



## Object affordances from the perspective of an avatar

Matthias G. Arend<sup>a,\*</sup>, Jochen Müsseler<sup>b</sup><sup>a</sup> Institute of Industrial Engineering and Ergonomics, RWTH Aachen University, Aachen, Germany<sup>b</sup> Work and Cognitive Psychology, RWTH Aachen University, Germany

## ARTICLE INFO

## Keywords:

Avatar  
Perspective taking  
Object affordances  
Spatial alignment effect  
Object-based Simon effect

## ABSTRACT

Humans often interact with avatars in video gaming, workplace, or health applications, for instance. The present research studied object affordances from an avatar's perspective. In two experiments, participants responded to objects with a left/right keypress, indicating whether the objects were upright or inverted. Task-irrelevant objects' handles were aligned with either the left or right hand of the actor and/or avatar. We hypothesized that actors respond faster when the handles are aligned, as compared to non-aligned, with the respective avatar hand (spatial alignment effect or object-based Simon effect). In Experiment 1, the spatial alignment effect was increased through the presentation of avatar hands as compared to when no hands were presented. In Experiment 2, the avatar perspective was rotated by 90° to the right and left of the actor's view. Here, the spatial alignment effect was guided by the avatar, suggesting that the actors took its perspective when perceiving objects' affordances.

Avatars, virtual representations of a user, operate in contexts such as video gaming, workplace activities (e.g., videoconferencing; Bessiere, Ellis, & Kellog, 2009; Inkpen & Sedlins, 2011) and augmented reality (e.g., Wang, Sato, Otsuki, Kuzuoka, & Suzuki, 2019). A classic feature of avatars is that—in contrast to virtual agents—they are typically guided by a human actor and do not operate on their own (Balcisoy, Torre, Ponder, Fua, & Thalmann, 2000). Hence, an actor completing (virtual) tasks through an avatar characterizes the prototypic user-avatar interaction. A feature of some of these interactions is that the actor's perspective differs from the avatar's perspective (such as third vs. first person view; e.g., Petkova, Khoshnevis, & Ehrsson, 2011). This can go as far as to both perspectives offering conflicting spatial information (e.g., Böffel & Müsseler, 2018; Müsseler, Ruhland, & Böffel, 2019). In these conflicting situations, successful task completion depends on whether actors take their own or the avatar's perspective (von Salm-Hoogstraeten & Müsseler, 2021). The goal of the present research was to contribute to the understanding of how actors integrate their own and the conflicting avatar's perspective for completing a task. For this purpose, a spatial compatibility paradigm based on the affordances of real-world objects was applied.

The notion of affordance arose under the assumption that vision and action are coupled, such that visual information is perceived through the lens of a potential action. According to Gibson (1977), affordances are features of the environment with relevance for (planned) actions of individuals. Tucker and Ellis (1998) give the example of a saucepan, whose affordance particularly lies in its graspability (see also Lopresti-Goodman, Turvey, & Frank, 2011). While the initial evidence for object affordances mainly stems from behavioral studies (Couth, Gowen, & Poliakoff, 2014; Symes, Ellis, & Tucker, 2005; Tucker & Ellis, 1998; 2001; Poliakoff, Galpin, Dick, Moore, & Tipper, 2007), there has also been evidence from neuropsychological studies indicating that the mere perception of objects

\* Corresponding author at: Institute of Industrial Engineering and Ergonomics, RWTH Aachen University, Eilfschornsteinstr. 18, 52062 Aachen, Germany.

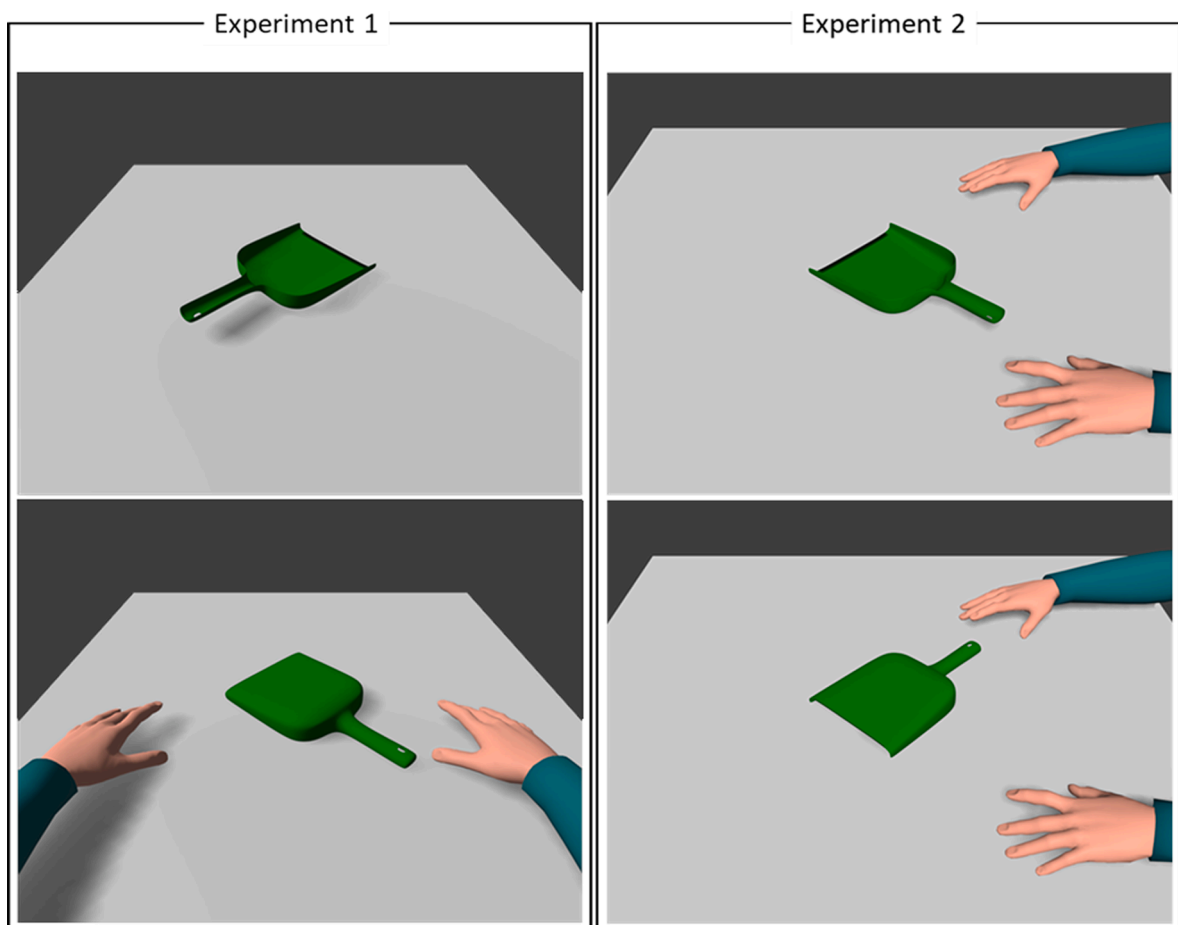
E-mail address: [m.arend@iaw.rwth-aachen.de](mailto:m.arend@iaw.rwth-aachen.de) (M.G. Arend).

already activates motor representations of the actions typically executed with this object (e.g., Grezes, Tucker, Armony, Ellis, & Passingham, 2003; Buccino, Sato, Cattaneo, Roda, & Riggio, 2009).

While Gibson's (1977) original approach of affordances was rather radical in rejecting any mental transformations and cognitive operations, other researchers by no means denied cognitive representations and processes underlying affordance-based compatibility effects (e.g., Ellis & Tucker, 2000; Riggio et al., 2008; Tipper, Paul, & Hayes, 2006; Tucker & Ellis, 1998). Still other researchers avoided the term affordance in this context and instead preferred the more neutral term object-based compatibility effects (e.g., Cho & Proctor, 2011, 2013). Although Cho and Proctor (2010) coined the advantages of this term, we stick to the notion of affordance in the present paper in appreciation of Gibson's work, who, to our knowledge, was the first to stress the graspable visual properties of real objects for action planning. The underlying cognitive processes of affordances when an actor-controlled avatar handles graspable objects will be elaborated in more detail in the General Discussion.

The present research used the seminal paradigm of Tucker and Ellis (1998) to study graspability affordances. In this study, objects with handles were presented to the participants, who had to indicate with a right- or left-hand keypress whether an object was shown upright or inverted. Although the orientation of the handle itself was task-irrelevant, the results showed that the participants responded faster and were less error-prone when the handle was aligned with their response hand than when it was aligned with their other hand (see also Symes et al., 2005). This effect is also referred to as spatial alignment effect (Costantini, Committeri, & Sinigaglia, 2011) or object-based Simon effect (e.g., Cho & Proctor, 2013).

That is, because in spatial aspects, the findings of Tucker and Ellis (1998) and others are comparable to the Simon effect. In a typical Simon task, participants respond with the left-hand key to one color and with the right-hand key to another color of a stimulus that is presented on the left or right side of a screen. Although the stimulus position is task-irrelevant, spatially corresponding conditions (e.g.,



**Fig. 1.** The setting of the experiments. In Experiment 1 (left panel), an upright or inverted (task-relevant feature) household object (here a dustpan) was presented with its handle on the left or right (task-irrelevant feature). This scene was presented with and without the avatar's hands. In Experiment 2 (right panel), the avatar was presented either to the left or right of the participant (here to the right). Consequently, the perspective of the avatar and that of the actor elicited different response tendencies under the assumption of the spatial alignment effect in some conditions. For instance, in the right panel, the handle is oriented to the right from the perspective of the actor and therefore should facilitate a right response. However, when the actor takes the perspective of the avatar, the handle is oriented to the left in the top-right picture (to the right in the bottom-right picture) and object affordance should thus facilitate a left (right) response.

stimulus left, response left) produce faster responses and fewer errors than spatially non-corresponding conditions (e.g., left stimulus, right response; Simon & Berbaum, 1990; for reviews see Lu & Proctor, 1995; Hommel, 2011). Recent studies in our lab demonstrated that the Simon effect can also be observed when an avatar is added to the scene (e.g., Böffel & Müsseler, 2019b; von Salm-Hoogstraeten, Bolzius, & Müsseler, 2020). By rotating the stimulus positions and the avatar by  $\pm 90^\circ$  from the actor's view, the stimulus does not contain spatial information on the horizontal (left-right) dimension from the actor perspective, but only from the avatar perspective. The results of the experiments indicated that actors take the avatar's perspective since they reacted in accordance with the Simon effect from the avatar's perspective (avatar-Simon effect; Böffel & Müsseler, 2019a, 2019b, 2020b; von Salm-Hoogstraeten et al., 2020).

Based on these observations, the idea of the present study was to use the spatial alignment paradigm to study avatar perspective taking by object affordances. There has been initial research considering the role of avatars for object affordances: One central finding of the spatial alignment effect—and most different to the Simon effect—is that it diminishes when the objects are presented outside the actor's reaching space (Costantini, Ambrosini, Tieri, Sinigaglia, & Committeri, 2010; Costantini, Ambrosini, et al., 2011). Yet, when an avatar could reach the objects that were outside the actor's reaching space, the spatial alignment effect reappeared (Costantini, Committeri, et al., 2011). This supports the notion that the presence of an avatar impacts the perception of object affordances. However, the question remains whether actors also take the perspective of an avatar when perceiving object affordances, even when it conflicts with their own perspective. Alternatively asked: Whether the spatial alignment effect can also be found from the avatar's perspective when it conflicts with the actor's perspective.

## 1. Present research and hypotheses

The objective of the present research was to test whether actors take an avatar's perspective when perceiving graspable objects by applying a spatial compatibility paradigm with real-world objects as introduced by Tucker and Ellis (1998). Therefore, two experiments were conducted in which participants had to respond to upright or inverted household objects placed in front of an avatar by pressing a left or right response key with their left or right index finger. Hence, the task-relevant feature of the stimuli (object upright or inverted) was varied independently of the task-irrelevant feature, the handle orientation, that captures the impact of object affordances. Experiment 1 was based on the assumption that presenting an avatar with the same perspective on the object as the actor increases the spatial alignment effect compared to when no avatar is presented. In Experiment 2, we examined whether the spatial alignment effect goes along with the avatar perspective even if it conflicts with the actor's perspective (see Fig. 1).

To examine the general assumption that actors take the avatar's perspective when perceiving object affordances and, thus, respond faster to handles that are aligned with the avatar hand that corresponded to their response hand, a set of hypotheses were derived. For the first experiment, participants were presented with objects whose handles were rotated for  $\pm 45^\circ$  from their body midline. Thus, the handles were aligned either with their left or right hand. Because this corresponds to the paradigm of the general spatial alignment effect, we hypothesized that this effect will be shown regardless of whether the avatar hands were additionally presented or not (H1). Second, it was hypothesized that this spatial alignment effect would be increased when the avatar hands (and an action effect) were additionally presented (H2). That is, compared to the condition without the presence of an avatar, the difference between spatially aligned and spatially non-aligned handles and response sides should be bigger when an avatar was present.

