frontiers | Frontiers in Virtual Reality

TYPE Original Research
PUBLISHED 04 November 2022
DOI 10.3389/frvir.2022.976849

# Action-control mappings of interfaces in virtual reality: A study of embodied interaction

Martin Lachmair[1]*, Martin H. Fischer[2] and Peter Gerjets[3,4]

[1]Baden-Wuerttemberg Cooperative State University, Villingen-Schwenningen, Germany, [2]Division of Cognitive Sciences, University of Potsdam, Potsdam, Germany, [3]Leibniz-Institut für Wissensmedien, Tübingen, Germany, [4]LEAD Graduate School and Research Network, Tübingen, Germany

OPEN ACCESS

EDITED BY
Doug A. Bowman,
Virginia Tech, United States

REVIEWED BY
Andrea Stevenson Won,
Cornell University, United States
Mina C. Johnson-Glenberg,
Arizona State University, United States

*CORRESPONDENCE
Martin Lachmair,
lachmair@dhbw-vs.de

SPECIALTY SECTION
This article was submitted to Virtual Reality and Human Behaviour, a section of the journal Frontiers in Virtual Reality

RECEIVED 23 June 2022
ACCEPTED 24 October 2022
PUBLISHED 04 November 2022

CITATION
Lachmair M, Fischer MH and Gerjets P (2022), Action-control mappings of interfaces in virtual reality: A study of embodied interaction.
*Front. Virtual Real.* 3:976849.
doi: 10.3389/frvir.2022.976849

COPYRIGHT
© 2022 Lachmair, Fischer and Gerjets. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

The development of interface technologies is driven by the goal of making interaction more positive through natural action-control mappings. In Virtual Reality (VR), the entire body is potentially involved for interaction, using such mappings with a maximum of degrees of freedom. The downside is the increase in interaction complexity, which can dramatically influence interface design. A cognitive perspective on detailed aspects of interaction patterns is lacking in common interface design guidelines, although it can be helpful to make this complexity controllable and, thus, make interaction behavior predictable. In the present study, the distinction between grounding, embodiment, and situatedness (the GES framework) is applied to organize aspects of interactions and to compare them with each other. In two experiments, zooming into or out of emotional pictures through changes of arm span was examined in VR. There are qualitatively different aspects during such an interaction: i) perceptual aspects caused by zooming are fundamental for human behavior (Grounding: closer objects appear bigger) and ii) aspects of gestures correspond to the physical characteristics of the agents (Embodiment: little distance of hands signals little or, in contrast, "creating more detail"). The GES-framework sets aspects of Grounding against aspects of Embodiment, thus allowing to predict human behavior regarding these qualitatively different aspects. For the zooming procedure, the study shows that Grounding can overrule Embodiment in interaction design. Thus, we propose GES as a cognitive framework that can help to inform interaction guidelines for user interface design in VR.

KEYWORDS

embodied interaction, grounded cognition, virtual reality, action-control mapping, zooming, valence, user interface, embodiment

## Introduction

The development of interface technology is driven by the goal to provide a positive user experience. To achieve this goal, a common method is to simplify interfacing through access to functions and operations with natural gestures (Wigdor and Wixon, 2011). This is also reflected in human computer interface (HCI) guidelines for interface design in Virtual Reality (VR). Typically, such guidelines consist of heuristics for graphical user interfaces (Nielsen, 1994; Sutcliffe and Gault, 2004). For example, the first heuristic of VR guidelines describes that

> "Interaction should approach the user's expectation of interaction in the real world as far as possible. Ideally, the user should be unaware that the reality is virtual. Interpreting this heuristic will depend on the naturalness requirement and the user's sense of presence and engagement." (Sutcliffe and Gault, 2004, p. 833).

Such heuristics are mostly derived from arbitrary use cases to distill common features as a kind of best practice across not necessarily similar situations. At the same time, the technology of course is still developing with more and more sophisticated controllers, ranging from a one-handed gyroscope controller with three degrees of freedom (3-DOF; e.g., Garcia-Bonete et al., 2019) to both hands of the user with six degrees of freedom (6-DOF; e.g., Johnson-Glenberg, 2018). This has of course an impact on the appropriate design principles and maybe one reason why current guidelines resemble a patchwork rather than a systematic framework. In addition, with this practice only little is said about the underlying cognitive aspects and principles. However, we believe that developing those aspects and principles could widen the view to provide such guidelines with more structure. This view is shared by a recent study where fundamental HCI guidelines for VR were tested in two use cases: in an application for marine archaeology and a car driving simulation (Sutcliffe et al., 2019). In a user-centered design process, the interfaces were designed along the experiences of the respective peer-group, in line with the aforementioned heuristic. Interestingly, the study notes that cognitive psychological background knowledge receives little consideration in the design process for interfaces in VR, and further, that the guidelines for traditional graphical user interfaces are difficult to apply. The authors describe this as "a dilemma between presenting easy-to-assimilate HCI and giving designers sufficient background knowledge to interpret high-level heuristics" and state that "the lack of HCI/cognitive psychology background in the design teams was a major factor in restricting its influence (Sutcliffe et al., 2019, p. 10)." According to their experiences, the authors of the study derived a new heuristic for designing virtual environments (VE), which is as follows:

> "Natural action-control mapping: to maximize plausibility. Object manipulation and controls related to the user's task need to leverage natural affordances for any known interaction metaphors. Actions are realized *via* a variety of devices (e.g., data glove, 6- DOF controllers) mapped to VE [virtual environment] elements (hands, arms, and artifacts) combined with a choice of feedback modalities (audio, haptic, and visual). [. . .]" (Sutcliffe et al., 2019, p. 9).

This heuristic is better suited for VR by acknowledging a natural action-control mapping. However, from a cognitive perspective, this formulation suggests a high level of situational flexibility for mapping natural actions with arbitrary control, allowing for a high degree of freedom in designing user interfaces for such a complex environment as a virtual space in VR. As such, this flexibility is also a potential source of ambiguity, leaving doubts that the guidance of the given heuristic is sufficient as we will show below.

Referring to the known interaction metaphors in Sutcliffe and colleagues' new heuristic, a great example is the pinch-gesture in the touchscreen world. It provides seemingly natural access to zooming functions through opening and closing gestures with thumb and index finger (e.g., Bay et al., 2013) and is today implemented on all touch-sensitive devices like smartphones or tablets. In the immersive virtual world, however, the pinch-gesture has typically another function, for example to confirm input. For the zooming function, other action-control mappings are in use, such as using the stick on the controller or sliders ranging from "+" to "-", but also using controllers or hands to first grasp an object and then increasing or decreasing its size by moving arms. In all these cases, the magnitude of the zooming function is not restricted to a given screen size or the size of a single hand. Instead, it is limited, for instance, by the maximum arm separation between the two hands of the user in the latter action-control mapping. For example, imagine a museum context in VR: Here it is possible, in contrast to reality, to grasp a picture in an exhibition with both hands and to manipulate its size by using large arm movements so that opening one's arms enlarges the size of the picture and closing one's arms shrinks the size of the picture.

