Plausibility and Perception of Personalized Virtual Humans between Virtual and Augmented Reality

Erik Wolf*
HCI Group
University of Würzburg

David Mal†
HCI Group
University of Würzburg

Viktor Frohnapfel‡
HCI Group
University of Würzburg

Nina Döllinger§
PIIS Group
University of Würzburg

Stephan Wenninger¶
Computer Graphics Group
TU Dortmund University

Mario Botsch‖
Computer Graphics Group
TU Dortmund University

Marc Erich Latoschik**
HCI Group
University of Würzburg

Carolin Wienrich††
PIIS Group
University of Würzburg

Figure 1: The figure shows a virtual human within each of our three realized conditions. The left image shows the VR condition, the middle one the video see-through AR condition, both taken from Varjo XR-3 HMD screen view. The right image shows a composition of the optical see-through AR condition, originally realized by a Microsoft HoloLens 2 and captured by a DSLR camera.

Abstract

This article investigates the effects of different XR displays on the perception and plausibility of personalized virtual humans. We compared immersive virtual reality (VR), video see-through augmented reality (VST AR), and optical see-through AR (OST AR). The personalized virtual alter egos were generated by state-of-the-art photogrammetry methods. 42 participants were repeatedly exposed to animated versions of their 3D-reconstructed virtual alter egos in each of the three XR display conditions. The reconstructed virtual alter egos were additionally modified in body weight for each repetition. We show that the display types lead to different degrees of incongruence between the renderings of the virtual humans and the presentation of the respective environmental backgrounds, leading to significant effects of perceived mismatches as part of a plausibility measurement. The device-related effects were further partly confirmed by subjective misestimations of the modified body weight and the measured spatial presence. Here, the exceedingly incongruent OST AR condition leads to the significantly highest weight misestimations as well as to the lowest perceived spatial presence. However, similar effects could not be confirmed for the affective appraisal (i.e., humanness, eeriness, or attractiveness) of the virtual humans, giving rise to the assumption that these factors might be unrelated to each other.

*e-mail: erik.wolf@uni-wuerzburg.de
†e-mail: david.mal@uni-wuerzburg.de
‡e-mail: viktor.frohnapfel@stud-mail.uni-wuerzburg.de
§e-mail: nina.doellinger@uni-wuerzburg.de
¶e-mail: stephan.wenninger@tu-dortmund.de
‖e-mail: mario.botsch@tu-dortmund.de
**e-mail: marc.latoschik@uni-wuerzburg.de
††e-mail: carolin.wienrich@uni-wuerzburg.de

Keywords: Mixed reality, immersion, coherence, presence, body weight perception, body image, serious application, uncanny valley

Index Terms: Human-centered computing—Empirical studies in HCI; Mixed / augmented reality; Virtual reality;

1 Introduction

Virtual representations of human beings, often called virtual humans, virtual alter egos (when user-personalized), or avatars (when user-controlled) [4], are integral to various mixed, augmented, and virtual reality applications (MR, AR, VR – XR for short). Examples can be found in various domains like mental health [34, 40], entertainment [30, 61], and education [31, 46]. In the area of mental health, the application of virtual humans for working on body perception in directions like body weight management or body image intervention is particularly promising [15, 16, 22, 59]. In this regard, an accurate and plausible representation and perception of a virtual human is beneficial [64]. Recent work has explored the use of personalized, photorealistic virtual alter egos for this purpose with great success [16, 56, 63]. However, not only the virtual human or virtual alter ego itself but also the mediating display technology and the resulting XR experience can influence the perception of a virtual human. For example, Wolf et al. [67] recently compared perception of embodied but non-personalized avatars between VR, video see-through AR (VST AR), and optical see-through AR (OST AR) and revealed significant perceptual differences in body weight estimations. It is a crucial finding since already a wide variety of display systems (e.g., various VR or AR see-through HMDs, AR projectors, or CAVE-like systems) are used in the research of body perception [66], and as consumer technology advances rapidly, the heterogeneity is expected to rise. While research on the perception between different types of virtual humans themselves [5, 18, 32, 38] or between desktop and VR applications [18–20, 45] have been studied intensively, investigations between different XR display systems seem rather scarce and require more attention [66].

To address this gap, our work explores the plausibility and perception of virtual alter egos between different XR display systems. In our experiment, 42 participants observed their animated, personalized, photorealistic virtual alter ego in a controlled environment using (1) a VR system, (2) a VST AR system, and (3) an OST AR system (see Fig. 1). For elaborating on the display-specific perceptual differences recently discovered by Wolf et al. [67], the body weight of the virtual alter ego was repeatedly changed while the participants had to estimate it. After XR exposure, participants judged the spatial presence felt during the XR experience and rated the virtual alter ego's appearance on further body perception-related measures (i.e., virtual human plausibility and affective appraisal). In the following, we define the characteristics of our display device-specific XR experiences based on recently introduced novel theoretical models [27, 49, 65] and compare the results of our measures between the experiences. We further explore the modification and estimation of virtual humans' body weight as a method for determining display-related differences in virtual human perception. Hence, our work contributes first empirical data to the verification of the theories, applies them to the important application field of virtual human perception, and provides further understanding of different XR experiences differ from each other.

2 Related Work

2.1 Spatial Presence and Virtual Human Plausibility

Skarbez et al. [49] recently introduced a revised version of Milgram's reality-virtuality continuum [36] as a taxonomy for describing XR experiences. While the initial version of Milgram related mainly to the used visual display, the revised version extends the continuum to all external senses and consists of the dimensions (1) immersion, (2) coherence, and (3) extent of world knowledge.

(1) *Immersion* (or rather system immersion as defined by Slater [50]) is determined by a system's objective hardware device specifications. The user's subjective reaction to a display's degree of immersion is the feeling of spatial presence (or place illusion) [49]. In other words, the higher the immersion of a display, the more the user feels really being in a virtual environment [50]. Table 1 summarizes the specifications of our used devices.

(2) *Coherence* refers to the conformity of different sensory information a user perceives during an XR experience. An example of our displays is the realism coherence of the content, which differs significantly between VR (rendered content and rendered environment) and AR (rendered content and real environment) [49]. The user's subjective judgment of coherence leads to the feeling of plausibility of the XR experience (or plausibility illusion) [47, 49].

(3) *Extent of world knowledge* describes the degree of reality a system incorporates (e.g., by remodeled environments or see-through functionality) into an XR experience [36]. The user's subjective reaction to world knowledge could be described as the user's real-world awareness [49]. Since our work focuses on immersion and congruence, we consider Skarbez et al. [48] to control important environmental cues influencing the world knowledge (e.g., constant room-scale, physically plausible objects, similar lighting).

Table 1: The table compares the specifications of our HMDs used. The Varjo XR-3 contains a separate display for the foveal area (27°×27°).

	Varjo XR-3	HoloLens 2
Horizontal FoV	115°	43°
Vertical FoV	80°	29°
Foveal Res.	1920 × 1920 px, 70 PPD	1440 × 936 px, 30 PPD
Peripheral Res.	2880 × 2720 px, 30 PPD	1440 × 936 px, 30 PPD
Refresh Rate	90 Hz	60 Hz
Luminosity	High	Low
Transparency	No	Yes

An alternative theoretical model by Latoschik and Wienrich [27] extends the first two dimensions of Skarbez et al. [49]. It similarly centers on *congruence* but argues that all congruence activations between cognitive, perceptual, and sensory layers (e.g., expectations, experiences, and habits) contribute to the plausibility of an XR experience reflected in various XR-related qualia, such as spatial presence. Therefore, device specification differences, like a larger field of view (FoV) or a higher resolution, lead to a stronger device-specific sensory congruence, while higher content transparency of virtual objects would cause incongruence. The device-specific congruencies impact the plausible generation of spatial cues and ultimately the XR qualia spatial presence. Mal et al. [33] further elaborated on the plausibility of virtual humans and highlighted (1) the virtual human's appearance and behavior congruence, which defines whether a virtual human appears and behaves plausibly within the environment, and (2) the virtual human's congruence with the virtual environment, which defines whether the virtual human's appearance and behavior are plausible in relation to the environment, as two factors to consider.

