Virtual Reality (2023) 27:1697–1715
https://doi.org/10.1007/s10055-023-00765-4

ORIGINAL ARTICLE

From virtual to physical environments when judging action opportunities: are diagnostics and trainings transferable?

Milena S. Gölz[1,2] · Lisa Finkel[1,2] · Rebecca Kehlbeck[3] · Anne Herschbach[1,2,4] · Isabel Bauer[1,2] · Jean P. P. Scheib[1,2] · Oliver Deussen[3] · Jennifer Randerath[1,2,5]

Received: 2 March 2022 / Accepted: 28 January 2023 / Published online: 15 February 2023
© The Author(s) 2023

Abstract

The proper evaluation of whether our given bodily capabilities and environmental properties allow particular actions is indispensable for pertinent decisions, so-called affordance judgments. These can be impacted by older age or brain damage. Virtual Environments (VEs) may provide an efficient opportunity to offer trainings. But do people make affordance judgments in VEs in the same way that they do in Physical Environments (PEs)? And are these decisions trainable by use of VEs? We investigated 24 healthy young adults' judgment performance of whether or not they could fit their hand into a given aperture. They were presented with a set of opening-increments and indicated their judgments by pressing a yes- or no-button. The stimuli were presented in PE using an aperture apparatus and in VE displayed by use of Oculus Rift goggles. Our results demonstrated the level of equivalence to be specific to the variable: While we found equivalence between VE and PE for the accuracy parameter, results were uncertain or non-equivalent for perceptual sensitivity and for judgment tendency, respectively. When applying training in VE, judgment accuracy improved significantly when tested subsequently within VE. Improvement appeared detectable in PE only on a descriptive level. Furthermore, equivalence testing post-training revealed that perceptual sensitivity performance in VE approached a PE-level. Promisingly, the VE training approach appeared applicable and efficacious within the VE. Future studies need to specify factors that enhance equivalence for detection theory variables and that facilitate transfer from VEs to PEs when judging action opportunities.

Keywords Virtual environments · Affordance judgment training · Equivalence testing · Detection theory · Affordances

1 Introduction

When moving purposefully in everyday life, we constantly are confronted with action opportunities. In such situations, we need to make decisions whether to pursue a certain action or not. For example, when grabbing something out of a drawer, the decision has to be made whether the hand fits through the opening. When crossing the street, it is important to accurately judge whether the street is passable. In this type of decisions, it is necessary to capture the interplay between environmental properties and an individual's own physical capabilities. Gibson (1977) defined the term *affordances* as the opportunities for action provided by the environment. The individual's ability to accurately perceive these opportunities for actions in relation to their capabilities is essential for interacting with the environment. The variety of affordances available to humans across different situations is vast. Therefore, researchers have investigated a wide range of settings and types of action, including,

Milena S. Gölz and Lisa Finkel have contributed equally to this work.

✉ Jennifer Randerath
J_Randerath@hotmail.com

1 Department of Psychology, University of Konstanz, Konstanz, Germany

2 Lurija Institute for Rehabilitation Science and Health Research, Allensbach, Germany

3 Department of Computer and Information Science, Centre for the Advanced Study of Collective Behavior, University of Konstanz, Konstanz, Germany

4 Department of Psychosomatic Medicine and Psychotherapy, Medical University Hospital Tuebingen, Tuebingen, Germany

5 Outpatient Unit for Research, Teaching and Practice, Faculty of Psychology, University of Vienna, Vienna, Austria

for example, reaching (Gabbard et al. 2006; Thomas et al. 2017; Wagman et al. 2019), stepping and leaping (Day et al. 2015) or fitting the hand into an aperture (Ishak et al. 2008; Randerath & Frey 2016). However, physical settings that study affordance judgment performance are very costly and elaborate. Virtual settings could be an alternative to investigate scenarios that may be impractical to implement in the physical world. Thus, several motor cognitive paradigms including sensorimotor illusions have been demonstrated to be realizable using a Virtual Environment (VE) design (e.g., Buckingham 2019). But, VE may impact basic motor-cognitive processes due to fewer binocular cues to depth, conflicting depth information and limited haptic feedback (Harris et al. 2019). In the present research, this raises the question of whether VE trainings can be an adequate replacement for physical world settings. In their review, Bohil et al. (2011) state that in applied domains such as rehabilitation, VE methods continue to accumulate validating results. As these methods become widely adopted, the studies become extended to different neuroscience areas and to a wider range of therapies. Software and hardware components become ubiquitous and VEs are increasingly viewed as an ordinary part of neuroscience research and therapy (Bohil et al. 2011). More than two decades ago, Waller, et al. (1998) examined the variables that mediate the transfer of spatial knowledge (maze environment) acquired in a VE to the implementation in a Physical Environment (PE) situation, using the concept of fidelity. Fidelity typically is defined by the extent to which behavior, observation and/or representation in the VE and PE are indistinguishable (Jerald 2015; Waller et al. 1998). It was shown that tasks requiring route knowledge presented in an immersive VE may surpass map training and be indistinguishable from training in the PE (Waller et al. 1998).

In their review, Jensen and Konradsen (2018) identified a number of situations in education and training in which head-mounted displays (HMDs) were useful for skill acquisition, including understanding spatial and visual information, knowledge and psychomotor skills related to head-movement. Accordingly, VE training has been applied in many fields: ranging from training in neurosurgery (Choudhury et al. 2013) to social skill development training (Howard & Gutworth 2020) or as a training tool in the mining industry (Zhang 2017). The study of affordance judgments in VE has the potential to expand the knowledge about how well the experience in VEs matches the one of the physical world (Creem-Regehr et al. 2019; Geuss, et al. 2010; Lin et al. 2015). Bliss et al. (1997) investigated the transfer of training from VE to PE in a spatial navigation task. The results showed that the VE and blueprint training groups were equally effective for teaching spatial navigation skills. In their review, Cooper et al. (2005) emphasized with regard to wheelchair users another advantage: VEs offer a training tool in varying risk-free environments without any indoor (e.g., walls, furniture, and stairs) and outdoor (e.g., curb cuts, uneven terrain, and street traffic) physical constraints. However, while being promising, not all variables appear transferable from VE to PE and vice versa. For example, previous study results suggested that people judge egocentric distances (from self to target) differently in PEs than in VEs (Geuss et al. 2010; Grechkin et al. 2010). According to Creem-Regehr et al. (2019), over the past 15 years, the body of research studying human performance in VEs has strongly grown including the study of affordances (Bhargava et al. 2020; Bodenheimer & Fu 2015; Creem-Regehr et al. 2019; Creem-Regehr et al. 2015; Geuss et al. 2010; Jun et al. 2015). For example, in a very recent study by Bhargava et al. (2020) passability judgments were compared between a VE and a PE using a between subjects design. Participants judged whether they could pass through various widths of a doorway. Half of them performed judgments in the PE condition and half of them in a VE condition. If uncertain of their ability to pass, participants were asked to walk towards the door until they were certain of their response. Results showed that, overall, participants in the VE did not differ from participants in the PE with respect to aperture passability judgments (measured by accuracy of judgments, and certainty of judgments). The authors conclude that the information which is needed to determine individual affordances (size and distance of the aperture relative to one's own geometric and dynamic properties) is available and salient in VE. However, the authors conceded that to reach a physical world level of judgment accuracy in VE, participants needed additional exposure to dynamic information in VE by walking closer to the door. Interestingly, the study mentioned that even though participants never received explicit feedback about the accuracy of their judgments, participants in the PE improved in accuracy over time, while participants in VE did not. Thus, with the aim to improve judgments, it remains unclear whether the environments may potentially deliver comparable results.

In physical settings, investigating training options to modulate and improve the capability of decision making based on affordances (affordance judgments) has gained relevance for rehabilitation purposes. While young healthy adults are able to make quick and adequate affordance judgments (Finkel et al. 2019b; Oxley et al. 2005; Zivotofsky et al. 2012), it has been demonstrated that affordance judgments can be altered due to sudden bodily changes with older age (Finkel et al. 2019b; Zivotofsky et al. 2012) or be impacted by brain damage such as stroke (Randerath et al. 2021, 2018). The existing research on training of affordance judgments demonstrated promising findings: In young adults, affordance judgments appeared trainable subsequent to action practice (Finkel et al. 2019a, b; Franchak et al. 2010; Randerath & Frey 2016). Franchak et al. (2010) for example found that participants benefited from action feedback in

judging whether doorways allowed passage. In our previous work we piloted a training approach with two studies, each testing a group of healthy young adults solving one PE task (Randerath & Frey 2016). In the first session, all participants solved a judgment task without executing the action (study 1: whether the hand could fit into an aperture (adapted from Ishak et al. 2008); study 2: whether an object was within reach (adapted from Gabbard et al. 2005). In the second session a training condition was introduced to half of the participants. These subjects were allowed to first judge and then perform the task for each trial and thereby received feedback (study 1: fitting the hand into the aperture; study 2: reach forward and touch the presented object). The other half of subjects solved the tasks without experience. Only participants obtaining experience demonstrated increased accuracy and improvements in discriminating a doable from a not doable action (perceptual sensitivity). A later training study by Finkel et al. (2019a) implemented the Aperture Task across different age groups. Participants judged whether their hand could fit into a given opening that was presented with varying sizes. The authors found an improvement when comparing performance in pre-training to post-training and after a one-week follow-up. Younger adults showed improved perceptual sensitivity and judgment accuracy. The trainability of affordance judgments within this short time appears promising for rehabilitation research.

Similar to our previous PE study (Randerath and Frey 2016) a recent training study presented at a conference investigated the trainability of affordance judgments for reachability in VE (Gagnon et al. 2021). Participants viewed targets that were farther or closer than their actual reachability and decided whether the target was within reach. In feedback trials, they judged and then reached out to the target to receive feedback. Participants received visual feedback from a handheld controller. Gagnon et al. (2021) found that reachability was initially overestimated, but became more accurate over feedback blocks in the VE.

Using an aperture paradigm, the present work examined the performance equivalence in a PE versus a VE condition for different performance variables: judgment accuracy as well as detection theory variables: perceptual sensitivity and judgment tendency (Fox 2004; Green and Swets 1966; Macmillan and Creelman 2004). Detection theory is a general psychophysical approach to measuring performance of decision processes. In our case it measures the skill in distinguishing a fit from a non-fit of one's hand in a given aperture (perceptual sensitivity), and it measures a particular setting of a decision threshold (criterion or bias or judgment tendency). Perceptual sensitivity captures the capability to discriminate between a fit and a non-fit (discriminability index d'). Judgment tendency reflects an individual's response bias (criterion c) that can for example be rather liberal (<0, responding with *yes it fits* in most trials) or conservative (>0, responding with *no it does not fit* in most trials) when deciding whether a certain action is possible (see section "2.1.4.1. Performance variables" in "2.1.4. Data analysis" section).