## Space-magnitude associations and zooming in immersive VR

From a psychological perspective, this kind of arm-gesture is bound to its functionality by following a similar cognitive principle as the pinch-gesture on touch-devices (cf. Mauney and Le Hong, 2010; Bay et al., 2013). For example, showing the size of an elephant is universally accompanied by opening the span of both arms to signals its sheer size by keeping the hands as far apart as possible. In contrast, showing the size of a bee is related to a closing-arm gesture or a small separation between thumb and index finger of one hand, showing its small size by a small distance between the body parts used to signal size. Similarly, small or large numbers (such as 1 or 9 in the single digit range) are spontaneously associated with precision and power grasps, respectively (Andres et al., 2004; Andres et al., 2008; Terrizzi et al., 2019; see also Fischer and Campens, 2009; and Woodin et al., 2021, for recent evidence from politicians' gesturing about small and large quantities).

Relying on such universal space-magnitude associations is a general principle of human cognition that links sensorimotor experiences and mental concepts. By interacting with the real world, we acquire concrete knowledge that we can metaphorically generalize to related but more abstract concepts, such as those found in the digital world (cf. Lakoff and Johnson, 1980; Casasanto, 2014; Winter and Matlock, 2017). This cognitive mechanism of analogizing provides the foundation of the described mapping of zoom-functionality and arm-gesture in VR.

## Emotional dimensions during zooming-interaction

Given this fundamental cognitive mechanism, one could assume that this kind of action-control-mapping would support, or at least would not disturb, cognition while zooming in or out from any digital object (e.g., Hoggan et al., 2013). However, affective evaluations inform and support behavioral tendencies as well (cf. Bailey et al., 2016). Thus, the valence dimension of a stimulus must also be considered in interface design (Picard, 2000; cf. De Gelder, 2006). Consider again the museum context in VR: First, a picture with negative valence, such as the picture of a grieving mother, is viewed and its valence is instantly determined and mentally represented by the viewer. Then the viewer zooms into the picture: Zooming into the picture should increase its perceived negativity due to the visual illusion of moving closer to that picture. In contrast, zooming out from that picture should be preferred because it reduces its perceived negativity due to the visual illusion of getting away from the picture (cf. Bamford and Ward, 2008). Conversely, people should prefer to approach rather than avoid pleasant stimuli. Given the supportive character of zooming gestures and their functional binding as described above, one would expect an approach-avoidance effect to be related to positive and negative evaluations of pictures when zoomed in compared to zoomed out, consistent with results of studies concerning approach-avoidance behavior (e.g., Chen and Bargh, 1999; Zech et al., 2020).

## The ambiguity of action-control-mappings

But would the same effect be expected with other functional mappings - that is, for instance, when bringing the hands together would magnify a picture and separating them would shrink a picture? Note that this kind of mapping could be a plausible option for an interface where zooming-in means to dive into more detail of a selected part of a picture by bringing one's hands together. Moreover, the visual effect of zooming in and out is then still the same; only the mapping between the zoom gesture and the visual effect is different. In other words, there is an interaction between the visual and the motor system.

The important point here is that Sutcliffe et al. (2019) heuristic gives no orientation regarding the ambiguity of the action-control-mapping since this mapping should not be of cognitive relevance. And indeed, many psychological studies that have investigated approach-avoidance behavior suggest a certain flexibility and context sensitivity of the preferred mapping (e.g., Bamford and Ward, 2008; Cannon et al., 2010; Carr et al., 2016; Cervera-Torres et al., 2019). Therefore, it is not clear how to derive systematic predictions regarding approach-avoidance behavior. This is especially true for heuristically motivated guidelines referring to interface design in VR. Interestingly, a closer look suggests hierarchically arranged components contributing to approach-avoidance behavior (cf. Winkielman et al., 2018). But for testable predictions, a more elaborate theoretical framework is needed that systematically relates the participating cognitive and sensori-motor components. We describe such a hierarchical framework in the next section.

## The GES-Framework and metaphorical ambiguity

In a series of papers, Fischer and colleagues (Fischer and Brugger, 2011; Pezzulo et al., 2011; Fischer, 2012; Pezzulo et al., 2013; Myachykov et al., 2014) described the GES framework for understanding cognition. GES is an abbreviation for the synthetic words Grounding, Embodiment and Situatedness. The purpose of this framework is to introduce a unified terminology that helps structuring cognitive processes hierarchically according to their fundamentality and flexibility. More specifically, grounding refers to fundamental cognitive processes that reflect physical laws such as gravity, object impermeability or causality. Such processes are rather stable, contribute directly to human survival and are therefore not only part of our early cognitive repertoire, but also difficult to unlearn or to relearn (cf. Lachmair et al., 2019). Embodiment, a term that is increasingly used as an umbrella term with a wide variety of meanings (in VR for example, it describes the ability to change one's character or perspective; e.g., Slater, 2017), is used in the GES framework in a narrower sense, specifically describing cognitive processes that take individual physical characteristics of an actor into account and attempt to exploit its advantages. These processes are inherently more flexible and variable across people, as they must adapt to physical characteristics of their bodies (e.g., right-handedness). Finally, situatedness addresses the issue that situational dependencies strongly influence cognitive processes because the purpose of those processes is to solve the task at hand in the current situation. Situated processes are even more flexible and also more powerful than grounded or embodied processes, due to the pressure to rapidly adapt to specific and arbitrary situations (cf. Myachykov et al., 2014). For example, situation-specific signals can achieve dominance over other factors, as in the higher relevance given to objects near the body (Abrams et al., 2008; see also Sui and Humphreys, 2015, for the case of self relevance).

Here, we focus on grounding and embodiment alone since we aim at identifying principles for HCI design in the context of VR that generalize across arbitrary situations. Our proposed hierarchy postulates the pervasiveness of grounding over embodiment. This is, for example, reflected in the universal association "more is up", according to which larger object accumulations must be taller as a result of object impermeability. On the other hand, habitual sensorimotor experiences can occasionally overrule these associations, as in the cultural convention of assigning smaller numbers to higher places in league tables (Holmes and Lourenco, 2011; Fischer, 2012; Lachmair et al., 2019). Nevertheless, GES predicts that this type of design decisions that are in conflict with the grounding principle of cognition impose unnecessary cognitive conflict and effort, thereby generally reducing performance.

