrated that the visual perspective of the avatar was not the determining factor for taking avatar's spatial position, but that its hand position (as the reference point) was decisive for the spatial coding of objects. Experiment 2 showed, however, if the participant's hand posture was not corresponding with the avatar's hand postures, the spatial referencing by avatar's hands expired, thereby demonstrating the limits of referential coding. Still, the present findings indicated that referential coding may be at the base when taking avatar's perspective. Accordingly, any study in perspective taking needs to consider and evaluate possible mechanisms of referential coding.

Keywords: visual perspective taking, referential coding, spatial SR compatibility, avatar

Public Significance Statement

The paper investigates the mechanisms which allow participants to take the spatial perspective of others, operationalized as their virtual avatar. The own knowledge about the perspective of another person is decisive in many forms of social interactions. Likewise, factors which determine spatial perception and action are of relevance to many scientists in cognitive psychology, informatics and robotics. Our experiments show that the presence of a virtual avatar has a regular influence on spatial compatibility relations. Knowledge of such spatial relations is equally important for applied scientists with interests in human factors.

Seeing the World through the Eyes of an Avatar?
Comparing Perspective Taking and Referential Coding

In social interactions it is necessary to understand how aspects of the world appear to others. For a long time, mechanisms have been assumed which make it possible to take the other person's spatial perspective (e.g. Flavell, Everett, Croft, & Flavell, 1981; Tversky & Hard, 2009). Similar to how humans seem to be able to put themselves in the perspective of others they take the position of a virtual avatar (Samson, Apperly, Braithwaite, Andrews, & Bodley Scott, 2010; Surtees, Apperly, & Samson, 2013; Furlanetto, Becchio, Samson, & Apperly, 2016; Cavallo, Ansuini, Capozzi, Tversky, & Becchio, 2017). For example, it has been shown recently that the presentation of an avatar on the opposite side of the participant causes a remapping of spatial relation to an object when left-right decisions are made (Cavallo et al., 2017).

Changes in own behavioral tendencies by the presence of (virtual) others have been considered to reflect a spatial perspective taking mechanism. Generally perspective taking is the ability to mentally rotate oneself in the position of another person (Flavell et al., 1981). Evidences for the existence of perspective taking are reported for the perspective adopted by the user towards virtual avatars or stick figures (Jancyk, 2013; Franz, Sebastian, Hust, & Norris, 2008; Samson et al., 2010) and human confederates (Freundlieb, Kovács, & Sebanz, 2016; Freundlieb, Sebanz, & Kovács, 2017).

Further, findings associated with perspective taking have been interpreted as a sign of a mechanism by which humans are able to spontaneously mentalize the spatial perspective of another and to 'see' the world through that person's eyes. This automatic processing of another's visual perspective is assumed as part of a broader ability to implicitly mentalize what another person is experiencing (see Heyes, 2014 for review and critical discussion). The strongest empirical evidence for the existence of an implicit adoption of another's visual perspective are findings showing that the own response tendencies of the participant towards an object are only changed by an avatar or by a human if the other has visual access to the scenario (Furlanetto et al., 2016; Freundlieb et al., 2017). The latter study showed that left-right coding of objects relative to the spatial perspective of a human task-partner was only adopted by the participant when the task partner saw the objects. But no changes in response tendencies of the participant were observed when the task partner wore a blindfold (Freundlieb et al., 2017).

In contrast, Quesque, Chabanat, and Rossetti (2018) reported that participants spontaneously encoded ambiguous digits (6 and 9) from the perspective of the other person sitting opposite them and not from their own position, even if the other person wore a blindfold and thus, had no visual access to the stimuli. In line with this observation, an increasing number of investigations is questioning the existence of a specific mechanism to process the visual perspective of another (virtual) person (Conway, Lee, Ojaghi, Catmur, & Bird, 2017; Santiesteban, Catmur, Hopkins, Bird, & Heyes, 2014). Thus, questions remain regarding which underlying mechanisms are responsible for incorporating the spatial perspective of other persons into our own actions.

The avatar-Simon task

Especially when avatars carry out one's own actions it is necessary to consider their spatial position. A series of recent experiments (Müsseler, Ruhland, & Böffel, 2019; Böffel & Müsseler, 2018, 2019a, b) showed that participants spontaneously included the spatial position of an avatar in their behavioral tendencies when solving spatial compatibility tasks. Generally, spatial stimulus-response compatibility (SR compatibility) describes the phenomenon that corresponding stimulus and response positions lead to better performance than non-corresponding conditions. For instance, left and right responses to stimuli on the same side of the own body (e.g. left stimulus, left response) are performed faster and with less errors than responses to stimuli on the other side (e.g. left stimulus, right response, see Proctor & Vu, 2006 for an overview).

Importantly, the spatial relationship between stimuli and responses still affects performance when the position of the stimulus is actually irrelevant to the task (Simon & Small, 1969). If, for example, participants respond with left and right key presses to the colors of the stimuli (e.g. a left-sided response to one color, a right-sided response to another color) and these stimuli appear randomly to the left or right of the body's midline, spatially corresponding responses show better performance than spatially non-corresponding responses (for reviews see Hommel, 2011; Lu & Proctor, 1995). The so-called Simon effect is often explained using dual-route accounts (e.g. Hommel, 1997; Kornblum, Hasbroucq, & Osman, 1990). Accordingly, the activation of a response by the spatial position of the stimulus can either take place automatically (via the direct route) or is determined by the SR mappings defined by the current task instructions (the indirect route). It is assumed that the performance losses in incompatible conditions are the result of a response conflict produced by non-matching activations of the direct and indirect route. In compatible conditions, the SR mappings assigned by the direct and indirect route match, resulting in faster and more accurate reactions.

Since users and avatars with different spatial positions are expected to have different spatial SR compatibilities to the same stimulus, the modulation of SR compatibility by an avatar can be used to investigate the extent to which users acquires the position of *their* avatar. For example, Böffel and Müsseler (2019a) showed that the perspective of an avatar is taken by the participant even if it is irrelevant to the task (the avatar-Simon effect). In an avatar-Simon procedure, participants typically categorize the color of a disc appearing in an upper or lower position of the central vertical axis of a computer screen and to press left or right response keys[1]. In addition, an avatar is presented on the screen, which is positioned either on the left or on the right side of the central horizontal axis. Thereby, the avatar is arranged in such a way that each arm is aligned with one stimulus position (avatars A and B in the upper panel of Fig. 1; Böffel & Müsseler, 2019a, Exp. 1). In contrast to the participant, to whose position the stimuli are orthogonally arranged, the avatar has a horizontal perspective towards the stimuli: In conditions in which the avatar is on the left side, the upper (lower) stimulus is on its left (right) side. When the avatar is positioned in the right position, the lower (upper) stimulus is arranged on the left (right) side of the avatar's body center.

With this arrangement, the spatial reference frame of the avatar towards the stimuli can be varied systematically, while the spatial reference frame of the participant (orthogonal alignment to the stimuli) is kept constant. The result of the avatar-Simon task is that spatial compatibility effects between the response positions of the participant and the stimulus positions are determined by the position of the avatar (Böffel & Müsseler, 2019a). When the avatar's position is on the left side (Fig. 1, avatar A) and the avatar has the upper stimulus to its left side and the lower stimulus to its right side, participant's responses with the left (right) hand to upper (lower) stimuli are facilitated compared to responses with the right (left) hand to upper (lower) stimuli. The SR compatibility effect turns in the exact opposite direction when the avatar is in the right position (Fig. 1, avatar B). In this case, right (left) responses to upper (lower) stimuli are facilitated compared to right (left) responses to lower (upper) stimuli.