We further examined the potential effects of a VE training for judgment performance in the VE as well as in the PE setting.

Thus, the present investigation was divided into two parts using a within-subjects design: In study 1, we compared initial affordance judgment performance in a PE versus a VE condition. In study 2, we examined within the same participants whether an intervention with visual feedback in the virtual setting had an effect on the judgment performance in the VE and PE condition.

2 Study 1

In study 1, we compared initial affordance judgment performance in a PE versus a VE. Participants were asked to indicate a fit or a non-fit of their hand into a given aperture (Aperture Task). In the PE condition, a physical aperture apparatus with varying opening widths was presented. In the VE condition, the task was presented via a virtual setting using Oculus Rift goggles. Prior studies assessing affordance judgments in PEs and VEs have demonstrated similarly accurate judgments for example during a passability task or while judging crossable gaps (e.g., Bhargava et al. 2020; Creem-Regehr et al. 2019). Therefore, we expected judgment behavior to be equivalent between conditions. Since traditional comparisons fail to reject a no-difference hypothesis we instead applied equivalence testing (Da Silva et al. 2009; Mascha & Sessler 2011; Rogers et al. 1993). We used the TOST (two-one-sided-tests) procedure to test statistical equivalence (Lakens et al. 2018). For details, see section "2.1.4.2. Equivalence testing" in "2.1.4. Data analysis".

2.1 Study 1 methods

The methods and materials were mainly adapted from previous work (Randerath & Frey 2016) and were implemented in both, study 1 and study 2.

2.1.1 Participants

Participants were recruited by announcement and through the online recruitment software SONA. They were tested at the University of Konstanz. Based on the inclusion criteria, all participants were right-handed according to the Edinburgh Handedness Inventory (Salmaso & Longoni 1983), had normal or corrected-to-normal vision, and reported no history of psychiatric or neurologic disorders. Stereo vision

was tested by the Lang II Stereo card, a revised version of the Lang Stereotest (Lang 1983; Lang & Lang 1988).

In total, data sets of 24 individuals between 20 and 33 years of age ($M = 24.1$ years, $SD = 3.3$; 17 females) were analyzed across two study parts. Please note, for the equivalence testing (TOST procedure), statistical power (1-β) was computed post hoc using the web page "http://powerandsamplesize.com/" (HyLown Consulting LLC, 2013–2022). Power analysis for accuracy revealed a power (1-β) of > 0.99 [$N = 24$, type I error rate (α) = 5%, $\text{Mean}_{\text{Group 'A'}} = 74.69$, $\text{Mean}_{\text{Group 'B'}} = 70.53$, $SD = 9.87$, $\delta = 29.88$].

2.1.2 Material

The Aperture Task was performed in both a PE and a VE. In Fig. 1 both experimental settings are depicted.

2.1.2.1 Physical aperture condition (PE) The aperture apparatus used in the PE condition was the same as used in previous studies (e.g., Finkel et al. 2019a, b). It was custom-built by the scientific workshops at the University of Konstanz and made of PVC (black board: 1000 mm length × 850 mm height) and aluminum. Centrally placed, at participant's eye level, there was a rectangular opening. Its height and width could be manipulated in millimeter steps. During the experiment, the protocol-related trial adjustments of horizontal openings were regulated by a computer-controlled step motor. In order to prevent visual feedback when necessary and to control vision in the physical condition, participants wore Plato-goggles (Translucent Technologies Inc.) that could be switched between opaque and transparent. Judgments were indicated by pressing a specified "yes" or "no" button on a response button-pad (Cedrus, RB540, see Fig. 1, bottom figure).

2.1.2.2 Virtual aperture condition (VE) First, the virtual scene was modelled and visualized by use of Unreal Engine (EpicGames). The apparatus was modelled with standard objects considering the implementation of lighting and texture. Afterwards, the experimental sequence and the interaction with Cedrus devices were implemented.

In order to keep the participants' positioning, button presses and surrounding sounds (e.g., motor sound) consistent across conditions, participants sat at the same table with the physical apparatus and the button-pad (see Fig. 1B, bottom figure). They wore Oculus Rift Virtual Reality goggles instead of the Plato-goggles. By use of the Oculus Rift Software, standard calibration of the Occulus Rift goggles was executed utilizing remote control. To ensure that the opening in the VE condition was centered in front of the participant and precisely matched the physical opening, height and orientation of the apparatus were calibrated for each test person. Therefore, the apparatus with an equilateral rectangular opening and additionally an equally sized transparent rectangle were presented in Oculus Rift goggles. The participant was then asked to indicate verbally in which direction the rectangle should be moved (up/down/left/right/front/back), until opening and rectangle overlapped exactly. Based on the subject's verbal command, the shifting of the rectangle was adjusted by the investigator using the arrow keys on the computer keyboard. In addition, the participant was asked to palpate the apparatus in front of them with both hands to get used to the VE. The calibration process was carried out until the participant confirmed that haptic feedback matched the visual impression.

2.1.3 Task and procedure

Participants sat in a straight position at a defined distance (20 cm from belly to table) in front of a height-adjustable table. They either wore Plato-goggles or Oculus Rift goggles, depending on the condition. The apparatus was controlled automatically by use of the SuperLab software (provided by Cedrus).

2.1.3.1 Measurements Hand measurements were taken without vision to avoid visual feedback. Both in PE and VE, the session started out, with the measurement of the actual maximum width and height of the participants' hands (maximum fitting size of the opening). Participants put their flat hand with fingers closely spaced through the opening and hand size measurements were taken by closing the opening tightly around the hand's widest part. In the PE condition the Plato-goggles were shut, and its glasses turned into an opaque grey. In the VE condition a grey background was presented via the Oculus Rift goggles.

2.1.3.2 Aperture Task In the affordance perception task, subjects judged whether they could fit the widest part of their right hand through a given horizontal aperture located at eye level which varied in width between trials. The presented aperture sizes varied relative to the participants' actual hand width using nine fixed increments: -16, -8, -4, -2, $0, +2, +4, +8, +16$ mm. The 0-trial reflected the actual participants' hand width. In order to achieve a balance in the number of yes-trials (5 openings with $0, +2, +4, +8, +16$ mm) and no-trials (4 openings with -16, -8, -4, -2 mm), a corresponding number of filler trials per block was added. These filler trials presented smaller openings for which the answer would be "no" as well (-20, -30, -40 mm). In other words: to avoid an imbalance toward more frequent yes-trials, we added filler trials for which the correct answer would be "No" (smaller than—16 mm). Please note, because the negative filler trials are much narrower than the 0-trial, correct answers become more likely. For this reason, filler trials were excluded from



Fig. 1 Experimental setting in the PE (**A**) and the VE (**B**) condition **A** Experimental setting (PE condition) including aperture apparatus, Plato-goggles (here: transparent), and the response button-pad placed under a cover plate. **B** Experimental setting (VE condition) including aperture apparatus and Oculus Rift Virtual Reality goggles. The response button-pad is also placed under the cover plate here. At the bottom picture: Cedrus response button-pad used in both the PE and the VE condition

further analysis. Participants indicated their judgments by pressing a specified "yes" or "no" button on a button-pad (Cedrus, RB540). They were instructed to use their dominant right hand to make the decision. To withdraw additional information about the hand in the PE condition, participants had to place their hand under a cover plank (see Fig. 1) in both experimental conditions. Since a previous study (Randerath and Frey 2016) showed that a stable judgment tendency seems to be formed during the first few judgment trials, a familiarizing introductory block with 20 trials (2×9 openings and 2 filler trials) was added for each condition and each study. An experimental block with the same

environmental condition with 30 trials (3×9 openings, 3 filler trials) followed.

In study 1, we aimed to examine whether the participants' initial performance in the Aperture Task measured by accuracy, perceptual sensitivity, and judgment tendency was equivalent in the PE compared to the VE condition. The study lasted approximately one hour. In this study, participants performed 6 blocks of the Aperture Task. Trials were blocked per environmental condition. To familiarize with both the VE and PE condition, participants started with an introductory (20 trials) and a subsequent experimental block (30 trials) with one randomly assigned condition. Afterwards, an introductory and an experimental block were presented with the respective alternative condition. Half of the group ($N = 12$) started with the PE condition, the other half with the VE condition ($N = 12$). After the two familiarizing blocks, participants performed another experimental block (30 trials) per environmental condition. The latter was used for data analysis. The exact procedure is shown in Fig. 2.

2.1.4 Data analysis

Experimental data were coded with SuperLab 5 Software. Behavioral data were analyzed with SPSS Statistics 27 (IBM).

2.1.4.1 Performance variables Firstly, general judgment accuracy (percent of correct judgments) was assessed. Second, to enable a more precise interpretation of participants' judgment behavior, detection theory variables were additionally calculated: False-alarm rates were determined by the ratio of the number of negative events wrongly categorized as positive (False-alarms, i.e., indicating "yes" in trials the hand actually does not fit through the given opening) and the total number of actual negative events (i.e., total amount of "no" trials with openings smaller than 0). The additionally calculated Hit rates depicted the ratio of the number of positive events successfully categorized as positive (Hits, i.e., indicating "yes" in trials the hand fits through the given opening) and the total number of actual positive events (i.e., total amount of "yes" trials with openings larger or equal to 0). Based on these False-Alarm and Hit rates, two independent detection theory variables, discriminability index and judgment tendency, were calculated: Perceptual sensitivity reflects the participants' ability to discriminate a fit from a non-fit. It is represented by means of the discriminability index d' which is assessed by the following formula: $d' = Z(\text{Hit rate}) - Z(\text{False - Alarm rate})$, Z reflects the z-standardization of the hit rate or the false alarm rate, respectively. This means the more sensitive the participant is at discriminating a fit from a non-fit, the larger the d'-value will be. Based on the idea that participants make decisions (i.e., respond yes or no) by comparing their observations with an experienced-based judgment criterion, we further assessed participants' judgment tendency represented by the criterion (c). It was calculated using the following formula: $c = -.5*[Z(\text{Hit rate}) + Z(\text{False} - \text{Alarm rate})]$, for a detailed derivation of the formula, please see Macmillan and Creelman (2004, p 27–31). A positive value is associated with a rather conservative judgment tendency. On the contrary, negative values reflect increasingly liberal judgments. By considering d-prime as a perceptual measure and