For our study, we realized the VR and VST AR conditions by a Varjo XR-3, while we used a HoloLens 2 for OST AR (see Table 1). In all immersion-related properties except peripheral resolution, the XR-3 can be considered to provide a higher device-specific congruence, which leads in both introduced models to a higher spatial presence. For the virtual human plausibility, we argue that the differences between VR and AR are more crucial and expect that, for example, a more congruent rendering of the content will lead to a higher virtual human plausibility. With regard to the practical relevance of the presented models, Wienrich et al. [65] also recently highlighted spatial presence and plausibility as two major dimensions for quantifying the overall quality of an XR experience. For the practical application of XR in the areas like mental health, a high overall quality of an XR experience has been introduced as a necessary hygiene factor for achieving desired effects [64]. Therefore, we consider spatial presence and plausibility as important factors for our work and formulate the following research question:

RQ1: How do differently immersive and congruent displays affect spatial presence and virtual human plausibility?

2.2 Virtual Human Perception

We examine the influence of different immersive and congruent displays on the virtual human perception-related measures of body weight perception and affective appraisal. Following the introduced theoretic model of Latoschik and Wienrich [27], the participants' cognitive and perceptual experiences regarding a virtual human can influence the interpretation of its plausibility and its perception. Hence, the kind of virtual human presentation is crucial when measuring virtual human perception. For example, the presentation of a generic virtual human could lead to a subjective interpretation regarding its generic appearance. To avoid participants' subjective interpretations and to achieve better comparability between display conditions, it suggests using personalized virtual humans as stimuli.

In addition, the observation perspective on a virtual human can affect its perception. For example, Neyret et al. [39] compared the impact of the perspective on virtual human perception and emphasized the importance of a third-person presentation for a more unbiased judgment. This includes the embodiment, which involves a change of observational perspective of a virtual human and implicitly manipulates its assessment. Recent work showed, for example, that the embodiment of a generic virtual human can lead to an altered body weight perception [68] or recognition of body weight changes [25]. Using embodied avatars could further lead to an uncontrolled exposition with the avatar, as the observation perspective on the body changes according to the participants' movements, leading to highly individual impressions. Here, prior work has clearly highlighted the importance of providing a holistic picture of the body in body perception research [13,58]. Hence, to keep the presentation of virtual humans as stable as possible, it suggests presenting them from a third-person perspective without embodiment as various prior works did [22, 42, 56–58]. Through the use of non-embodied virtual alter egos, we expect to decrease interindividual differences in body weight perception by providing all participants a controlled reference template for their judgments [12].

2.2.1 Body Weight Perception

Applications in the area of mental health like body image interventions can benefit greatly from the use of virtual humans in XR [16, 22, 59]. They also show huge potential in the research of body perception [55]. We further suggest that body weight estimates can also serve as a measure for evaluating display-specific perceptual differences. Wolf et al. [66] recently summarized that systems used in body weight perception-related works differed widely in their implementation, including the display type and the conveyed XR experience and raised the question of whether a system's implementation might influence results of investigations and their interpretation. Indeed, prior work compared body perception of a photorealistic but generic embodied virtual human between a VR, a VST AR, and an OST AR display and found highly significant differences of up to 8.5 % in body weight estimations between the display conditions [67]. What appears to be a disadvantage in the practical application of XR in the area of mental health, namely system-related differences, may prove to be advantageous when investigating the effects of different display types on the perception of virtual humans. Therefore, we are investigating the use of body weight estimates in this direction and formulating the following research question for our work:

RQ2: How do differently immersive and congruent displays affect the perception of a virtual alter ego's body weight?

2.2.2 Affective Appraisal

Another part of the virtual human perception to be considered is their affective appraisal, especially regarding the so-called uncanny valley effect [37]. The effect describes the paradoxical reaction that the perception of a virtual human changes from pleasantness to eeriness as soon as the virtual human's appearance approaches but does not fully reach a convincing human-like appearance [38]. The feeling of uncanniness towards a virtual person is thereby determined by its human likeness and the affinity towards its observer. Since our and similar work [16] implements photorealistic and personalized virtual alter egos, which should be fairly close to a human-like appearance, the question arises whether the differences between our XR experiences and their presumed effects on the congruence of the virtual alter ego could also influence its affective appraisal. While there has been a great deal of research on different types of virtual humans (e.g., depending on their anthropomorphism [8, 32], reconstruction method [5], stylism [18, 19], and many more) presented by the same display type, there seems to be far less research on effects triggered by differently immersive displays and incongruent presentations. In a work comparing the perception of virtual humans between desktop and VR, Roth and Wienrich [45] could not find differences in the uncanny valley relevant measures of humanness, attractiveness, and eeriness [21]. Hepperle et al. [19] found that an uncanny valley effect is more likely to appear in VR than on a desktop screen by employing a similar comparison. However, they could only find greater differences for virtual humans judged to be within the uncanny valley. To our knowledge, there is no work comparing the affective appraisal of virtual humans regarding the uncanny valley between a VR, an VST AR, and an OST AR display. Hence, we are formulating the following research question:

RQ3: How do differently immersive and congruent displays affect the affective appraisal of a virtual alter ego?

3 METHOD

A detailed ethics proposal following the Declaration of Helsinki was submitted to the ethics committee of the Institute Human-Computer-Media (MCM) of the University of Würzburg and found to be ethically unobjectionable. During the acquisition and evaluation process of the study, freely available support offers from the Anorexia Nervosa and Associated Disorders organization (ANAD) were explicitly highlighted in case participants would feel uncomfortable regarding their body weight after the study.

3.1 Participants

A total of 42 participants (23 female, 19 male) were recruited from the university's participant management system and either received 15 € or student credit points equal to the participation time. Individuals could not register if they (1) had no normal or corrected to normal vision and hearing, (2) had less than ten years of experience with the German language, (3) currently suffered from a diagnosed mental, psychosomatic, or body weight disorder, or (4) had a known sensitivity to simulator sickness. Additionally, one participant had to be excluded after participation due to technical problems during the experiment. 31 participants were students of the local university. Eight of the participants had less than one hour of XR experience, 32 used XR for one to twenty hours, and two used XR already for more than twenty hours. More descriptive data can be found in Table 2.

Table 2: The table shows age and body measurements of our sample.

	Range	*M* (*SD*)
Age	19 – 64	26.21 (10.03)
Body height (m)	1.56 – 1.91	1.73 (0.09)
Body weight (kg)	45.1 – 123.8	71.64 (18.28)
BMI	16.56 – 35.79	23.81 (3.42)

3.2 Experimental Task

The participants' experimental task was to observe their previously generated personalized virtual alter ego moving in the virtual environment while sitting in a fixed position within the laboratory. The observation phase consisted of nine cycles in which participants had to judge the virtual alter ego concerning our measures explained below. In each observation cycle, the virtual alter ego walked with a different (modified) body weight about 1.2 m into the room, posed from the front and both sides, and left the room again. Body weight modifications ranged ±20%, split into 5% intervals, and were performed in a counterbalanced manner. In total, the virtual alter ego was visible for 32 s per cycle and provided a holistic picture of itself during this time, as suggested by prior work [13, 16, 58].

3.3 Design

Our study followed a 3×1 within-subjects design with *display type* being the independent variable. Hence, each participant performed the experimental task using the *VR*, *VST AR*, and *OST AR* display. The order was counterbalanced. As dependent variables, we captured the participants' feeling of *spatial presence*, perceived *virtual human plausibility*, *body weight perception*, and *affective appraisal*. On an exploratory basis and without prior hypotheses, we investigated the influence of the performed body weight modifications and the participants' gender on body weight perception. Additionally, we monitored simulator sickness-related symptoms before and after exposure. The operationalization of the variables will be explained below.