Böffel and Müsseler (2019a) saw the avatar-Simon effect as the tendency to spontaneously take the perspective of the avatar. The rotated avatars (Fig. 1, upper panel) used in this study have the stimuli on the left or right side of their visual open space. If participants take avatar's visual perspective, they show response tendencies that facilitate responses on its left (right) side to stimuli in the left (right) field of vision of the avatar.

However, in the avatar scenario used by Böffel and Müsseler (2019a) the left (right) hand of the avatar might have formed already a frame of reference to the stimuli without the participants having to actually take the avatar's perspective visually. If we assume that the left hand of the avatar encodes the near stimulus as in left position and the right hand encodes the other stimulus position as in right position, this would also create a spatial reference frame towards the stimuli, which is independent of the visual perspective of the avatar.

That the spatial coding of objects can arise in reference to other objects is an idea postulated by the referential coding account which was originally proposed to explain spatial compatibility effects in the standard Simon effect (Hommel, 1993) and was recently elaborated with regards to the joint Simon effect (Dolk, Hommel, Prinz, & Liepelt, 2013; Dolk et al., 2014). According to the referential coding account, the basic spatial representation develops from the actor's perspective, which, however, contains already all spatial relationships between objects in the visual space (cf. the visual sensory map of van der Heijden, Müsseler & Bridgeman, 1999). As a consequence, the actor does not need to construct a new visual representation from the other's field of vision when taking the perspective of another person (as is suggested by the perspective taking assumption), she or he rather recodes the existing spatial relationships with regard to any other new reference point. Thus, instead of 'seeing' the world through the avatar's eyes, the existing spatial positions are simply recoded.

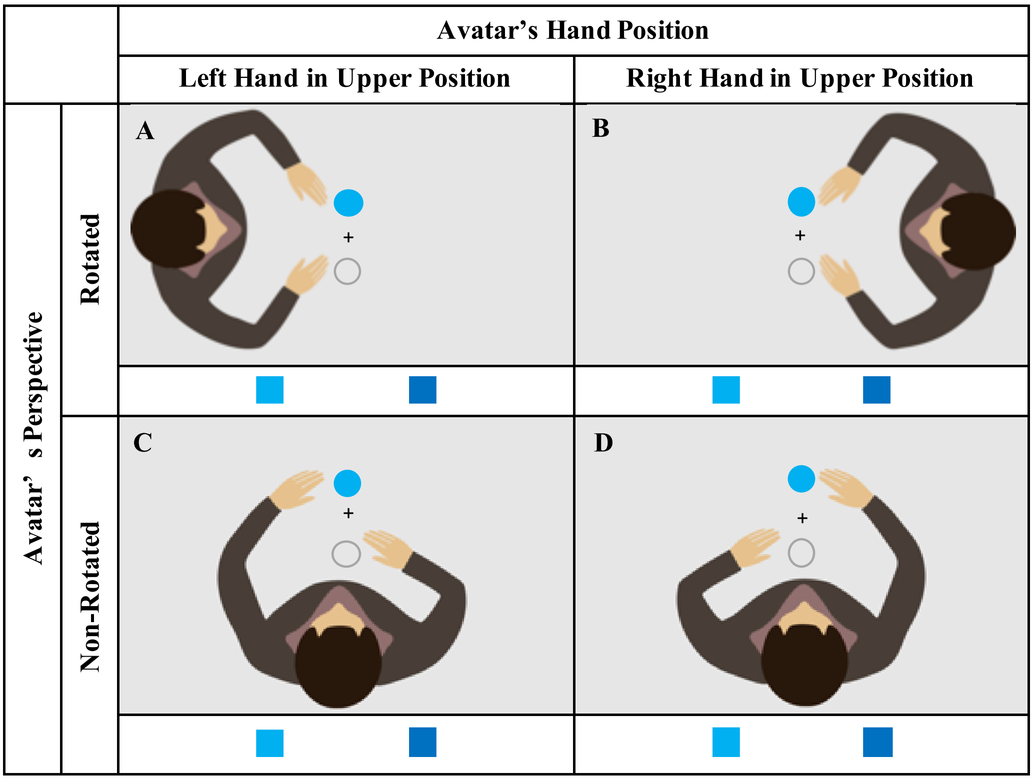


Figure 1. Four different avatars (A-D) used in Experiment 1 and 2 as a function of avatar's hand positions and avatar's perspective. The avatars A and B "see" the stimuli (dark or light blue discs, here light blue) to their left and right side (the other possible stimulus position is shown as a black circle, not presented in the experiments). From the perspective of the avatars C and D the stimuli are arranged vertically. The avatars A and C have their left hand aligned with the upper stimulus position and their right hand aligned with the lower stimulus position. Contrary, the right hands of the avatars B and D were aligned with the upper stimulus position while the left hands were aligned with the lower stimulus position. The response-button positions of the participants were horizontally arranged in Experiment 1 (shown under each avatar). In the figure, the light blue disk calls for left-sided response and the dark blue disk (not depicted) calls for a right-sided response. This mapping is illustrated by the coloring of the response buttons, which were not colored in the experiments.

The present study

The question whether a perspective taking or referential coding mechanism is responsible for the SR compatibility effects from avatar's position can be broken down to the question whether the viewing angle or the hand positions of the avatar modulate SR compatibility. Therefore, the present study compares two different avatar scenarios in which the visual perspective and hand positions of the avatar are disentangled (Fig. 1). In the following experiment we investigated whether the hand positions or the visual perspective of the avatar form the spatial frame of reference towards the stimuli.

Therefore, we introduce a new avatar scenario in which the avatar takes a non-rotated, orthogonal position to the stimuli positions (same visual perspective as the participants, but different hand position, Fig. 1 lower panel). In the non-rotated avatar scenario, only the hand position of the avatar was varied. The avatar had either its left hand in the upper stimulus position and its right hand in the lower position (avatar C) or its right hand in upper position and its left hand in the lower position (avatar D). In the non-rotated avatar scenario, the visual perspective of the avatar cannot modulate SR compatibility because it always corresponds to the perspective of the participant. However, if the hand position of the avatar is responsible for taking spontaneous behavioral tendencies from the position of the avatar, here quantified as the avatar-Simon effect, we expect a modulation of SR compatibility also in the non-rotated scenario.

Experiment 1

Experiment 1 investigated if a perspective taking or a referential coding mechanism is capable to explain the modulation of the SR compatibility effect by avatar's position. We compared two scenarios (Fig. 1, upper vs. lower panel) that disentangled the visual perspective and the hand positions of the avatar, both of which may be responsible for shaping the spatial frame of reference.

Method

Participants. Twenty-four students (3 male and 21 female[2], mean [*M*] age of 21.75 years, standard deviation [*SD*] 2.55 years) participated and received course credit for participation. The sample size was determined prior to the data collection and is equal to the number of participants tested in our previous study (Böffel & Müsseler, 2019a; Exp. 1). All participants reported normal or corrected to normal vision. For the reported series of experiments all participants provided informed consent before the testing. Further, all procedures of the study were in accordance with the ethical principles of the Declaration of Helsinki, except for the study not being pre-registered.