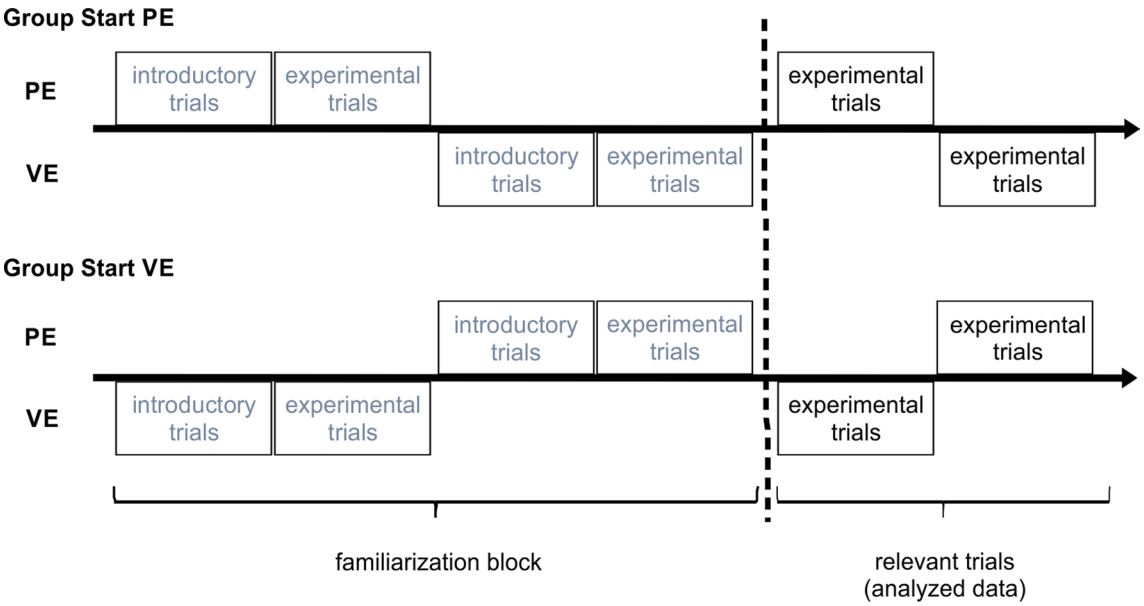


Fig. 2 Experimental procedure in study 1 Participants either started with the PE condition or with the VE condition

criterion c as a measure of conservative versus liberal judgment tendencies, a more precise interpretation of participants' judgment behavior was possible.

Normality was assessed by screening normal probability plots and with the Shapiro–Wilk Test. Since the direction of the hypotheses was determined a priori, exact instead of asymptotic p-values were reported one-tailed ($p < 0.05$). For the sake of completeness, we also provide adjusted p-values using the stepwise Holm–Bonferroni procedure (p_{adj}) to correct for family-wise error rate.

2.1.4.2 Equivalence testing Since the mere absence of significant differences cannot be interpreted as equivalence, a method from biometrics, the so-called equivalence test (Da Silva et al. 2009; Mascha & Sessler 2011; Rogers et al. 1993), was used to test the similarity of the two conditions (performance in the Aperture Task in a PE and a VE). Following this approach, to assume equivalence, the difference in mean performance between the PE and the VE setting needs to be significantly within an a priori specified equivalence region ranging from—δ (*minus delta*) to $+\delta$ (*plus delta*). In the present study, the following null hypothesis (H_0) was formulated: the difference between the conditions (e.g., performance in VE minus performance in PE) is below or above the equivalence region $\delta\left(H_0 : \mu_{VE} - \mu_R \leq -\delta \text{ or } \mu_{VE} - \mu_R \geq +\delta\right)$. The corresponding alternative hypothesis (H_1) stated: the already mentioned difference between the conditions is within the delta-range ($H_1 : \mu_{VE} - \mu_R > -\delta$ and $\mu_{VE} - \mu_R < +\delta$). The equivalence bounds ($+\delta$ and $-\delta$) are determined based on FDA guidelines with the reference product (here: mean value of the variables in the PE condition) * 0.2 (Machin et al. 2011).

Based on this principle, equivalence between VE and PE is proposed for judgment accuracy (hypothesis 1), perceptual sensitivity (hypothesis 2) and judgment tendency (hypothesis 3).

To test the hypotheses, two one-sided tests were applied (TOST). If the observed confidence interval falls within the predicted range, both tests become significant ($p < 0.05$) and H_0 can be discarded (Williams et al. 2002). If the confidence interval is within the predicted differential range, equivalence is suggested. In Fig. 3, possible results and the resulting implications are illustrated.

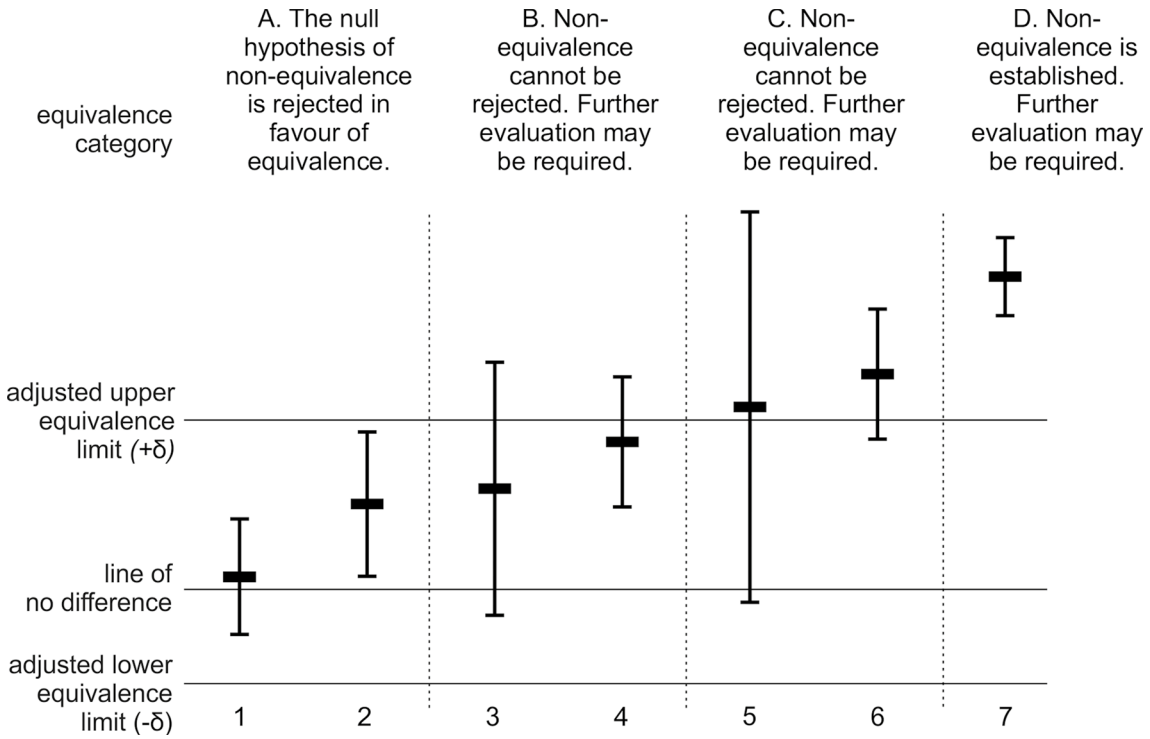


Fig. 3 Illustration of possible equivalence outcomes by use of the confidence interval approach Possible case outcomes are expressed as mean difference between PE and VE (square) conditions including confidence intervals (whiskers). The prespecified range of equivalence is defined by the limiting lines above ($+\delta$) and below ($-\delta$) the line of no difference. Equivalence categories A-D indicate the according recommendation per case. For example, case outcomes 1 and 2 clearly fall within the equivalence range and would indicate equivalent performance between the VE and the PE setting, while case 7 would represent an example for non-equivalence. For simplicity reasons the figure shown here only illustrates cases exceeding the upper limit. For similar illustrations see for example EFSA Panel on Genetically Modified Organisms: European Food Safety Authority (2010)

2.2 Study 1 results

Affordance judgment performance in the PE and the VE was compared for equivalence considering accuracy and detection theory values by use of the above-mentioned procedure. Descriptive data for the three variables are shown in Table 1. Moreover, Fig. 4 shows the results of the equivalence testing. Descriptive and inferential equivalence statistics are listed in Table 2.

Table 1 Descriptive data for the three variables examined per condition

Variable	Mdn	Mean	SD
Accuracy (%)			
PE	75.93	74.69	9.87
VE	74.07	70.53	13.88
Perceptual sensitivity (d')			
PE	1.75	1.69	0.60
VE	1.47	1.48	0.90
Judgment tendency (c)			
PE	−0.40	−0.27	0.93
VE	0.56	0.14	0.98

2.2.1 Accuracy

Concerning accuracy, equivalence testing revealed significance. A significant difference was found between the mean condition difference (VE minus PE performance) and the upper equivalence bound on the one hand, and between the mean condition difference and the lower equivalence bound on the other hand. This means that H_0 can be discarded. Instead, our hypothesis of performance equivalence in the VE and the PE condition is supported for accuracy.

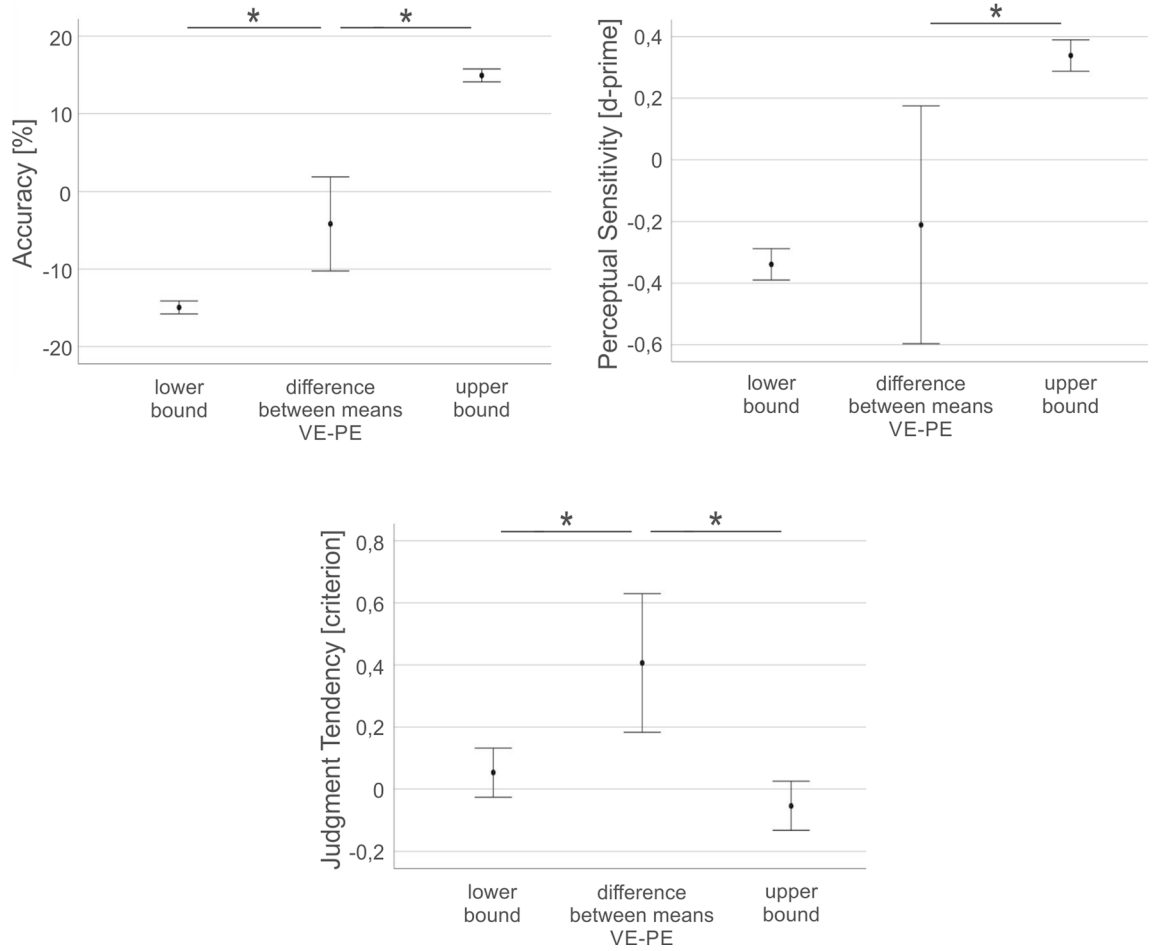


Fig. 4 Results of equivalence testing The confidence interval of the upper and lower equivalence bounds (Epsilon* ± 0.2) as well as of the difference between the VE and PE condition is shown. The lower bound and upper bound in the judgment tendency appear reversed due to the algebraic sign of the mean value of the criterion in the PE condition (−0.27). *$p < .05$