3.4 Measures

Spatial Presence We used the *ITC-Sense of Presence Inventory* (ITC-SOPI) [28] to test whether and to what extent our manipulation of the display type affected participants' feeling for spatial presence. The questionnaire was developed to capture differences cross-media and consists of four presence-related sub-dimensions, from which we only took spatial presence (SP). The averaged scores were taken from 19 items and range from 1 to 5 (*5 = highest SP*).

Virtual Human Plausibility We assessed the presented virtual alter egos' plausibility using the *Virtual Human Plausibility Questionnaire* (VHPQ) [33]. It captures (1) the virtual humans' appearance and behavior plausibility (ABP) and (2) the virtual humans' match to the virtual environment (MVE) using 11 different items, which are all rated on a scale from 1 to 7 (*7 = highest plausibility*).

Affective Appraisal We measured the participants' affective appraisal of the virtual alter ego using the revised version of the *Uncanny Valley Index* (UVI) [21]. It includes the three sub-dimensions of *humanness* (H), *eeriness* (E), and *attractiveness* (A). The scales examine anthropomorphic properties of the virtual alter ego and range from 1 to 7 (*7 = highest H, E, A*).

Body Weight Perception Our body weight estimations followed the idea of prior work [16, 66–68] and served as the operationalization of participants' perception of the virtual alter egos' body weight. Participants had to numerically estimate the presented virtual alter ego's body weight in kg in each observation cycle. We used the estimations to calculate the misestimation M for each body weight modification as $M = \frac{e-p}{p}$, where e is the estimated body weight, and p is the presented body weight. A negative value states an underestimation of the body weight and a positive value an overestimation. Based on the misestimations M, we calculated the average body weight misestimation (BWM) $\overline{M} = \frac{1}{n}\sum_{k=1}^{n} M_k$ over all observation cycles n. As an indicator of the difficulty and uncertainty of participants' individual body weight estimations, we further calculated the average percentage of absolute misestimation as $\overline{A} = \frac{1}{n}\sum_{k=1}^{n} |M_k|$.

3.4.1 Simulator Sickness

To control whether our used displays systematically provoked simulator sickness (e.g., by latency of the VST cameras or general latency jitter [53, 54]), we captured whether and to what extent participants experienced simulator sickness-associated symptoms. Participants assessed the 16 items of the *Simulator Sickness Questionnaire* (SSQ) [6, 26] before performing the first condition of the study and after the last one. The total score of the questionnaire ranges from 0 to 235.62 (*236 = strongest simulator sickness*). An increase in the score by 20 between a pre- and post-measurement indicates the occurrence of simulator sickness [51].

3.5 Apparatus

3.5.1 Hard- and Software

The virtual environment was implemented using Unity version 2020.3.11f1 [60]. It ran on a powerful workstation that consisted of an Intel Core i7-9700K CPU, a NVIDIA GeForce RTX 2080 Ti, and 32 GB RAM. We further provided participants with an ordinary office workstation equipped with keyboard, mouse, and 24-inch LCD screen, which they used to answer questionnaires outside of XR presented using LimeSurvey 4 [29].

VR and VST AR Our study's VR and VST AR conditions were realized using a Varjo XR-3 HMD [62]. The technical specification of the display can be found in Table 1, left. The absolute position of the HMD was tracked by four SteamVR Base Stations 2.0. In the VST AR display mode, the real environment was captured by two 12 Mpx VST low latency cameras running at 90 Hz. According to the manufacturer, the recorded content is displayed on the HMD with a latency of < 20 ms. Since this low latency is achieved by using hardware-accelerated integration on the device directly, we could not verify the latency using trivial latency measurement methods [52, 54]. To integrate the XR-3 into our application, we used Varjo's Unity XR plugin [1] in conjunction with Varjo Base, both in version 3.2.0.

OST AR The OST AR condition of our study was realized using a Microsoft HoloLens 2 [35]. The technical specification of the display can be found in Table 1, right. The absolute position was tracked using the built-in inside-out tracking. In our evaluation, the HoloLens 2 was connected via 100 MBit/s ethernet to the previously mentioned high-end PC used to render the content via Holographic Remoting. For integrating the HoloLens 2 into our application, we used Microsoft's Mixed Reality Toolkit (MRTK) [2] in version 2.7.0.

3.5.2 Environment

To keep the extent of world knowledge on our dependent variables across VR and AR as constant as possible, we aimed for a virtual environment similar to the real environment in the VR condition [48]. To this end, we created a 3D model of the real-world laboratory in which the experiment was conducted (see Fig. 1). To ensure that the virtual and real environments were properly aligned, an environmental calibration using a predefined anchor point in both environments was performed by putting the HMD on this point. In both the virtual and real environment, we left the laboratory door open during the study, allowing the virtual alter ego to leave the room during a weight change to increase the overall plausibility. By masking the room's door in the aligned virtual environment as a pass-through object[3], we could realize that the real door also occluded the virtual alter ego in the AR conditions. Since the experimenter, who was also positioned in the laboratory during the study, could have been observed in the AR conditions, we concealed him and the used PC with two fabric walls. In consequence, participants could only see the static objects of the environment. While the whole virtual environment was rendered in the VR condition, only the virtual alter ego was shown in the AR conditions (see Fig. 1).

3.5.3 Virtual Alter Ego

Generation To generate the personalized virtual counterpart of the participants, we use the method proposed by Achenbach et al. [1]. A custom-built multi-DSLR-camera setup (see Bartl at al. [5], Figure 1, top) produces the input images for a multi-view stereo reconstruction step resulting in a dense point cloud of the scanned subject. Pose and shape parameters of a statistical model of human shape variation are optimized to fit the scanner data. A final non-rigid deformation step ensures a closer match to the scanner data, as the statistical model parameters alone cannot completely explain the observation. The model is based on a fully rigged template mesh from the Autodesk Character Generator [3], resulting in a virtual alter ego fully compatible with the common XR engines like Unity.

Animation We imported the generated virtual alter egos into Unity using a custom FBX-based runtime loader. It automatically generates a fully rigged, humanoid virtual character object immediately ready for animation. During our study, the virtual alter egos were animated using Unity's built-in character animation system playing pre-recorded humanoid animations. The animations were recorded using the system of Wolf et al. [66]. By using FinalIK's Humanoid Baker [44], the movements were created directly as Unity-compatible animations.

Modification For modifying the body weight of the virtual alter ego at runtime, we follow the method described by Döllinger et al. [16]. They build a model of human shape variation based on Principal Component Analysis (PCA) by non-rigidly registering a template mesh to a subset of the CAESAR database [43], which consists of 3D scans with corresponding anthropometric measurements.

[1] https://github.com/varjocom/VarjoUnityXRPlugin

[2] https://github.com/microsoft/MixedRealityToolkit-Unity

[3] https://developer.varjo.com/docs/unity-xr-sdk/masking-with-varjo-xr-plugin

Figure 2: The figure shows an exemplary generated virtual alter ego with modified body weight of $\mathrm{BMI} = 22$ (left) and $\mathrm{BMI} = 32$ (right) on the same pose within our virtual environment.

Learning the correlation between the measurements and the PCA subspace allows expressing the desired change in body weight as a change in the subspace, which can be used to reconstruct a modified mesh for the virtual alter ego. Improving on similar approaches for body weight modification of virtual alter egos [42], Döllinger et al. [16] additionally keep a small area of the face region fixed to preserve the virtual alter ego's identity better. Fig. 2 compares an exemplary generated virtual alter ego with a modified body weight.