Apparatus and stimuli. The experimental program was run on a Macintosh computer using Matlab software (Mathworks) with the Psychtoolbox-3 extension (Kleiner, Brainard, & Pelli, 2007). Stimuli were displayed on a 22" CRT monitor (Iiyama Visionmaster Pro 514, with a resolution of 1,024 × 768 pixels at 100 Hz). Participants sat approximately 60 cm in front of the display and responded by using their index fingers to press a left or right response key (distance in between 13 cm). The target was a dark blue (RGB 36 115 254) or a light blue (RGB 98 193 254) disc (Ø 50 pixel) which appeared 45 pixels above or below a centered fixation cross in front of a grey background (RGB 231 230 230). The four different avatar images that flanked the target in every trial are illustrated in Figure 1. The avatar was positioned in a way that each hand was aligned with one of the two stimulus positions.

Procedure. The experiment consisted of four parts in each of which one of the four different avatars A to D was continuously presented. Participants were instructed to categorize light and dark blue discs by pressing a left or the right key. They were instructed to take the avatar's perspective and control the avatar with quick and accurate responses. Every part of the experiment consisted of nine blocks with twenty trials each, whereby the first block of every part was considered as practice and excluded from the analysis. In every block both target positions and target colors were repeated five times and the trials were presented in random order. The order within and between both rotated and non-rotated conditions as well as the color to key mapping was counterbalanced across participants.

During each block the avatar and fixation cross were continuously displayed. Every trial started with the presentation of the target which remained visible until any response was made. After a response, the next trial started after an interval of 1,500 ms. When an erroneous response was made or the response interval was lower than 100 ms or exceeded 1,500 ms, a beep feedback was given (two beeps with a frequency of 720 Hz with a duration of 50 ms each and separated by 50 ms). However, the time feedback was implemented solely to improve and speed up participants' responses. After the feedback, the next trial started with an additional delay of 1,500 ms. Participants were tested in one single session which took approximately 60 minutes.

Design. The factors avatar's perspective (left/right rotated or non-rotated), avatar's hand position (left hand in upper position or right hand in upper position), response position of the participant (left or right keypress to the corresponding color of the disc) and stimulus position (upper or lower position) were included in the analysis and formed a 2x2x2x2 within-subject design. The dependent variables were reaction times (RT) and percentages of errors (PE).

Results

The focus of the analysis was to examine whether the visual perspective of the avatar (perspective taking) or its hand posture (referential coding) is responsible for shaping the spatial frame of reference from avatar's position.

Data treatment. Errors were defined as wrong keypresses (2.3 % in total). Error probabilities were arcsine-transformed for inferential statistical analyses with regard to Kirk (1996, p. 106). However, for an easier interpretation untransformed error rates are reported in the figures and in the main text. For RT analysis, only correct responses were considered. Further, individual RT outliers were identified for each participant and each cell of the experimental design by applying Tukey's criterion (i.e., RTs 1.5 times the interquartile range below the first quartile or above the third quartile; Tukey, 1977) and omitted from each individual's RT distribution. This resulted in an exclusion of 4.5% of all trials. Mean RT and PE were submitted separately to two repeated-measures analysis of variance (ANOVAs). We calculated post-hoc pairwise comparisons for all differences relevant to our hypotheses (*t*-tests, two-tailed).

Reaction times. A main effect of avatar's hand position was observed, $F(1, 23) = 6.79$, $p = .016$, $\eta_p^2 = .23$, reflecting faster RT for trials in which avatar's left hand was in upper stimulus position compared to when its right hand was in upper position (434 vs. 441 ms, cf. Fig. 2). Further, there was an interaction between avatar's perspective and the stimulus position, $F(1, 23) = 4.83$, $p = .038$, $\eta_p^2 = .17$. When the avatar was unrotated, responses were faster to upper than to lower stimuli (440 vs. 442 ms). However, when the avatar was rotated responses were slower to upper than to lower stimuli (436 vs. 432 ms). Also, the stimulus position and avatar's hand position interacted, $F(1, 23) = 5.42$, $p = .029$, $\eta_p^2 = .19$. When the avatar had its left hand in upper and its right hand in lower position, responses were faster to lower than to upper stimuli (432 vs. 436 ms). When avatar's right hand was in upper and its left hand in lower position, responses were slower to lower than to upper stimuli (442 vs. 440 ms).

More importantly, a three-way interaction between the factors avatar's hand position, response position and stimulus position was observed, $F(1, 23) = 35.18$, $p < .001$, $\eta_p^2 = .61$. In trials in which the avatar had the left hand in the upper and the right hand in lower position, responses to upper targets were faster with the left than with the right hand (424 vs. 448 ms, $t[23] = -4.08$, $p < .001$). Responses to lower stimuli were faster with the right than with the left hand (441 vs. 422 ms, $t[23]= 2.24$, $p = .035$). But when the avatar had the right hand in upper and the left hand in lower position, responses to upper stimuli were faster with the right than with the left hand (427 vs. 453 ms, $t[23] = 3.70$, $p = .001$) and responses to lower stimuli were faster with the left than with the right hand (456 vs. 428 ms, $t[23] = -4.92$, $p < .001$).

The four-way interaction between the factors avatar's perspective, avatar's hand position, response position and stimulus position was also significant, $F(1, 23) = 6.60$, $p = .017$, $\eta_p^2 = .22$. To further analyze this interaction, we calculated two separate ANOVAs for both visual perspectives of the avatar (rotated vs non-rotated), which both showed significant three-way interactions (non-rotated perspective: $F[1, 23] = 46.50$, $p < .001$, $\eta_p^2 = .67$, rotated perspective: $F[1, 23] = 25.94$, $p < .001$, $\eta_p^2 = .53$). As can be seen from Figure 2, the difference between the rotated and non-rotated avatar was only the quantitative size of the compatibility effect determined by avatar's corresponding hand position. The SR compatibility effects for the rotated avatar was 19 ms on average, whereas for the non-rotated avatar it was even 29 ms.

Percentages of errors. The analysis of PE revealed a significant interaction between avatar's hand position, stimulus position and response position, $F(1, 23) = 9.77$, $p = .005$, $\eta_p^2 = .30$. In trials where the left hand of the avatar was aligned to the upper stimulus position and the right hand to the lower stimulus position, responses to upper stimuli were more accurate with the left hand than with the right hand (1.5 vs. 3.3 % errors, $t[23] = -2.54$, $p = .018$). To lower stimuli right-sided responses were numerically more accurate than left-sided responses (3.0 vs. 1.8 % errors, $t[23] = 2.38$, $p = .238$).

When avatar's right hand was in upper its left hand in lower position, participants responses to upper stimuli were numerically more accurately with the right than with the left hand (1.3 vs. 2.7 % errors, $t[23] = 1.60$, $p = .122$). Responses to lower stimuli were more accurate with the left hand than with the right hand (1.4 vs. 3.7 % errors, $t[23] = -3.57$, $p = .002$).