Returning to the initial zooming example of this study, we were interested to set principles of grounding against principles of embodiment regarding two different action-control-mappings. GES allows a prediction of specific contributions from visual approach-avoidance and from gestural activities to the observable cognitive effect of valence perception. More specifically, we expected an approach-avoidance effect resulting in more extreme evaluations when positive and negative pictures are zoomed in compared to zoomed out with a grounded mapping (a closer object appears bigger and an open gesture signals more; grounding-hypothesis). We also expected, however, that this effect is affected when using a mapping based on a plausible metaphor "creating more detail" with a closing gesture (embodiment-hypothesis). Given that, according to GES, grounding-constraints on cognition should be more potent than embodiment-constraints, we also expected that the grounding-pattern would merely be diluted, but not reversed with a mapping as for the embodiment-hypothesis. If so, this would help to solve the ambiguity towards an effective user-interface design.

# Methods

We conducted two experiments. In the first experiment, we tested the grounding-hypothesis in VR. In the second experiment the embodiment-hypothesis was tested. A further post-hoc analysis with the data of both experiments together tested the dominance of grounding over embodiment.

# Experiment 1

In this experiment, participants zoomed into or out of high- or low-valence pictures. Subsequently, these pictures were rated on a scale from very negative to very positive. The zooming procedure was performed with a grounding-mapping of zoom (closer object is bigger) and gesture (large distance between hands signals more), i.e., an opening-gesture to zoom in and a closing-gesture to zoom out. We expected to find an approach-avoidance effect, showing more negative evaluations for negative pictures and more positive evaluations for positive pictures after zooming into compared to zooming out of the presented pictures.

## Participants

The sample consisted of 24 participants who were acquired with a database with students and other participants (14 females, mean age = 26.33 years, SD = 12.23 years, 1 left-handed). They all had normal or corrected vision. All participants were naive regarding the research question and gave written consent. They had also the option to delete their data until fourteen days after acquisition. As expense allowance, participants were paid 8 EUR. The experiment was approved by the local ethics committee of the Leibniz-Institut für Wissensmedien Tübingen (LEK 2016/039).

## Experimental setup and material

The experiment was conducted under Unity, a game framework that supports the use of Virtual Reality (VR) systems, on a high-performance gaming desktop PC with dedicated graphic card (Nvidia gtx 980ti). As VR system, we used an HTC Vive headset (i.e., a head-mounted display or HMD), including its two hand-held controllers and its tracking system. The affective pictures we used were part of the international affective picture system (IAPS; Lang et al., 1997). For this study we used the same pictures as the recent study by Cervera-Torres et al. (2020). Participants were presented with 20 positive and 20 negative pictures which were controlled for arousal. An ANOVA on the pictures' valence means confirmed significant differences between the valence categories ($M_{positive}$ = 7.22, $SD_{positive}$ = 0.53; $M_{negative}$ = 2.77, $SD_{negative}$ = 0.53; $F$ (1, 38) = 595, $p < 0.001$). Pictures' arousal means, on the contrary, did not show significant differences ($M_{positive}$ = 4.86, $SD_{positive}$ = 0.39; $M_{negative}$ = 5.03, $SD_{negative}$ = 0.49; $F$ (1,38) = 1.46, $p$ = 0.15). All pictures had the same size (width = 1,024 pixels, height = 768 pixels) with a resolution of 72 dots per inch.

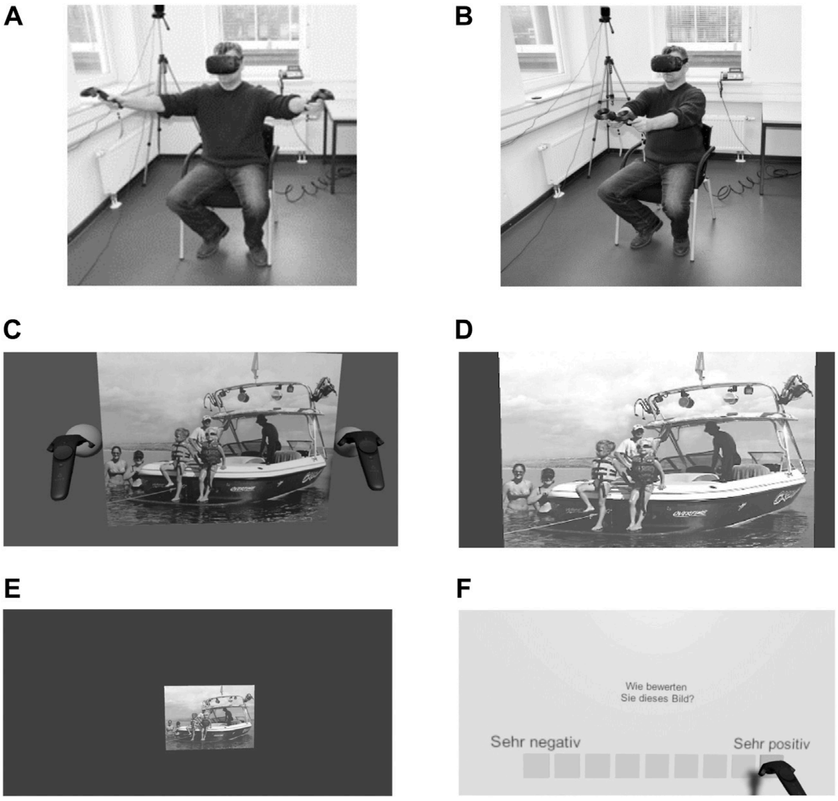

FIGURE 1
VR Setup. **(A)** Open-Arms Gesture; **(B)** Closed-Arms Gesture; **(C)** The controllers touch the 3D-knobs near the colored 2D-picture immediately before performing the Zoom Gesture; **(D)** The picture after zooming in; **(E)** The picture after zooming out; **(F)** The rating panel after performing the Zoom Gesture.

## Procedure

First, participants rated the pictures' valence and arousal in an on-line survey. Afterwards, the HMD was strapped to their head and the controllers were handed over. The participants had enough space to perform full-span arm movements with a controller held in each hand. In VR, the participant was in an empty space without visuo-spatial references and without a bodily representation, except the instruction texts and the presented pictures.