3.6 Hypotheses

Considering the above-presented literature and the concrete implementation of our experimental conditions, we formulate operationalized hypotheses for each of our variables. As detailed in the last paragraph of Sect. 2.1, we expect for spatial presence and avatar plausibility (RQ1) that our VR and VST AR (using the XR-3) conditions have a similar degree of immersion while the OST AR (using the HoloLens 2) provides a lower degree of immersion. We further expect the VR display to provide a more congruent experience than both of our AR displays. Supported by empirical works comparing the feeling of presence between different XR experiences [9, 14, 63], we propose the following operationalized hypotheses:

- H1.1: Participants will report a higher ITC-SOPI SP score when using the more immersive VR and VST AR displays than when using the less immersive OST AR display.
- H1.2: Participants will report no different VHPQ ABP scores when using the VR, VST AR, and OST AR display.
- H1.3: Participants will report a higher VHPQ MVE score when using the more congruent display (VR) than when using the less congruent ones (VST AR and OST AR).

As highlighted in Sect. 2.2.1, we presume for body weight perception (RQ2) an influence of the display used based on previous work [66, 67] and formulate the following undirected hypothesis:

- H2.1: Participants' body weight misestimations $\overline{M}$ of the observed virtual alter egos will differ between the used VR, VST AR, and OST AR displays.

Following our argumentation in Sect. 2.2.2, we formulate for the affective appraisal of the virtual alter egos (RQ3) our hypothesis based on the performed comparisons between desktop and VR systems [19, 45] and expect:

- H3.1: Participants will report no different UVI H scores when using the VR, VST AR, and OST AR display.
- H3.2: Participants will report no different UVI A scores when using the VR, VST AR, and OST AR display.
- H3.3: Participants will report no different UVI E scores when using the VR, VST AR, and OST AR display.

3.7 Procedure

Our study followed the procedure visualized in Fig. 3. It was divided into a body scan and exposure session, which were performed on two different appointments. The time between the two sessions ranged from 10 minutes to a maximum of seven days. If the sessions were performed on two different days, participants were asked to wear the same clothing for both sessions.

In the body scan session, participants were first informed about the local COVID-19 regulations, received information about the scans process, gave their consent, and generated a personal pseudonymization code to store the captured data. Afterwards, participants answered demographic questions and further questions about their prior VR and AR experiences. Lastly, the experimenter measured the participant's body height and body weight and performed the body scan as explained in Sect. 3.5.3 without shoes. The whole body scan session took on average 23 min.

In the exposure session, participants first received information about the following exposure, gave their consent, generated a personal pseudonymization code for storing the data collected during the study and answered the pre-SSQ. Afterwards, the exposure phase for each display type followed in a counterbalanced order. The experimenter explained the corresponding HMD, made sure the participants wore it correctly, and started a test sequence that presented all relevant information using pre-recorded audio and text instructions. It further triggered the animations and body weight modifications of the virtual alter ego. Participants performed the experimental task explained in Sect. 3.2. Hence, they estimated the body weight nine times and answered the ITC-SOPI, UVI, and VHPQ directly afterwards. After finishing all three conditions, the participants answered the post-SSQ and could leave further comments on their body weight estimation strategy and the study itself. The entire exposure session took on average 58 min.

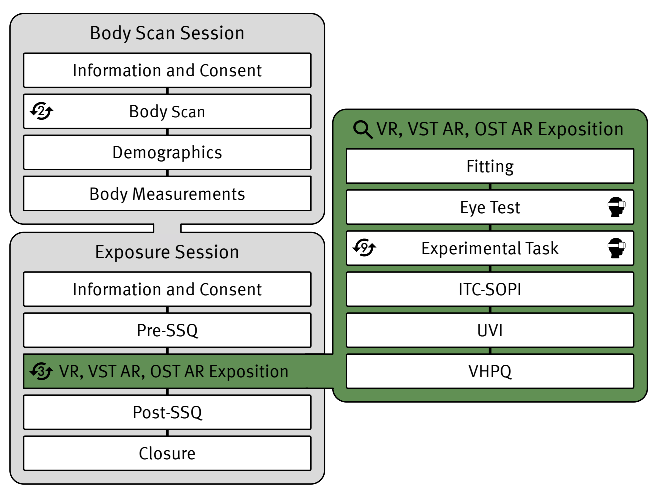

Figure 3: The figure shows the experimental procedure of our study.

Table 3: The table shows the descriptive values together with the test statistics for each measurement compared between the different display types. Single-asterisks indicate significant and double-asterisks highly significant p-values.

	Range	VR	VST AR	OST AR	Test statistics
		M (SD)	M (SD)	M (SD)	
ITC-SOPI SP	[1 – 5]	2.58 (0.65)	2.61 (0.64)	2.38 (0.68)	$F(1.583, 82) = 3.768, p = .038, \eta_p^2 = 0.084^*$
VHPQ ABP	[1 – 7]	5.23 (0.80)	5.13 (0.77)	4.93 (0.87)	$\chi^2(2) = 3.768, p = .152, W = 0.045$
VHPQ MVE	[1 – 7]	5.82 (0.78)	5.26 (0.98)	4.81 (1.21)	$\chi^2(2) = 21.319, p < .001, W = 0.254^{**}$
BWM $\overline{\text{M}}$		-0.11 (4.86)	-0.20 (6.05)	2.06 (5.08)	$F(2, 82) = 9.956, p < .001, \eta_p^2 = 0.195^{**}$
UVI H	[1 – 7]	3.61 (1.17)	3.45 (1.11)	3.35 (1.15)	$F(1.575, 82) = 2.646, p = .091, \eta_p^2 = 0.061$
UVI A	[1 – 7]	4.53 (0.87)	4.49 (0.86)	4.55 (0.89)	$F(1.540, 82) = 0.452, p = .587, \eta_p^2 = 0.011$
UVI E	[1 – 7]	3.61 (0.78)	3.57 (0.82)	3.60 (0.75)	$F(1.592, 82) = 0.127, p = .834, \eta_p^2 = 0.003$

4 Results

The statistical analysis of our data was performed using SPSS version 27.0.1.0 [24]. Before comparing our conditions, we performed tests for normality and sphericity of our dependent variables to check whether the prerequisites for parametric testing were met. While normality and sphericity were given for BWM $\overline{\text{M}}$, the assumption of sphericity was violated for ITC-SOPI SP, UVI H, UVI A, and UVI E. For each of the priorly mentioned variables, we calculated a repeated-measures ANOVA using Greenhouse–Geisser adjustment where necessary to test for differences between our three groups. Due to a violation of normality, we calculated a Friedman-test for each of ITC-SOPI EVN, VHB ABP, and VHB MVE. For all variables discovering significant differences between groups, we decided to calculate separate post-hoc tests. All tests were performed against an α of .05. For directed hypotheses we calculated one-sided tests, for undirected two-sided tests [17]. The descriptive values and the statistical tests of the comparisons are summarized in Table 3. Any calculated post-hoc tests or further exploratory analyses can be found in the corresponding sections of the measurements below. On an exploratory basis, we examined differences between gender and VR experience for all measures and found no significant differences.

4.1 Spatial Presence and Plausibility

Confirming hypothesis H1.1, the comparison of ITC-SOPI SP data showed significant differences between our conditions. The performed one-tailed paired-sample post-hoc t-tests revealed significant differences between VR and OST AR, $t(41) = 1.80, p = .040, d_z = 0.28$, and between VST AR and OST AR, $t(41) = 2.66, p = .006, d_z = 0.41$. No significant difference was found in a two-tailed comparison between VR and VST AR, $t(41) = 0.42, p = .676, d_z = 0.06$. All results for spatial presence and plausibility are shown in Fig. 4.

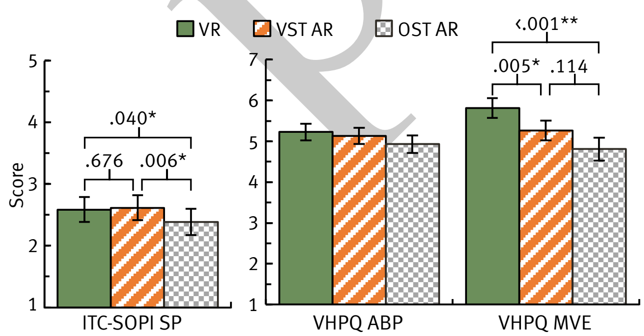

Figure 4: The chart shows the ITC-SOPI and the VHPQ scores for each condition. Error bars represent 95 % confidence intervals. Single-asterisks indicate significant and double-asterisks highly significant p-values.