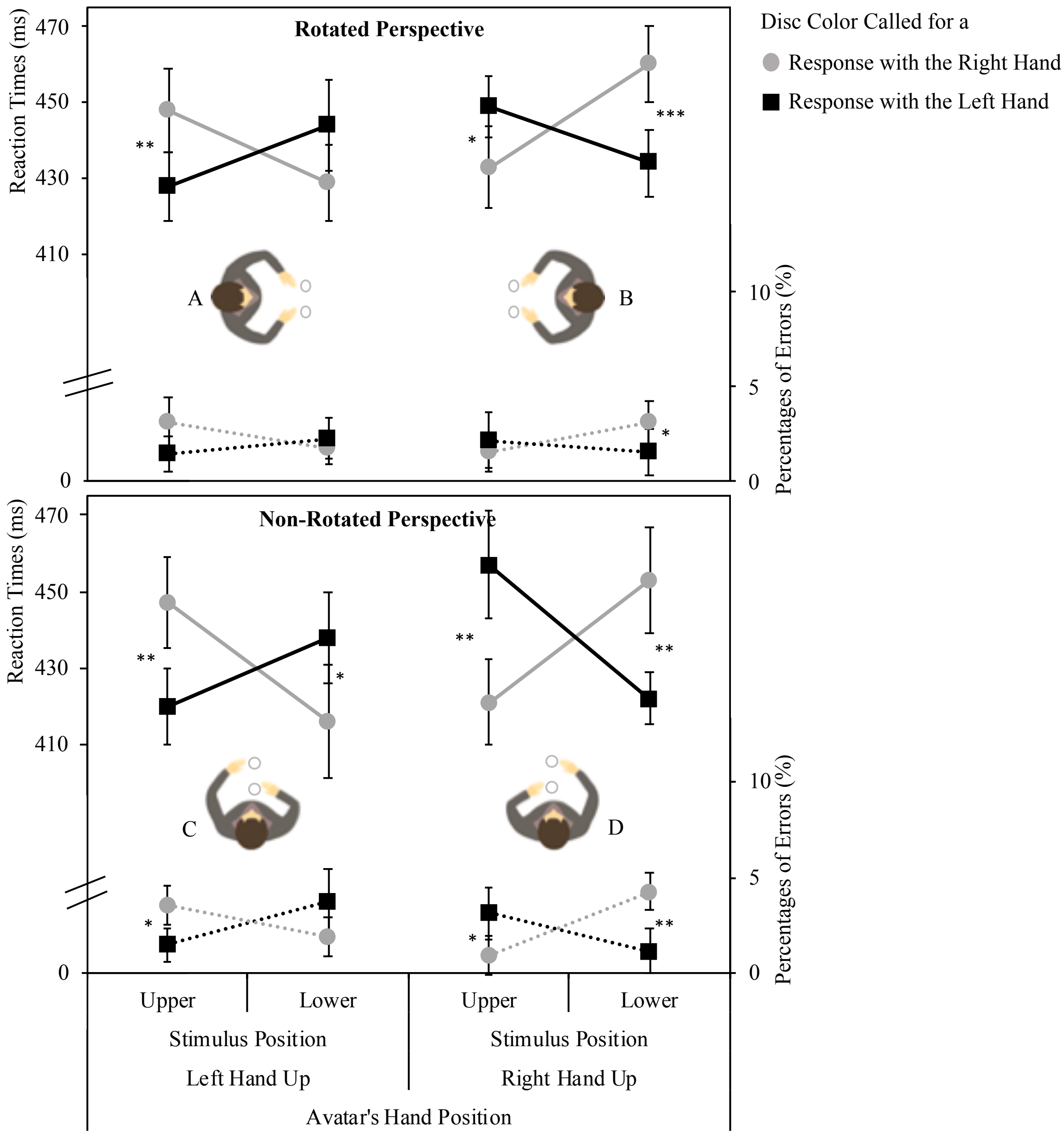


Figure 2. Mean reaction times (solid lines, left axis) and percentage of errors (dotted lines, right axis) as a function of avatar's perspective (upper panel: rotated, lower panel: non-rotated), avatar's hand position, stimulus position and response position from Experiment 1. The corresponding avatars (A-D) are displayed between RT and PE data points. Grey borders next to the avatar's hands show both positions where the disc could appear. Asterisks indicate significant pairwise comparison between right and left responses (t-tests, two-tailed, *: $p < .05$, **: $p < .01$, ***: $p < .001$). Error bars show within-subject 95% confidence intervals from normalized data (Cousineau, 2005).

Discussion

The findings are clear cut. Participants took avatar's position and showed response tendencies that can only be explained from the position of the virtual avatar. Thereby, the experiment successfully replicated the study of Böffel and Müsseler (2019a) according to which SR compatibility in the rotated scenario is determined by the left and right position of the avatar (Fig. 2, upper panel, avatars A and B).

Further, we observed a pronounced avatar-Simon effect also in the non-rotated avatar scenarios (Fig. 2, lower panel, avatars C and D). This effect cannot originate from avatar's visual perspective, but only from the reference codes of avatar's hands. Moreover, SR compatibility effects of the non-rotated avatars were more pronounced compared to the rotated avatars. Maybe the less pronounced compatibility effects in the rotated scenario originated from the need for a mental rotation to take avatar's position. The additional mechanism could have superimposed the compatibility process and thus reduced the effect.

In sum, Experiment 1 shows the central role of the avatar's hand positions in adopting avatar's spatial position. The assignments of upper/lower stimulus positions to left/right response positions were indicated by the avatar's hand positions, which the participants adopted. Experiment 2 should clarify how non-corresponding hand positions between avatar and participant affected the results.

Experiment 2a and 2b

Experiment 2 investigated whether differences in the spatial hand positions of participant and avatar have an influence on the spatial reference coding. The question is whether participants could completely neglect their own proximal hand positions in favor of the distal avatar's hands. We created conditions in which participants' and avatars' hand postures were corresponding or non-corresponding. To achieve this, the response keys were arranged vertically (Figure 3). Now the hand postures of the avatar and the participant can correspond (e.g. both left hands in the upper position and both right hands in the lower position) or not (e.g. avatars left but participants right hand in upper [lower] position and vice versa). As in Experiment 1, we expected pronounced compatibility effects when the hand positions of the avatar and the participant correspond.

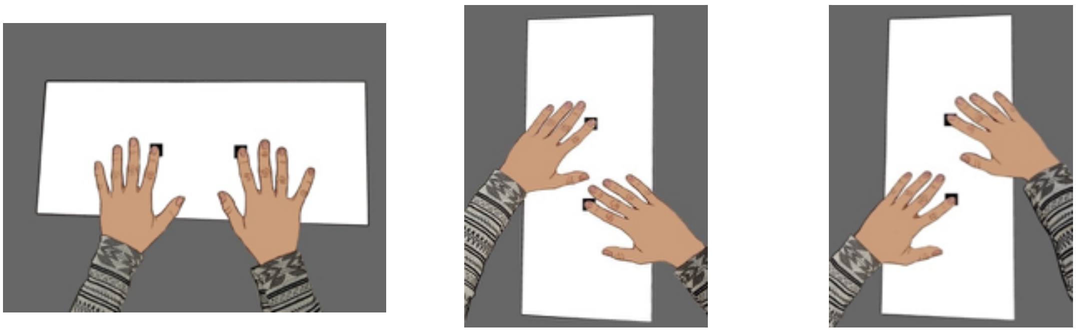
The situation is more complicated if the hand postures of avatar and participant do not match. In this case, participant's proximal left (right) hand in the upper (lower) position controlled avatar's distal left (right) hand in the lower (upper) position. Maybe the left (right) code is weakened by the contradictory proximal and distal coding. Another possibility is that the non-matching hand postures increase the tendency to ignore the avatar. In any case, when the hand postures of the participant and avatar do not match (non-corresponding condition), a reduction or even an elimination of the avatar-Simon effect was predicted.