Next, a calibration procedure was conducted to find the optimal distance and zooming range for the pictures regarding arm length. Therefore, the arms were stretched to the front, followed by pulling the triggers on both controllers. Then the arms were moved to each side, left arm to the left and right arm to the right, again pulling the triggers. Finally, the arms were closed again until touch, confirmed with the triggers (cf. Figures 1A,B).

Afterwards, participants read the instructions on a text display and performed one of two zooming actions in separate blocks on pictures with a subsequent valence and arousal evaluation following each action. In one block the participants performed a "zoom in"-action by opening their arms whereas in the other block participants performed a "zoom-out"-action by closing their arms. The block order was counterbalanced across participants. The position of the picture was centrally in front of the participant, with a calibrated distance according to individual arm length (see above). Two knobs were presented half-way along the left and right edges of the picture, respectively. The initial coordinates of these knobs were in both blocks and for each picture the same, resulting in a constant viewing angle of 90 degrees for each participant and picture at the beginning of each trial (cf. Figure 1C). The knobs were the interaction points for the Vive controllers; once the knobs had been touched, a short vibrating feedback was given to the participant. Then, they pulled and held their index fingers on the triggers of both controllers and performed the zoom gesture to a maximum or minimum size to either enlarge or shrink the picture (cf. Figures 1D,E, respectively). An endpoint was not prescribed. After releasing the index fingers from the triggers, the picture disappeared, and the evaluation procedure began. For each of the recently manipulated pictures, its valence was evaluated on a Likert-scale from 1 (very negative) to 9 (very

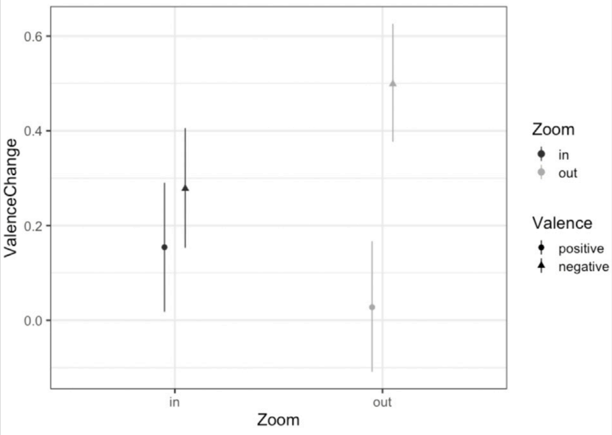

FIGURE 2
Mean Values of Valence Change as a function of Zoom and Valence. Bars represent the 95% Confidence Interval. The "Zoom in" condition led to a positive valence change of ~0.3 for the negative pictures and a positive valence change of ~0.15 for positive pictures, while the "Zoom out" condition resulted in an increase in positive valence of 0.4 for negative pictures and almost no change for positive pictures.

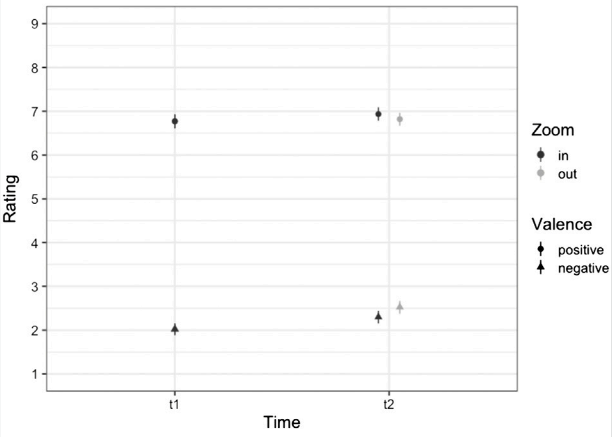

FIGURE 3
Mean Values of Ratings as a function of Time (t1 = pretest, t2 = posttest), Zoom and Valence. Bars represent the 95% Confidence Interval. The valence for the negative pictures increased in both the "Zoom in" and "Zoom out" conditions but increased more in the "Zoom out" condition. For the positive pictures, in the "Zoom in" condition valence increased at t2 compared to t1, but for "Zoom out" it remained about the same.

positive) and then its arousal also from 1 (not at all) to 9 (very much). The two scales were represented as horizontally arranged rectangular boxes below the questions "How do you evaluate the valence of this picture?" and "How do you evaluate the arousal of this picture?", respectively (see Figure 1F). The entire experiment lasted 60 min, including a break between the first evaluation and the VR procedure and after the first block in VR.

## Design

The 40 pictures were presented randomly during the 2 counterbalanced blocks, resulting in a total number of 80 trials per participant. The study was a full factorial 2 × 2 design with the within-factors "zooming gesture" (in vs. out) and "valence-category of the pictures" (positive vs. negative). As dependent variable, we analyzed the change of valence-ratings as the difference between the ratings of each picture after performing the zooming gesture (Time = t2) and the ratings of each picture of the initial rating procedure (Time = t1).

## Results

All analyses were conducted with R (R core team, 2020) and the packages "lmer" and "lmerTest" for obtaining *p*-values. Two participants were excluded for performing the wrong gesture in the second block. Another two participants were excluded due to high error-rates (> 20%). Poorly performed zooming gestures were also excluded from further analysis (< 1.88%). With the remaining data we conducted analyses of variance (ANOVA) with the random factor "participant" and the fixed factors "zooming gesture" (in vs. out), "valence-category of the pictures" (positive vs. negative) and their interaction. As dependent variable, we analyzed the change of valence-ratings.

First, no main effects were found (both *Fs* < 1.92, both *ps* > 0.18). However, the results showed a significant interaction between the zoom gesture and the valence-category of pictures [*F* (1,19) = 9.40, *p* = 0.006]. This interaction is displayed in Figure 2. Please note, we also displayed the ratings of the pre-test (t1) and the post-test (t2) in Figure 3 as a direct reference to the valence change. Accordingly, the results show first that the post-test showed less negative ratings for negative pictures and quite similar ratings for positive pictures compared to the pre-test ratings. This is in line with studies concerning habituation. For example, Dijksterhuis and Smith (2002) found in their study that extreme positive and negative stimuli became less extreme after repeated presentation. They also showed that this effect is often not symmetrical for both valence dimensions. Leventhal et al. (2007) emphasize that habituation even takes place after the first presentation of the stimuli. Both works describe the effect of habituation as crucial for human behavior to adapt to unexpected exposures to negative but also positive stimuli in our environment. It is even claimed that this behavior is fundamental for desensitization for example in therapy of phobia Dijksterhuis and Smith (2002). Against this background we interpret the results of Experiment 1 that zooming into the pictures with a grounding-mapping intensifies their perceived valence compared to zooming out, relatively to an initial valence evaluation. We consider this as evidence for the grounding-hypothesis, according to which interactions with the implemented mapping are non-demanding and bring out established or grounded behavioral patterns. But so far, we cannot tell how the sensorimotor influences originating from the motor system were involved in this effect. To shed more light on this, we conducted a second experiment where grounding (visual perception of approach and avoidance) and embodiment (motor-driven judgment bias) were set against each other.