In line with our hypothesis H1.2, the calculated comparison of the VHPQ ABP data showed no significant differences between our conditions (see Table 3). Hence, we calculated no post-hoc tests.

Confirming hypothesis H1.3, the calculated comparison of the VHPQ MVE data showed significant differences between our conditions. Pairwise one-tailed post-hoc comparisons using Dunn's test revealed significant differences between VR and VST AR, $z = 2.56, p = .005, r = 0.40$, and between VR and OST AR, $z = 4.15, p < .001, r = 0.64$. No significant difference was found in a two-tailed comparison between VST AR and OST AR, $z = 1.58, p = .114, r = 0.24$.

4.2 Body Weight Perception

Regarding our hypothesis H2.1, the calculated comparison of the participants' misestimations of their virtual alter ego's body weight ($\overline{M}$) showed significant differences between our conditions. The calculated two-tailed paired-sample post-hoc t-tests revealed no significant differences between VR and VST AR, $t(41) = 0.17, p = .865, d_z = 0.03$, but between VR and OST AR, $t(41) = 3.63, p < .001, d_z = 0.56$, and between VST AR and OST AR, $t(41) = 3.67, p < .001, d_z = 0.57$. As we found no significant differences between VR and VST AR, we did not accept our hypothesis H2.1. A further calculated post-hoc t-test showed that body weight estimations differed significantly from zero in the OST AR condition, $t(41) = 2.62, p = 0.012, d = 0.40$.

On an exploratory basis, we further investigated the body weight misestimations $\overline{M}$ with regard to the performed body weight modifications and gender differences. To this end, we added the modification level (±20% in 5% intervals) as a within-subject factor and gender (female and male) as a between-subject factor to the repeated measures ANOVA calculated for H3.1. Test results showed no significant main effect for the modification level, $F(8, 320) = 1.33, p = .230$ (see Fig. 5). With regard to gender differences, no significant

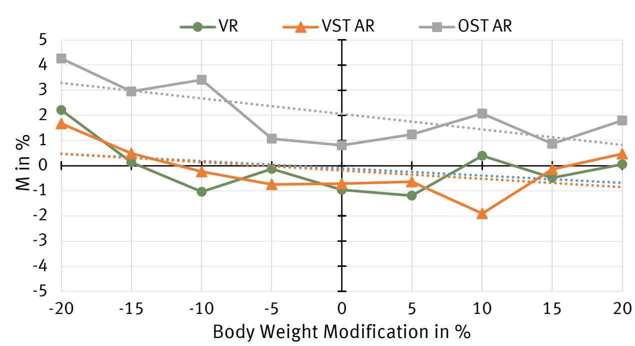

Figure 5: The chart shows the body weight misestimations M depending on the performed body weight modification of the virtual alter ego for each condition.

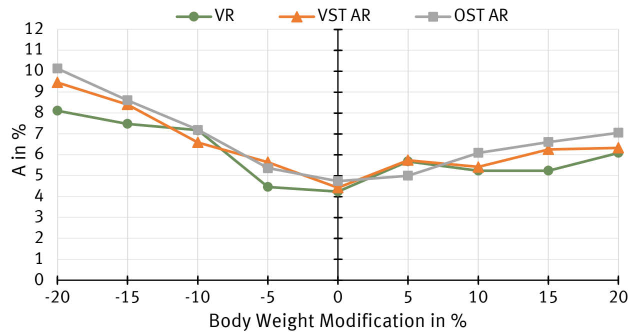

Figure 6: The chart shows the absolute body weight misestimations A depending on the performed body weight modification of the virtual alter ego for each condition.

main effect was found, $F(1,40) = 0.51, p = .638$. Furthermore, no significant interaction effects were found within the entire model.

In addition, we explored absolute body weight misestimations $\overline{A}$ with regard to the performed body weight modification and gender differences. Following the exploration approach of $\overline{M}$, we calculated a repeated-measures ANOVA with the within-subject factors display type (VR, VST AR, and OST AR) and modification level ($\pm 20\%$ in 5% intervals) and the between-subject factor gender (female and male). Results showed no significant main effect for display type, $F(2,80) = 2.05, p = .135$, but a significant main effect for the modification level, $F(8,328) = 8.65, p < .001$ (see Fig. 6). With regard to gender differences, no significant main effect was found, $F(1,40) = 0.34, p = .562$. Furthermore, no significant interaction effects were found within the entire model.

To further explore the significant differences between modification levels, we averaged the absolute body weight misestimations $\overline{A}$ across the display type and split the modification levels in a high negative modification group ($-10\%, -15\%, -20\%$), a low modification group ($+5\%, 0\%, -5\%$), and a high positive modification group ($+20\%, +15\%, +10\%$). The calculated two-tailed paired-sample post-hoc t-tests revealed significant differences between the high negative modification ($M = 8.13, SD = 4.07$) and the low modification group ($M = 5.02, SD = 2.43$), $t(41) = 5.57, p < .001$, and between the high negative modification and the high positive modification group ($M = 6.03, SD = 3.44$), $t(41) = 2.60, p = .013$, but not between the low modification and the high positive modification group, $t(41) = 1.89, p = .066$.

4.3 Affective Appraisal

In line with H2.1, H2.2, and H2.3 on the affective appraisal of the virtual alter egos, we could not find significant differences between VR, VST AR, and OST AR for the virtual alter egos' humanness (H2.1), attractiveness (H2.2), and eeriness (H2.3). Hence, we calculated no post-hoc tests.

4.4 Simulator Sickness

To control the influence of our different XR displays on simulator sickness-related symptoms, we compared SSQ pre- and post-measurements (descriptive values in Table 2) using a two-tailed Wilcoxon signed-rank test as the normality pre-requirement for parametric testing was violated. The SSQ ratings did differ significantly between pre-measurement ($M = 11.30, SD = 15.85$) and post-measurement ($M = 16.56, SD = 17.27$), $Z = 2.61, p = .009$. However, the observed increase in SSQ scores of 5.26, as well as the absolute post-SSQ score, were below the 20 points indication threshold for the occurrence of simulator sickness [51].

5 Discussion

5.1 Spatial Presence and Virtual Human Plausibility

Our study presented participants a content-like XR experience using three different displays. While the VR and VST AR experience was realized using a Varjo XR-3, we used a Microsoft HoloLens 2 for the AR OST experience. We assumed that the device with a potentially higher level of immersion (XR-3) would also induce a higher spatial presence (H1.1). Our results fully confirm our assumption since we showed a significantly higher spatial presence for VR and VST VR than for OST AR, but no differences between VR and VST AR. However, when comparing our results to similar prior work, it does not seem to be necessarily the case that the same device provides a similar level of spatial presence in VR and VST AR mode. In Wolf et al. [66], the researchers used an HTC Vive Pro for both conditions and reported a significantly higher spatial presence for VR than for VST AR. A explanation could be the quality of the video see-through implementation. While the XR-3 uses two high-resolution and low-latency see-through RGB cameras (12 Mpx, 90 Hz), the Vive Pro has a lower resolution and refresh rate (0.3 Mpx, 60 Hz). Hence, the technical implementation of the see-through functionality might affect the spatial presence a device provides regardless its display properties. This shows that determining the degree of immersion of a multi-XR device is not straightforward. Therefore, the concept of device-specific congruence [27] should continue to be specified and operationalized as an extension of immersion.