With the second experiment, the spatial alignment effect was studied for avatar perspectives rotated for  $90^\circ$  towards the left and right from the participant's body midline. The handle itself was rotated for  $\pm 45^\circ$  and  $\pm 135^\circ$  from the left and right of the participant's body midline, pointing towards the avatar's hands. In this way, when the handle was rotated for  $\pm 45^\circ$ , it was oriented towards the avatar's right (for an avatar rotated towards the actor's left;  $-45^\circ$ ) and the participant's left hand or towards the avatar's left (for an avatar rotated towards the actor's right;  $+45^\circ$ ) and the participant's right hand. For the conditions where the handle was rotated for  $\pm 135^\circ$ , the avatar and participant's hands both corresponded with the response hand. Hence, when participants perceive object affordances from their own perspective, they would show a classic stimulus alignment effect for both, handle orientations towards non-corresponding ( $\pm 45^\circ$ ) and corresponding ( $\pm 135^\circ$ ) avatar and participant hands. Yet, if participants took the avatar's perspective, they would respond with the pattern of an inversed spatial alignment effect for the non-corresponding ( $\pm 45^\circ$ ) conditions: Responses would be faster when the handle was misaligned with their hands, because of the alignment with the avatar hands, whose perspective they took. Hence, the third hypothesis assumed that participants respond to object affordances from the avatar's perspective—even if it was misaligned with their own (H3)<sup>1</sup>.

As for the original paradigm (Tucker & Ellis, 1998), a sample of participants responded to a sample of household objects in both experiments, with both samples having been drawn from their populations of human individuals and household objects with handles. Therefore, systematic between-participant and between-object differences would be expected which have to be accounted for in the statistical analysis. Cross-classified models constitute an approach to model data structures in which each case is nested within one participant and one object (Judd, Westfall, & Kenny, 2017), while the effects of experimental factors are estimated similar to an ANOVA approach. Within this multilevel modelling approach, the clusters participant and object constitute the random part and the effects of experimental variables the fixed part of the model. Additionally, between-cluster differences in the fixed effects of

<sup>1</sup> Although we hypothesized that all the effects would occur in the reaction times (in accordance with most research applying the object-based Simon effect paradigm; Azaad, Laham, & Shields, 2019), we also wanted to investigate the error rates in order to account for potential speed-accuracy tradeoffs. Hence, each hypothesis would receive support, if the corresponding effect was found in the reaction times or error rates. To address this multiple testing problem, the significance level  $\alpha$  was Bonferroni-adjusted for the statistical analysis of each experiment.

experimental variables can be modelled as random slopes. The present research seeks to quantify the between-participant variability in the fixed effects which indicate how the spatial alignment effect is influenced by the degree of perspective taking.

## 2. Experiment 1

The objective of Experiment 1 was to establish the spatial alignment effect in a 3-dimensional scene in which an avatar was present or not. Therefore, ten household objects were presented inverted or upright (task-relevant feature) on a table, with their handles pointing to the left or right (task-irrelevant feature, Fig. 1, top left). Participants responded to the task-relevant feature with a left or right keypress. The hands of an avatar were added to this scene (Fig. 1, bottom left), with the index fingers imitating the finger movements of the participant when pressing the respective response key. Our hypotheses were (H1) that the spatial alignment effect is found for conditions without and with avatar hands, and (H2) that it is stronger for conditions in which avatar hands are present.

### 2.1. Method

#### 2.1.1. Power analysis

Prior to both experiments, a power analysis was conducted to determine the appropriate sample size for the random factor participant. Since the fixed and random parts of the cross-classified model, on which the power analysis had to be based, was very complex, no formulae existed to derive the statistical power (Judd et al., 2017 provide formulae for comparing two-conditions but no interaction effects in the fixed part). Thus, we relied on Monte Carlo simulation for the power analysis (cf. Arend & Schäfer, 2019; Claus, Arend, Burk, Kiefer, & Wiese, 2020). Based on the considerations by Judd et al. (2017), the population model for the random factors participant and object was set up (i.e., proportions of variance were estimated for each random factor). For deriving estimates for the fixed part of the population model (which is set up similar to an ANOVA), information provided in the seminal study by Tucker and Ellis (1998) and in the study by Costantini, Committeri, et al. (2011) was used. By the end of this process, a population model that included estimates for the hypothesized fixed effects as well as variances for the random factors was set up, from which repeated samples ( $N = 10,000$ ) would be drawn as part of the Monte Carlo procedure (cf. Arend & Schäfer, 2019). Appendix A and D describe the complete process of how the power analysis was conducted.

We were seeking to study ten objects together with at least 24 participants, since this number of participants has successfully been used in recent studies on avatar perspective taking (Böffel & Müsseler, 2019a, 2020a; von Salm-Hoogstraeten, Bolzius, & Müsseler, 2020). In accordance with the hypotheses, the power was estimated for reaction times. Furthermore, for each of eight experimental conditions, each trial should be repeated four times per combination of participant and object. The first step in the power analysis was to determine whether this participant sample size would yield sufficient statistical power (i.e., power  $> 0.80$ , see Cohen, 1988) for testing the two hypotheses in Experiment 1 based on the population model derived. For the single statistical tests associated with both hypotheses, power was sufficient with 24 participants, 10 objects, and 4 repetitions per object (1.00 for H1; 0.82 for H2) at the significance level  $\alpha = 0.05$ . Hence, 24 participants were sampled in the first experiment. Appendix D provides the full code.

The Bonferroni correction later applied during the statistical analysis reduced this power (due to the corrected significance level  $\alpha = 0.025$  for single tests) to 0.73 for H2 (the power for H1 remained at 1.00).

#### 2.1.2. Participants

Twenty-four participants (21 female and three male), all psychology students at RWTH Aachen University, were recruited via an E-Mail distribution list of the Institute of Psychology. They received course credit for their participation. Participants were between 18 and 29 years old ( $M = 20.67$ ;  $SD = 3.21$ ) and reported normal or corrected to normal vision.

#### 2.1.3. Ethics approval

Before the start of each experiment, all participants provided written and informed consent. Neither physical nor psychological stress by participating in the present study was anticipated and no data were obtained to identify specific properties of individual participants, but only to examine general cognitive processes. Furthermore, there was no deception (through a cover-story or comparable) and no direct risk of physical injury through the participation in the study. Hence, the Ethics Committee of the Faculty of Philosophy at RWTH Aachen University has not identified any ethical concerns with the study (ref. 2020\_002\_FB7\_RWTH Aachen University).

#### 2.1.4. Apparatus and stimuli

The experiment was conducted in a dimly lit laboratory with the following technological equipment: An Apple Macintosh computer, controlled by the software MATLAB (MathWorks) and the Psychtoolbox-3-Extension (Brainard, 1997; Pelli, 1997). The stimuli were presented on a 22" CRT monitor with a resolution of  $1024 \times 768$  pixels at a refresh-rate of 100 Hz. The participants were seated in front of the monitor, with approximately 60 cm distance. They put their index fingers on separate left and right response keys (distance in between 13 cm) having approximately 50 cm distance to the monitor. Participants' hands were not covered.

The avatar was created with the open-source tool MakeHuman v1.1.1 (<http://www.makehumancommunity.org/>) and then imported into the 3D-modelling software Blender v2.8 (Blender Foundation, 2018; <http://www.blender.org>). Only the avatar hands were visible in the experiment (as depicted in Fig. 1). The avatar's posture was designed to be similar to the actor's posture, with its hands placed to the right and left with similar distance from its body midline. The three-dimensional scene was illuminated by one light from the left of the avatar and participant perspective.

To create a realistic relation between actor and avatar, the avatar's index fingers mimicked the actor's finger movement when pressing the response key (action effect). Before the participant's response, the avatar's left and right index finger were lifted. When the participant pressed a key, the corresponding index finger of the avatar was lowered to the level of the table surface. It was lifted again when the participant released the key. This was implemented through the presentation of two pictures: On the first picture, the avatar had its fingers lifted. When the participant pressed a key, the second picture was presented with the next vertical retrace of the monitor. For a right-hand (left-hand) response, the picture showed the avatar's lowered right (left) index finger. When the participant lifted the index finger again, the initial picture was displayed again with the next vertical retrace of the monitor.

Objects were chosen that had a clearly identifiable handle which could be seen from different angles. All objects utilized in this study belonged to the household category: can, coffee dripper, cup, dustbrush, dustpan, iron, kettle, frying pan, pot, sieve (see Fig. 2). All objects were created from scratch with Blender (Blender Foundation, 2018).

### 2.1.5. Procedure

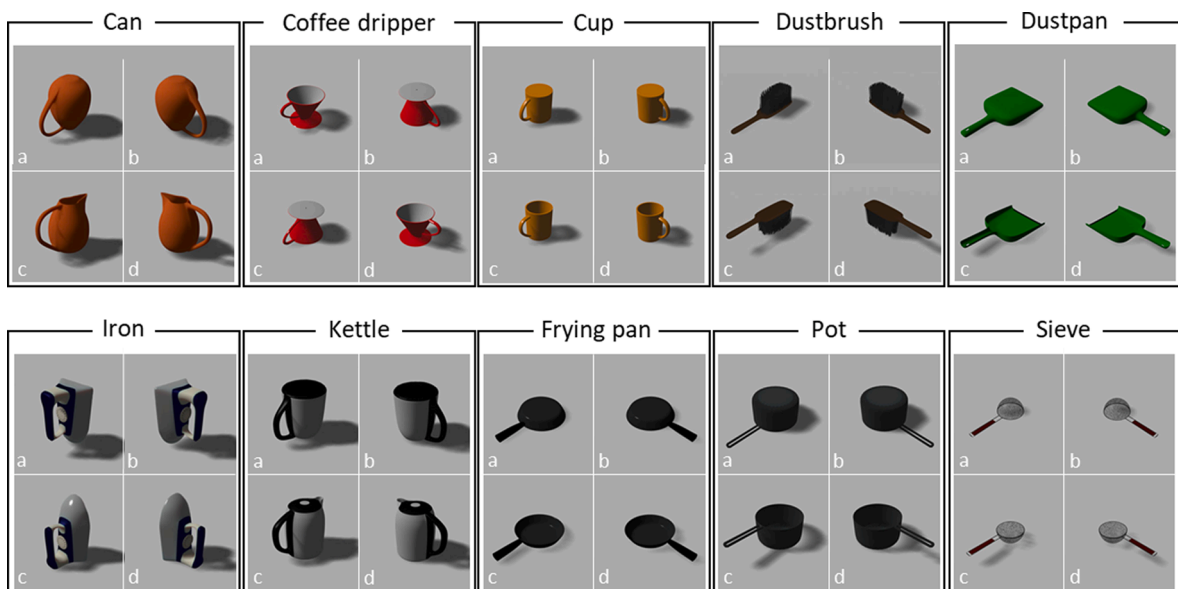
Participants received verbal and written instructions (see Appendix B and C). In all conditions where an avatar was present, participants were asked to take the perspective of it. When no avatar was present, this phrase was skipped. Furthermore, half of the participants received an instruction that mapped the left response-key to inverted objects and the right response-key to upright objects. For the other half of participants, this mapping was reversed.

The procedure for each trial was as follows: An experimental block started with the presentation of the avatar's hands, which remained visible until the end of the block (or with an empty table in the without-avatar condition of Experiment 1). After 1500 ms, one of the household objects appeared in the middle of the table, upright or inverted. The handle of this object was pointing to the right or left hand of the avatar. The participants reacted by pressing the left or right response key with their index finger, accompanied by the action effect of the avatar's corresponding index finger. Then, the next trial was initiated after 1500 ms. When an erroneous response was made, a beep feedback was given (440 Hz with a duration of 50 ms). When the response occurred less than 100 ms or more than 1000 ms after the object appeared, two beeps were given (720 Hz with a duration of 50 ms each and separated by 50 ms). This time feedback was implemented only to speed up participants' responses. After an error and/or time feedback, the next trial was delayed by an additional interval of 1500 ms.

### 2.1.6. Design

The experimental factors were *avatar* (without vs. with avatar), *handle orientation* (handle to the left vs. right), and *response side* (left vs. right response, mapped to the upright vs. inverted objects). The two conditions of the factor avatar were presented consecutively: One half of the participants started with the conditions without an avatar, followed by the conditions with an avatar. This sequence was reversed for the other half of participants.

The ten upright and inverted objects were randomly presented with the handle on the left or right. Thus, for the two avatar conditions, each block, which consisted of 40 trials, was repeated 5 times. The first block in each avatar condition was a practice block and, thus, not included in the analyses. Hence, each participant went through 320 experimental trials, with each experimental condition repeated 32 times per object and 40 times per participant.



**Fig. 2.** The objects as used in the first experiment. The objects were inverted (a, b) or upright (c, d), with their handles rotated for  $-45^\circ$  (a, c) and  $+45^\circ$  (b, d). All objects were created with colors that should ensure discriminability from the table and avatar, and a handle that should be easily visible from the actor's perspective.