## Experiment 2

In this experiment, participants performed the same task as before but now with a mapping based on the metaphor "create more detail" with a closing arm gesture. Thus, they had to close their arms to zoom in and to open their arms to zoom out of the pictures. The GES framework predicts that the effect observed in Experiment 1 will be diluted, in line with the embodiment-hypothesis, but not reversed due to the dominance of the grounding-effect in Experiment 1.

### Participants

The sample consisted of 19 participants who were again acquired with a database containing students and other participants (12 female, mean age = 23.21 years, SD = 3.49 years, 2 left-handed). All particpants had normal or corrected vision and were naive regarding the research question and gave written consent. They had also the option to delete their data until fourteen days after acquisition. As expense allowance, participants were paid 8 EUR. The experiment was approved by the local ethics committee of the Leibniz-Institut für Wissensmedien Tübingen (LEK 2016/039).

### Experimental setup and material

The experimental Setup and the Material were the same as in Experiment 1.

### Procedure and design

The Procedure and the design were the same as in Experiment 1, except that this time zoom-in was performed with a closing-arms gesture and zoom-out with an open-arms gesture.

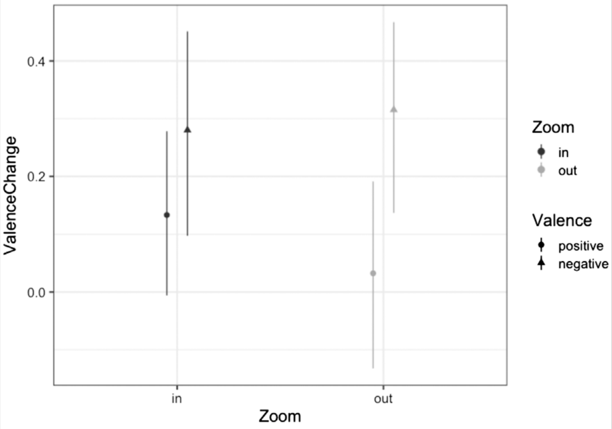

FIGURE 4
Mean Values of Valence Change as a function of zoom and valence categories of the pictures for Experiment 2. Bars represent the 95% Confidence Interval.

## Results

Data of two participants were excluded due to performing wrong arm movements in the second block of the experiment (i.e., 50% of the trials). Like in Experiment 1, incorrect response gestures were also excluded from further analysis (< 1%). With the remaining data, the valence change of picture evaluations was computed as in Experiment 1 and then again analyzed with ANOVAs.

The results, displayed in Figure 4, showed no main effects (*Fs* < 1, *ps* > 0.37). Interestingly, there was also no significant interaction between zoom and valence categories of the pictures [*F* (1,17) = 0.58, *p* = 0.46; cf. Figure 4]. Keeping in mind controversies regarding the confirmation of a null hypothesis using traditional statistical inference, we employed a Bayesian method. The method described in detail by Masson, 2011 enables calculating graded evidence for a null hypothesis (i.e., no difference between groups) and alternative hypothesis (i.e., difference between groups). In this analysis, the sum of squares and number of observations from ordinal analysis of variance (ANOVA) are used to calculate Bayesian factors, which then can be used to calculate posterior probabilities. Based on our data, the posterior probability of the null hypothesis for a non-significant interaction was 0.75 (alternative hypothesis 0.25). Applying the criteria suggested by Masson, 2011; see also Raftery, 1995, this is positive evidence for the null hypothesis, assuming no difference between zoom gestures and valence-category of pictures.

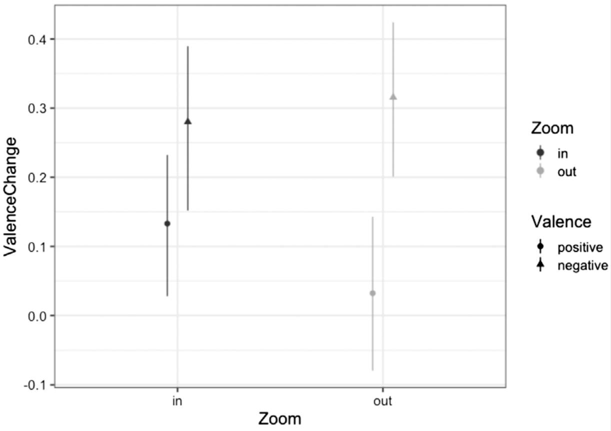

FIGURE 5
Mean Values of ValenceChange as a function of Zoom and Valence of the Post-hoc analysis. Bars represent the 95% Confidence Interval.

## Post-hoc analysis for experiment 1 + experiment 2

To further investigate the relation between grounding and embodiment, we conducted another post-hoc analysis on the data of Experiment 1 and 2 together by introducing the between-factor “Experiment”. According to the dominance of grounding over embodiment, we still expected a grounding effect, which means a significant interaction between zoom and valence category of the pictures. The results showed no main effect for Zoom [$F$ (1,37) < 1, $p$ = 0.73] nor for Valence [$F$ (1,37) = 2.76, $p$ = 0.11] or for Experiment [$F$ (1,36) < 1, $p$ = 0.65]. Also, the two-way interactions between zoom and experiment and between valence category of the pictures and experiment were not significant, as well as the three-way interaction between zoom, valence category and experiment (all $Fs$ < 1.54, all $ps$ > 0.22). But again, the interaction between zoom and valence category of the pictures *was* significant [$F$ (1,37) = 5.95, $p$ = 0.02; cf. Figure 5]. We consider this as further evidence for the grounding-hypothesis.

## Discussion

In this study we were concerned with general psychological principles of natural user interaction in VR. We tested the situated flexibility suggested by a recent heuristic formulated by Sutcliffe et al. (2019). The heuristic proposes a “natural action-control mapping” but does not consider deeper cognitive aspects related to stimulus valence. In two experiments conducted in VR, participants manipulated pictures with positive or negative valence by zooming in or out. We were interested in studying the foundation of this gesture-based user interaction by mapping action and control, i.e., large arm gestures and the zooming functions, in two different ways. In Experiment 1, zooming-in (a closer object is bigger) was mapped to an opening arm gesture (large distance between hands signals “more”) and zooming-out (object far away is smaller) to a closing arm gesture (small distance signals “little”). We argued that this kind of mapping follows general cognitive foundations (grounding principle). In Experiment 2, the mapping was reversed, following a rather metaphorical but also plausible mapping of zoom-in (“creating or diving into more detail”) with a closing arm gesture and zoom-out (“step back for overview”) with an opening arm gesture (embodiment principle). We asked whether these different mappings would affect basic approach-avoidance behavior of human actors.

