From 2D-Screens to VR: Exploring the Effect of Immersion on the Plausibility of Virtual Humans

David Mal
HCI Group
University of Würzburg
Würzburg, Germany
david.mal@uni-wuerzburg.de

Erik Wolf
HCI Group
University of Würzburg
Würzburg, Germany
erik.wolf@uni-wuerzburg.de

Nina Döllinger
PIIS Group
University of Würzburg
Würzburg, Germany
nina.doellinger@uni-wuerzburg.de

Mario Botsch
Computer Graphics Group
TU Dortmund University
Dortmund, Germany
mario.botsch@tu-dortmund.de

Carolin Wienrich
PIIS Group
University of Würzburg
Würzburg, Germany
carolin.wienrich@uni-wuerzburg.de

Marc Erich Latoschik
HCI Group
University of Würzburg
Würzburg, Germany
marc.latoschik@uni-wuerzburg.de

Figure 1: Participants observed realistic or abstract virtual humans, either with low (2D-screen) or high immersion (VR). Realistic virtual humans engaged in idle motion are shown on the left, and abstract virtual humans performing wave motion on the right.

ABSTRACT

Virtual humans significantly contribute to users' plausible XR experiences. However, it may be not only the congruent rendering of the virtual human but also the degree of immersion having an impact on virtual humans' plausibility. In a low-immersive desktop-based and a high-immersive VR condition, participants rated realistic and abstract animated virtual humans regarding plausibility, affective appraisal, and social judgments. First, our results confirmed the factor structure of a preliminary virtual human plausibility questionnaire in VR. Further, the appearance and behavior of realistic virtual humans were overall perceived as more plausible compared to abstract virtual humans, an effect that increased with high immersion. Moreover, only for high immersion, realistic virtual humans were rated as more trustworthy and sympathetic than abstract virtual humans. Interestingly, we observed a potential uncanny valley effect for low but not for high immersion. We discuss the impact of a natural perception of anthropomorphic and realistic cues in VR and highlight the potential of immersive technology to elicit distinct effects in virtual humans.

CHI EA '24, May 11–16, 2024, Honolulu, HI, USA

CCS CONCEPTS

• **Human-centered computing → Virtual reality**; *Empirical studies in HCI*; *HCI theory, concepts and models.*

KEYWORDS

Avatar, Agent, Realism, Anthropomorphism, Immersion, Presence

ACM Reference Format:
David Mal, Erik Wolf, Nina Döllinger, Mario Botsch, Carolin Wienrich, and Marc Erich Latoschik. 2024. From 2D-Screens to VR: Exploring the Effect of Immersion on the Plausibility of Virtual Humans. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '24), May 11–16, 2024, Honolulu, HI, USA.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3613905.3650773

1 INTRODUCTION

Virtual humans play a central role in various relevant applications of mixed, augmented, and virtual reality (MR, AR, VR: XR for short), encompassing beneficial areas such as mental health [3, 11, 50], motor rehabilitation [17], or education [35, 51]. Aiming for high efficacy, these application domains necessitate high credibility in the system and a thorough understanding of perceptive and cognitive processes in conjunction with utilizing virtual humans. In this context, recent theories and models have drawn attention towards exploring experiences and effects in XR concerning the notions of plausibility, i.e., "what is apparently happening is really happening" [59, p. 3553], and congruence (coherence respectively), i.e., the objective match between processed and expected information contributing to plausibility [30, 57]. Being an essential entity of many XR scenarios, virtual humans and the congruence of their appearance and behavior, as well as their congruence with the virtual environments (VEs), have been considered important contributing to a user's plausible XR experience [38, 58]. Striving for life-likeness, typical cues shaping these congruencies have been related to their degree of realism [15, 29, 70] and anthropomorphism [36, 45].

However, it may be not only the congruent rendering of the virtual human but also the intermediary display technology having a significant impact on the perception and plausibility of virtual humans [67]. Aiming towards a cross-platform Metaverse [9], social VE's have recently gained considerable popularity by fostering pro-social interactions [9, 43, 54]. Thereby, they can support a broad range of XR technologies as well as devices with 2D screens, like desktop computers or mobile handhelds [9, 37, 71]. In comparisons to desktop environments, related studies have demonstrated VR's capacity to augment the sense of social or co-presence [16, 69] with virtual humans, and associated social judgments [53]. Yet, higher levels of immersion may also lead to increased eeriness ratings linked to the potential emergence of the uncanny valley effect [19]. While there is a remarkable body of research concerning the role of different types of virtual humans, in general, (see Weidner et al. [65] for an overview) and a solid base of related work comparing virtual humans between desktop and VR, this work is particularly concerned with the research question of:

RQ: How does the degree of immersion (low/high) affect the perceived plausibility of virtual humans showing dissimilar cues concerning realism and anthropomorphism?

To address the research question, we conducted a user study comparing the plausibility of two types of animated virtual humans presented on a 2D screen or in VR. The virtual humans were distinct in realism and anthropomorphism: (1) realistic virtual humans striving for a life-like appearance created by a 3D-reconstruction photogrammetry process and (2) abstract virtual humans with generic anthropomorphic features. Participants assessed the virtual humans' plausibility, affective appraisal, sympathy, and trust. Our evaluation is accompanied by validating the factor structure of a preliminary virtual human plausibility questionnaire in VR and exploring a potential uncanny valley effect and social judgments associated with virtual humans. This work applies empirical evidence to the significant domain of the plausibility of virtual humans and enhances our understanding of the disparities between low- and high-immersive experiences.