### 2.1.7. Statistical analysis

To address the challenge of adequately analyzing reaction time data, we decided to follow [Lo and Andrews \(2015\)](#) by analyzing the complete and untransformed reaction time distribution (instead of only relying on measures of central tendency which are often biased; [Speelman & McGann, 2013](#); [Whelan, 2008](#)). Their approach applies generalized mixed-effects models (GLMMs) that can model non-normal response distributions, such as the typically right-skewed reaction time distributions. Following [Lo and Andrews \(2015\)](#), models with normal and right-skewed response distributions (Gaussian, inverse Gaussian, Gamma) were fitted to the data and compared by the Akaike Information Criterion (AIC, lower values indicate better fit). To avoid losing type 1 error control, null models (i.e., models that do not include predictor variables) were used for selecting the response distribution that fitted the reaction time data gathered in the experiment best.

Second, by using GLMMs with cross-classified random factors (i.e., CCGLMMs; see e.g., [Judd et al., 2017](#)), the different random factors could be accounted for (in contrast to an ANOVA, in which systematic variance between objects cannot be accounted for; see also [Judd et al., 2017](#)). Each of 24 participants (first random factor) repetitively went through each of eight experimental conditions (fixed factors) for each of ten objects (second random factor). Hence, the resulting dataset constituted of fully crossed data, because each case was clustered in one participant and one object ([Judd et al., 2017](#)). Cross-classified models account for the sources of variance attributable to random factors (participant, object, and participant-per-object combination) that, hence, do not contribute to the residual (error) variance. This allows for a more adequate estimation of the coefficients and standard errors for the fixed effects (i.e., the experimental conditions; [Judd et al., 2017](#)). Therefore, the CCGLMM approach was favored over an ANOVA, which would not be suitable for this type of data.

The present study consisted of a  $2 \times 2 \times 2$ -within-subject design with the experimental factors *orientation*, *response*, and *avatar*. With these experimental factors, full models were estimated for the fixed part (like an ANOVA, with all main effects, two-way and three-way interactions, denoted as regression coefficients  $\beta_0 - \beta_{ora}$ ). In addition, CCGLMMs also estimate the random part with the two crossed random factors participant and object  $\sigma_{pt}^2, \sigma_{ob}^2, \sigma_{pt \times ob}^2$ . The random part consisted of the participant random intercepts (i.e., variable of mean differences between participants, with variance  $\sigma_{pt}^2$ ), the object random intercepts (i.e., variable of mean differences between objects, with variance  $\sigma_{ob}^2$ ) and the participant-per-object random intercepts (i.e., variable of mean differences between individual combinations of participants and objects, with variance  $\sigma_{pt \times ob}^2$ ). Furthermore, in order to study participant's individual perspective taking of the avatar's perspective, random slopes were included for the three-way interaction on the random factor participant. This three-way interaction indicates the degree to which the spatial alignment effect is increased (for positive values) or decreased (negative values) when avatar hands are additionally presented. Hence, through estimating random slopes for the random factor participant, the decrease/increase in the spatial alignment effect when an avatar was additionally presented can be estimated for each participant. The variance of the random slope is denoted as  $\sigma_{pt \times ora}^2$ .

We hypothesized that the basic spatial alignment effect (as found by [Tucker & Ellis, 1998](#)) would be replicated. When the object's handle was rotated to the actors' left-hand side, they were assumed to react faster and, perhaps (in case of a potential speed-accuracy tradeoff), less error-prone with their left than with their right hand; similarly, when the objects' handle was rotated to the actors' right-hand side, they were assumed to react faster and, perhaps (in case of a potential speed-accuracy tradeoff), less error-prone with their right than with their left hand. This corresponded to a negative coefficient for the two-way interaction (denoted  $\beta_{or}$ , thus  $H_0: \beta_{or} \geq 0$ ). Furthermore, we assumed that the spatial alignment effect was enhanced through the presentation of avatar hands. Thus, the two-way interaction between orientation and response should be stronger with than without the presentation of avatar hands. This hypothesis corresponded to a positive three-way interaction coefficient in the CCGLMM (denoted  $\beta_{ora}$ , thus  $H_0: \beta_{ora} \leq 0$ ).

## 2.2. Results

The data was analyzed with CCGLMMs as described in the Method section. The models were estimated in R (v3.5.3), based on the R-package lme4 ([Bates, Mächler, Bolker, & Walker, 2015](#)). All the reported values were derived from the models (with  $M$  denoting the expected value derived from the model and *not* sample means). The significance level for the general null hypothesis was  $\alpha = 0.05$ . Bonferroni correction was applied in order to account for multiple testing (since although the hypotheses targeted reaction times, significance tests were also applied for the error rates), reducing the significance level of a single test to  $\alpha = 0.025$ . For the directional hypotheses, one-tailed significance tests were applied. Since there were no hypotheses for the further main effects, two- and three-way interactions, two-tailed significance tests were applied for those. For being able to test the directional hypotheses,  $t$ -tests were used (in contrast to the  $F$ -tests, which are typically used to test non-directional omnibus hypotheses. Following the recommendation by [Ulrich and Miller \(1994\)](#), no outliers were excluded. Incorrect responses (341; 4.44% of the total of 7680 responses) were removed from the reaction time distribution. Participants had error rates between 1.88% and 8.75% ( $Mdn = 4.38\%$ ). The objects, correspondingly, evoked error rates between 1.30% and 17.19% ( $Mdn = 2.67\%$ ).

### 2.2.1. Reaction times

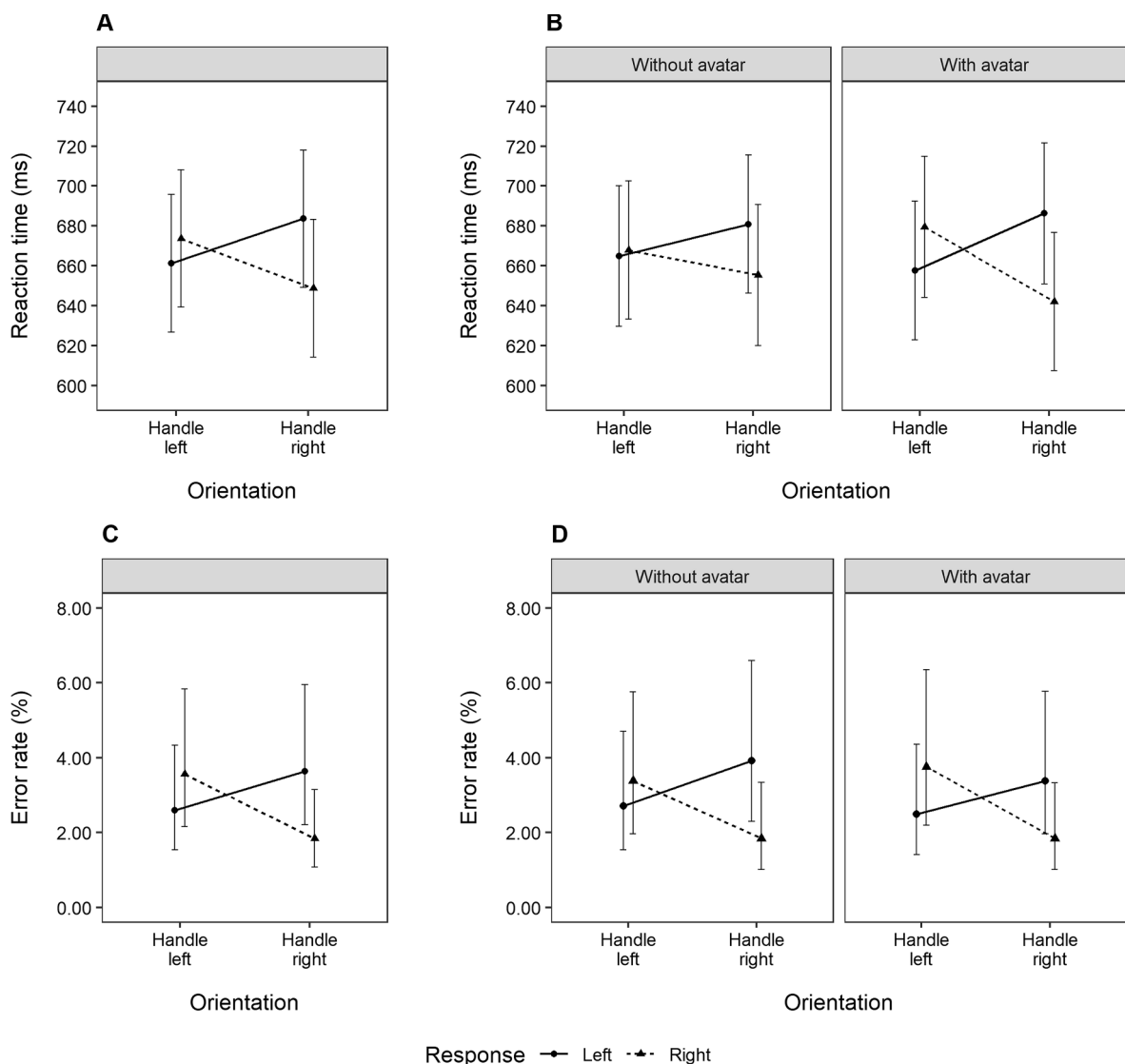
To first identify the most adequate response distribution for the reaction times (in the subscript: RT), the null model (which, contrary to the full experimental model, does not include any experimental factors) was fitted with (i) a normal distribution, (ii) an inverse gaussian distribution, and (iii) a gamma distribution (all with the identity link-function, as recommended by [Lo & Andrews, 2015](#)). The models were compared based on the AIC (Akaike Information Criterion). The model with the inverse gaussian distribution ( $AIC = 90,635$ ) provided a better fit than the one with the gaussian ( $AIC = 92,991$ ) and gamma ( $AIC = 91,077$ ) distributions. Hence, the reaction time analysis was based on a CCGLMM with an inverse gaussian response distribution and an identity link function (cf. [Lo](#)

& Andrews, 2015).

The results from this model indicated that among the main effects only the one of response side was significant ( $\beta_r = -5.62$ ;  $t = -4.23$ ;  $p < .001$ ), indicating that right-hand responses ( $M_{RT} = 672.41$ ) took significantly longer than left-hand responses ( $M_{RT} = 661.17$ ;  $M_{Diff} = 11.24$ ). The main effects of handle orientation ( $\beta_o = -0.66$ ;  $t = -0.50$ ;  $p = .620$ ) and avatar ( $\beta_a = 0.47$ ;  $t = 0.35$ ;  $p = .724$ ) were not significant.

There was a significant two-way interaction between handle orientation and response ( $\beta_{or} = -11.82$ ;  $t = -8.94$ ;  $p < .001$ ): When the handle was oriented to the left-hand side of the actor, their left-hand responses ( $M_{RT} = 661.25$ ) were faster than their right-hand responses ( $M_{RT} = 673.64$ ;  $M_{Diff} = -12.39$ ). For handles oriented to the right-hand side of the actor, participants reacted slower with their left hand ( $M_{RT} = 683.57$ ) than with their right hand ( $M_{RT} = 648.70$ ;  $M_{Diff} = 34.88$ ). The effect is illustrated in Fig. 3A. The two-way interaction between orientation and avatar was not significant ( $\beta_{oa} = 1.52$ ;  $t = 1.15$ ;  $p = .252$ ), as well as the interaction between response and avatar ( $\beta_{ra} = -0.01$ ;  $t = -0.01$ ;  $p = .992$ ). In summary, responses were faster when an object's handle was oriented towards the actors' and avatar's response hand.

Furthermore, the three-way interaction between the factors orientation, response and avatar was significant ( $\beta_{ora} = 4.71$ ;  $t = 2.24$ ;  $p = .012$ ), indicating that the handle orientation effect was increased when the avatar was presented compared to the condition when it was not presented. For handles oriented towards the left, the difference between left-hand ( $M_{RT} = 657.57$ ) and right-hand ( $M_{RT} = 679.42$ ) responses was higher when avatar hands were presented ( $M_{Diff} = -21.85$ ) than when no avatar hands were presented (left



**Fig. 3.** Reaction times (A, B) and error rates (C, D) in Experiment 1 as function of the two-way interaction between handle orientation and response (A, C) and the three-way interaction between handle orientation, response and with/without avatar (B, D). Handle orientation left/right is depicted from the perspective of the avatar. Error bars represent 95%-Confidence Intervals (CIs).

hand  $M_{RT} = 664.92$ ; right hand  $M_{RT} = 667.87$ ;  $M_{Diff} = -2.94$ ). Similarly, for handles oriented towards the actor's right-hand side, the difference between left-hand ( $M_{RT} = 686.29$ ) and right-hand responses ( $M_{RT} = 642.01$ ) was higher when an avatar was presented ( $M_{Diff} = 44.28$ ) than when no avatar was presented (left hand:  $M_{RT} = 680.86$ ; right hand  $M_{RT} = 655.38$ ;  $M_{Diff} = 25.48$ ). Hence, the spatial alignment effect was stronger when avatar hands, mimicking the actor's finger movement, were additionally presented (Fig. 3B).