By using the VHPQ, we confirmed our hypothesis H1.2 and found no significant differences in the virtual alter ego's appearance and behavior plausibility. It suggests that the incongruence of having a rendered virtual alter ego within a real environment not clearly affects the internal consistency of a virtual alter ego's appearance and behavior. We further could confirm our hypothesis H1.3 by showing a significant impact of the virtual alter ego's match to the virtual environment between our VR and AR conditions. Hence, the priorly described mismatch in rendering realism is likely to affect the alter ego's plausibility. In VR, the congruence between the generation of virtual human and environment (both synthetic) resulted in a higher avatar plausibility, while the higher incongruence in the AR conditions (synthetic avatar vs. captured environment) led to a significant lower avatar plausibility. Considering the descriptive differences between VST AR and OST AR, we even assume that device-specific congruencies (captured environment vs. real environment) could have impacted the avatar plausibility. However, there was no significant statistical difference.

To summarize on RQ1, we can fully confirm our previous assumptions about the influence of different congruencies on spatial presence and avatar plausibility. Thus, we also can confirm the respective parts of Latoschik and Wienrich's novel theoretical model. Our results further are in line with Skarbez et al.'s [49] recently introduced taxonomy for describing XR experiences. However, during our study design, we noted that an update and revalidation of relevant measurement tools in line with recently updated concepts [27,49,65], like Mal et al. [33] already did for avatar plausibility, seems timely. Here, it seems worth considering the work of Brühbach et al. [7], which was published after the conception of our study.

5.2 Virtual Human Perception

5.2.1 Body Weight Perception

Although we could not fully confirm our initial hypothesis H2.1, our results on body weight perception have various valuable implications. We could confirm prior findings of Wolf et al. [67] and show significant differences between VR and VST AR in comparison to OST AR. While our initial hypothesis was formulated under the assumption that both the displays' immersion and the provided congruence of the XR experience might impact body weight perception, our findings point to immersion as the main moderator. Our work and Wolf et al. [67] found significant differences between the used HMD devices (here XR-3 vs. HoloLens 2 and in Wolf et al. Vive Pro vs. HoloLens 2). It suggests that explicitly hardware-specific factors (e.g., built-in lenses, FoV, resolution of the display and its luminosity, or transparency of the rendered content) contribute to the differences in perception. Adams et al. [2] suggested similar as they found differences in distance perception between XR-3 and HoloLens 2. It raises the question of whether distorted distance perception might lead to a distorted body weight perception. Future research needs to address this question.

Exploring body weight perception regarding the performed body weight modification and gender provided further valuable insights. First, we showed that the body weight modification of the virtual alter ego had no significant effect on the estimators' body weight misestimations M. This is in line with Wolf et al. [68], who showed a significant moderation of body weight misestimations by the body weight difference between estimator and avatar only when the estimator embodied the avatar. Therefore, we consider our decision not to employ full-body illusions to be justified.

Second, while we observed differences in body weight misestimations $\overline{M}$ (averaging over- and underestimations across all nine modifications) between displays, we could not find differences in the absolute body weight misestimations $\overline{A}$ (considering only the absolute magnitude per misestimations). In other words, the inaccuracy of individual estimations appears constant between displays, whereas the overall impression of body weight seems to differ. Hence, we assume that display-related body weight misperceptions are unlikely dependent on the stimulus's body weight (modification).

Third, the absolute body weight misestimations $\overline{A}$ follow a pattern independent of the used display, as they become descriptively less accurate with increasing body weight deviation between stimulus and observer. This observation is consistent with the theory of contraction bias [11, 12], which states that body weight estimates are the most accurate around a subjective mental reference model of a body. For personalized, photorealistic virtual alter egos, it could be the own body since estimates are most accurate around there. Considering Weber's Law [12], which assumes that changes in body weight are more difficult to detect when body weight increases, it is surprising that the contraction bias is significantly more pronounced in the case of a negative body weight modification.

Fourth, we showed no significant gender differences in the body weight estimations. Prior work repeatedly reported gender differences generally [10, 23, 41] but also in the context of virtual alter egos [57]. As the prior work often refers to samples with eating- or body weight disorders, we attribute the differences to our sample.

5.2.2 Affective Appraisal

Regarding the affective appraisal of the virtual alter egos depending on our differently immersive and congruent XR experiences, we confirm our hypotheses H3.1, H3.2, and H3.3. Although the expected manipulation of the spatial presence of the XR experience and the plausibility of the virtual alter ego was successful, we found no influence of the manipulation on the perceived humanness, attractiveness, and eeriness. Therefore, we assume that the differences in immersion and congruence between our XR experiences are unlikely to affect the affective appraisal of our virtual alter egos. These results are not unexpected since previous work has found no differences between even more differently immersive systems for virtual humans perceived to be outside the uncanny valley [19]. When considering the absolute judgments of affective appraisal in our study, especially for uncanniness, we would not locate the perception of our virtual alter ego in the uncanny valley. Nevertheless, our comparison offers novel insights regarding the perception of virtual alter egos since we are not aware of any other work investigating their affective appraisal between differently congruent XR experiences. Further research needs to confirm our findings.

5.3 Limitation and Future Work

Our work has several limitations that need to be addressed in future work. First, we could confirm our assumptions regarding the virtual alter ego's match to the virtual environment between our conditions (H1.2). However, since the avatar's shadow was only rendered in the VR conditions, we had a confound between VR and AR, which was not directly due to the display type. Therefore, future work will need to address the role of shadow casting in avatar plausibility.

Second, we controlled the extent of world knowledge as best as possible between conditions. Nevertheless, there still were differences, such as the lack of a representation of the own body in VR while it was visible in AR, that might have affected the congruence of the XR experiences further. Future work should therefore include the influence of the extent of world knowledge.

Third, we decided to use personalized virtual alter egos to control for various factors influencing virtual human perception. However, using alter egos instead of generic virtual humans may also impact perception due to the individual's body image. While it seems almost impossible to control all possible influencing factors in a single experimental design, future work should continue to address the influence of personalization.

Fourth, we assessed the affective appraisal of the virtual alter ego on the sub-dimensions of humanness, eeriness, and attractiveness and found no differences between displays. However, the appraisal of an experience is certainly not limited to these three factors but may also include factors such as emotional response. Here, previous work provides indications of differences in dependence on the display technology used [63]. Hence, future work should also consider other factors like the emotional appraisal.

Finally, we identified immersion-related factors as possible reason causing differences in body weight perception but could not pinpoint a specific factor with our device-based manipulation. Hence, our results are bound to the properties of our tested devices (see Table 1). Future work should investigate the influence of distinct properties of immersion. Furthermore, it should be ruled out that immersion affects body weight perception via differently pronounced distance compression effects between different display devices.

6 Contribution and Conclusion

The rapid technical development of XR HMDs leads to a variety of differently immersive and congruent XR experiences. As a consequence, research faces the challenge that gained knowledge from one kind of experience might not be transferable to another one. Our work has addressed this challenge for the perception of virtual humans. To the best of our knowledge, we are the first to explore the plausibility and affective appraisal of virtual humans between different XR experiences. Furthermore, we confirm assumptions that body weight perception can be highly distorted between different XR HMDs by adding a comprehensive study design addressing the limitations of prior work. Hence, we suggest body weight perception as an interval scaled measure offering a nuanced perspective on possible display distortion effects beyond item-based questionnaires like those used for spatial presence and avatar plausibility. Although we worked with virtual alter egos in our study, we expect that the findings are largely transferable to the perception and plausibility of virtual humans and avatars.

Acknowledgments

The authors thank the Embodiment Lab of the University of Würzburg as well as Marie Fielder and Nico Erdmannsdörfer for their help with body scanning during the data collection. This research has been funded by the German Federal Ministry of Education and Research in the project ViTraS (project number 16SV8219).

References

[1] J. Achenbach, T. Waltemate, M. E. Latoschik, and M. Botsch. Fast generation of realistic virtual humans. In *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology*. Association for Computing Machinery, New York, NY, USA, 2017. doi: 10.1145/3139131.3139154

[2] H. Adams, J. Stefanucci, S. Creem-Regehr, and B. Bodenheimer. Depth perception in augmented reality: The effects of display, shadow, and position. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 792–801, 2022. doi: 10.1109/VR51125.2022.00101

[3] Autodesk. Character Generator. https://charactergenerator.autodesk.com, 2014.