Two parallel experiments were carried out. In Experiment 2a, the participants had their left hand on the upper response key and their right hand on the lower response key. In Experiment 2b it was the other way around.

Method

Participants. A new sample of 48 students (8 male and 40 female, *M* age of 21.44 years, *SD* 2.67 years) participated in the experiments for course credits. All participants reported normal or corrected to normal vision. Participants were assigned randomly to two equally sized groups who participated in Experiment 2a (4 male and 20 female, *M* age of 22.17 years, *SD* 2.76 years) or Experiment 2b (4 male and 20 female, *M* age of 20.71 years, *SD* 2.42 years).

Procedure. Unlike Experiment 1 we arranged the response board vertically in Experiment 2 (Fig. 3, middle and right). In Experiment 2a participants responded with their left hand in the upper response position and with their right hand in the lower position (Fig. 3, middle). This corresponds to the hand position of avatars A and C and is opposite to the hand position of avatars B and D. In Experiment 2b participants responded with their left hand in lower position and their right hand in upper position (Fig. 3, right) what corresponds with the hand position of avatars B and D. Participants were instructed to respond the light (dark) blue disc with the left (right) hand, again the mapping was balanced between the participants. The experimental program used in Experiment 2a and b was identical to Experiment 1, only the response positions were arranged vertically and varied between Experiment 2a and 2b with regard to the hand positions as illustrated in Figure 3. The lower key was placed 18 cm and the upper key was placed 31 cm from participant's seating position.



Experiment 1 **Experiment 2a** **Experiment 2b**

Figure 3. Response positions of participant's hands in Experiment 1, 2a and 2b.

Design and data treatment. ANOVAs were calculated for Experiment 2a and b separately, including the factors avatar's perspective (rotated or non-rotated), avatar's hand position (left or right hand in upper position), response position of the participant (upper or lower position) and stimulus position (upper or lower position). In each experiment the factors formed a 2x2x2x2 within-subject design with RT and PE as dependent variables. Again, we calculated post-hoc pairwise comparisons for all differences relevant to our hypotheses (*t*-tests, two-tailed). The data were pre-processed the same way as in Experiment 1. In 2.7 % of the trials in Experiment 2a and 2.2 % of trials in Experiment 2b an erroneous response was given. 4.7 % of RTs were additionally excluded from the analysis in Experiment 2a (4.6 % in Experiment 2b) as they failed Tukey's criterion (Tukey, 1977).

Results of Experiment 2a

As in Experiment 2a results showed neither a main effect nor an interaction with the factor avatar's perspective (rotated or non-rotated), RT and PE were collapsed across this factor in Figure 4.

Reaction times. An interaction between response position and stimulus position was observed with $F(1, 23) = 62.00$, $p < .001$, $\eta_p^2 = .73$. Participants responded faster to upper stimuli with the left than with the right hand (424 vs. 453 ms). Responses to lower stimuli were faster with the right than the left hand (443 vs. 427 ms).

The predicted interaction between the factors avatar's hand position, stimulus position and response position was significant with $F(1, 23) = 16.76$, $p < .001$, $\eta_p^2 = .42$. When the hand postures of the participant and the avatar corresponded (left hand in upper and right hand in lower position, Fig. 4 left panels), responses with the left hand to upper stimuli were faster than responses with the right hand (421 vs. 461 ms, $t[23] = - 6.81$, $p < .001$) while responses to lower stimuli were faster with the right hand than with the left hand (448 vs. 420 ms, $t[23] = 4.31$, $p < .001$). For avatars holding their right hand in upper and their left hand in lower position (non-corresponding postures, Fig. 4, right panels) responses to upper stimuli were faster with the left than with the right hand (428 vs. 445 ms, $t[23] = - 2.67$, $p = .014$). For responses to lower stimuli, there was no significant difference between responses with the left and right hand (433 vs. 438 ms, $t[23] = .70$, $p = .490$).

Percentages of errors. The interaction of stimulus position and response position was significant, $F(1, 23) = 24.17$, $p < .001$, $\eta_p^2 = .51$. Responses to upper stimuli were more accurate with the left than with the right hand (1.9 vs. 3.4 % errors) while responses to lower stimuli were more accurate with the right than with the left hand (3.6 vs. 1.9 % errors).

The interaction between avatar's hand position, stimulus position and response position was observed, $F(1, 23) = 8.48$, $p = .008$, $\eta_p^2 = .27$. In conditions where avatar's left hand was aligned with the upper and the right hand with lower position (corresponding hand postures), responses to upper stimuli were more accurate with the left than with the right hand (1.7 vs. 4.0 % errors, $t[23] = -3.78$, $p = .001$). Responses to lower stimuli were more accurate with the right than with the left hand (3.4 vs. 1.4 % errors, $t[23] = 3.98$, $p = .001$). When avatar's right hand was in upper position (non-corresponding hand postures) there was neither a significant difference between responses with the left and right hand for upper stimuli (2.1 vs. 2.92 % errors, $t[23] = -1.34$, $p = .195$) nor for lower stimuli (3.7 vs. 3.0 % errors, $t[23] = 1.57$, $p = .129$).