The random part of the model further decomposed the variances into random intercepts (for the factors participant, object, and participant-per-object) and random slopes (for the factors participant). The intercepts of the reaction times varied substantially within each of the three factors participant ( $\sigma_{Pt}^2 = 470.58$ ;  $SD = 21.69$ ), object ( $\sigma_{Ob}^2 = 371.97$ ;  $SD = 19.29$ ) and participant-per-object ( $\sigma_{Pt \times Ob}^2 = 749.03$ ;  $SD = 27.37$ ). Furthermore, there was also variance in the random slopes of the three-way interaction ( $\sigma_{Pt \times ORA}^2 = 30.88$ ;  $SD = 5.56$ ). The correlation between the random intercepts and slopes for the random factor participant was low ( $Cor = 0.15$ ). As can be seen in Fig. 4, for the majority of participants, the presence of avatar hands increased the spatial alignment effect (as indicated by a positive value). That is, of the twenty-four participants, only two had an individual coefficient for the three-way interaction of zero or smaller. For the further twenty-two participants, the presence of an avatar increased the spatial alignment effect.

### 2.2.2. Error rates

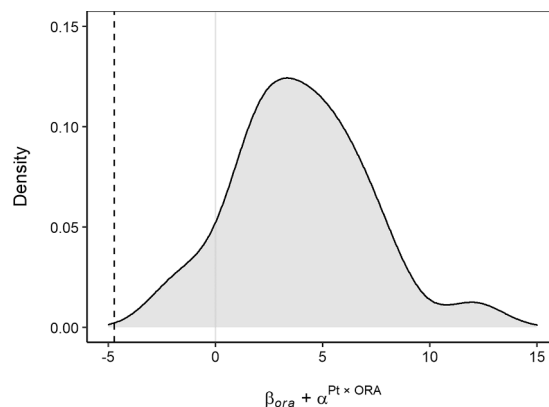
The full experimental model was also estimated for the error rates (in the subscript: ER). For error rates, the adequate response distribution is a binomial distribution (since each response can be either correct or incorrect). Furthermore, a logit-link function was used. Since no  $t$ -test was available, a  $z$ -test was used for significance testing.

The only significant effect in this model was the two-way interaction between orientation and response ( $\beta_{or} = -0.26$ ;  $z = -4.50$ ;  $p < .001$ ). When the object handle was oriented to the actors' left-hand side, they responded less error-prone with their left ( $M_{ER} = 2.60\%$ ) than with their right hand ( $M_{ER} = 3.57\%$ ;  $M_{Diff} = -0.97\%$ ). Accordingly, when the object handle was oriented to the actors' right-hand side, their left-hand responses were more error-prone ( $M_{ER} = 3.64\%$ ) than their right-hand responses ( $M_{ER} = 1.85\%$ ;  $M_{Diff} = 1.79\%$ ; see Fig. 3C). In contrast to the reaction time analysis, the three-way interaction did not yield a significant effect ( $\beta_{ora} = 0.01$ ;  $z = 0.10$ ;  $p = .461$ ): The spatial alignment effect was not significantly increased when avatar hands were additionally presented. Thus, the results of this model generally supported the spatial alignment effect, which did not increase with the presentation of avatar hands.

## 2.3. Discussion

The objective of Experiment 1 was to test the object affordances paradigm for avatar perspective taking by examining (i) if the spatial alignment effect is generally found and (ii) if the presence of avatar hands showing an action effect corresponding to the actor's button press increased the effect. The results indicated that participants reacted faster and less error-prone to objects whose handles were aligned with their response hand, compared to handles that were pointing towards the opposite (non-responding) hand. This implies that the object's affordances elicited corresponding responses in the participants and thus are suitable to study how actors take an avatar's perspective by object affordances. Furthermore, the avatar hands affected the actors' responses, increasing the spatial alignment effect (as hypothesized).

Note, however, that in a recent study, [Chen et al. \(2021\)](#) observed that hand visibility plays no role in the original Simon task (i.e., with no avatar). Why, then, does the presentation of avatar hands in the present experiment produced an increased object compatibility effect? We attribute this mainly to two aspects: (i) the avatar hands were presented immediately adjacent to the imperative objects on the screen, making them far more intrusive; (ii) the participants' keypresses were accompanied by action effects of the corresponding avatar hand. In a similar experimental set-up, we were able to demonstrate that consistent action effects of an avatar strengthened compatibility ([Böffel & Müsseler, 2019a](#)). Likewise, [Pfister, Weller, Dignath, and Kunde \(2017\)](#) observed that



**Fig. 4.** The inter-individual differences in the three-way interaction between orientation, response and avatar of Experiment 1. The fixed effect ( $\beta_{ora}$ ) is indicated by the dashed line. Higher values indicate a stronger increase in the spatial alignment effect when avatar hands were additionally presented.



participants' responses were facilitated when they were imitated by an avatar (cf. also the role of action effects in joint tasks with other humans; Pfister, Dolk, Prinz, & Kunde, 2014; Müller, 2016). Because of these differences, we do not think that the present results contradict the findings of Chen et al. (2021).

### 3. Experiment 2

While in Experiment 1, the spatial alignment effect was facilitated through the presentation of avatar hands that were aligned with the participant's perspective, the main question for Experiment 2 was whether participants perceive object affordances from the avatar's perspective, even if it was misaligned with their own. For this purpose, the avatar and objects were rotated by  $\pm 90^\circ$  from the actor's perspective, so that the avatar would be positioned at the virtual table to either the left or right of the participant (Fig. 1, right panel). The misalignment of actor and avatar perspectives created a situation in which the direction of the object handle could afford conflicting responses from the actor and avatar perspectives. For instance, in the top-right picture of Fig. 1 ( $\pm 45^\circ$  object rotation from participant's perspective), the handle is oriented towards the right from the perspective of the participant and therefore should facilitate a right response. However, when the participant takes the perspective of the avatar, object affordance should facilitate a left-hand response since the handle is oriented towards the avatar's left. In the bottom-right picture of Fig. 1 ( $\pm 135^\circ$  object rotation from participant's perspective), the handle is oriented to the right from the perspective of the avatar and the participant (although for the latter it is also pointing upwards).

Our hypothesis (H3) was that participants take the avatar perspective and thus, the spatial alignment effect corresponds with the avatar perspective. Hence, with a left-hand response, participants should react faster to an object with its handle directed to the avatar's left-hand side than to an object handle directed to the avatar's right-hand side (and vice versa). This should be independent of the rotation of the avatar's perspective (i.e., whether the avatar was located to the right or left of the participant). To sum up, participants should show a normal spatial alignment effect when the handle is oriented towards corresponding actor and avatar hands (i.e., for the  $\pm 135^\circ$  conditions) and an inverted spatial alignment effect when the handle is oriented towards non-correspondent actor and avatar hands (i.e., for the  $\pm 45^\circ$  conditions).

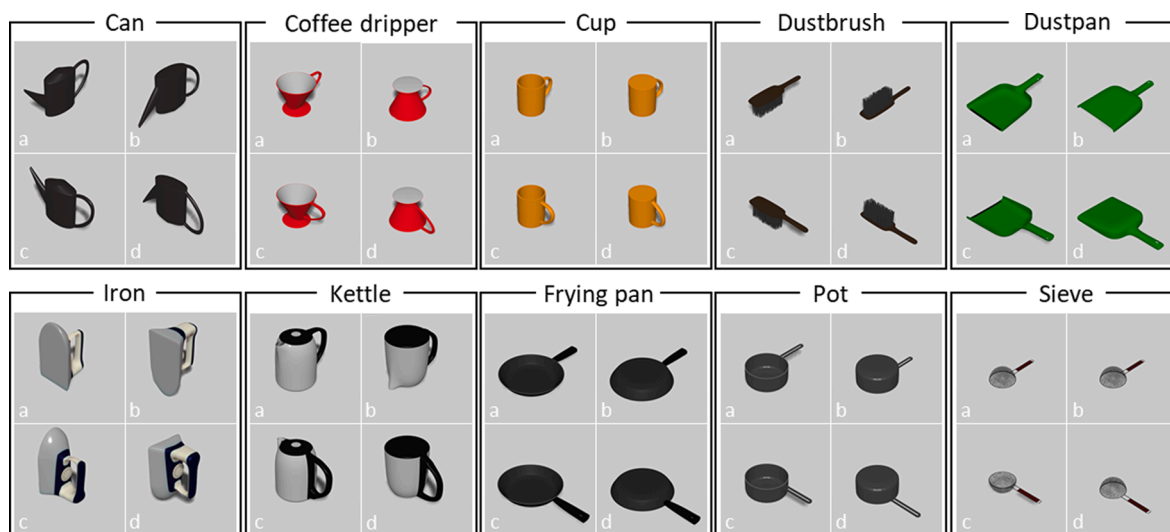
#### 3.1. Method

##### 3.1.1. Power analysis

Again, a power analysis was conducted based on the parameters from the seminal study by Tucker and Ellis (1998), with the procedures described by Judd et al. (2017) as well as Arend and Schäfer (2019) and Claus et al. (2020). Appendix D provides the full code. A power of 1.00 for H3 was achieved with 24 participants, 10 objects and 4 repetitions per condition, object, and participant. The Bonferroni correction later applied during the statistical analysis did not reduce power for H3.

##### 3.1.2. Participants and objects

Twenty-four new participants (20 female and four male) took part in the study, all but two (mathematics and social sciences) psychology students at RWTH Aachen University. Again, the students were recruited via the Institute of Psychology's E-Mail



**Fig. 5.** The objects as used in Experiment 2. The objects were upright (a, c) or inverted (b, d), with their handles rotated for  $+135^\circ$  (a, b) or  $+45^\circ$  (c, d) from the participants point of view. The same objects as in Experiment 1 were used for Experiment 2. However, as the objects were rotated additionally by  $\pm 135^\circ$ , their appearance had to be slightly adapted to guarantee visibility of the handle for these rotations. For example, the handle of the can used in Experiment 1 was not clearly visible after the rotation, and thus another type of can was created with Blender.

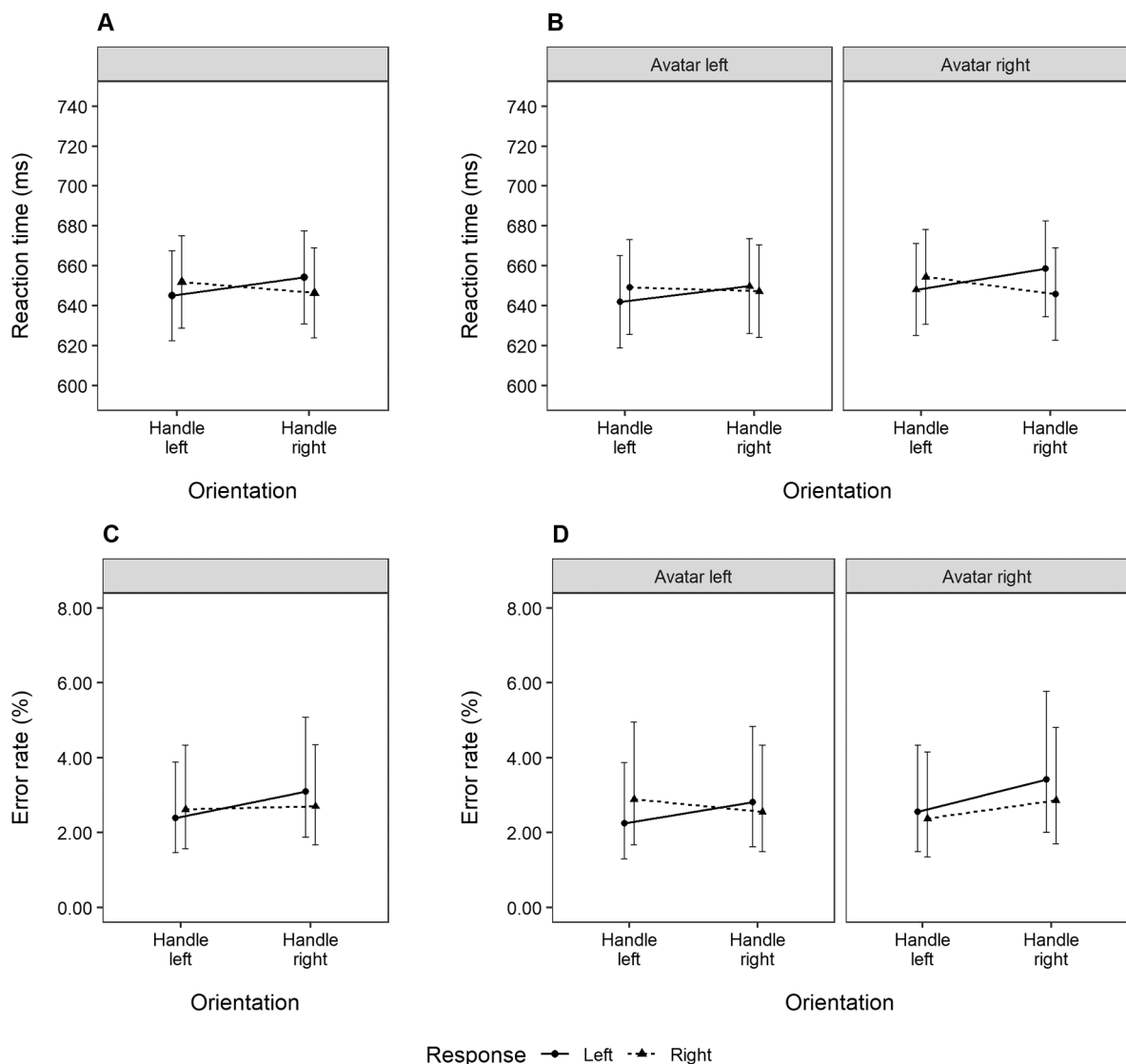
distribution list and psychology students received course credit for their participation. The participants were between 18 and 27 years old ( $M = 21.33$ ;  $SD = 3.00$ ) and reported normal or corrected to normal vision. The same ethics considerations and approval (ref. 2020\_002\_FB7\_RWTH Aachen University) also applied for Experiment 2. Fig. 5 provides an overview of the objects used. The procedure was the same as in Experiment 1.