Table 2 Descriptive and inferential statistics of the equivalence testing

Variable	Mean	*T*	*df*	*p*	p_{adj}	d_{Cohen}
Accuracy (%)						
Difference between means	− 4.17					
Lower equivalence bound	− 14.94	3.85	23	< .001*	.002*	.79
Upper equivalence bound	14.94	− 6.13	23	< .001*	< .001*	− 1.25
Perceptual sensitivity (*d'*)						
Difference between means	− .21					
Lower equivalence bound	− .34	.72	23	.241	.241	.15
Upper equivalence bound	.34	− 2.80	23	.005*	.010*	− .57
Judgment tendency (*c*)						
Difference between means	.41					
Lower equivalence bound	.05	3.29	23	.002*	.005*	.67
Upper equivalence bound	− .05	3.80	23	< .001*	.002*	.78

Equivalence of the PE and the VE condition is tested using one-tailed t-test analyses for accuracy, perceptual sensitivity and judgment tendency

*$p < .05$

2.2.2 Detection theory values

With regard to perceptual sensitivity (*d*-prime), results revealed a significant difference between the mean condition difference and the upper equivalence bound. The difference between performance and the lower equivalence bound did not reach significance. As a result, the equivalence of perceptual sensitivity (*d*-prime) between the PE and the VE condition must be described as uncertain. When considering the judgment tendency (*c*), a different picture emerged: the difference in the performance between conditions varied significantly from the upper equivalence bound as well as from the lower equivalence bound. However, the graphical check (see Fig. 4) reveals that the confidence interval is beyond both boundaries. Thus, there is no equivalence between the two conditions.

2.3 Study 1 discussion

In study 1, we aimed at evaluating the equivalence of a PE and a VE setting by use of an Aperture Task. Besides accuracy, detection theory variables served as measures and were analyzed. Equivalence between conditions could be shown for accuracy, but not for judgment tendency. For perceptual sensitivity, it remained uncertain. Interestingly, descriptive results revealed that in the VE condition, participants decided more conservatively than in the PE condition (see Table 1).

The lack of equivalence in judgment tendency could be explained by a more conservative approach due to increased safety behavior in VE. Possible reasons for a lack of equivalence in the detection theory variables might include the influence of familiarity which may lead to more conservative behavior as well as differences in the quality of perceptual cues which may lead to worse perceptual sensitivity in the VE condition. In terms of a more global explanation for the lack of equivalence in judgment tendency, the degree of immersion or the sense of presence could be questioned, both of which are important factors for eliciting realistic behavior in VEs (Bowman and McMahan 2007; Sanchez-Vives and Slater 2005). While some studies have discussed that virtual cues like holographic objects may facilitate object-related action planning (Rohrbach et al. 2021), it is currently unknown to which extent and quality objects in the virtual world provide affordances for action (Harris et al. 2019). On the one hand, in the literature the high ecological validity of VEs for cognitive and motor neuroscience has been stressed (Parsons 2015; Tieri et al. 2018), On the other hand, it also has been discussed (Gerig et al. 2018; Wilson & Soranzo 2015) that important perceptual cues (e.g., depth cues) are lacking in VE.

While haptic feedback was held constant between conditions in our study, the perceptual system was still provided with fewer binocular cues to depth in the VE condition. Furthermore, previous studies found that size and distance to objects is consistently underestimated in VE (Interrante et al. 2008). This finding is in line with our finding of a more conservative judgment tendency in VE, as participants might have underestimated the opening size in the VE setting in our study in a similar way. Another reason for more conservative behavior could be a lack of familiarity. Typically, participants are less familiar with virtual settings (Lindner et al. 2019), and it could be speculated that judgment tendency is mediated by familiarity with more insecure and conservative behavior in the unfamiliar VE. Similar arguments have been raised within PE conditions for the retrieval of a stable judgment tendency (criterion), i.e., the criterion-instability hypothesis proposed by Finkel et al. (2019a) when adapting

to new conditions. It suggests a prolonged habituation period for unfamiliar conditions introducing a phase of instability during which a new criterion needs to be built or encoded, respectively. It is assumed that this process causes variability in the judgment tendency.

Future studies should address a possible influence of these factors, e.g., by carefully manipulating variables such as familiarity. While the observed equivalence between conditions in regard to accuracy emphasizes the potential of virtual settings to assess judgment performance, the current equivalence study also stresses that different variables should be considered to provide a more complete picture. We can only speculate about the underlying reasons for non-equivalence or uncertainty for the detection theory variables. The identification of influencing factors causing non-equivalence is beyond the scope of the current manuscript. Future studies could for example pursue the presented approach and specify potential factors and manipulations that might establish enhanced equivalence for detection theory variables. Furthermore, the investigated task included a static environment only. It would be of interest to implement the current study approach in more dynamic settings, including for example moving stimuli.

3 Study 2

The literature suggests that affordance judgments can be trained in PE (Finkel et al. 2019a; Randerath & Frey 2016), and most recently training effects were reported in a reachability design in VE (Gagnon et al. 2021). The major aim of study 2 was to examine whether participants volunteering in study 1 would benefit from a training using a VE. Study 2 took place one or two days after study 1 and lasted approximately 90 min. It was analyzed whether potential improvements in judgment performance due to feedback provided via the VE setting could be measured in a subsequent VE assessment (without feedback), and whether a potential performance improvement could be detected in the PE assessment as well. Obviously, variables demonstrating at least uncertain or given equivalence between PE and VE should be considered for this purpose. For variables that can be considered to evoke equivalent performance in PEs and VEs we hypothesized to find transfer effects of VE training leading to improved performance in the physical world behavior as well. Based on the equivalence results of study 1 this pertains to the variables judgment accuracy and perceptual sensitivity. Finkel et al. (2019a) and Randerath and Frey (2016) have demonstrated that in healthy young adults these two variables were trainable using our physical aperture setting. In these previous PE training studies participants tried to fit their hand into the opening for each trial. Feedback included visual (goggles were open during the action) and haptic (for fits they were able to touch a back panel that elicited a noise) information. In the current study, participants received a virtual training that provided visual feedback only.

3.1 Study 2 methods

The methods of study 2 were mainly adapted from study 1. Altered parts are described below.

3.1.1 Participants

To estimate a minimum sample size (to detect a training effect if present), we a priori ran a power analysis based on prior training data in PE only (Randerath and Frey 2016). For pair-wise group comparisons (Wilcoxon signed rank tests), statistical power (1-β; as the complement of Type II error magnitude) was computed by use of G-Power (Faul et al. 2007). Test power was calculated one-tailed and with

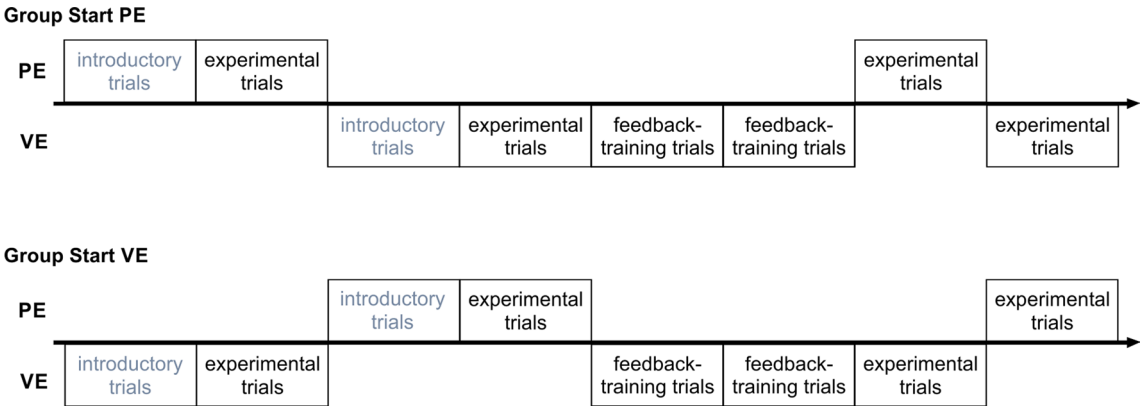


Fig. 5 Experimental procedure in study 2 Participants either started with the PE assessment or with the VE assessment. Aligned with study 1, the same half of the participants started with the PE assessment ($N=12$) and half started with the VE assessment ($N=12$). Subsequent to the assessments, feedback was introduced via the VE. After the training, judgment performance was assessed again for each condition