[4] J. N. Bailenson and J. Blascovich. Avatars. In *Encyclopedia of Human-Computer Interaction*, pp. 64–68. Berkshire Publishing Group, 2004.

[5] A. Bartl, S. Wenninger, E. Wolf, M. Botsch, and M. E. Latoschik. Affordable but not cheap: A case study of the effects of two 3D-reconstruction methods of virtual humans. *Frontiers in Virtual Reality*, 2, 2021. doi: 10.3389/frvir.2021.694617

[6] P. Bimberg, T. Weissker, and A. Kulik. On the usage of the simulator sickness questionnaire for virtual reality research. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pp. 464–467. IEEE, 2020. doi: 10.1109/VRW50115.2020.00098

[7] L. Brübach, F. Westermeier, C. Wienrich, and M. E. Latoschik. Breaking plausibility without breaking presence – Evidence for the multi-layer nature of plausibility. *IEEE Transactions on Visualization and Computer Graphics*, 28(5):2267–2276, 2022. doi: 10.1109/TVCG.2022.3150496

[8] T. Chaminade, J. Hodgins, and M. Kawato. Anthropomorphism influences perception of computer-animated characters' actions. *Social cognitive and affective neuroscience*, 2(3):206–216, 2007.

[9] I. A. Chicchi Giglioli, C. Bermejo Vidal, and M. Alcañiz Raya. A virtual versus an augmented reality cooking task based-tools: A behavioral and physiological study on the assessment of executive functions. *Frontiers in Psychology*, 10:2529, 2019. doi: 10.3389/fpsyg.2019.02529

[10] P. A. Connor-Greene. Gender differences in body weight perception and weight-loss strategies of college students. *Women & Health*, 14(2):27–42, 1988. doi: 10.1300/J013v14n02_03

[11] K. K. Cornelissen, A. Bester, P. Cairns, M. J. Tovée, and P. L. Cornelissen. The influence of personal bmi on body size estimations and sensitivity to body size change in anorexia spectrum disorders. *Body Image*, 13:75–85, 2015. doi: 10.1016/j.bodyim.2015.01.001

[12] K. K. Cornelissen, L. J. Gledhill, P. L. Cornelissen, and M. J. Tovée. Visual biases in judging body weight. *British Journal of Health Psychology*, 21(3):555–569, 2016. doi: 10.1111/bjhp.12185

[13] P. L. Cornelissen, K. K. Cornelissen, V. Groves, K. McCarty, and M. J. Tovée. View-dependent accuracy in body mass judgements of female bodies. *Body Image*, 24:116–123, 2018. doi: 10.1016/j.bodyim.2017.12.007

[14] J. J. Cummings and J. N. Bailenson. How immersive is enough? A meta-analysis of the effect of immersive technology on user presence. *Media Psychology*, 19(2):272–309, 2016. doi: 10.1080/15213269.2015.1015740

[15] N. Döllinger, C. Wienrich, E. Wolf, and M. E. Latoschik. ViTraS – Virtual reality therapy by stimulation of modulated body image – project outline. In *Mensch und Computer 2019 - Workshopband*. Gesellschaft für Informatik e.V., Bonn, 2019. doi: 10.18420/muc2019-ws-633

[16] N. Döllinger, E. Wolf, D. Mal, S. Wenninger, M. Botsch, M. E. Latoschik, and C. Wienrich. Resize me! Exploring the user experience of embodied realistic modulatable avatars for body image intervention in virtual reality. *Frontiers in Virtual Reality*, 2022. doi: 10.3389/frvir.2022.935449

[17] A. Field. *Discovering statistics using IBM SPSS statistics*. SAGE, 2013.

[18] D. Hepperle, C. F. Purps, J. Deuchler, and M. Wölfel. Aspects of visual avatar appearance: Self-representation, display type, and uncanny valley. *The Visual Computer*, 38(4):1227–1244, 2022. doi: 10.1007/s00371-021-02151-0

[19] D. Hepperle, H. Ödell, and M. Wölfel. Differences in the uncanny valley between head-mounted displays and monitors. In *2020 International Conference on Cyberworlds (CW)*, pp. 41–48, 2020. doi: 10.1109/CW49994.2020.00014

[20] F. Herrera, J. Bailenson, E. Weisz, E. Ogle, and J. Zaki. Building long-term empathy: A large-scale comparison of traditional and virtual reality perspective-taking. *PLOS ONE*, 13(10):e0204494, 2018. doi: 10.1371/journal.pone.0204494

[21] C.-C. Ho and K. F. MacDorman. Measuring the uncanny valley effect. *International Journal of Social Robotics*, 9(1):129–139, 2017. doi: 10.1007/s12369-016-0380-9

[22] M. Horne, A. Hill, T. Murells, H. Ugail, Irving, R. Chinnadorai, and M. Hardy. Using avatars in weight management settings: A systematic review. *Internet Interventions*, 19:100295, 2020. doi: 10.1016/j.invent.2019.100295

[23] L. K. G. Hsu. The gender gap in eating disorders: Why are the eating disorders more common among women? *Clinical Psychology Review*, 9(3):393–407, 1989. doi: 10.1016/0272-7358(89)90063-9

[24] IBM. SPSS statistics, version 27.0.1.0. https://www.ibm.com/products/spss-statistics, 2022.

[25] M. Jung, J. Kim, and K. Kim. Measuring recognition of body changes over time: A human-computer interaction tool using dynamic morphing and body ownership illusion. *PLOS ONE*, 15(9):1–14, 09 2020. doi: 10.1371/journal.pone.0239322

[26] R. S. Kennedy, N. E. Lane, K. S. Berbaum, and M. G. Lilienthal. Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The International Journal of Aviation Psychology*, 3(3):203–220, 1993. doi: 10.1207/s15327108ijap0303_3

[27] M. E. Latoschik and C. Wienrich. Congruence and plausibility, not presence: Pivotal conditions for XR experiences and effects, a novel approach. *Frontiers in Virtual Reality*, 3, 2022. doi: 10.3389/frvir.2022.694433

[28] J. Lessiter, J. Freeman, E. Keogh, and J. Davidoff. A cross-media presence questionnaire: The ITC-sense of presence inventory. *Presence: Teleoperators and Virtual Environments*, 10(3):282–297, 2001. doi: 10.1162/105474601300343612

[29] LimeSurvey GmbH. Limesurvey 4. https://www.limesurvey.org, 2020.

[30] J.-L. Lugrin, M. Ertl, P. Krop, R. Klüpfel, S. Stierstorfer, B. Weisz, M. Rück, J. Schmitt, N. Schmidt, and M. E. Latoschik. Any "body" there? Avatar visibility effects in a virtual reality game. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 17–24, 2018. doi: 10.1109/VR.2018.8446229

[31] J.-L. Lugrin, M. E. Latoschik, M. Habel, D. Roth, C. Seufert, and S. Grafe. Breaking bad behaviors: A new tool for learning classroom management using virtual reality. *Frontiers in ICT*, 3, 2016. doi: 10.3389/fict.2016.00026

[32] J.-L. Lugrin, J. Latt, and M. E. Latoschik. Anthropomorphism and illusion of virtual body ownership. In *Proceedings of the 25th International Conference on Artificial Reality and Telexistence and 20th Eurographics Symposium on Virtual Environments*, pp. 1–8. Eurographics Association, 2015.

[33] D. Mal, E. Wolf, N. Döllinger, M. Botsch, C. Wienrich, and M. E. Latoschik. Virtual human coherence and plausibility – Towards a validated scale. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pp. 788–789, 2022. doi: 10.1109/VRW55335.2022.00245

[34] M. Matamala-Gomez, A. Maselli, C. Malighetti, O. Realdon, F. Mantovani, and G. Riva. Virtual body ownership illusions for mental health: A narrative review. *Journal of Clinical Medicine*, 10(1):139, Jan. 2021. doi: 10.3390/jcm10010139

[35] Microsoft. HoloLens 2. https://www.microsoft.com/hololens, 2019.