### 3.1.3. Design

The experimental factors were *avatar position* (left vs. right), *handle orientation* (handle to the left vs. right), and *response side* (left vs. right, mapped to the upright vs. inverted objects). The two conditions of the factor avatar were presented consecutively: One half of the participants started with the conditions in which the avatar was rotated to the left, followed by the conditions in which the avatar was rotated to the right. This sequence was reversed for the other half of participants. Furthermore, the mapping of left vs. right responses to upright vs. inverted objects inverted for one half of the participants). The presentation of objects, blocks, and number of trials was the same as in Experiment 1.

### 3.1.4. Statistical analysis

To address the perspective taking hypothesis, two (nearly identical) analyses were performed. In the first analysis, the experimental factors were orientation, response hand, and avatar position, with a random slope for the two-way interaction between orientation and



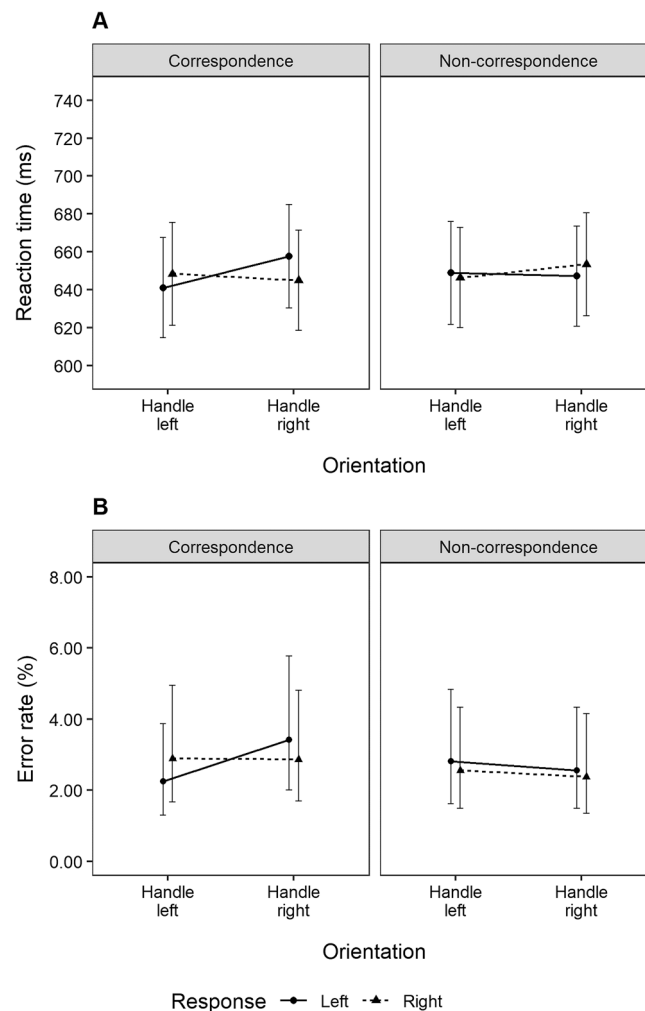
**Fig. 6.** Findings from the avatar perspective in Experiment 2: Reaction times (A, B) and error rates (C, D) as function of the two-way interaction between orientation and response (A, C) and the three-way interaction between orientation, response, and avatar left/right (B, D). Error bars represent 95%-CIs.

response side (i.e., the spatial alignment effect) for the random factor participant. With this model, the presence/absence of the spatial orientation effect from the *avatar perspective* was modelled. Hypothesis H3 claimed that, due to taking the avatar's perspective, when the handle was oriented to the avatar's left hand, actors would react faster and, perhaps (in case of a potential speed-accuracy tradeoff), less error-prone with their left than with their right hand. Accordingly, when the handle was oriented to the avatar's right hand, participants would react faster and, perhaps (in case of a potential speed-accuracy tradeoff), less error-prone with their right than with their left hand. This corresponded to a negative coefficient for the two-way interaction between orientation and response (denoted  $\beta_{or}$ , thus  $H_0: \beta_{or} \geq 0$ ).

Furthermore, to depict the results also from the actor perspective, handle orientation was recoded as left (i.e., handle orientation of  $-45^\circ$  and  $-135^\circ$ ) and right (i.e., handle orientation of  $+45^\circ$  and  $+135^\circ$ ) from the actor's body midline. Furthermore, the variable hand-correspondence (C) was introduced, indicating whether the handle was spatially oriented towards the same hand of actor and avatar (correspondent,  $\pm 135^\circ$ ) or towards opposing hands (non-correspondent,  $\pm 45^\circ$ ). When actors take the avatar's perspective, an inverted spatial alignment effect would be expected for the condition with non-correspondent hands, but a classic spatial alignment effect for the condition when the hands corresponded. A corresponding second model was set up to model individual three-way interaction estimates (i.e., the estimates of individual spatial alignment effects), with negative estimates indicating that actors rather preferred the avatar's perspective and positive estimates indicating that they rather preferred their own perspective.

### 3.2. Results

The data were analyzed as in Experiment 1. Again, Bonferroni correction was applied because H3 was tested for both, the reaction times and error rates. Hence, the significance level of  $\alpha = 0.05$  was reduced to  $\alpha = 0.025$  for each of the two single tests.



**Fig. 7.** Findings from the actor perspective in Experiment 2: Reaction times (A) and error rates (B) as function of the three-way interaction between orientation, response, and hand correspondence. Error bars represent 95%-CIs.

### 3.2.1. Reaction times

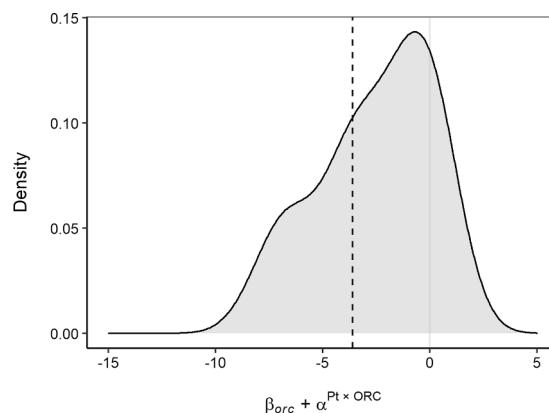
Again, the null model was fitted with (i) a normal distribution, (ii) an inverse gaussian distribution, and (iii) a gamma distribution (all with the identity link-function) to identify the most adequate response distribution for the reaction times (cf. Lo & Andrews, 2015). The models were compared based on the AIC (Akaike Information Criterion). As in Experiment 1, the model with the inverse gaussian distribution ( $AIC = 91,013$ ) provided a better fit than the one with the gaussian ( $AIC = 93,356$ ) and gamma ( $AIC = 91,472$ ) distributions. Hence, the reaction time analysis was based on a CCGLMM with an inverse gaussian response distribution and an identity link function.

A full model with all experimental variables as fixed effects, the random factors participant, object, and participant-per-object as well as the random slope of the two-way interaction between orientation and response for the random factor participant was estimated. The results from the model indicated that no main effects were significant. Yet, there was a significant two-way interaction between handle orientation and response ( $\beta_{or} = -3.63$ ;  $t = -1.99$ ;  $p = .023$ ): When the handle was oriented to the left-hand side of the avatar, the actors' left-hand responses ( $M_{RT} = 644.96$ ) were faster than their right-hand responses ( $M_{RT} = 651.84$ ;  $M_{Diff} = -6.89$ ). In contrast, when the handle was oriented to the right-hand side of the avatar, the actors took longer to respond with their left ( $M_{RT} = 654.13$ ) than with their right hand ( $M_{RT} = 646.47$ ;  $M_{Diff} = 7.66$ ; see Fig. 6).

Furthermore, neither the two-way interaction between orientation and avatar position ( $\beta_{oa} = -0.50$ ;  $t = -0.38$ ;  $p = .705$ ) nor between response and avatar position ( $\beta_{ra} = -1.40$ ;  $t = -1.056$ ;  $p = .291$ ) were significant, nor was the three-way interaction between all variables ( $\beta_{ora} = -1.13$ ;  $t = 0.86$ ;  $p = .392$ ). Hence, the results from this model generally support the hypothesis that participants react to object affordances from the avatar's perspective.

Additionally, the conditions were recoded to contain the factors handle orientation from the participant perspective and hand-correspondence, indicating whether the handle was oriented towards non-corresponding or corresponding actor and avatar hands. For conditions in which the handle was oriented towards the corresponding actor and avatar hand (i.e., rotated for  $\pm 135^\circ$ ), a classic spatial alignment effect occurred (see Fig. 7): For handles oriented towards their left, participants reacted faster with their left hand ( $M_{RT} = 641.03$ ) than with their right hand ( $M_{RT} = 648.38$ ;  $M_{Diff} = 7.34$ ), whereas for handles oriented towards their right, participants reacted slower with their left ( $M_{RT} = 644.91$ ) than with their right hand ( $M_{RT} = 657.56$ ;  $M_{Diff} = 12.64$ ). In contrast, for conditions in which the handle was oriented towards non-corresponding actor and avatar hands ( $\pm 45^\circ$ ), participants reacted faster to handles that were misaligned than to handles that were aligned with their hands. For handles oriented towards their left, participants reacted slower with their left ( $M_{RT} = 648.87$ ) than with their right hand ( $M_{RT} = 646.35$ ;  $M_{Diff} = 2.52$ ). For handles oriented towards their right, participants reacted faster with their left ( $M_{RT} = 647.19$ ) than with their right hand ( $M_{RT} = 653.47$ ;  $M_{Diff} = -6.29$ ). This result illustrates that the spatial alignment effect was reversed for conditions in which the avatar perspective provided conflicting information.

Furthermore, a second CCGLMM which contained the recoded experimental factors was estimated to consider the distribution of the inter-individual three-way interaction effects. The random part of the model further decomposed the variances into random intercepts (for the factors participant, object, and participant-per-object) and random slopes (for the factor participant). The intercepts of the reaction times varied substantially within each of the three factors participant ( $\sigma_{Pt}^2 = 558.56$ ;  $SD = 23.63$ ), object ( $\sigma_{Ob}^2 = 376.36$ ;  $SD = 19.40$ ) and participant-per-object ( $\sigma_{Pt \times Ob}^2 = 701.71$ ;  $SD = 26.49$ ). Furthermore, there was also variance in the random slopes of the three-way interaction ( $\sigma_{Pt \times ORA}^2 = 20.34$ ;  $SD = 4.51$ ). The correlation between the random intercepts and slopes for the random factor participant was low ( $Cor = -0.21$ ). As indicated in Fig. 8, most of the participants showed a three-way interaction effect close to zero. Yet, half of the participants had an individual three-way interaction coefficient of  $-2$  or lower, indicating that a non-negligible inversion of the spatial alignment effect occurred since these participants seem to have taken the avatar's perspective. Furthermore, the distribution is skewed to the left (i.e., towards more negative individual coefficients), suggesting a stronger perspective taking mechanism in this group. The random slopes were further only slightly correlated with the random intercepts in the factor participant ( $cor = -0.21$ ).



**Fig. 8.** The inter-individual differences in the three-way interaction between orientation, response, and correspondence of Experiment 2. The fixed effect ( $\beta_{ora}$ ) is indicated by the dashed line. Negative values indicate an inversion of the spatial alignment effect for conflicting spatial information from the avatar and actor perspective.

### 3.2.2. Error rates

For the error rates, the same analyses were performed. First, the model from the avatar perspective was estimated. In this model, no effect was significant. The critical two-way interaction between orientation and response ( $\beta_{or} = -0.06$ ;  $z = -0.85$ ;  $p = .199$ ) showed the following values: When the handle was oriented to the left-hand side of the avatar, the actors' left-hand responses were less error-prone ( $M = 2.39\%$ ) than their right-hand responses ( $M = 2.62\%$ ;  $M_{Diff} = -0.23\%$ ). In contrast, when the handle was oriented to the actors' right-hand side, they took longer to respond with their left ( $M = 3.10\%$ ) than with their right hand ( $M = 2.70\%$ ;  $M_{Diff} = 0.40\%$ ; see Fig. 7). Hence, although there was a small spatial alignment effect in the error rates for this sample, this effect was not significant.