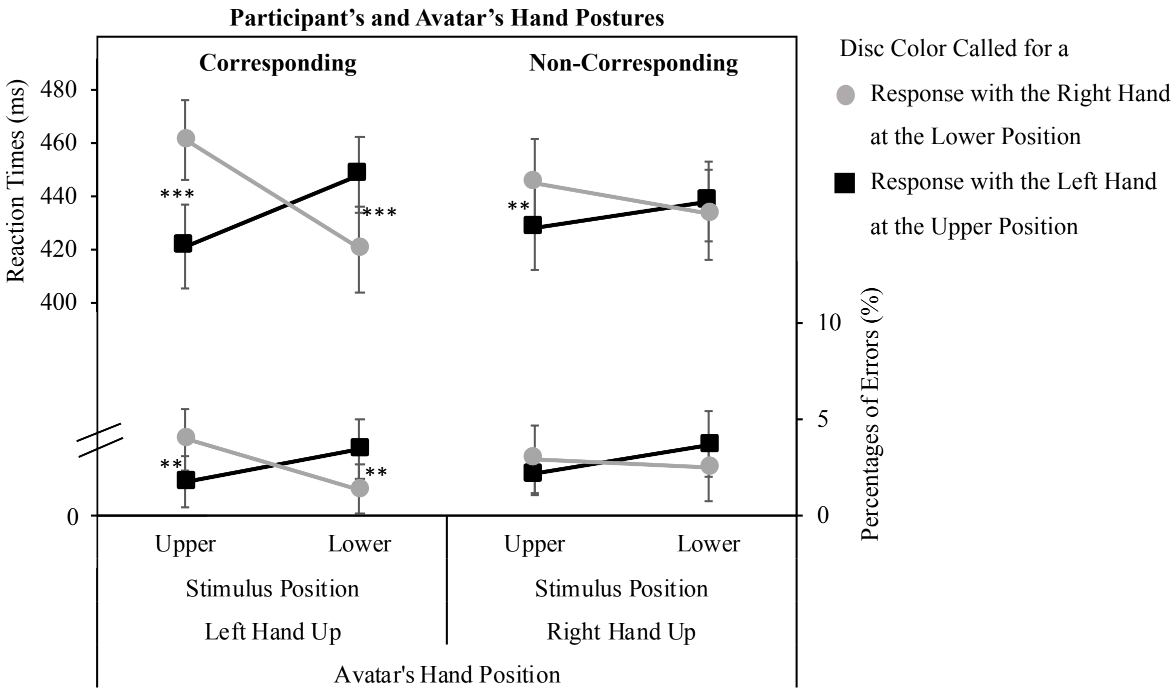


Figure 4. Mean reaction times (solid lines, left axis) and percentages of errors (dotted lines, right axis) as a function of avatar's hand position, stimulus position and response position from Experiment 2a. The hand postures of the participant and the avatar could correspond (corresponding hand postures, left side of the figure) or not correspond (non-corresponding hand postures, right side of the figure). Asterisks indicate significant pairwise comparison between right and left responses (t-tests, two-tailed, *: $p < .05$, **: $p < .01$, ***: $p < .001$). Error bars show within-subject 95% confidence intervals from normalized data (Cousineau, 2005).

Results of Experiment 2b

As in Experiment 2b results showed neither a main effect nor an interaction with the factor avatar's perspective (rotated or non-rotated) RT and PE were collapsed across this factor in Figure 5.

Reaction times. We observed a main effect of response position $F(1, 23) = 18.42$, $p < .001$, $\eta_p^2 = .45$, indicating faster responses with the right than with the left hand (424 vs. 440 ms) and a main effect of stimulus position, $F(1, 23) = 9.20$, $p = .006$, $\eta_p^2 = .29$, with faster responses to upper than to lower stimuli (429 vs. 435 ms). Further, the interaction between response position and stimulus position was significant, $F(1, 23) = 72.46$, $p < .001$, $\eta_p^2 = .76$. Responses to upper stimuli were faster with the right than with the left hand (410 vs. 449 ms) and responses to lower stimuli were faster with the left than with the right hand (438 vs. 432 ms).

Importantly, there was an interaction between avatar's hand position, stimulus position and response position with $F(1, 23) = 19.20$, $p < .001$, $\eta_p^2 = .46$. In trials where avatar's left hand was in upper and the right hand was in lower position (non-corresponding hand postures; Fig. 5, left panel), responses to upper stimuli were faster with the right than with the left hand (429 vs. 438 ms, $t[23] = 2.91$, $p = .008$). There was no significant difference between left and right responses to lower stimuli (442 vs. 430 ms, $t[23] = 1.79$, $p = .087$).

When the right hand of the avatar was aligned with the upper and the left hand with the lower position (corresponding hand postures; Fig. 5, right panel) responses to upper stimuli were faster with the right than with the left hand (400 vs. 459 ms, $t[23] = 8.10$, $p < .001$) and responses to lower stimuli were faster with the left than with the right hand (446 vs. 422 ms, $t[23] = -3.90$, $p = .001$).

Percentages of errors. A main effect of response position was observed with $F(1, 23) = 11.34$, $p = .003$, $\eta_p^2 = .33$. Fewer errors were made in responses with the left hand than with the right hand (1.6 vs. 2.8%). Further, there was an interaction between stimulus position and response position, $F(1, 23) = 32.34$, $p < .001$, $\eta_p^2 = .58$. Responses to upper stimuli were more accurate with the right than with the left hand (2.1 vs. 1.9 %) while responses to lower stimuli were more accurate with the left than with the right hand (1.1 vs. 3.6 %).

There was an interaction between the hand position of the avatar, stimulus position and response position with $F(1, 23) = 14.12$, $p = .001$, $\eta_p^2 = .38$. When avatar had its left hand in upper position and its right hand in lower position, there was no significant difference between responses with the left or right hand for upper stimuli (1.3 vs. 2.7 % errors, $t[23] = -1.85$, $p = .077$) and also not for lower stimuli (2.4 vs. 1.9 % errors, $t[23] = -.90$, $p = .379$). However, when avatar's right hand was in upper and its left hand in lower position, responses to upper stimuli were more accurate with the right than with the left hand (1.1 vs. 2.9 %, $t[23] = 2.74$, $p = .012$) and responses to lower stimuli were more accurate with the left than with the right hand (4.8 vs. 0.2 %, $t[23] = -7.08$, $p < .001$).

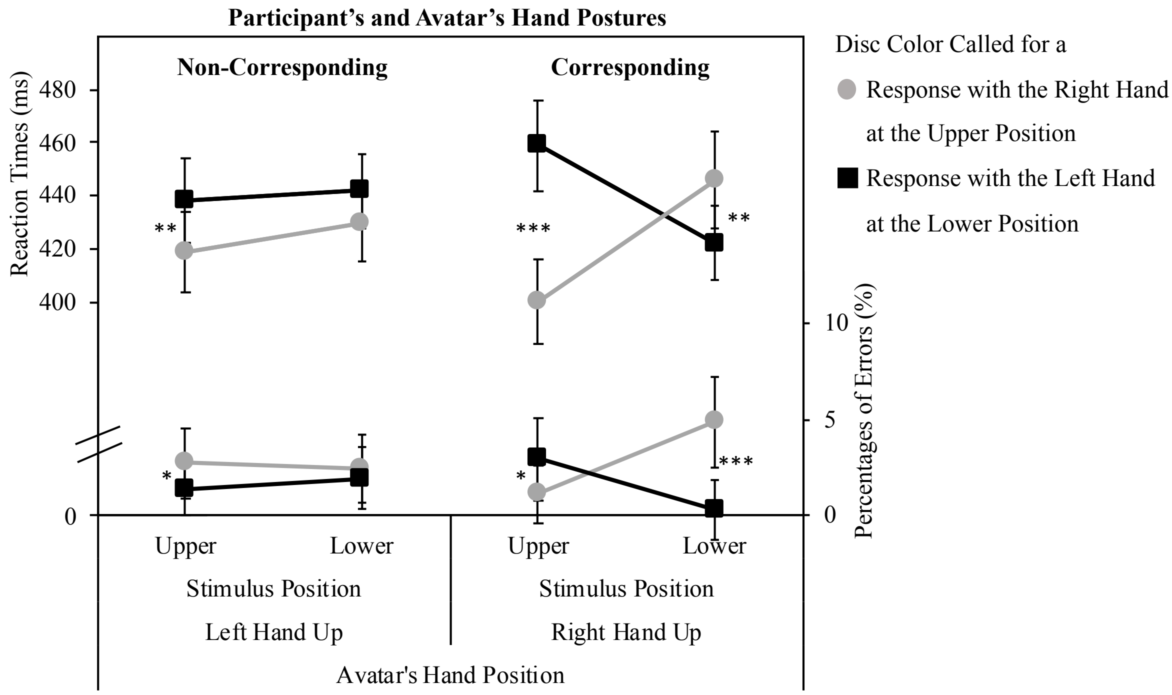


Figure 5. Mean reaction times (solid lines, left axis) and percentages of errors (dotted lines, right axis) as a function of avatar's hand position, stimulus position and response position from Experiment 2a. The hand postures of the participant and the avatar could correspond (corresponding hand postures, left side of the figure) or not correspond (non-corresponding hand postures, right side of the figure). Asterisks indicate significant pairwise comparison between right and left responses (t-tests, two-tailed, *: $p < .05$, **: $p < .01$, ***: $p < .001$). Error bars show within-subject 95% confidence intervals from normalized data (Cousineau, 2005).