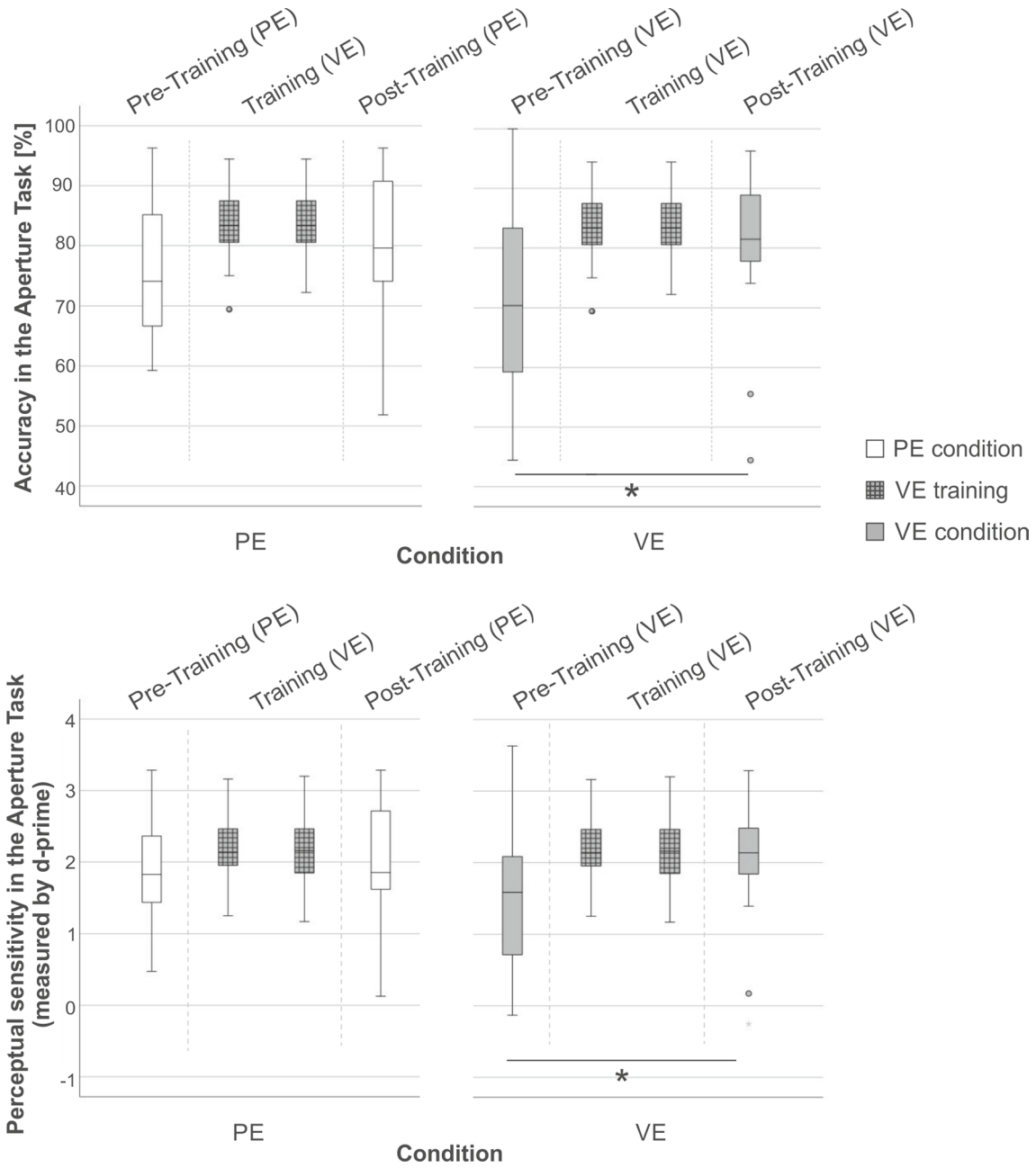


Fig. 6 Accuracy [%] and perceptual sensitivity measured by d-prime for condition (PE or VE training) White boxplots refer to the assessment conducted in the PE condition. Boxplots in grey indicate the assessment in the VE condition. Boxplots with black grid refer to performance while solving the VE training (block 1 and block 2). *$p < .05$

an alpha level of .05. Power analysis revealed a minimal sample-size-proposal of $N = 21$ [Power: 0.96, $d = 0.79$]. In total, data sets of the same 24 individuals (between 20 and 33 years of age ($M = 24.1$ years, $SD = 3.3$; 17 females) as in Study 1 were subjected.

3.1.2 Procedure

In Fig. 5, the procedure of study 2 is displayed. Subjects started with 20 introductory trials either in the PE or in the VE and 30 experimental trials in the same mode. Afterward, the other mode followed with 20 introductory and 30 experimental trials. Subsequently, two VE feedback-training blocks (40 trials each) were conducted and the subjects again

executed an experimental block in each mode (PE and VE), in the same order as in the beginning. As in study 1, they executed each trial with their right hand.

3.1.2.1 Training VE setting During the two training blocks participants were first asked to judge whether their hand could fit into the opening. They indicated their answer by pressing a yes- or no-button. If the action was actually feasible, a sound was presented immediately after the answer was given, and the entire image turned green. In contrast, if the actability was not given, the whole image turned yellow.

3.1.3 Data analysis

For the following data analysis, we only used data of the second study. In addition to judgment accuracy, we included perceptual sensitivity in order to evaluate the training effect in detail. Since judgment tendency was clearly shown to be non-equivalent between the PE and the VE, this measure was excluded from analysis.

Parallel to study 1, normality was assessed by screening normal probability plots and with the Shapiro–Wilk Test. As part of the data were not normally distributed (Accuracy VE post-training, $p < .001$, perceptual sensitivity (*d*-prime) VE post-training, $p = .002$), behavioral data were analyzed non-parametrically. To evaluate whether there was a main effect of VE training in the second study within each environmental condition (PE and VE assessment) and per variable (accuracy and perceptual sensitivity) we ran Wilcoxon signed rank tests. As in Study 1, exact instead of asymptotic *p*-values were reported one-tailed ($p < .05$). Adjusted *p*-values using the stepwise Holm–Bonferroni procedure (p_{adj}) are provided. Corresponding z-values reported by these tests were used to calculate the effect size *r* as proposed by Cohen (Cohen 1988) by dividing *z* by the square root of *N*. Please note that *N* corresponds to the number of observations ($N = 48$ (24×2); Field, 2013).

3.1.3.1 Post hoc equivalence analysis To further evaluate replicability of the equivalence results from study 1, we ran equivalence analyses for PE and VE pre-training for all variables. Furthermore, to explore potential effects by training, equivalence tests were repeated post-training. The statistical procedure was the same as described in study 1 ("2.1.4 Data analysis", section "2.1.4.2. Equivalence testing").

3.2 Study 2 results

Descriptive and inferential statistics (exact and adjusted *p*-values) are summarized in Table 3. Figure 6 displays variable values of study 2 in boxplots. Please note, to evaluate potential repetition effects, we ran a control analysis which is provided in the supplementary material (S1—Table 4). It suggests no relevant improvement by mere repetition for the variables accuracy and perceptual sensitivity.

3.2.1 Accuracy

On a descriptive level, participants improved in judgment accuracy in both PE and VE from pre-training to post-training. However; the statistical pre-post comparison revealed that participants' performance only improved significantly within the VE. In the PE, participants' accuracy performance did not improve significantly.

3.2.2 Perceptual sensitivity

Concerning perceptual sensitivity, a similar picture emerged: The difference between pre-training and post-training assessment revealed observable effects in VE, but not in the PE condition (Table 3).

Post hoc we further investigated equivalence between the PE and the VE pre-training as well as between the PE and the VE post-training for accuracy and detection theory values (descriptive data are depicted in Table 4.). In addition, Fig. 7 shows the graphical results of the equivalence testing post-training. Descriptive and inferential equivalence statistics are listed in Table 5.

Table 3 Descriptive and inferential statistics of accuracy and perceptual sensitivity

Variable	Pre-training Mdn	Post-training Mdn	*Z*	p_{exact}	p_{adj}	*r*
Accuracy (%)						
PE	74.07	79.63	−1.14	.133	.266	.16
VE	70.37	81.48	−2.48	.006*	.022*	.36
Perceptual sensitivity (*d'*)						
PE	1.83	1.85	−.68	.258	.258	.10
VE	1.59	2.14	−2.10	.017*	.052	.30

Wilcoxon-Test results of comparisons between pre-training versus post-training affordance judgment assessments are shown.

*$p < .05$, one-tailed

Table 4 Descriptive data for the three variables examined per condition (pre-training and post-training)

Variable	Pre-training			Post-training		
	Mdn	Mean	SD	Mdn	Mean	SD
Accuracy (%)						
PE	74.07	76.70	10.73	79.63	79.94	12.59
VE	70.37	71.91	14.97	81.48	81.17	11.13
Perceptual sensitivity (d')						
PE	1.83	1.89	0.68	1.85	2.02	0.85
VE	1.59	1.53	0.99	2.14	2.02	0.75
Judgment tendency (c)						
PE	−0.04	−0.14	0.92	−0.13	−0.21	0.75
VE	0.01	−0.09	1.01	−0.16	−0.06	0.51

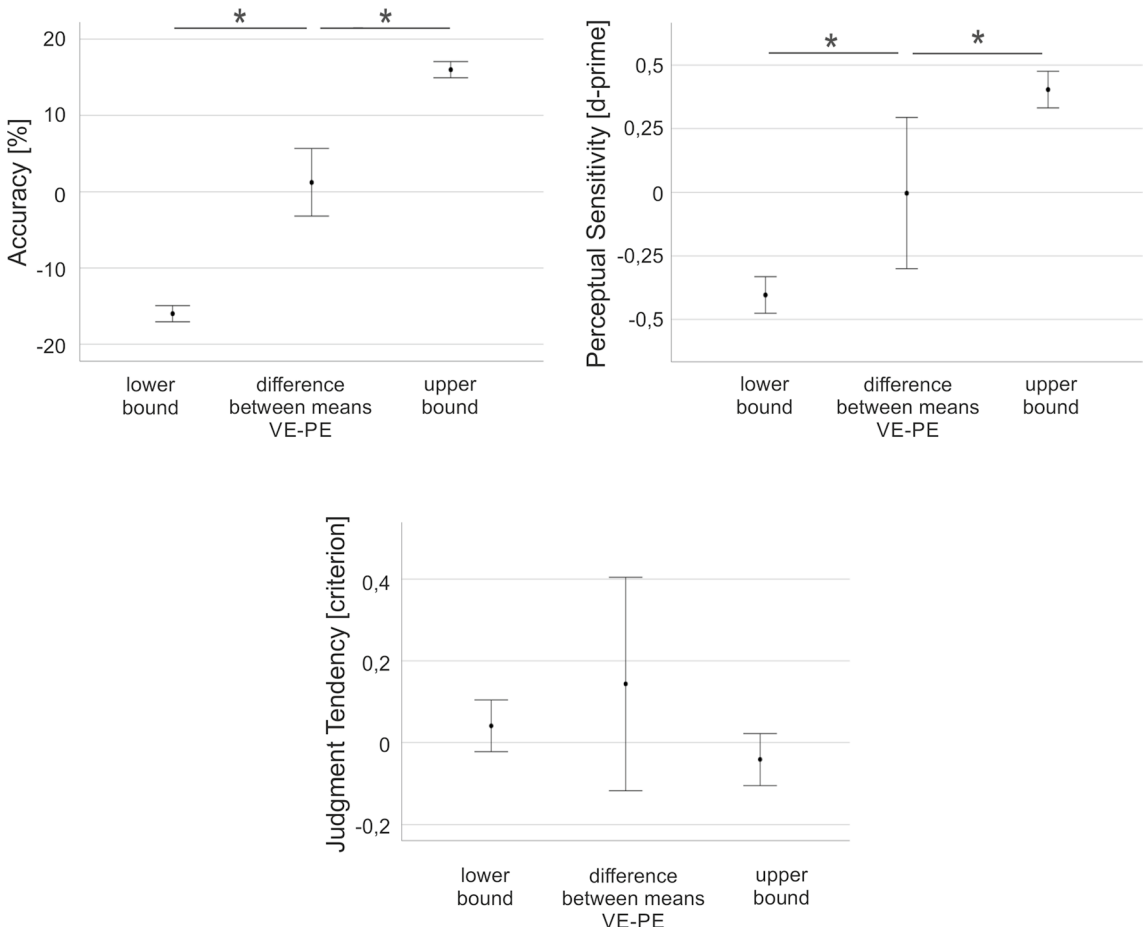


Fig. 7 Results of equivalence testing after feedback training (post-training) The confidence interval of the upper and lower equivalence bounds (Epsilon* ± 0.2) as well as of the difference between the VE and PE condition is shown. The lower bound and upper bound in the judgment tendency appear reversed due to the algebraic sign of the mean value of the criterion in the PE condition (− 0.21). * $p < .05$

3.2.2.1 Accuracy Concerning accuracy, equivalence testing for pre-training revealed significance and showed a similar picture as in study 1. This implies that equivalence in accuracy performance can be assumed in study 2 as well. In addition, post-training the equivalence testing also showed significant results.