According to the GES framework for cognition (Pezzulo et al., 2011; Fischer, 2012; Pezzulo et al., 2013; Myachykov et al., 2014), we formulated two hypotheses: First, according to the grounding-hypothesis, we should observe a classic approach-avoidance effect of valence evaluations in both experiments, since the visual zooming-in is associated with approach, resulting in a stronger change of valence evaluations; and the visual zooming-out with avoidance, resulting in a weaker change of valence evaluations. However, according to the embodiment-hypothesis the data should show an effect of mapping. This should lead to a reduced or diluted effect in Experiment 2. According to the GES framework, grounding should nevertheless dominate embodiment, preventing a complete reversal of the initially observed effect. This should be reflected when analyzing the data of both experiments together.

Our results support these considerations. First, the results of Experiment 1 showed a significant interaction between zoom direction and the valence category of the pictures: Approaching both negative and positive pictures induced subsequent valence evaluations with more negative ratings of negative pictures and more positive ratings of positive pictures, compared to initial evaluations. The reverse held for pictures that were moved away from the viewer. This confirms the grounding-hypothesis. Second, the results of Experiment 2 showed no such interaction after reversing the mapping of arm movements and zooming functions, thus confirming that embodied sensorimotor habits do influence the grounded mechanisms of cognition.

However, another explanation for this outcome might be that the unfamiliarity of the mapping distracts participants from observing details of the perceived picture causing the null-effect in Experiment 2. This seems plausible if we assume that most participants are more familiar with the mapping of Experiment 1 when considering the daily use of the pinch-gesture with two fingers on touch-devices. However, we do not consider this very likely for two reasons. First, although the gestures in this study with two arms are also strongly connected to the same metaphor than the pinch-gesture, it seems that they are physically different enough. Thus, we can assume that the mapping in this study is probably not overlearned according to usage in daily life. Second, looking at the procedure of the experiment, the picture was first displayed and then the participant had to detect and grasp the knobs on both sides of the picture with the controller. At this stage it is highly likely that the participant processed and recognized the relevant parts of the picture to determine its valence category as expected. This is supported by the main effect of valence category in the analysis of valence ratings of the pictures (cf. Figure 3). Thus, it is plausible to assume that, not distracting participants from observing details but rather the valence category of the pictures together with their movement through an unexpected action-control mapping affected the ratings afterwards causing the null-effect in Experiment 2.

Finally, a post-hoc analysis across the data of both experiments with an additional factor for experiment still showed a significant interaction between zoom and valence category of the pictures in line with the grounding-hypothesis, although no effect in favor of the embodiment-hypothesis.

The value of a theoretical framework like GES arises when compared to predictions of other cognitive frameworks. For example, the Theory of Event Coding (TEC) is a framework postulating that perception and action share a common cognitive representation. Accordingly, “perceived events (perceptions) and to-be-produced events (actions) are equally represented in the brain by integrated, task-tuned networks of feature codes, the so-called event codes” (Hommel et al., 2001, p. 849; cf. Hommel, 2015). Arguing along these lines would mean that the zoom-effect of the first experiment would also appear in the second experiment, because according to TEC the feature binding between action and perception does not depend on different experiential sources and thus, no cognitive conflict between the perceptual zooming and the performed gesture should appear. However, this is not supported by either the null-effect of Experiment 2 or the overall dominance of the grounding-hypothesis.

The null-effect of the second experiment can be explained within the framework of GES in analogy to the reasoning of cf. Wood et al. (2006); Lachmair et al. (2019) about conflicting experiential reference frames. In a tone discrimination task with cello players and non-musicians, Lachmair and colleagues found that the association of high-pitch of tones with higher spatial positions and low-pitch of tones with lower spatial positions is eliminated, not reversed, for cello players in the context of playing tones with cello timbre, but not in the context of piano timbre and not for non-musicians. The authors argued that two different reference frames are activated simultaneously, one referring to the general experiences of high pitch with higher positions and low pitch with lower positions as in non-musicians (a grounded experience, cf. Parise et al., 2014) and an opposing one referring to the experiences of well-practiced cello players (low pitch—high position, high pitch—low position, an embodied experience). Applied to the interpretation of the null-effect in Experiment 2 of the present study, the approach-avoidance effect in the zoom procedure would activate a grounded reference frame along general cognitive principles. The mapping of gesture and zooming would activate an opposing embodied reference frame along the metaphorical mapping of creating or diving into more detail. These conflicting reference frames could have caused the elimination of the zoom-effect. Nevertheless, when putting data of both experiments together the post-hoc analysis showed that the grounded zoom-effect prevailed. This is in line with the prediction of GES, postulating the dominance of grounding over embodiment.

Similar effects of conflicting reference frames might also arise with other VR interaction patterns, for example teleporting. The beam of a teleport is often visualized by an arc. The arc is ascending at the beginning. Following PC guidelines, teleporting further away is often bound to pulling the lever of the joystick, i.e., a movement towards the agent’s body (e.g., in flight simulators). I.e., a position further away in virtual space would be associated with a movement of the arm or stick on the VR controller towards the agent’s body to increase the height of the arc for a longer teleporting distance. Thus, the reasoning behind this mapping is not “teleport me far away”. It is rather “bring this far place nearer to me”. This ambiguity could lead to potential cognitive conflict for example with a guiding system of a VE that might affect the decision of the agent, where to teleport (Norouzi et al., 2021). In line with the approach-avoidance reasoning in the present study, it is plausible to assume, that interesting but positive connoted places or objects have a higher probability to be visited with this kind of action-control mapping compared to interesting but negative connoted places or objects. On the other hand, teleporting is, from a cognitive perspective, functionally fundamental different to zoom. While the first is about moving the agent through virtual space, the latter is about operating on an object. Here, we have further aspects that could play important roles, like aspects of hand position near the object (e.g., Brockmole et al., 2013) or hand dominance (e.g., Cervera-Torres et al., 2020). However, these should be investigated in further studies.