Discussion

The experiment showed that the modulation of SR compatibility by avatar's position was affected by the match of the hand positions of participant and avatar. In conditions where the hand positions of the participant and the avatar corresponded, we observed the interaction effect in RTs and PEs (left side of Fig. 4 and right side of Fig. 5) similar to the finding of Experiment 1. Conforming our hypothesis, no interaction effect was found in the non-corresponding condition of Experiment 2b (Fig. 5, left side). In Experiment 2a, we also observed an interaction of stimulus and response position in the non-corresponding condition, which, however, does not reflect the avatar-Simon effect (Fig. 4, right side).

Taken together the results showed that objects were only coded relative to avatar's hand positions when the spatial code of participant's hand position corresponded. Thereby, Experiment 2a and 2b confirmed the results of Experiment 1 that with corresponding hand postures, avatar's hands form a spatial frame of reference to the objects, which is taken by the participant. However, the spatial codes relative to participant's hand positions cannot be ignored, but interfere if they do not correspond with the coding relative to avatar's hands.

General Discussion

Which mechanisms allow participants to take the spatial perspective of others, here operationalized as their virtual avatar? The present study was conducted to determine whether perspective taking or referential coding is capable to elicit spontaneous response tendencies of humans from the (virtual) person's perspective. Our study showed that, regardless of (virtual) person's perspective, stimuli are coded as left and right with regard to the hand positions of the avatar (Experiment 1). Further, we showed that the spatial reference frame of the avatar is neglected when it is contrary to the participant's reference frame. The avatar's spatial perspective was only adopted if the avatar's hand posture matched with the participant's hand posture. Otherwise the perspective did not seem to be adopted (Experiment 2).

Experiment 1 showed that participants linked avatar's position to their own behavioral tendencies which is reflected in a reversal of the SR compatibility depending on avatar's position. Results revealed that for the rotated avatar scenario SR compatibility relations emerged that can only be expected from avatar's spatial position (the avatar-Simon effect). The modulation of SR compatibility by the presence of (virtual) persons has been proposed to reflect a visual perspective taking mechanism (Freundlieb et al., 2016, 2017; Müsseler et al., 2018; Böffel & Müsseler, 2018, 2019a, b). In our study, the perspective taking account predicts that the visual field of the rotated left or right avatar shapes the frame of reference towards the disks. Contrary to this assumption we demonstrated that the interaction pattern indicative for spatial coding relative to avatar's position was present and even more pronounced in conditions in which the unrotated avatar had the same visual field towards the discs, but took different hand positions. The participants coded the upper (lower) stimulus as 'right' and the lower (upper) stimulus as 'left' if avatar's hand determined so.

Our interpretation is that this finding reflects a referential coding mechanism according to which spatial relations among objects or events in the visual field are formed in reference to any other object or event in the visual field (cf. van der Heijden, Müsseler & Bridgeman, 1999). In the current study, both the participant and the avatar created potential spatial references to the stimuli what could cause a selection problem. According to the referential coding account, the salience of the potential referent is decisive whether it is used by the actor as a spatial reference. As participants were instructed throughout the study to take avatar's position it is plausible that in Experiment 1 spatial codes where formed relative to avatar's hand positions and not to their own hand positions.

In our study, the avatar's hand positions provided the reference frame for coding the stimulus positions as left and right. But is there anything unique about the hands compared to other referents? According to the referential coding account any event can potentially produce a reference frame to another object. Referential coding is thereby proposed as a general and not exclusively social mechanism and no qualitative difference is made whether the own person or another living or non-living object acts as a spatial reference point to a stimulus. In accordance, studies of Hommel and Lippa (1995), Proctor and Pick (1999), and Pick, Specker, Vu, and Proctor (2014) showed that stimuli were coded as left and right relative to the eye positions of a 90° rotated face stimulus. Pick et al. (2014) replicated the spatial coding effects using the two headlights of an automobile and provided evidence for the assumption that spatial coding relative to objects is not dependent on the aliveness or human appearance of the reference object. In conclusion, although in our experiments the hand positions have proved to be determinant, we do not exclude other reference points than hands[3].

Experiment 2 showed that even though participants transferred their action tendencies to the distal avatar, their own proximal hand positions have not been completely neglected (for related results in tool use see Ladwig, Sutter & Müsseler, 2012; Sutter, Sülzenbrück, Rieger, & Müsseler, 2013). Experiment 2a and 2b showed that position taking towards an avatar depended on the match between the hand postures of the avatar and the participant. We observed a modulation of the SR compatibility effect indicative for the spatial coding of avatar's position only in conditions where the hand postures of the avatar and the participant corresponded. But no evidence of a position taking towards the avatar was observed in conditions in which the hand postures of the participant and the avatar did not match. We have predicted the latter result based on the assumption that the left/right code is weakened by the opposing coding and/or by the tendency to ignore the avatar under these conditions. The opposing coding might have created a conflict. To overcome this conflict, increasing cognitive control and more controlled actions are necessary. With regard to dual-route accounts, cognitive control is reflected by the indirect mapping route which is driven by the current task requirements. If the task difficulty increases, the controlled processes and thus the indirect route is strengthened, which is at the expense of the automatic route (for a comparable interpretation see also Proctor, Yamaguchi, Dutt, & Gonzalez, 2013). In any case, our findings indicate that the spatial perspective of the avatar is only adopted automatically if it is not conflicting with the own spatial reference frame.

The present study compared mechanisms of visual perspective taking with referential coding, when controlling an avatar. At first glance, our results seem to contradict the view that when taking avatar's perspective, a representation is created that enables the observer to see the world through the eyes of the avatar. Instead, the findings seem to favor a referential (re-)coding of the existing spatial relationships, here with regard to the avatar's hands. However, both approaches might not be mutually exclusive. Consider the case where referential coding is seen as a mechanism that enables perspective taking. How could it be otherwise? A viewer perceives the surrounding visual space only through the information provided by her/his eyes. Any conclusion about the perceived space is based on this information. In this sense, referential coding may be at the base of any perspective taking that occurs.

Conclusion

In conclusion our study shows that people are able to link avatar's position to their own behavioral tendencies demonstrated in an avatar-Simon task. Further, our study gives an insight into the mechanisms that make position taking possible: It is not the (re-)constructed left and right visual field of the avatar, but her/his left and right hand positions that formed the spatial frame of reference for imperative stimuli. We were also able to demonstrate a limit on the position taking: The spatial reference frame shaped by avatar's hand posture was only selected by the participant when it corresponded with participant's hand posture. Otherwise the spatial referencing by avatar's hands expires.

References

Bauer, D. W., & Miller, J. (1982). Stimulus–response compatibility and the motor system. *Quarterly Journal of Experimental Psychology, 34A,* 367–380. https://doi.org/10.1080/14640748208400849

Böffel, C. & Müsseler, J. (2018). Perceived ownership of avatars influences visual perspective taking. *Frontiers in Psychology, 9*, 1-9. doi: 10.3389/fpsyg.2018.00743

Böffel, C., & Müsseler, J. (2019a). Visual perspective taking for avatars in a Simon task. *Attention, Perception, & Psychophysics, 81(1)*, 158-172. doi:10.3758/s13414-018-1573-0

Böffel, C. & Müsseler, J. (2019b). Action effect consistency and body ownership in the avatar-Simon task. *PLoS ONE, 14(8),* e0220817.