3.2.2.2 Detection theory values In terms of perceptual sensitivity, for pre-training, uncertainty in the equivalence between VE and PE was replicated. However, post-training, the picture changed: the difference between means was significantly different from both bounds. Visual inspection confirmed that the equivalence of perceptual sensitivity

Table 5 Descriptive and inferential statistics of the equivalence testing pre-training and post-training

Variable	Pre-training						Post-training					
	Mean	*T*	*df*	*p*	p_{adj}	d_{Cohen}	Mean	*T*	*df*	*p*	p_{adj}	d_{Cohen}
Accuracy (%)												
Difference between means	−4.78						1.23					
Lower equivalence bound	−15.34	4.11	23	< .001*	.001*	.84	−15.99	9.05	23	< .001*	< .001*	1.85
Upper equivalence bound	15.34	−7.33	23	< .001*	< .001*	−1.50	15.99	−6.01	23	< .001*	< .001*	−1.23
Perceptual sensitivity (*d'*)												
Difference between means	−0.35											
Lower equivalence bound	−0.38	0.13	23	.448	.448	.03	− .40	1.23	23	.002*	.010*	.64
Upper equivalence bound	0.38	−3.93	23	< .001*	.001*	−.80	.40	3.12	23	.011*	.032*	−.50
Judgment tendency (*c*)												
Difference between means	0.06						.14					
Lower equivalence bound	0.03	0.22	23	.413	.826	.05	.04	0.97	23	.171	.171	.20
Upper equivalence bound	−0.03	0.61	23	.273	.818	.13	− .04	1.23	23	.115	.230	.25

Equivalence of the PE and the VE condition is tested using one-tailed t-test analyses for accuracy, perceptual sensitivity and judgment tendency
*$p < .05$

after training can be assumed. Similar to study 1 for judgment tendency equivalence cannot be assumed. The equivalence testing showed (as in study 1) both, pre-training and post-training, uncertain results.

3.3 Study 2 discussion

In the past two decades evidence has accumulated showing that feedback in PE has an advantageous effect on performing affordance judgment tasks (Finkel et al. 2019a; Franchak et al. 2010; Randerath and Frey 2016). First attempts to implement trainings within VEs show promising effects. For example, Gagnon et al. (2021) found that reachability was initially overestimated, but became more accurate over feedback blocks in the VE. With study 2 we add to this line of research. Our training study delivers mixed results. On the group level, VE training only improves accuracy within the same mode. One reason for a lack of transfer of the training effect from the VE to the PE could partly be explained by the overall high performance from the beginning, which is quite common in healthy young samples (Kwok et al. 2011; Mahncke et al. 2006). Thus, a lack of improvement-transfer could be explained by the fact that there is low potential to improve.

Another possible explanation for hampered training transfer from the VE to the PE in accuracy and perceptual sensitivity measures could be that participants in the VE feedback training make use of learning strategies that may not apply to physical settings. They may have stored the correct response in semantic memory (concept-based knowledge unrelated to specific experiences), but not the underlying meaning (Kumar 2021). The association of a presented opening size and the correct response could have been merely learned by heart for the VE, and not include any action-related information, i.e., action procedure or a simulation of considering whether the person's hand fits, which may be relevant for performing the correct and perceptual sensitive judgments in physical settings. This is conceivable as the participants were given feedback in the VE for one opening after the other without actually trying to fit through the opening. Haptic feedback and actually moving as implemented in PE studies may lead to an essential link of environmental properties and the person's actual capabilities: a presented opening and the individual body proportions as well as involved action production processes that may deliver the basics for involving action-perception networks and perhaps train action simulation processes. For example, Franchak et al. (2010) demonstrated that action feedback aided perceptual judgments by facilitating scaling to body dimensions as compared to a group that had to rely on perception only. For walking through doorways of varying widths in the action (vs. mere perception) group, judgments were more accurate and strongly related to height, weight, and torso size of the participants. However, to maintain the benefits of VE training, haptic feedback or learning by doing is not an option. Possibilities such as additional verbal or visual cues could be considered to further strengthen the associative link.

In post hoc equivalence analyses, results from study 1 were replicated before training (pre-training) for accuracy (equivalent between PE and VE) and perceptual sensitivity (equivalence unsecure between PE and VE). Judgment tendency revealed pre-training as well as post-training uncertain equivalence results (in study 1: not equivalent). After training (post-training), accuracy appeared still equivalent. Interestingly, perceptual sensitivity reached an equivalent level for PE and VE post-training. From these results we

conclude that mere exposure to the task appeared to have no influence on improving equivalence as has been shown by the replication of equivalence results for accuracy and perceptual sensitivity. However, the VE-training led to a convergence towards PE-levels by enhancing perceptual sensitivity performance in VE.

Therefore, the utility of VE trainings should be further explored by including participant-groups with high variability in performance to make more precise statements about the potential of VE trainings in affordance judgments. In addition, it may be of interest to explore the effect of differential feedback mechanisms in VEs, and perhaps come up with settings that allow to increasingly converge to action-related learning mechanisms. The hope would be to achieve increasingly better results in performing affordance judgments in PEs that were trained within a virtual setting.

4 General discussion

Judging action opportunities requires an estimation of environmental conditions and our own physical abilities. When navigating through our everyday life, we constantly make decisions about whether or not to perform certain actions which requires the evaluation of the fit between perceived environmental properties and our own physical capabilities, so-called affordance judgments. For example, when crossing a street, we need to match the speed of the approaching cars with our own walking speed. Earlier studies using PEs showed that young and healthy adults are able to make quick and adequate affordance decisions (Finkel et al. 2019b; Franchak et al. 2012). With older age or after stroke, impairment of this ability was observed which may lead to serious consequences (Finkel et al. 2019b; Luyat et al. 2008; Muroi et al. 2017; Pereira et al. 2020; Randerath et al. 2018). Therefore, it would be desirable to develop training approaches for advancing age or neuropsychological rehabilitation. Controlled training settings in PEs are, unfortunately, often elaborate and cost-intensive. In some other cases, these settings are also limited by safety issues (e.g., traffic situations). The use of VE may offer a feasible alternative for implementing trainings. The current work aimed at investigating whether equivalent performance can be demonstrated when using VEs versus PEs to analyze judgments of action opportunities in a static Aperture Task (study 1). In addition, we examined whether visual feedback training in the VE can improve judgment performance in both the VE and the PE assessment (study 2). In our VE and PE assessment participants judged whether their hand would fit into a presented opening that varied in width. Participants' accuracy (percentage of correct answers) was analyzed. To provide a more detailed picture, we applied a detection theory approach to evaluate perceptual sensitivity and judgment tendencies in both conditions.

Data analysis in study 1 revealed equivalence between the VE and PE condition for the accuracy of decisions in the Aperture Task. This finding is in line with Bhargava et al. (2020) and Geuss et al. (2015) who also found a similarity in accuracy between a physical and virtual setting. Thus, the present results at first glance provide support for the hypothesis that the information which is needed to perform affordance judgments is available and salient in VEs. However, the analysis of detection theory variables and the lack of equivalence between conditions therein presents a more nuanced picture. In terms of perceptual sensitivity, equivalence was shown to be uncertain and for judgment tendency no equivalence between the VE and the PE could be found. Concerning the uncertainty of equivalence for perceptual sensitivity as well as the lack of equivalence for the judgment tendency measures, our findings are in line with Grechkin et al. (2010) who found evidence that not all variables appear transferable from VEs to PEs and vice versa. For example, the study by Grechkin et al. (2010) proposed an underestimation of distance using a head mounted display compared to a physical setting. Also Geuss et al. (2010) demonstrated a difference in perceiving ego- and exocentric distances between virtual and physical conditions. Thus, similarity between environments may pertain to the type of evaluated variable. Prior studies in stroke patients performing affordance judgments indicate different neuroanatomical systems to be essential for judgment tendency on the one hand and perceptual sensitivity on the other hand. By use of voxel-based lesion-symptom mapping an association has been shown between damage in left or right dorsal routes and diminished perceptual sensitivity, whereas impaired judgment tendency was associated with primarily left ventro-dorsal brain damage (Randerath et al. 2018). Thus, the current and previous research seems to suggest that affordance judgment performance can vary depending on the assessed variable, and study outcomes reflect different neuroanatomical systems to be involved. Speculating, it seems as if parts of the affordance judgment systems (e.g., executive functions) may have highly overlapping involvement in virtual and physical settings. Others currently show uncertainty (perceptual information processing) and even non-equivalence (retrieval of response biases). It is plausible that insufficient equivalence hinders the transfer of training effects between modes. Study 2 showed that in healthy young students visual feedback training in the VE led to a significant post-training improvement in the VE in regard to accuracy. Perceptual sensitivity performance improved on a descriptive level in VE but did not reach statistical significance. Furthermore, in study 2 no significant training transfer effect was found in PE performance.

Neither accuracy nor perceptual sensitivity performance improved in PE after the VE training. Remarkably, we found equivalence in perceptual sensitivity between VE and RE after training. This result suggests that our training approach in VE was sufficient to improve perceptual sensitivity performance in such a way that it reaches the level achieved in PE.