Accordingly, the spatial alignment effect was not reversed in the error rates for the recoded factors: For conditions in which the handle was oriented towards the corresponding actor and avatar hand (i.e., rotated for  $\pm 135^\circ$ ), participants reacted less error-prone with their left hand ( $M = 2.24\%$ ) than with their right hand ( $M = 2.89\%$ ;  $M_{Diff} = 0.65\%$ ) to handles that were oriented towards their left, whereas for handles oriented towards their right, participants reacted more error-prone with their left ( $M = 3.42\%$ ) than with their right hand ( $M = 2.86\%$ ;  $M_{Diff} = 0.56\%$ ). For conditions in which the handle was oriented towards non-corresponding actor and avatar hands ( $\pm 45^\circ$ ), participants reacted more error-prone with their left ( $M = 2.81\%$ ) than with their right hand ( $M = 2.55\%$ ;  $M_{Diff} = 0.26\%$ ) for handles oriented towards their left. For handles oriented towards their right, participants reacted more error-prone with their left ( $M = 2.55\%$ ) than with their right hand ( $M = 2.37\%$ ;  $M_{Diff} = 0.18\%$ ).

### 3.3. Discussion

The goal of Experiment 2 was to study whether and how actors take an avatar's perspective by object affordances. Therefore, participants were asked to take an avatar's perspective for completing the experimental task. Although the task-relevant information (i.e., objects inverted or upright) was the same from both perspectives, the task-irrelevant spatial cue (the handle orientation) provided conflicting information from the participant and avatar perspective. For reaction times, the results indicated that participants perceived the object affordances from the avatar's instead of their own perspective. However, this effect was comparably small.

## 4. General discussion

The objective of the present research was to extend the understanding of whether and how actors' take the avatar's perspective in an object-affordance paradigm. In this paradigm, participants were presented with objects whose graspable handles were aligned with their own hands or the avatar's hands. It was expected that, when taking the avatar's perspective, actors react faster when the responding hand corresponded to the avatar's hand that was aligned with the object's handle.

In Experiment 1, the typical spatial alignment effect without an avatar was observed (Tucker & Ellis, 1998; Costantini, Committeri, et al., 2011; Symes et al., 2005). Furthermore, the spatial alignment effect increased as predicted when the hands of an avatar were additionally presented close to the objects handles. In Experiment 2, the perspective taking mechanism was tested by rotating the avatar perspective for  $90^\circ$  towards the left or right from the actor perspective. Although smaller than in Experiment 1, a significant spatial alignment effect from the perspective of the avatar was found. This was also confirmed from the actor's perspective, from which participants reacted with the pattern of an inverse spatial alignment effect when the objects' handles were oriented towards the non-corresponding hands of the actor and avatar: Participants reacted faster to handles that were aligned with their non-responding hand that corresponded to the avatar's hand to which the handle was oriented. In other words, in conditions in which the object's affordance through its graspable handle was orthogonal for the participant and avatar perspectives, most participants showed a spatial alignment effect in accordance with the avatar's—and not their own—perspective. Hence, we can conclude that the present study contributes to and extends the evidence that object affordances also have an effect from the perspective of others.

Beyond our contribution to the literature studying affordances in spatial compatibility paradigms (Symes et al., 2005; Tucker & Ellis, 1998), our results extend the findings on spatial coding of affordances in the presence of an avatar. In the study by Costantini, Committeri, et al. (2011), a virtual individual (a non-agent, static avatar) was also placed to the left and right of the participant with its perspective rotated for  $\pm 90^\circ$  from the actor's perspective. Yet, the objects were not placed in the middle of the table (i.e., not with similar distance from the participant and avatar point of view) but near (reachable distance) or far (non-reachable distance) from the actor, and—most differently to the present study—with the objects handle always pointing towards the avatar. The results of this study indicated that, when the objects were presented in reachable space, participants showed a spatial alignment effect from their own perspective, disregarding the avatar. Hence, these results remained behind those of the present Experiment 2, where the spatial alignment effect was inverted from the participants' perspective in favor of the avatar's perspective.

By applying a mixed-effects modelling approach, the present study provides initial evidence that there are substantial inter-individual differences in the degree to which participants take the avatar's perspective. For Experiment 1, the presence of an avatar increased the spatial alignment effect for all but two participants. For Experiment 2, there was substantial variance in the inverse spatial alignment effect between participants: Whereas for many participants, this effect was very close to zero, there was also another large group of participants who had coefficients that indicated that they strongly took the avatar's perspective for perceiving and reacting to object affordances. This research thus adds to the plausible evidence for inter-individual differences in the degree of perspective taking (Kessler & Wang, 2012). Hence, this modelling approach could further be applied in order to extend and strengthen the understanding of inter-individual differences in avatar perspective taking and perceived body ownership.

Furthermore, it is important to note that in our experiments, the avatar acted as an agent of the participants by showing action effects in accordance with their finger movements. These action effects were explicitly introduced to facilitate perspective taking and to enable identification with the avatar (for recent evidence regarding this assumption see Pfister et al., 2017; Böffel & Müsseler,

2019a; von Salm-Hoogstraeten & Müsseler, 2021). We therefore assume that action effects contributed to the present results as expected. It is also possible that, thereby, actors accepted the avatar's hands as their own hands, but this remains speculative (although subjective data from body-ownership questionnaires pointed in this direction, cf. Böffel & Müsseler, 2018, 2019a).

Although, as intended, the action effects shown by the avatar should facilitate the identification with the virtual person, they might create an interpretation problem. The increased spatial alignment effect with the avatar obtained in Experiment 1 and the spatial alignment effect from the perspective of the avatar obtained in Experiment 2 could have been due to a stimulus-stimulus compatibility between the handle of an object and avatar's (anticipated) finger movement (cf., Kornblum, Stevens, Whipple, & Requin, 1999; Shin, Proctor, & Capaldi, 2010). For the avatar-Simon effect, this interpretation cannot be used in this form, since the effect was also observed without presenting action effects at the avatar (Böffel & Müsseler, 2019; von Salm-Hoogstraeten et al., 2020). Unless one assumes that imagining the action effects is sufficient to produce the findings reported there (as in tools use, cf. Müsseler, Wühr, & Ziessler, 2014). This remains to be clarified in future studies.

So, what else can account for the underlying cognitive representations and processes that gave rise to the present spatial alignment effect (or referring to Cho & Proctor, 2010, the object-based Simon effect)? In this paper, the discussion favoring the affordance account over the object-based account (or vice versa) will not be continued and clarified, as our findings do not contribute to this debate—and especially since this has already been done in detail elsewhere (see Cho & Proctor, 2010, 2011, 2013). Instead, the role of perspective taking for the (inverted) spatial alignment effect needs to be illuminated here.

Visual perspective taking is usually understood as a process that brings about the ability to see the world from another person's perspective, taking into account what they see and how they see it (Flavell, 1977). To make this possible, the mechanism of perspective taking could create a visual spatial representation from another reference frame. So, if the participants take the view of the avatar, they “see” the objects' handles in the perspective-created spatial representation on the left or right side (see also: Costantini, Committeri, et al., 2011), yielding left or right codes which underly the compatibility effects (for the different accounts of stimulus-response compatibility see Proctor & Vu, 2006).

Recent studies from our lab cast doubt on this simplification of perspective taking. Von Salm-Hoogstraeten, Bolzius and Müsseler (2020) compared two avatar scenarios: The first scenario was similar to the one used in the present Experiment 2. An avatar sat either to the left or to the right of a table and participants performed a Simon color-classification task to left/right stimuli from the viewpoint of the avatar. Note, that from the participants' point of view, the stimuli were arranged one above the other (i.e., no spatial information on the horizontal dimension). The second scenario was similar to the one used in the present Experiment 1, that is, the participant took the ego perspective of the avatar, but the avatar's right and left hand were now at the top and bottom stimulus position. In the latter scenario, only the avatar's hands formed the left and right relation to the stimulus positions. A perspective-created visual representation could only account for effects in the first scenario while the avatars' hands could produce a left-right frame of reference in both scenarios. The results showed pronounced avatar-Simon effects in both scenarios.

We interpreted this result as evidence for the view that the position of another person, and also the spatial positions of any other object in the scene, could be selected as a new spatial reference point from which the spatial relationships of the objects to each other can be redefined. That the spatial coding of objects can arise in reference to other objects is an idea postulated by the referential coding account that was originally proposed to explain spatial compatibility effects in the standard Simon effect (Hommel, 1993), and then was applied to the orthogonal compatibility effect (Lippa, 1996; Lippa & Adam, 2001; Cho & Proctor, 2005) and the object-based Simon effect (Cho & Proctor, 2010, 2011, 2013). Recently the referential coding account was also extended with regard to the joint Simon effect (e.g., Dolk, Hommel, Prinz, & Liepelt, 2013).

According to the referential coding account of perspective taking, the basic spatial map develops from the actor's perspective, which, however, already contains all spatial relationships between objects in the visual space (cf. the visual sensory map of van der Heijden, Müsseler, & Bridgeman, 1999). Consequently, the actor does not need to create a new visual spatial map from the avatar's perspective, but rather recodes the existing coordinates with regard to the new reference point. Thus, there is little visual in visual perspective taking (see also Cole & Millett, 2019).

## 5. Conclusion

The goal of the present research was to examine object affordances as a paradigm for studying whether and how actors take an avatar's perspective. A spatial alignment effect from the avatar perspective was found in two experiments: Participants responded faster to objects whose handles were aligned with the avatar hand that corresponded to the participant's response hand. This effect persisted when the handle was aligned with non-corresponding actor and avatar hands, inverting the spatial alignment effect from the actor's in favor of the avatar's perspective, from which a normal spatial alignment effect was found. Our interpretation of the results favored a referential coding account of perspective taking.

## Author note

This study was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation; project number MU 1298/11-1) and was associated with the DFG Priority Program “The Active Self” (DFG SPP 2134).

## CRediT authorship contribution statement

**Matthias G. Arend:** Conceptualization, Formal analysis, Investigation, Methodology, Project administration, Visualization,



Writing - original draft, Writing - review & editing. **Jochen Müsseler**: Conceptualization, Funding acquisition, Methodology, Resources, Software, Supervision, Writing - original draft, Writing - review & editing.

## Appendix A. Power analyses

Generally, our study was designed following the seminal object-affordances study by [Tucker and Ellis \(1998\)](#), the avatar study applying their paradigm by [Costantini, Committeri, et al. \(2011\)](#), as well as studies on avatar perspective taking conducted in our lab. Prior to the power analysis, it was decided that at least 10 objects should be studied since those mirrored the general categories of household objects used by [Tucker and Ellis \(1998\)](#) and constitute the minimum number of levels for a random factor to receive unbiased parameter estimates in multilevel modelling contexts. Furthermore, we wanted to repeat each condition (each participant with each object and each of eight experimental conditions) at least four times, in order that each participant had 40 repetitions per experimental condition (for all objects together). Similar to an ANOVA approach, eight experimental conditions result of all combinations of the three within-participant/object factors handle orientation (*o*), response side (*r*), and avatar (*a*). Furthermore, we estimated the power for reaction times, since our hypotheses assumed that the experimental variations would result in reaction time differences in the spatial alignment effect (see section Present Research and Hypotheses).

Furthermore, since the appropriate way to analyze the resulting dataset was cross-classified multilevel or mixed-effects modelling, a power analysis had to be performed for this type of model. To our knowledge, no procedure for power analyses of cross-classified multilevel models with two crossed random factors and eight experimental conditions (three factors; see [Judd et al., 2017](#) for a basic approach with one dichotomous explanatory variable) existed. Hence, we relied on Monte Carlo simulation for the power analysis ([Arend & Schäfer, 2019](#); [Claus et al., 2020](#)).