Cavallo, A., Ansuini, C., Capozzi, F., Tversky, B., & Becchio, C. (2017). When far becomes near: Perspective taking induces social remapping of spatial relations. *Psychological Science, 28,* 69-79. doi:10.1177/0956797616672464

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.

Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorials in Quantitative Methods for Psychology, 1*(1), 42-45.

Dolk, T., Hommel, B., Prinz, W., & Liepelt, R. (2013). The (not so) social Simon effect: a referential coding account. *Journal of Experimental Psychology: Human Perception and Performance, 39,* 1248–1260. doi: 10.1037/a0031031

Dolk, T., Hommel, B., Colzato, L. S., Schütz-Bosbach, S., Prinz, W., & Liepelt, R. (2014). The joint Simon effect: A review and theoretical integration. *Frontiers in Psychology, 5,* 974.

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,* 99-103. doi:10.1037/0012-1649.17.1.99

Franz, E. A., Sebastian, A., Hust, C., & Norris, T. (2008). Viewer perspective affects central bottleneck requirements in spatial translation tasks. *Journal of Experimental Psychology: Human Perception and Performance, 34,* 398-412. doi:10.1037/0096-1523.34.2.398

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

Freundlieb, M., Sebanz, N., & Kovács, Á. M. (2017). Out of your sight, out of my mind: Knowledge about another person's visual access modulates spontaneous visuospatial perspective-taking. *Journal of Experimental Psychology: Human Perception and Performance, 43,* 1065–1072. doi:10.1037/ xhp0000379

Furlanetto, T., Becchio, C., Samson, D., & Apperly, I. (2016). Altercentric interference in Level 1 visual perspective taking reflects the ascription of mental states, not submentalizing. *Journal of Experimental Psychology: Human Perception and Performance, 42,* 158-163.

Heyes, C. (2014). Submentalizing: I am not really reading your mind. *Perspectives on Psychological Science, 9,* 131–143. http://dx.doi.org/10 .1177/1745691613518076

Hommel, B. (1993). The role of attention for the Simon effect. *Psychological Research*, *55*, 208-222.

Hommel, B. (1997). Toward an action-concept model of stimulus-response compatibility. *Advances in Psychology, 118*, 281-320.

Hommel, B., & Lippa, Y. (1995). SR compatibility effects due to context-dependent spatial stimulus coding. *Psychonomic Bulletin & Review, 2*(3), 370-374.

Hommel, B. (2011). The Simon effect as tool and heuristic. *Acta Psychologica, 136*, 189-202.

Janczyk, M. (2013). Level 2 perspective taking entails two processes: Evidence from PRP experiments. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39,* 1878-1887. doi:10.1037/a0033336

Kirk, R. E. (1995). *Experimental Design* (3rd ed.). Pacific Grove, CA: Brooks/Cole.

Kornblum, S., Hasbroucq, T., & Osman, A. (1990). Dimensional overlap: Cognitive basis for stimulus-response compatibility. A model and taxonomy. *Psychological Review, 97,* 253–270.

Ladwig, S., Sutter, C., & Müsseler, J. (2012). Crosstalk between proximal and distal action effects during tool use. *Zeitschrift für Psychologie / Journal of Psychology, 220*, 10-15. doi: 10.1027/2151-2604/a000085

Lu C.-H., & Proctor R. W. (1995). The influence of irrelevant location information on performance: A review of the Simon and spatial Stroop effects. *Psychonomic Bulletin & Review, 2*(2), 174-207. doi:10.3758/BF03210959

Müsseler, J., Ruhland, L., & Böffel, C. (2018). Reversed effect of spatial compatibility when taking avatar's perspective. *Quarterly Journal of Experimental Psychology*. Advance online publication. doi:10.1177/1747021818799240

Pick, D. F., Specker, S., Vu, K. P. L., & Proctor, R. W. (2014). Effects of face and inanimate-object contexts on stimulus–response compatibility. *Psychonomic Bulletin & Review, 21(2)*, 376-383.

Proctor, R. W., & Pick, D. F. (1999). Deconstructing Marilyn: Robust effects of face contexts on stimulus—response compatibility. *Memory & Cognition, 27*(6), 986-995.

Proctor, R. W., & Vu, K.-P. L. (2006). *Stimulus-response compatibility principles: Data, theory, and application*. Boca Raton, FL: CRC Press.

Proctor, R. W., Yamaguchi, M., Dutt, V., & Gonzalez, C. (2013). Dissociation of S-R compatibility and Simon effects with mixed tasks and mappings. *Journal of Experimental Psychology: Human Perception and Performance, 39*(2), 593-609.

Quesque, F., Chabanat, E., & Rossetti, Y. (2018). Taking the point of view of the blind: spontaneous level-2 perspective-taking in irrelevant conditions. *Journal of Experimental Social Psychology, 79,* 356-364.

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.

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.

Simon, J. R., & Small, A. M., Jr. (1969). Processing auditory information: Interference from an irrelevant cue. *Journal of Applied Psychology*, *53*(5), 433-435. doi:10.1037/h0028034

Stoet, G. (2017). Sex differences in the Simon task help to interpret sex differences in selective attention. Psychological Research, 81(3), 571-581.

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

Sutter, C., Sülzenbrück, S., Rieger, M., & Müsseler, J. (2013). Limitations of distal effect anticipation when using tools. *New Ideas in Psychology, 31*, 247–257. doi: 10.1016/j.newideapsych.2012.12.001

Tarampi, M. R., Heydari, N., & Hegarty, M. (2016). A tale of two types of perspective taking: Sex differences in spatial ability. *Psychological Science, 27*(11), 1507-1516.

Tukey, J. W. (1977). *Exploratory Data Analysis*. Reading, MA: Addison-Wesley.

Tversky, B., & Hard, B. M. (2009). Embodied and disembodied cognition: Spatial perspective-taking. *Cognition, 110,* 124–129. doi: 10.1016/j.cognition.2008.10.008

Van der Heijden, A.H.C., Müsseler, J., & Bridgeman, B. (1999). On the perception of position. *Advances in Psychology, 129*, 19-37. doi: 10.1016/S0166-4115(99)80005-3

Footnotes

1. Apart from the presentation of the avatar, the setting of Böffel und Müsseler (2019a) was used to explore the orthogonal Simon effect, reflecting performance advantages for left-sided responses to lower stimuli and right-sided responses to upper stimuli (Bauer & Miller, 1982). The result pattern in the condition in which the avatar is in the right position corresponds to the orthogonal Simon effect. However, please note that the positioning of the avatar on the left side leads to an exact reversal of the orthogonal Simon effect, which is clear evidence for the modulation of SR compatibility by the avatar (see Böffel & Müsseler, 2019a, for a more detailed discussion).

2. Both the experiments included mostly females. As gender differences are reported for a large range of spatial abilities (e.g. Tarampi, Heydari, & Hegarty, 2016) this could be a limitation of our study. However, in the case of spatial compatibility research, the differences between the genders are usually comparably small and there are pronounced SR compatibility effects for both women and men (e.g. Stoet, 2017).

3. We are grateful to Robert Proctor for the valuable reference to existing studies on referential coding relative to faces and inanimate objects.