Taken together, these important insights have general implications for guidelines of interface design in VR. The practice of deriving heuristics from certain use cases of interaction is of course valid and valuable but does not consider deeper cognitive mechanisms. This practice supports the view of arbitrary flexibility and context sensitivity of cognitive processes during user interaction. But not considering the limits of this flexibility bears a potential source for ambiguity, lowering the effectiveness of user interface designs. The present study shows indeed that this flexibility has its limits and is rather relied on and restricted by fundamental cognitive aspects due to physical properties and laws. The approach-avoidance paradigm is such a fundamental cognitive principle and thus need to be considered. GES allows a systematic prediction of that paradigm’s impact on user interaction and clearly recommend following the grounding principle when zooming into or out from objects with a clear emotional content.

Thus, we propose to consider GES in future guidelines of user interface design in VR, since user interfaces can be built with high degrees of freedom and, moreover, in contexts with just a small or even without metaphorical reference to reality. In such cases, what remains are the characteristics of the human body together with general cognitive principles stemming from physical laws that might potentially interact. So, a closer look at these principles from a cognitive perspective might help to give general orientation and to supplement existing guidelines. In this regard, GES as a cognitive framework can help to inform interaction guidelines for user interface design in VR for a better cognitive performance during interaction.

## Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.researchgate.net/publication/346034043_Data_ZoomVR.

## Ethics statement

The studies involving human participants were reviewed and approved by local ethics committee of the Leibniz-Institut für Wissensmedien Tübingen (LEK 2016/039). The patients/participants provided their written informed consent to participate in this study.

## Author contributions

ML and PG contributed to conception and design of the study. ML analyzed the data. MF wrote sections of the manuscript. All authors contributed to manuscript, read, and approved the submitted version.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## References

Abrams, R. A., Davoli, C. C., Du, F., Knapp, W. H., III, and Paull, D. (2008). Altered vision near the hands. *Cognition* 107 (3), 1035–1047. doi:10.1016/j.cognition.2007.09.006

Andres, M., Davare, M., Pesenti, M., Olivier, E., and Seron, X. (2004). Number magnitude and grip aperture interaction. *Neuroreport* 15 (18), 2773–2777.

Andres, M., Ostry, D. J., Nicol, F., and Paus, T. (2008). Time course of number magnitude interference during grasping. *Cortex* 44 (4), 414–419. doi:10.1016/j.cortex.2007.08.007

Bailey, J. O., Bailenson, J. N., and Casasanto, D. (2016). When does virtual embodiment change our minds? *Presence. (Camb).* 25 (3), 222–233. doi:10.1162/pres_a_00263

Bamford, S., and Ward, R. (2008). Predispositions to approach and avoid are contextually sensitive and goal dependent. *Emotion* 8 (2), 174–183. doi:10.1037/1528-3542.8.2.174

Bay, S., Brauner, P., Gossler, T., and Ziefle, M. (2013). "Intuitive gestures on multi-touch displays for reading radiological pictures," in *International conference on human interface and the management of information* (Berlin, Heidelberg: Springer), 22–31.

Brockmole, J. R., Davoli, C. C., Abrams, R. A., and Witt, J. K. (2013). The world within reach: Effects of hand posture and tool use on visual cognition. *Curr. Dir. Psychol. Sci.* 22 (1), 38–44. doi:10.1177/0963721412465065

Cannon, P. R., Hayes, A. E., and Tipper, S. P. (2010). Sensorimotor fluency influences affect: Evidence from electromyography. *Cognition Emot.* 24 (4), 681–691. doi:10.1080/02699930902927698

Carr, E. W., Rotteveel, M., and Winkielman, P. (2016). Easy moves: Perceptual fluency facilitates approach-related action. *Emotion* 16 (4), 540–552. doi:10.1037/emo0000146

Casasanto, D. (2014). "Experiential origins of mental metaphors: Language, culture, and the body," in *The power of metaphor: Examining its influence on social life*. Editors M. Landau, M. D. Robinson, and B. P. Meier (American Psychological Association), 249–268.

Cervera Torres, S., Ruiz Fernández, S., Lachmair, M., and Gerjets, P. (2020). Coding valence in touchscreen interactions: Hand dominance and lateral movement influence valence appraisals of emotional pictures. *Psychol. Res.* 84 (1), 23–31. doi:10.1007/s00426-018-0971-1

Cervera-Torres, S., Ruiz Fernández, S., Lachmair, M., Riekert, M., and Gerjets, P. (2019). Altering emotions near the hand: Approach–avoidance swipe interactions modulate the perceived valence of emotional pictures. *Emotion* 21, 220–225. doi:10.1037/emo0000651

Chen, M., and Bargh, J. A. (1999). Consequences of automatic evaluation: Immediate behavioral predispositions to approach or avoid the stimulus. *Pers. Soc. Psychol. Bull.* 25 (2), 215–224. doi:10.1177/0146167299025002007

De Gelder, B. (2006). Towards the neurobiology of emotional body language. *Nat. Rev. Neurosci.* 7 (3), 242–249. doi:10.1038/nrn1872

Dijksterhuis, A. P., and Smith, P. K. (2002). Affective habituation: Subliminal exposure to extreme stimuli decreases their extremity. *Emotion* 2 (3), 203–214. doi:10.1037/1528-3542.2.3.203

Fischer, M. H., and Brugger, P. (2011). When digits help digits: Spatial-numerical associations point to finger counting as prime example of embodied cognition. *Front. Psychol.* 2, 260. doi:10.3389/fpsyg.2011.00260

Fischer, M. H., and Campens, H. (2009). Pointing to numbers and grasping magnitudes. *Exp. Brain Res.* 192 (1), 149–153. doi:10.1007/s00221-008-1622-3

Fischer, M. H. (2012). "The spatial mapping of numbers – its origin and flexibility," in *Language and action in cognitive neurosciences*. Editors Y. Coello and A. Bartolo (London: Psychology Press), 225–242.

Garcia-Bonete, M. J., Jensen, M., and Katona, G. (2019). A practical guide to developing virtual and augmented reality exercises for teaching structural biology. *Biochemistry and Molecular Biology Education* 47 (1), 16–24.

Hoggan, E., Nacenta, M., Kristensson, P. O., Williamson, J., Oulasvirta, A., and Lehtiö, A. (2013). “Multi-touch pinch gestures: Performance and ergonomics,” in Proceedings of the 2013 ACM international conference on Interactive tabletops and surfaces, St. Andrews, Scotland, (New York, NY: Association for Computing Machinery), 219–222.