For Monte Carlo simulation, a population model had to be estimated from which all parameters that are needed for setting up the cross-classified multilevel model must be derived. Thereby, these parameters can be distinguished into the fixed and random part of the model. Like for ANOVA, the fixed part contains all combinations of predictors, resulting in main effects, two- and three-way interactions of the factors *o*, *r*, and *a*. The random part of the model contains the random factors participant, object, and participant-per-object. Furthermore, since we wanted to derive parameters for the subject-specific spatial alignment effects, population effects for a random slope for the three-way interaction (Experiment 1), or the two-way interaction between orientation and response (Experiment 2) were estimated. The following formulae of the full population models resulted for Experiment 1 and 2:

$$\text{Experiment 1: } Y_{ijk} = \beta_I + \beta_o O_{ijk} + \beta_r R_{ijk} + \beta_a A_{ijk} + \beta_{or} O_{ijk} R_{ijk} + \beta_{oa} O_{ijk} A_{ijk} + \beta_{ra} R_{ijk} A_{ijk} + \beta_{ora} O_{ijk} R_{ijk} A_{ijk} + \alpha_i^{Pt} + \alpha_j^{Ob} + \alpha_{ij}^{Pt \times Ob} + \alpha_i^{Pt \times ORA} O_{ijk} R_{ijk} A_{ijk} + \epsilon_{ijk}$$

$$\text{Experiment 2: } Y_{ijk} = \beta_I + \beta_o O_{ijk} + \beta_r R_{ijk} + \beta_a A_{ijk} + \beta_{or} O_{ijk} R_{ijk} + \beta_{oa} O_{ijk} A_{ijk} + \beta_{ra} R_{ijk} A_{ijk} + \beta_{ora} O_{ijk} R_{ijk} A_{ijk} + \alpha_i^{Pt} + \alpha_j^{Ob} + \alpha_{ij}^{Pt \times Ob} + \alpha_i^{Pt \times OR} O_{ijk} R_{ijk} + \epsilon_{ijk}$$

To derive population parameters, we first set up the unique design matrix, which contains all variables for the coefficients of the fixed part of the model. Therefore, each factor was assumed to be contrast-coded with  $-1$  and  $+1$ . For orientation (*o*; handle oriented to the left  $[-1]$  vs. right  $[+1]$ ) and response (*r*; response with the left  $[-1]$  and right  $[+1]$  hand), coding was similar in Experiment 1 and 2. For avatar (*a*), Experiment 1 coded the with avatar  $[-1]$  and the without avatar  $[+1]$  conditions, and Experiment 2 coded the avatar left  $[-1]$  vs. avatar right  $[+1]$  conditions. With all combinations of the values of these three factors, each of the eight conditions can be coded (column two to four in the unique design matrix). Furthermore, the vectors for each of the further regression coefficients can be created by multiplication:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \\ Y_6 \\ Y_7 \\ Y_8 \end{bmatrix} = \begin{bmatrix} 1 & -1 & -1 & -1 & +1 & +1 & +1 & -1 \\ 1 & +1 & -1 & -1 & -1 & -1 & +1 & +1 \\ 1 & -1 & +1 & -1 & -1 & +1 & -1 & +1 \\ 1 & +1 & +1 & -1 & +1 & -1 & -1 & -1 \\ 1 & -1 & -1 & +1 & +1 & -1 & -1 & +1 \\ 1 & +1 & -1 & +1 & -1 & +1 & -1 & -1 \\ 1 & -1 & +1 & +1 & -1 & -1 & +1 & -1 \\ 1 & +1 & +1 & +1 & +1 & +1 & +1 & +1 \end{bmatrix} \times \begin{bmatrix} \beta_I \\ \beta_o \\ \beta_r \\ \beta_a \\ \beta_{or} \\ \beta_{oa} \\ \beta_{ra} \\ \beta_{ora} \end{bmatrix}$$

The next part of the power analysis was to derive plausible population values for the model. For more basic analyses, this is typically based on standardized effect sizes, which are derived from prior research. Yet, since there are —to our knowledge— no proven standardized effect sizes for our context and analyses, we first used all information available from [Tucker and Ellis' \(1998\)](#) seminal study, and complemented it with information from [Costantini et al. \(2011\)](#) as well as [Arend and Schäfer \(2019\)](#) to create the population model. From [Tucker and Ellis' \(1998\)](#) study, we obtained the values for the general spatial alignment effect (without presentation of the avatar hands). They reported that left-hand responses were faster when the handles were also oriented towards the left ( $M = 628.2$  ms) than to the right ( $M = 638.8$  ms) hand; whereas right-hand responses were faster when the handles were oriented to the right ( $M = 627.3$  ms), as compared to the left ( $M = 639.8$  ms) hand. These values were used to define the population model for the without-avatar-condition ( $a = +1$ ) in Experiment 1, as well as the hypothesized spatial alignment effect in Experiment 2 (which was not expected to differ for avatar left [ $a = -1$ ] and right [ $a = +1$ ]).

Furthermore, H2 hypothesized that the spatial alignment effect would increase when an avatar was presented. To derive plausible values for the population model for this hypothesis, the study of [Costantini, Committeri, et al. \(2011\)](#), who examined the difference between a without-avatar and a with-avatar condition, was used. In their study, the difference between compatible and incompatible

handle orientations and response hands for reachable objects in the no-avatar condition ( $M = 399.8$  ms for compatible;  $M = 420.7$  ms for incompatible conditions) was smaller than in the with-avatar condition ( $M = 402.0$  ms for compatible;  $M = 430.9$  ms for incompatible conditions). Hence, the spatial alignment effect was  $(420.7 - 399.8) - (430.9 - 402.0) = 8$  ms bigger in the with-avatar condition (i.e., the difference between compatible and incompatible conditions was stronger with avatar). We included this value in our population model as follows: For the with-avatar conditions ( $a = -1$ ), 4 ms were subtracted for compatible orientation-response pairs (i.e., left-left and right-right or  $[o = r = -1]$  and  $[o = r = 1]$  were assumed to yield 4 ms faster responses) and 4 ms were added for the incompatible orientation-response pairs (i.e., left-right  $[o = -1, r = +1]$  and right-left  $[o = +1, r = -1]$  were expected to yield 4 ms slower responses). From these condition means, the coefficient values for  $\beta_b$ ,  $\beta_o$ ,  $\beta_r$ ,  $\beta_a$ ,  $\beta_{or}$ ,  $\beta_{oa}$ ,  $\beta_{ra}$ , and  $\beta_{ora}$  can be derived by taking the mean of the multiplication of the coefficient vector with the expected response vector of the population model (i.e.,  $Y_1, Y_2, \dots, Y_8$ ). The following unique design matrices (i.e., only for the fixed part) for Experiment 1 and Experiment 2 resulted.

$$\begin{array}{lcl}
 \text{Experiment 1:} & \begin{bmatrix} 624.2 \\ 642.8 \\ 643.8 \\ 623.3 \\ 628.2 \\ 638.8 \\ 639.8 \\ 627.3 \end{bmatrix} & = \begin{bmatrix} 1 & -1 & -1 & -1 & +1 & +1 & +1 & -1 \\ 1 & +1 & -1 & -1 & -1 & -1 & +1 & +1 \\ 1 & -1 & +1 & -1 & -1 & +1 & -1 & +1 \\ 1 & +1 & +1 & -1 & +1 & -1 & -1 & -1 \\ 1 & -1 & -1 & +1 & +1 & -1 & -1 & +1 \\ 1 & +1 & -1 & +1 & -1 & +1 & -1 & -1 \\ 1 & -1 & +1 & +1 & -1 & -1 & +1 & -1 \\ 1 & +1 & +1 & +1 & +1 & +1 & +1 & +1 \end{bmatrix} \times \begin{bmatrix} 633.525 \\ -0.475 \\ 0.025 \\ 0.000 \\ -7.775 \\ 0.000 \\ 0.000 \\ 2.000 \end{bmatrix} \\
 \\ 
 \text{Experiment 2:} & \begin{bmatrix} 628.2 \\ 638.8 \\ 639.8 \\ 627.3 \\ 628.2 \\ 638.8 \\ 639.8 \\ 627.3 \end{bmatrix} & = \begin{bmatrix} 1 & -1 & -1 & -1 & +1 & +1 & +1 & -1 \\ 1 & +1 & -1 & -1 & -1 & -1 & +1 & +1 \\ 1 & -1 & +1 & -1 & -1 & +1 & -1 & +1 \\ 1 & +1 & +1 & -1 & +1 & -1 & -1 & -1 \\ 1 & -1 & -1 & +1 & +1 & -1 & -1 & +1 \\ 1 & +1 & -1 & +1 & -1 & +1 & -1 & -1 \\ 1 & -1 & +1 & +1 & -1 & -1 & +1 & -1 \\ 1 & +1 & +1 & +1 & +1 & +1 & +1 & +1 \end{bmatrix} \times \begin{bmatrix} 633.525 \\ -0.475 \\ 0.025 \\ 0.000 \\ -5.775 \\ 0.000 \\ 0.000 \\ 0.000 \end{bmatrix}
 \end{array}$$

Furthermore, the random part of the model, containing the between-participant [ $\text{var}(\alpha_i^{\text{Pt}})$ ], between-object [ $\text{var}(\alpha_j^{\text{Ob}})$ ], the between-participant-per-object variances [ $\text{var}(\alpha_{ij}^{\text{Pt} \times \text{Ob}})$ ], the random slope variances [ $\text{var}(\alpha_i^{\text{Pt} \times \text{ORA}}$ ) for Experiment 1;  $\text{var}(\alpha_i^{\text{Pt} \times \text{OR}}$ ) for Experiment 2], and the error variance [ $\text{var}(\varepsilon_{ijk})$ ] had to be estimated. Tucker and Ellis (1998) reported an error variance of 634.0 ms (mean squared error, MSE; p. 835) in their seminal study (i.e., the L1 variance after accounting for the effects of all explanatory variables). Hence, we used this value as estimator for the error variance in the present study. Furthermore, for cross-classified multilevel designs, the population parameters of the variance components for the random effects need to be estimated. Judd et al. (2017) recommend using variance partitioning coefficients (with values between 0 and 1), that indicate the percentage of the outcome variable that can be attributed to each random factor. In accordance with the literature review by Arend and Schäfer (2019) who found that the mean L1-variance for studies with within-participant designs was 0.58 (Mdn = 0.60), we assumed that 60% ( $\text{VPC} = 0.60$ ) of the outcome variance resided at L1 (with 1% [ $\text{VPC} = 0.01$ ] attributable to the random slope included in both experiments; small random slope variance, see Arend & Schäfer, 2019). Hence, the L1 variance was composed of the variance explained by all the fixed effects, the error variance, and the random slope variance. Furthermore, we assumed that of the remaining 40% outcome variance, 25% ( $\text{VPC} = 0.25$ ) reside between participants, 10% ( $\text{VPC} = 0.10$ ) between objects, and 5% ( $\text{VPC} = 0.05$ ) between participant-per-object combinations. It should be noted that for these 40%, the values were somewhat arbitrarily set because there were no good estimators available. Yet, since these variance components do not contribute to the standard errors of the hypothesized fixed effects (which were all at the lower level, i.e., clustered in participant-object pairs), they should not affect their power.

Hence, from these values the full population models for Experiment 1 and 2 could be estimated. Monte Carlo simulation was performed for both experiments with  $N = 10,000$  replications. For the coefficients that corresponded to our hypotheses (H1, H2, H3), one-sided  $P$ -values were computed from Wald  $Z$ -tests. As usual for Monte Carlo simulation studies, statistical power was defined as the percentage of significant results from all replications.

For Experiment 1, power for the two-way interaction (H1) was 1.00 (i.e., all replications yielded significant results), before and after Bonferroni correction, whereas the power for the three-way interaction (H2) was 0.82 before and 0.73 after Bonferroni correction. For Experiment 2, power for the two-way interaction (H3) was 1.00 before and after Bonferroni correction. Appendix D provides the full code for the power analyses.

## Appendix B. Verbal instruction (Translated)

Hello! Thank you for participating in this experiment.

The whole experiment takes about half an hour and is divided into 10 blocks, between which you can always take a break.

The first and sixth block is a practice block so that you can familiarize with the task. These practice blocks will not be included in the later analyses but serve as a means for you to familiarize with the task.

In case there is a problem or a question, you can always call me.

After the first practice block (block 1), you can always go forward to the next block by pressing a button.

In case you make a mistake, there are two different error-message tones.

When you react too slow, it will beep briefly two times and when you react on the wrong side, it will beep one time longer.

In case you will react wrong and too slow, there will be both tones.

Now, please place your index fingers on the two buttons on the response keyboard.

Please, read the instruction.

A total of ten different objects will be presented to you.

Generally, the upright and inverted objects are easy to recognize. The hand broom is upright when presented according to its usage (bristles down) and it is inverted when presented contrary to its conventional use (bristles up).

## Appendix C. Written instruction (Translated)

Thank you for deciding to participate in our experiment.

[Please take the perspective of the avatar]

Please respond as fast as possible to an inverted object with a [left/right] keypress, to an upright object with a [right/left] keypress.

So:

inverted -> [left/right]

upright -> [right/left]

Please respond as fast and accurate as possible.

Continue with a keypress.

## Appendix D. Supplementary material

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.concog.2021.103133>.