Thus, while previous studies (Finkel et al. 2019a; Randerath & Frey 2016) demonstrated improvement in PE variables after training in PE, we were not able to show improvement in PE performance based on training in VE. Non-equivalence in behavior between environmental conditions may contribute to missing transfer. In their recent review, Harris et al. (2019) outline the possibility that, due to the artificial presentation of egocentric distance cues in VEs, not only the execution of motor skills but also basic visual perception is affected. The authors express concerns as fewer binocular cues to depth, conflicting depth information and limited haptic feedback are available. Therefore, they assume a shift from dorsal (online) control of action in PEs to the ventral stream in VEs which may have important consequences on perception and action (decisions). Also, Goodale et al. (1994) found that a general lack of haptic information may further push users into a ventral mode of processing when solving basic reaching and grasping movements. This division would go along with our neuroanatomical working model for affordance judgments in PEs proposed by Randerath et al. (2021). Based on lesion data, the model attributes ventro-dorsal sites to be essential for detection theory variables: perceptual sensitivity (more dorsal and bilaterally) and judgment tendencies (left lateralized). From a more compensatory perspective this may be good news for stroke patients with impaired affordance judgments and dorsal lesions, as this may hypothetically imply that ventral regions could be used to train accuracy in VEs. For our current results this dorsal to ventral shift from a PE to a VE could explain both the uncertain equivalence in perceptual sensitivity and the absence of the transfer effect from VE training to the PE setting for this variable. It could be speculated that the enhancement in accuracy after VE training and its transfer potential from a VE to a PE is accomplished by learned associations (perceived opening x correct response) and handled via semantic routes in the ventral stream.

As mentioned before, one reason for non-equivalence for judgment tendencies could be a lack of familiarity in virtual environments. Participants may not be able to retrieve or reactivate any known criterion bias in the virtual setting as other studies confirmed that participants are less familiar with virtual settings (Lindner et al. 2019). Heightened insecurity may elicit conservative judgment behavior in the unfamiliar VE. The criterion-instability hypothesis proposed by Finkel et al. (2019a) when adapting to new conditions may apply: Variability in judgment tendency may be influenced by a prolonged habituation period for the unfamiliar VE condition during which a new criterion for judgments needs to be built or encoded, respectively.

In future studies it would therefore be useful to replicate the current results, to investigate the underlying mechanisms of affordance judgments, elaborate on differences versus equivalence in VEs and PEs in more detail and encompass determinant factors. For example, including versus excluding experience using virtual limbs within a VE could have an influence. Linkenauger et al. (2015) investigated affordance judgments in a reaching task in a VE and found a body scaling effect, where enacted reaching capability with a virtual arm influenced perceived distance in virtual settings. Merely having a long or short virtual arm, was not sufficient to influence distance perception. But minimal reaching experience with the virtual arm influenced perceived distance, with longer arm's reach resulting in shorter perceived distances and vice versa.

There are manifold forms of different types of affordance judgment tasks and a whole range of possible VEs and available displays with constantly improving technology. In this context it would be important to implement further studies incorporating other affordance judgments, e.g., passability judgments or stepping over obstacles, as implemented by Geuss et al. (2010) or Lin et al. (2015). Further, newer hardware for displaying VEs may provide progress in resolution, graphic fidelity, field of view and may therefore improve the transfer effect of VE training to the physical world.

Promisingly, our data suggests that training within VE can lift performance levels to reach equivalence between VE and PE. Thus, despite the illustrated limitations, the current study's results raise hope that this noninvasive, easy to use VE technique could support the rehabilitation process of individuals who show a weak performance in affordance decision-making in PE.

5 Conclusion

In two studies we investigated first the equivalence of affordance decision-making in a virtual and a physical setting by use of an Aperture Task. Second, we evaluated the potential of a virtual training to improve judgment behavior and transfer the performance enhancement within the VE or to the PE, respectively. Equivalence between VE and PE conditions was demonstrated for judgment accuracy. Equivalence for perceptual sensitivity appeared to be unsecure, and judgment tendency did not show equivalence. In terms of the virtual training, accuracy performance in the VE could be increased significantly, while perceptual sensitivity showed better performance on a descriptive level only. A transfer effect of the VE-training to the performance in the PE was not achieved. However, post-hoc analyses of equivalence in

performance between VE and PE showed comparable performance in perceptual sensitivity (d-prime) post-training. Thus, while prior to training perceptual sensitivity levels appeared non-equivalent between VE and PE, after training perceptual sensitivity in the VE approached PE performance levels. This raises hope for an effective implementation of VE trainings in persons with decreased affordance judgment performance.

Further, we discussed the idea of distinguished networks being involved while solving affordance judgments depending on the environmental mode. Distinguished networks may contribute to dissimilarities in judgment behavior while using physical versus virtual settings. While it presents a limitation in comparability between environments, this point could also emerge as a chance to embrace compensatory mechanisms by use of virtual settings for individuals with brain damage that affected affordance judgments in physical settings. The exact mechanisms and reasons that may tweak equivalence and heighten training transfer for affordance judgment performance from VEs to PEs should be explored in further studies. Improvements in resolution, graphic fidelity, and field of view offered by newer VE hardware may stepwise increase similarity between physical and virtual settings, thereby potentially improving equivalence in affordance judgment performance and perhaps enlarging the potential for training transfer among environments. The current paradigm presents one opportunity to evaluate the advancements. Eventually, VE trainings of affordance judgments should be considered to be tested in individuals suffering from deficits in affordance decision-making.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10055-023-00765-4.

Acknowledgements This work was supported by the EU FP7 Marie Curie Zukunftskolleg Incoming Fellowship Programme at the University of Konstanz (291784) and by the Deutsche Forschungsgemeinschaft (DFG, Grant No. 438470816) awarded to JR. Further, the project was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2117 – 422037984. We thank the reviewers and Prof. Klaus Willmes for their helpful comments on this manuscript.

Funding Open Access funding enabled and organized by Projekt DEAL.

Data availability The data for all experiments have been made available at the Zenodo Repository https://zenodo.org/record/7225320#.Y1DknZBxmM

Declarations

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.

Ethical approval The project was approved by the ethics committee of the University of Konstanz (#15/2020) and conducted in accordance with the Declaration of Helsinki. All participants provided informed written consent including information about the tasks. Their participation in the study was financially rewarded (10€ per hour) or compensated with study credits (1 per hour). After participation participants received a verbal debriefing. They learned that the goal was to investigate the potential of a "3D-Training" to improve cognitive action planning.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Bhargava A, Lucaites KM, Hartman LS, Solini H, Bertrand JW, Robb AC, Pagano CC, Babu SV (2020) Revisiting affordance perception in contemporary virtual reality. Virtual Real. https://doi.org/10.1007/s10055-020-00432-y

Bliss JP, Tidwell PD, Guest MA (1997) The effectiveness of virtual reality for administering spatial navigation training to firefighters. Presence Teleoperat Virtual Environ 6(1):73–86. https://doi.org/10.1177/1541931295039014 12

Bodenheimer B, Fu Q (2015) The effect of avatar model in stepping off a ledge in an immersive virtual environment. In: Paper presented at the proceedings of the ACM SIGGRAPH symposium on applied perception

Bohil CJ, Alicea B, Biocca FA (2011) Virtual reality in neuroscience research and therapy. Nat Rev Neurosci 12(12):752–762. https://doi.org/10.1038/nrn3122

Bowman DA, McMahan RP (2007) Virtual reality: how much immersion is enough? Computer 40(7):36–43. https://doi.org/10.1109/MC.2007.257

Buckingham G (2019) Examining the size-weight illusion with visuo-haptic conflict in immersive virtual reality. Quart J of Exp Psychol 72(9):2168–2175. https://doi.org/10.1177/1747021819835808

Choudhury N, Gélinas-Phaneuf N, Delorme S, Del Maestro R (2013) Fundamentals of neurosurgery: virtual reality tasks for training and evaluation of technical skills. World Neurosurg 80(5):e9–e19. https://doi.org/10.1016/j.wneu.2012.08.022

Cohen J (1988) Statistical power analysis for the behavioral sciences. L. Erlbaum Associates, Hillsdale

Cooper RA, Ding D, Simpson R, Fitzgerald SG, Spaeth DM, Guo S, Koontz AM, Cooper R, Kim J, Boninger ML (2005) Virtual reality and computer-enhanced training applied to wheeled mobility: an overview of work in Pittsburgh. Assistive Technol 17(2):159–170. https://doi.org/10.1080/10400435.2005.10132105s

Creem-Regehr SH, Gill DM, Pointon GD, Bodenheimer B, Stefanucci JK (2019) Mind the gap: gap affordance judgments of children, teens, and adults in an immersive virtual environment. Front Robot AI 6:96. https://doi.org/10.3389/frobt.2019.00096

Creem-Regehr SH, Stefanucci JK, Thompson,WB (2015).Perceiving absolute scale in virtual environments: How theory and application have mutually informed the role of body-based perception.