Holmes, K. J., and Lourenco, S. F. (2011). “Horizontal trumps vertical in the spatial organization of numerical magnitude,” in *Proceedings of the annual meeting of the cognitive science society*. Editors L. Carlson, C. Hoelscher, and T. F. Shipley (California, CA: Open Access Publications from the University of California Merced), 33.33

Hommel, B., Müsseler, J., Aschersleben, G., and Prinz, W. (2001). The theory of event coding (TEC): A framework for perception and action planning. *Behav. Brain Sci.* 24 (5), 849–878. doi:10.1017/s0140525x01000103

Hommel, B. (2015). The theory of event coding (TEC) as embodied-cognition framework. *Front. Psychol.* 6, 1318. doi:10.3389/fpsyg.2015.01318

Johnson-Glenberg, M. C. (2018). Immersive VR and education: Embodied design principles that include gesture and hand controls. *Front. Robot. AI* 5, 81. doi:10.3389/frobt.2018.00081

Lachmair, M., Cress, U., Fissler, T., Kurek, S., Leininger, J., and Nuerk, H. C. (2019). Music-space associations are grounded, embodied and situated: Examination of cello experts and non-musicians in a standard tone discrimination task. *Psychol. Res.* 83 (5), 894–906. doi:10.1007/s00426-017-0898-y

Lakoff, G., and Johnson, M. (1980). The metaphorical structure of the human conceptual system. *Cognitive Sci.* 4 (2), 195–208. doi:10.1207/s15516709cog0402_4

Lang, P. J., Bradley, M. M., and Cuthbert, B. N. (1997). International affective picture system (IAPS): Technical manual and affective ratings. *NIMH Cent. Study Emot. Atten.* 1, 39–58.

Leventhal, A. M., Martin, R. L., Seals, R. W., Tapia, E., and Rehm, L. P. (2007). Investigating the dynamics of affect: Psychological mechanisms of affective habituation to pleasurable stimuli. *Motiv. Emot.* 31 (2), 145–157. doi:10.1007/s11031-007-9059-8

Masson, M. E. (2011). A tutorial on a practical Bayesian alternative to null-hypothesis significance testing. *Behavior research methods* 43 (3), 679–690. doi:10.1111/tops.12024

Mauney, D., and Le Hong, S. (2010). “Cultural differences and similarities in the use of gestures on touchscreen user interfaces,” in Proc. UPA Atlanta, GA: Usability Professionals Association.

Myachykov, A., Scheepers, C., Fischer, M. H., and Kessler, K. (2014). Test: A tropic, embodied, and situated theory of cognition. *Top. Cogn. Sci.* 6, 442–460. doi:10.1111/tops.12024

Nielsen, J. (1994). “Enhancing the explanatory power of usability heuristics,” in Proceedings of the SIGCHI conference on Human Factors in Computing Systems, , Boston, Mass., (Association for Computing Machinery, New York, NY, United States), 152–158.

Norouzi, N., Bruder, G., Erickson, A., Kim, K., Bailenson, J., Wisniewski, P., et al. (2021). Virtual animals as diegetic attention guidance mechanisms in 360-degree experiences. *IEEE Trans. Vis. Comput. Graph.* 27 (11), 4321–4331. doi:10.1109/tvcg.2021.3106490

Parise, C. V., Knorre, K., and Ernst, M. O. (2014). Natural auditory scene statistics shapes human spatial hearing. *Proc. Natl. Acad. Sci. U. S. A.* 111 (16), 6104–6108. doi:10.1073/pnas.1322705111

Pezzulo, G., Barsalou, L. W., Cangelosi, A., Fischer, M. H., Spivey, M., and McRae, K. (2013). Computational grounded cognition: A new alliance between grounded cognition and computational modeling. *Frontiers in psychology*, tier 2 (invited after frequent downloads) article. *Front. Psychol.* 22, 612. doi:10.3389/fpsyg.2012.00612

Pezzulo, G., Barsalou, L. W., Cangelosi, A., Fischer, M. H., Spivey, M., and McRae, K. (2011). The mechanics of embodiment: A dialog on embodiment and computational modeling. *Front. Psychol.* 2, 5. doi:10.3389/fpsyg.2011.00005

Picard, R. W. (2000). *Affective computing*. Cambridge, Mass.: MIT press.

Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological methodology* 111, 163

R Core Team (2020). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Available at: https://www.R-project.org/.

Slater, M. (2017). “Implicit learning through embodiment in immersive virtual reality,” in *Virtual, augmented, and mixed realities in education* (Singapore: Springer), 19–33.

Sui, J., and Humphreys, G. (2015). The integrative self: How self-reference integrates perception and memory. *Trends Cognitive Sci.* 19 (12), 719–728. doi:10.1016/j.tics.2015.08.015

Sutcliffe, A., and Gault, B. (2004). Heuristic evaluation of virtual reality applications. *Interact. Comput.* 16 (4), 831–849. doi:10.1016/j.intcom.2004.05.001

Sutcliffe, A. G., Poullis, C., Gregoriades, A., Katsouri, I., Tzanavari, A., and Herakleous, K. (2019). Reflecting on the design process for virtual reality applications. *Int. J. Human–Computer. Interact.* 35 (2), 168–179. doi:10.1080/10447318.2018.1443898

Terrizzi, B. F., Brey, E., Shutts, K., and Beier, J. S. (2019). Children’s developing judgments about the physical manifestations of power. *Dev. Psychol.* 55 (4), 793–808. doi:10.1037/dev0000657

Wigdor; and Wixon (2011). *Brave NUI world*. Amsterdam: Elsevier.

Winkielman, P., Coulson, S., and Niedenthal, P. (2018). Dynamic grounding of emotion concepts. *Phil. Trans. R. Soc. B* 373 (1752), 20170127. doi:10.1098/rstb.2017.0127

Winter, B., and Matlock, T. (2017). *Primary metaphors are both cultural and embodied*. Metaphor: Embodied cognition and discourse, 99–115.

Wood, G., Nuerk, H. C., and Willmes, K. (2006). Crossed hands and the snarc effect: A failure to replicate dehaene, bossini and giraux (1993). *Cortex* 42 (8), 1069–1079. doi:10.1016/s0010-9452(08)70219-3

Woodin, G., Winter, B., Perlman, M., Littlemore, J., and Matlock, T. (2021). Tiny numbers’ are actually tiny: Evidence from gestures in the TV News Archive. *PLoS ONE* 15 (11), e0242142. doi:10.1371/journal.pone.0242142

Zech, H. G., Rotteveel, M., van Dijk, W. W., and van Dillen, L. F. (2020). A mobile approach-avoidance task. *Behav. Res. Methods* 52 (5), 2085–2097. doi:10.3758/s13428-020-01379-3