## References

- Arend, M. G., & Schäfer, T. (2019). Statistical power in two-level models: A tutorial based on Monte Carlo simulation. *Psychological Methods*, 24, 1–19. <https://doi.org/10.1037/met0000195>.
- Azaad, S., Laham, S. M., & Shields, P. (2019). A meta-analysis of the object-based compatibility effect. *Cognition*, 190, 105–127. <https://doi.org/10.1016/j.cognition.2019.04.028>.
- Balcisoy, S., Torre, R., Ponder, M., Fua, P., & Thalmann, D. (2000). Augmented reality for real and virtual humans. In *Proceedings Computer Graphics International 2000*, Geneva, Switzerland (pp. 303–307). IEEE. <https://doi.org/10.1109/CGI.2000.852346>.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67. <https://doi.org/10.1109/10.18637/jss.v067.i01>.
- Bessiere, K., Ellis, J.B., and Kellogg, W.A. (2009). Acquiring a professional “second life”: Problems and prospects for the use of virtual worlds in business. In *CHI '09 Extended Abstracts on Human Factors in Computing Systems*. CHI 2009, Boston, Massachusetts (pp. 2883–2898). ACM. <https://doi.org/10.1145/1520340.1520416>.
- Blender Foundation (2018). Blender (v2.8) [Computer software]. Retrieved from <http://www.blender.org>.
- Böffel, C., & Müsseler, J. (2018). Perceived ownership of avatars influences visual perspective taking. *Frontiers in Psychology*, 9. <https://doi.org/10.3389/fpsyg.2018.00743>.
- Böffel, C., & Müsseler, J. (2019a). Action effect consistency and body ownership in the avatar-Simon task. *PloS One*, 14. <https://doi.org/10.1371/journal.pone.0220817>.
- Böffel, C., & Müsseler, J. (2019b). Visual perspective taking for avatars in a Simon task. *Attention, Perception, & Psychophysics*, 81, 158–172. <https://doi.org/10.3758/s13414-019-01690-y>.
- Böffel, C., & Müsseler, J. (2020a). No evidence for automatic response activation with target onset in the avatar-compatibility task. *Memory & Cognition*, 48(7), 1249–1262. <https://doi.org/10.3758/s13421-020-01052-2>.
- Böffel, C., & Müsseler, J. (2020b). Taking time to take perspective? Rapidly changing reference frames in the avatar-Simon task. *Acta Psychologica*, 204, 103005. <https://doi.org/10.1016/j.actpsy.2020.103005>.
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, 10(4), 433–436. <https://doi.org/10.1163/156856897X00357>.
- Buccino, G., Sato, M., Cattaneo, L., Rodà, F., & Riggio, L. (2009). Broken affordances, broken objects: A TMS study. *Neuropsychologia*, 47(14), 3074–3078. <https://doi.org/10.1016/j.neuropsychologia.2009.07.003>.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Claus, A. M., Arend, M. G., Burk, C. L., Kiefer, C., & Wiese, B. S. (2020). Cross-classified models in I/O psychology. *Journal of Vocational Behavior*, 120, 103447. <https://doi.org/10.1016/j.jvb.2020.103447>.
- Chen, J., Seibold, J. C., Zhong, Q., Müsseler, J., & Proctor, R. W. (2021). Is effector visibility critical for performance asymmetries in the Simon task? Evidence from hand- and foot-press responses. *Attention, Perception and Psychophysics*, 83(1), 463–474. <https://doi.org/10.3758/s13414-020-02205-w>.
- Cho, Y. S., & Proctor, R. W. (2005). Representing response position relative to display location: Influence on orthogonal stimulus-response compatibility. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, 58(5), 839–864. <https://doi.org/10.1080/02724980443000359>.
- Cho, D. T., & Proctor, R. W. (2010). The object-based Simon effect: Grasping affordance or relative location of the graspable part? *Journal of Experimental Psychology: Human Perception and Performance*, 36(4), 853–861. <https://doi.org/10.1037/a0019328>.
- Cho, D. T., & Proctor, R. W. (2011). Correspondence effects for objects with opposing left and right protrusions. *Journal of Experimental Psychology: Human Perception and Performance*, 37(3), 737–749. <https://doi.org/10.1037/a0021934>.
- Cho, D. T., & Proctor, R. W. (2013). Object-based correspondence effects for action-relevant and surface-property judgments with keypress responses: Evidence for a basis in spatial coding. *Psychological Research*, 77(5), 618–636. <https://doi.org/10.1007/s00426-012-0458-4>.
- Cole, G. G., & Millett, A. C. (2019). The closing of the theory of mind: A critique of perspective-taking. *Psychonomic Bulletin & Review*, 26(6), 1787–1802. <https://doi.org/10.3758/s13423-019-01657-y>.
- Costantini, M., Ambrosini, E., Scorolli, C., & Borghi, A. M. (2011). When objects are close to me: Affordances in the peripersonal space. *Psychonomic Bulletin & Review*, 18(2), 302–308. <https://doi.org/10.3758/s13423-011-0054-4>.
- Costantini, M., Ambrosini, E., Tieri, G., Sinigaglia, C., & Committeri, G. (2010). Where does an object trigger an action? An investigation about affordances in space. *Experimental Brain Research*, 207, 95–103. <https://doi.org/10.1007/s00221010-2435-8>.

- Costantini, M., Committeri, G., & Sinigaglia, C. (2011). Ready both to your and to my hands: mapping the action space of others. *PLoS One*, 6. <https://doi.org/10.1371/journal.pone.0017923>.
- Couth, S., Gowen, E., & Poliako, E. (2014). Dissociating affordance and spatial compatibility effects using a pantomimed reaching action. *Experimental Brain Research*, 232(3), 855–864. <https://doi.org/10.1007/s00221-013-3798-4>.
- 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. <https://doi.org/10.1037/a0031031>.
- Ellis, R., & Tucker, M. (2000). Micro-affordance: The potentiation of components of action by seen objects. *British Journal of Psychology*, 91, 451–471. <https://doi.org/10.1348/000712600161934>.
- Flavell, J. H. (1977). The development of knowledge about visual perception. *Nebraska Symposium on Motivation*, 25, 43–76.
- Gibson, J. (1977). The theory of affordances. In: R. Shaw & J. Bransford (Eds.) *Perceiving, acting and knowing: toward an ecological psychology* (pp. 67–82). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Grèzes, J., Tucker, M., Armony, J., Ellis, R., & Passingham, R. E. (2003). Objects automatically potentiate action: An fMRI study of implicit processing. *European Journal of Neuroscience*, 17(12), 2735–2740. <https://doi.org/10.1046/j.1460-9568.2003.02695.x>.
- Hommel, B. (2011). The Simon effect as tool and heuristic. *Acta Psychologica*, 136(2), 189–202. <https://doi.org/10.1016/j.actpsy.2010.04.011>.
- Hommel, B. (1993). Inverting the Simon effect by intention. *Psychological Research*, 55(4), 270–279. <https://doi.org/10.1007/BF00419687>.
- Inkpen, K.M., & Sedlins, M. (2011). Me and my avatar: exploring users' comfort with avatars for workplace communication. In *Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work. CSCW '11, Hangzhou, 2011* (pp. 383–386). ACM. <https://doi.org/10.1145/1958824.1958883>.
- Judd, C. M., Westfall, J., & Kenny, D. A. (2017). Experiments with more than one random factor: Designs, analytic models, and statistical power. *Annual Review of Psychology*, 68(1), 601–625. <https://doi.org/10.1146/annurev-psych-122414-033702>.
- Kessler, K., & Wang, H. (2012). Spatial perspective taking is an embodied process, but not for everyone in the same way: Differences predicted by sex and social skills score. *Spatial Cognition & Computation*, 12(2–3), 133–158. <https://doi.org/10.1080/13875868.2011.634533>.
- Kornblum, S., Stevens, G. T., Whipple, A., & Requin, J. (1999). The effects of irrelevant stimuli: 1. The time course of stimulus–stimulus and stimulus–response consistency effects with Stroop-like stimuli, Simon-like tasks, and their factorial combinations. *Journal of Experimental Psychology: Human Perception and Performance*, 25(3), 688–714. <https://doi.org/10.1037/0096-1523.25.3.688>.
- Lippa, Y. (1996). A referential coding explanation for compatibility effects of physically orthogonal stimulus and response dimensions. *Quarterly Journal of Experimental Psychology*, 49(4), 950–971. <https://doi.org/10.1080/17470219608419687>.
- Lippa, Y., & Adam, J. J. (2001). An explanation of orthogonal S-R compatibility effects that vary with hand or response position: The end-state comfort hypothesis. *Perception & Psychophysics*, 63(1), 156–174. <https://doi.org/10.3758/BF03200510>.
- Lo, S., & Andrews, S. (2015). To transform or not to transform: Using generalized linear mixed models to analyse reaction time data. *Frontiers in Psychology*, 6. <https://doi.org/10.3389/fpsyg.2015.01171>.
- Lopresti-Goodman, S. M., Turvey, M. T., & Frank, T. D. (2011). Behavioral dynamics of the affordance “graspable”. *Attention, Perception, & Psychophysics*, 73(6), 1948–1965. <https://doi.org/10.3758/s13414-011-0151-5>.
- 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. <https://doi.org/10.3758/BF03210959>.
- Müller, R. (2016). Does the anticipation of compatible partner reactions facilitate action planning in joint tasks? *Psychological Research*, 80(4), 464–486. <https://doi.org/10.1007/s00426-015-0670-0>.
- Müsseler, J., Wühr, P., & Ziessler, M. (2014). Using tools with real and imagined tool movements. *Frontiers in Psychology*, 5, Article 515. <https://doi.org/10.3389/fpsyg.2014.00515>.
- Müsseler, J., Ruhland, L., & Böffel, C. (2019). Reversed effect of spatial compatibility when taking avatar's perspective. *Quarterly Journal of Experimental Psychology*, 72(6), 1539–1549. <https://doi.org/10.1177/1747021818799240>.
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, 10(4), 437–442. <https://doi.org/10.1163/156856897X00366>.
- Petkova, V. I., Khoshnevis, M., & Ehrsson, H. H. (2011). The perspective matters! multisensory integration in ego-centric reference frames determines full-body ownership. *Frontiers in Psychology*, 2. <https://doi.org/10.3389/fpsyg.2011.00035>.
- Pfister, R., Dolk, T., Prinz, W., & Kunde, W. (2014). Joint response–effect compatibility. *Psychonomic Bulletin & Review*, 21(3), 817–822. <https://doi.org/10.3758/s13423-013-0528-7>.
- Pfister, R., Weller, L., Dignath, D., & Kunde, W. (2017). What or when? The impact of anticipated social action effects is driven by action–effect compatibility, not delay. *Attention, Perception, & Psychophysics*, 79(7), 2132–2142. <https://doi.org/10.3758/s13414-017-1371-0>.
- Poliako, E., Galpin, A., Dick, J., Moore, P., & Tipper, S. P. (2007). The effect of viewing graspable objects and actions in parkinson's disease. *NeuroReport*, 18(5), 483–487. <https://doi.org/10.1097/WNR.0b013e32805867a1>.
- Proctor, R. W., & Vu, K.-P.-L. (2006). *Stimulus-response compatibility: Data, theory, and applications*. Boca Raton, FL: CRC Press.
- Riggio, L., Iani, C., Gherri, E., Benatti, F., Rubichi, S., & Nicoletti, R. (2008). The role of attention in the occurrence of the affordance effect. *Acta Psychologica*, 127(2), 449–458. <https://doi.org/10.1016/j.actpsy.2007.08.008>.
- Shin, Y. K., Proctor, R. W., & Capaldi, E. J. (2010). A review of contemporary ideomotor theory. *Psychological Bulletin*, 136(6), 943–974. <https://doi.org/10.1037/a0020541>.
- Simon, J. R., & Berbaum, K. (1990). Effect of conflicting cues on information processing: The ‘Stroop effect’ vs. the ‘Simon effect’. *Acta Psychologica*, 73(2), 159–170. [https://doi.org/10.1016/0001-6918\(90\)90077-S](https://doi.org/10.1016/0001-6918(90)90077-S).
- Speelman, C., & McGann, M. (2013). How mean is the mean? *Frontiers in Psychology*, 4. <https://doi.org/10.3389/fpsyg.2013.00451>.
- Symes, E.D., Ellis, R., & Tucker, M. (2005). Dissociating object-based and space-based affordances. *Visual Cognition*, 12(7), 1337–1361. <https://doi.org/10.1080/1350628044000445>.
- Tipper, S. P., Paul, M. A., & Hayes, A. E. (2006). Vision for action: The effects of object property discrimination and action state on affordance Simon effects. *Psychonomic Bulletin & Review*, 13, 493–498. <https://doi.org/10.3758/BF03193875>.
- Tucker, M., & Ellis, R. (1998). On the relations between seen objects and components of potential actions. *Journal of Experimental Psychology: Human Perception and Performance*, 24(3), 830–846. <https://doi.org/10.1037/0096-1523.24.3.830>.
- Tucker, M., & Ellis, R. (2001). The potentiation of grasp types during visual object categorization. *Visual Cognition*, 8(6), 769–800. <https://doi.org/10.1080/13506280042000144>.
- Ulrich, R., & Miller, J. (1994). Effects of truncation on reaction time analysis. *Journal of Experimental Psychology: General*, 123, 34–80. <https://doi.org/10.1037/0096-3445.123.1.34>.
- Van der Heijden, A. H. C., Müsseler, J., & Bridgeman, B. (1999). On the perception of position. *Advances in Psychology*, 129, 19–37. [https://doi.org/10.1016/S0166-4115\(99\)80005-3](https://doi.org/10.1016/S0166-4115(99)80005-3).
- von Salm-Hoogstraeten, S., & Müsseler, J. (2021). Perspective taking while interacting with a self-controlled or independently-acting avatar. *Computers in Human Behavior*, 118, 106698. <https://doi.org/10.1016/j.chb.2021.106698>.
- von Salm-Hoogstraeten, S., Bolzius, K., & Müsseler, J. (2020). Seeing the world through the eyes of an avatar? Comparing perspective taking and referential coding. *Journal of Experimental Psychology: Human Perception and Performance*, 46, 264–273. <https://doi.org/10.1037/xhp0000711>.
- Wang, T.Y., Sato, Y., Otsuki, M., Kuzuoka, H., & Suzuki, Y. (2019). Effect of full body avatar in augmented reality remote collaboration. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE VR, 2019, Osaka (pp. 1221–1222). IEEE. <https://doi.org/10.1109/VR.2019.8798044>.
- Whelan, R. (2008). Effective analysis of reaction time data. *The Psychological Record*, 58(3), 475–482. <https://doi.org/10.1007/BF03395630>.