[36] P. Milgram, H. Takemura, A. Utsumi, and F. Kishino. Augmented reality: A class of displays on the reality-virtuality continuum. In *Telemanipulator and Telepresence Technologies*, vol. 2351, pp. 282–293. International Society for Optics and Photonics, 1995. doi: 10.1117/12.197321

[37] M. Mori. Bukimi no tani [the uncanny valley]. *Energy*, 7:33–35, 1970.

[38] M. Mori, K. F. MacDorman, and N. Kageki. The uncanny valley [from the field]. *IEEE Robotics Automation Magazine*, 19(2):98–100, 2012. doi: 10.1109/MRA.2012.2192811

[39] S. Neyret, A. I. Bellido Rivas, X. Navarro, and M. Slater. Which body would you like to have? The impact of embodied perspective on body perception and body evaluation in immersive virtual reality. *Frontiers in Robotics and AI*, 7:31, 2020. doi: doi.org/10.3389/frobt.2020.00031

[40] S. O'Connor. Virtual reality and avatars in health care. *Clinical Nursing Research*, 28(5):523–528, 2019. doi: 10.1177/1054773819845824

[41] S. Paeratakul, M. A. White, D. A. Williamson, D. H. Ryan, and G. A. Bray. Sex, race/ethnicity, socioeconomic status, and BMI in relation to self-perception of overweight. *Obesity Research*, 10(5):345–350, 2002. doi: 10.1038/oby.2002.48

[42] I. V. Piryankova, J. K. Stefanucci, J. Romero, S. De La Rosa, M. J. Black, and B. J. Mohler. Can I recognize my body's weight? The influence of shape and texture on the perception of self. *ACM Transactions on Applied Perception (TAP)*, 11(3), 2014. doi: 10.1145/2641568

[43] K. M. Robinette, S. Blackwell, H. Daanen, M. Boehmer, and S. Fleming. Civilian american and european surface anthropometry resource (CEASAR). Technical report, Sytronics Inc, 2002. doi: 10.21236/ada406704

[44] Rootmotion. FinalIK. http://root-motion.com, 2019.

[45] D. Roth and C. Wienrich. Effects of media immersiveness on the perception of virtual characters. In *15th Workshop on Virtual and Augmented Reality of the GI Special Interest Group VR/AR*, 2018.

[46] A. Scavarelli, A. Arya, and R. J. Teather. Virtual reality and augmented reality in social learning spaces: A literature review. *Virtual Reality*, 25(1):257–277, 2021.

[47] R. Skarbez, J. Frederick P. Brooks, and M. C. Whitton. A survey of presence and related concepts. *ACM Computing Surveys*, 50(6):96, 2017. doi: 10.1145/3134301

[48] R. Skarbez, J. Gabbard, D. A. Bowman, T. Ogle, and T. Tucker. Virtual replicas of real places: Experimental investigations. *IEEE Transactions on Visualization and Computer Graphics*, pp. 1–1, 2021. doi: 10.1109/TVCG.2021.3096494

[49] R. Skarbez, M. Smith, and M. C. Whitton. Revisiting milgram and kishino's reality-virtuality continuum. *Frontiers in Virtual Reality*, 2, 2021. doi: 10.3389/frvir.2021.647997

[50] M. Slater. Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1535):3549–3557, 2009. doi: 10.1098/rstb.2009.0138

[51] K. M. Stanney, R. S. Kennedy, and J. M. Drexler. Cybersickness is not simulator sickness. In *Proceedings of the Human Factors and Ergonomics Society annual meeting*, vol. 41, pp. 1138–1142. SAGE Publications Sage CA: Los Angeles, CA, 1997. doi: 10.1177/107118139704100292

[52] J.-P. Stauffert, K. Korwisi, F. Niebling, and M. E. Latoschik. Ka-Boom!!! Visually exploring latency measurements for XR. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, pp. 1–9. Association for Computing Machinery, New York, NY, USA, 2021.

[53] J.-P. Stauffert, F. Niebling, and M. E. Latoschik. Effects of latency jitter on simulator sickness in a search task. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 121–127, 2018. doi: 10.1109/VR.2018.8446195

[54] J.-P. Stauffert, F. Niebling, and M. E. Latoschik. Latency and cybersickness: Impact, causes, and measures. A review. *Frontiers in Virtual Reality*, 1, 2020. doi: 10.3389/frvir.2020.582204

[55] A. Thaler. *The Role of Visual Cues in Body Size Estimation*, vol. 56 of *MPI Series in Biological Cybernetics*. Logos Verlag Berlin GmbH, 2019.

[56] A. Thaler, M. N. Geuss, S. C. Mölbert, K. E. Giel, S. Streuber, J. Romero, M. J. Black, and B. J. Mohler. Body size estimation of self and others in females varying in BMI. *PLOS ONE*, 13(2):e0192152, 2018. doi: 10.1371/journal.pone.0192152

[57] A. Thaler, I. V. Piryankova, J. K. Stefanucci, S. Pujades, S. de la Rosa, S. Streuber, J. Romero, M. J. Black, and B. J. Mohler. Visual perception and evaluation of photo-realistic self-avatars from 3D body scans in males and females. *Frontiers in ICT*, 5, 2018. doi: 10.3389/fict.2018.00018

[58] A. Thaler, S. Pujades, J. K. Stefanucci, S. H. Creem-Regehr, J. Tesch, M. J. Black, and B. J. Mohler. The influence of visual perspective on body size estimation in immersive virtual reality. In *ACM Symposium on Applied Perception 2019*, pp. 1–12. Association for Computing Machinery, New York, NY, USA, 2019. doi: 10.1145/3343036.3343134

[59] C. Turbyne, A. Goedhart, P. de Koning, F. Schirmbeck, and D. Denys. Systematic review and meta-analysis of virtual reality in mental health-care: Effects of full body illusions on body image disturbance. *Frontiers in Virtual Reality*, 2:39, 2021. doi: 10.3389/frvir.2021.657638

[60] Unity Technologies. Unity. https://unity3d.com, 2020.

[61] Valve. Half-life: Alyx. https://www.half-life.com/alyx, 2020.

[62] Varjo. XR-3. https://varjo.com/products/xr-3, 2020.

[63] T. Waltemate, D. Gall, D. Roth, M. Botsch, and M. E. Latoschik. The impact of avatar personalization and immersion on virtual body ownership, presence, and emotional response. *IEEE Transactions on Visualization and Computer Graphics*, 24(4):1643–1652, 2018. doi: 10.1109/TVCG.2018.2794629

[64] C. Wienrich, N. Döllinger, and R. Hein. Behavioral framework of immersive technologies (BehaveFIT): How and why virtual reality can support behavioral change processes. *Frontiers in Virtual Reality*, 2:84, 2021. doi: 10.3389/frvir.2021.627194

[65] C. Wienrich, P. Komma, S. Vogt, and M. E. Latoschik. Spatial presence in mixed realities – Considerations about the concept, measures, design, and experiments. *Frontiers in Virtual Reality*, 2, 2021. doi: 10.3389/frvir.2021.694315

[66] E. Wolf, N. Döllinger, D. Mal, C. Wienrich, M. Botsch, and M. E. Latoschik. Body weight perception of females using photorealistic avatars in virtual and augmented reality. In *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 583–594, 2020. doi: 10.1109/ISMAR50242.2020.00071

[67] E. Wolf, M. L. Fiedler, N. Döllinger, C. Wienrich, and M. E. Latoschik. Exploring presence, avatar embodiment, and body perception with a holographic augmented reality mirror. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 350–359, 2022. doi: 10.1109/VR51125.2022.00054

[68] E. Wolf, N. Merdan, N. Döllinger, D. Mal, C. Wienrich, M. Botsch, and M. E. Latoschik. The embodiment of photorealistic avatars influences female body weight perception in virtual reality. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*, pp. 65–74, 2021. doi: 10.1109/VR50410.2021.00027