In: Psychology of learning and motivation (Vol 62, pp 195–224), Elsevier
Da Silva GT, Logan BR, Klein JP (2009) Methods for equivalence and noninferiority testing. Biol Blood Marrow Transplant 15(1):120–127. https://doi.org/10.1016/j.bbmt.2008.10.004
Day BM, Wagman JB, Smith PJ (2015) Perception of maximum stepping and leaping distance: stepping affordances as a special case of leaping affordances. Acta Physiol (oxf) 158:26–35. https://doi.org/10.1016/j.actpsy.2015.03.010
EFSA Panel on Genetically Modified Organisms: European Food Safety Authority (2010) Statistical considerations for the safety evaluation of GMOs. EFSA J. https://doi.org/10.2903/j.efsa.2010.1250
EpicGames. Unreal Engine. https://www.unrealengine.com/en-US/
Faul F, Erdfelder E, Lang A-G, Buchner A (2007) G* Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods 39(2):175–191. https://doi.org/10.3758/BF03193146
Field A (2013) Discovering statistics using IBM SPSS statistics. Sage Publications, Los Angeles
Finkel L, Engler S, Randerath J (2019) Does it fit?–Trainability of affordance judgments in young and older adults. Plos one 14(2):e0212709. https://doi.org/10.1371/journal.pone.0212709
Finkel L, Schmidt K, Scheib JPP, Randerath J (2019) Does it still fit?–Adapting affordance judgments to altered body properties in young and older adults. Plos one 14(12):e0226729. https://doi.org/10.1371/journal.pone.0226729
Fox JR (2004) A signal detection analysis of audio/video redundancy effects in television news video. Commun Res 31(5):524–536
Franchak JM, van der Zalm DJ, Adolph KE (2010) Learning by doing: action performance facilitates affordance perception. Vision Res 50(24):2758–2765. https://doi.org/10.1016/j.visres.2010.09.019
Franchak JM, Celano EC, Adolph KE (2012) Perception of passage through openings depends on the size of the body in motion. Exp Brain Res 223(2):301–310. https://doi.org/10.1007/s00221-012-3261-y
Gabbard C, Ammar D, Rodrigues L (2005) Perceived reachability in hemispace. Brain Cogn 58(2):172–177. https://doi.org/10.1016/j.bandc.2004.10.001
Gabbard C, Ammar D, Lee S (2006) Perceived reachability in single- and multiple-degree-of-freedom workspaces. J Mot Behav 38(6):423–429. https://doi.org/10.3200/JMBR.38.6.423-429
Gagnon HC, Rohovit T, Finney H, Zhao Y, Franchak JK, Stefanucci J K, Creem-Regehr SH, Bodenheimer B (2021) The effect of feedback on estimates of reaching ability in virtual reality. In: *IEEE* virtual reality and 3D user interfaces (VR). https://doi.org/10.1109/VR50410.2021.00107
Gerig N, Mayo J, Baur K, Wittmann F, Riener R, Wolf P (2018) Missing depth cues in virtual reality limit performance and quality of three dimensional reaching movements. Plos one 13(1):e0189275. https://doi.org/10.1371/journal.pone.0189275
Geuss MN, Stefanucci JK, Creem-Regehr SH, Thompson WB, Mohler BJ (2015) Effect of display technology on perceived scale of space. Hum Factors 57(7):1235–1247. https://doi.org/10.1177/0018720815590300
Geuss, M. N., Stefanucci, J., Creem-Regehr, S., & Thompson, W. B. (2010). Can I pass? Using affordances to measure perceived size in virtual environments. In: Paper presented at the Proceedings of the 7th symposium on applied perception in graphics and visualization.
Gibson JJ (1977) The theory of affordances (Vol. 1),
Goodale M, Jakobson L, Keillor J (1994) Differences in the visual control of pantomimed and natural grasping movements. Neuropsychologia 32(10):1159–1178. https://doi.org/10.1016/0028-3932(94)90100-7
Grechkin TY, Nguyen TD, Plumert JM, Cremer JF, Kearney JK (2010) How does presentation method and measurement protocol affect distance estimation in real and virtual environments? ACM Trans Appl Percept 7(4):1–18. https://doi.org/10.1145/1823738.1823744
Green DM, Swets JA (1966) Signal detection theory and psychophysics, vol 1. Wiley, New York
Harris DJ, Buckingham G, Wilson MR, Vine SJ (2019) Virtually the same? How impaired sensory information in virtual reality may disrupt vision for action. Exp Brain Res 237(11):2761–2766. https://doi.org/10.1007/s00221-019-05642-8
Howard MC, Gutworth MB (2020) A meta-analysis of virtual reality training programs for social skill development. Comput Educ 144:103707. https://doi.org/10.1016/j.compedu.2019.103707
HyLown Consulting LLC (2013–2022) Power and Sample Size.com.
Interrante V, Ries B, Lindquist J, Kaeding M, Anderson L (2008) Elucidating Factors that Can Facilitate Veridical Spatial Perception in Immersive Virtual Environments. Presence Teleoper Virtual Environ 17(2):176–198. https://doi.org/10.1162/pres.17.2.176
Ishak S, Adolph KE, Lin GC (2008) Perceiving affordances for fitting through apertures. J Exp Psychol Hum Percept Perform 34(6):1501. https://doi.org/10.1037/a0011393
Jensen L, Konradsen F (2018) A review of the use of virtual reality head-mounted displays in education and training. Educ Inf Technol 23(4):1515–1529. https://doi.org/10.1007/s10639-017-9676-0
Jerald J (2015) The VR book: human-centered design for virtual reality. Morgan & Claypool, San Rafael
Jun E, Stefanucci JK, Creem-Regehr SH, Geuss MN, Thompson WB (2015) Big foot: using the size of a virtual foot to scale gap width. ACM Trans Appl Percept 12(4):1–12. https://doi.org/10.1145/2811266
Kumar AA (2021) Semantic memory: a review of methods, models, and current challenges. Psychon Bull Rev 28(1):40–80. https://doi.org/10.3758/s13423-020-01792-x
Kwok TC, Chau WW, Yuen KS, Wong AY, Li JC, Shiu RY, Ho FK (2011) Who would benefit from memory training? A pilot study examining the ceiling effect of concurrent cognitive stimulation. Clin Interv Aging 6:83. https://doi.org/10.2147/CIA.S16802
Lakens D, Scheel AM, Isager PM (2018) Equivalence testing for psychological research: a tutorial. Adv Methods Pract Psychol Sci 1(2):259–269
Lang JI, Lang TJ (1988) Eye screening with the Lang stereotest. Am Orthoptic J 38(1):48–50. https://doi.org/10.1080/0065955X.1988.11981769
Lang JI (1983) A new stereotest. In: SLACK incorporated thorofare, NJ
Lin Q, Rieser J, Bodenheimer B (2015) Affordance judgments in hmd-based virtual environments: stepping over a pole and stepping off a ledge. ACM Trans Appl Percept 12(2):1–21. https://doi.org/10.1145/2720020
Lindner P, Miloff A, Zetterlund E, Reuterskiöld L, Andersson G, Carlbring P (2019) Attitudes toward and familiarity with virtual reality therapy among practicing cognitive behavior therapists: a cross-sectional survey study in the Era of consumer VR platforms. Frontiers in Psychology. https://doi.org/10.3389/fpsyg.2019.00176
Linkenauger SA, Bülthoff HH, Mohler BJ (2015) Virtual arm′ s reach influences perceived distances but only after experience reaching. Neuropsychologia 70:393–401. https://doi.org/10.1016/j.neuropsychologia.2014.10.034
Luyat M, Domino D, Noël M (2008) Surestimer ses capacités peut-il conduire à la chute ? Une étude sur la perception des affordances posturales chez la personne âgée. Psychol Neuropsychiatr Vieil 6(4):287–297. https://doi.org/10.1684/pnv.2008.0149
Machin D, Campbell MJ, Tan SB, Tan SH (2011) Sample size tables for clinical studies. John Wiley & Sons, New York

Macmillan NA, Creelman CD (2004) Detection theory: a user's guide. Psychology press, London

Mahncke HW, Connor BB, Appelman J, Ahsanuddin ON, Hardy JL, Wood RA, Nicholas MJ, Tania B, Sharona MA, Merzenich MM (2006) Memory enhancement in healthy older adults using a brain plasticity-based training program: a randomized, controlled study. Proc Natl Acad Sci 103(33):12523–12528. https://doi.org/10.1073/pnas.0605194103

Mascha EJ, Sessler DI (2011) Equivalence and noninferiority testing in regression models and repeated-measures designs. Anesth Analg 112(3):678–687. https://doi.org/10.1213/ANE.0b013e318206f872

Muroi D, Hiroi Y, Koshiba T, Suzuki Y, Kawaki M, Higuchi T (2017) Walking through apertures in individuals with stroke. Plos One 12(1):e0170119. https://doi.org/10.1371/journal.pone.0170119

Oxley JA, Ihsen E, Fildes BN, Charlton JL, Day RH (2005) Crossing roads safely: an experimental study of age differences in gap selection by pedestrians. Accid Anal Prev 37(5):962–971. https://doi.org/10.1016/j.aap.2005.04.017

Parsons TD (2015) Virtual reality for enhanced ecological validity and experimental control in the clinical, affective and social neurosciences. Front Hum Neurosci. https://doi.org/10.3389/fnhum.2015.00660

Pereira C, Bravo J, Veiga G, Marmeleira J, Mendes F, Almeida G (2020) Stepping-forward affordance perception test cut-offs: Red-flags to identify community-dwelling older adults at high risk of falling and of recurrent falling. Plos one 15(10):e0239837. https://doi.org/10.1371/journal.pone.0239837

Randerath J, Frey SH (2016) Diagnostics and training of affordance perception in healthy young adults - implications for post-stroke neurorehabilitation. Front Hum Neurosci 9:674. https://doi.org/10.3389/fnhum.2015.00674

Randerath J, Finkel L, Shigaki C, Burris J, Nanda A, Hwang P, Frey SH (2018) Does it fit?–Impaired affordance perception after stroke. Neuropsychologia 108:92–102. https://doi.org/10.1016/j.neuropsychologia.2017.11.031

Randerath J, Finkel L, Shigaki C, Burris J, Nanda A, Hwang P, Frey S (2021) Is this within reach? Left but not right brain damage affects affordance judgment tendencies. Front Hum Neurosci 14:561. https://doi.org/10.3389/fnhum.2020.531893

Rogers JL, Howard KI, Vessey JT (1993) Using significance tests to evaluate equivalence between two experimental groups. Psychol Bull 113(3):553. https://doi.org/10.1037/0033-2909.113.3.553

Rohrbach N, Krewer C, Löhnert L, Thierfelder A, Randerath J, Jahn K, Hermsdörfer J (2021) Improvement of apraxia with augmented reality: influencing pantomime of tool use via holographic cues. Front Neurol. https://doi.org/10.3389/fneur.2021.711900

Salmaso D, Longoni AM (1983) Hand preference in an Italian sample. Percept Motor skills 57(3_suppl):1039–1042. https://doi.org/10.2466/pms.1983.57.3f.1039

Sanchez-Vives MV, Slater M (2005) From presence to consciousness through virtual reality. Nat Rev Neurosci 6(4):332–339. https://doi.org/10.1038/nrn1651

Thomas BJ, Wagman JB, Hawkins M, Havens M, Riley MA (2017) The independent perceptual calibration of action-neutral and -referential environmental properties. Perception 46(5):586–604. https://doi.org/10.1177/0301006616679172

Tieri G, Morone G, Paolucci S, Iosa M (2018) Virtual reality in cognitive and motor rehabilitation: facts, fiction and fallacies. Expert Rev Med Devices 15(2):107–117. https://doi.org/10.1080/17434440.2018.1425613

Wagman JB, Cialdella VT, Stoffregen TA (2019) Higher order affordances for reaching: perception and performance. Q J Exp Psychol 72(5):1200–1211. https://doi.org/10.1177/1747021818784403

Waller D, Hunt E, Knapp D (1998) The transfer of spatial knowledge in virtual environment training. Presence Teleoper Virtual Environ 7(2):129–143. https://doi.org/10.1162/105474698565631

Williams RL, Chen ML, Hauck WW (2002) Equivalence approaches. Clin Pharmacol Ther 72(3):229–237. https://doi.org/10.1067/mcp.2002.126705

Wilson CJ, Soranzo A (2015) The use of virtual reality in psychology: a case study in visual perception. Comput Math Methods Med. https://doi.org/10.1155/2015/151702

Zhang H (2017) Head-mounted display-based intuitive virtual reality training system for the mining industry. Int J Min Sci Technol 27(4):717–722. https://doi.org/10.1016/j.ijmst.2017.05.005

Zivotofsky AZ, Eldror E, Mandel R, Rosenbloom T (2012) Misjudging their own steps: why elderly people have trouble crossing the road. Hum Factors 54(4):600–607. https://doi.org/10.1177/0018720812447945

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.