2 RELATED WORK

2.1 Virtual Experiences and Plausibility

Plausibility and coherence have emerged as fundamental concepts in the realm of XR experiences and effects [30, 56, 61]. Thereby, plausibility has been described as the illusion of "what is apparently happening is really happening" [59, p. 3553] and coherence refers to the inherent properties of the virtual scenario, contributing to its plausibility [57, 61]. Latoschik and Wienrich [30] proposed an alternative theoretical congruence and plausibility model (CaP model), naming congruence as an ontological specification of coherence. The model assumes a state of plausibility to arise from the congruence of cues on sensory, perceptual, and cognitive levels influencing relevant qualia of XR, such as spatial presence. As virtual humans are a central component in numerous VEs, these virtual entities and the congruence of their cues have also been indicated to contribute to a user's plausible XR experiences [4, 7, 8, 25, 28, 39, 58, 67]. In this context, Mal et al. [38] referred to the CaP model, stating that *virtual human plausibility (VHP)* is the subjective feeling of how reasonable and believable a virtual human appears to a user. VHP, therefore, arises from the congruence of habitual sensory, proximal perceptual, or higher-order cognitive cues of a virtual human's appearance and behavior within a VE [30, 38].

2.1.1 *Realism and Anthropomorphism in Virtual Humans.*

Typical cues shaping a virtual human's appearance and behavior, as well as its match with the VE, are related to their degree of realism, i.e., the perception that something could realistically or possibly exist in a non-mediated context, and anthropomorphism, i.e., the perception or assignment of human-like properties or characteristics to entities that may or may not be human [36, 45]. Indeed, extensive research emphasizes the influence of diverse cues of virtual humans in striving for a life-like experience [15, 24, 29, 36, 45, 70], on relevant qualia of XR experiences. Thereby, the degree of realism or anthropomorphism of virtual humans may match the perceived plausibility of virtual humans in a specific VE; however, they can also diverge depending on priming and habituation to environments that simulate an alternative reality following rules not coherent with the physical world. In our study, we compare the plausibility of two types of virtual humans rendering dissimilar cues concerning realism and anthropomorphism in form and material: (1) realistic virtual humans created by a 3D-reconstruction photogrammetry process striving for a life-like appearance and (2) abstract virtual humans of lower realism and anthropomorphism resembling a wooden mannequin with generic human anatomy.

2.1.2 *Virtual Human Plausibility Questionnaire.*

To assess the plausibility of virtual humans, in previous work, we presented a *Virtual Human Plausibility Questionnaire* (VHPQ) [38], identifying and interpreting the underlying structure of the 11 items in an online study. Yet, the reliability and confirmation of the factor structure in an immersive context remain open. We consider understanding the mechanisms behind the perception of virtual humans, their contribution to users' plausible XR experiences, and the development of a standardized scale to assess VHP as essential for developing relevant applications with virtual humans. We, therefore, further evaluate the factor structure and reliability of the VHPQ in VR.

2.1.3 Media Immersion. Numerous studies have highlighted disparities in presenting content on a VR head-mounted display (HMD) versus a traditional 2D screen [26, 46, 55, 62]. Also, how we perceive virtual humans has been shown to depend on the intermediary display technology and its degree of immersion. Related work indicated increased sensitivity to virtual agents' footing cues in VR [47], and enhanced sense of social or co-presence in VR towards a human-like virtual agent [5, 16, 69]. According to Slater [59], the immersion of an interactive medium is shaped by its objective technical properties and the extent to which they support valid sensorimotor contingencies, allowing users to perceive and interact with the provided content naturally. In that sense, VR applications are considered to be more immersive media compared to desktop applications [60]. Eventually, we hypothesize a more natural perception of virtual humans in immersive VR would lead to an increased focus on the congruence of their anthropomorphic and realistic cues. A virtual human showing more human or life-like cues would then be perceived as more human in VR compared to a lower immersive medium, and (in)congruencies in these cues would have an intensified effect on the perceived VHP. We follow the introduced concept of VHP and conclude with the following hypotheses.

H1.1: Realistic virtual humans will be rated significantly higher in VHP than abstract virtual humans.
H1.2: Differences in VHP between realistic and abstract virtual humans will be intensified in VR.

2.2 Uncanny Valley Effect and Social Judgments

We further exploratively evaluate the virtual human's affective appraisal, as well as social judgments in sympathy and trust. The affective appraisal of virtual humans has been considered crucial concerning the uncanny valley effect [67]. It describes a shift from initial affinity to a sense of eeriness as an anthropomorphic character approaches but falls short of achieving convincing human-likeness [44]. Latoschik et al. [29] suggested the potential presence of an uncanny valley effect when comparing realistic virtual humans and an abstract virtual human similar to the virtual humans in our study. Interestingly, Hepperle et al. [19, 20] indicated a more pronounced uncanny valley effect in VR compared to a desktop environment, while Wolf et al. [67] did not find display dependent differences in humanness or eeriness in virtual humans between VR and AR. Further, we evaluate social judgments of trust and sympathy. These aspects are highly relevant in everyday social interactions [53] having a favorable impact on mediated social interactions [34]. While Weidner et al. [65] summarized more realistic avatars to benefit likability and trust, Latoschik et al. [29] showed comparable trust ratings between abstract and realistic virtual humans in VR. Interestingly, rather realistic virtual humans were perceived as more sympathetic and trustworthy in VR compared to a desktop condition [53].

3 METHODS

3.1 Design

We employed a 2×2 mixed design with the independent variables *degree of immersion* (low/high) as a between-subject factor and *virtual human realism* (abstract/realistic) as a within-subject factor. Participants either joined an online desktop (low immersion) or an on-site VR (high immersion) study and were consecutively presented with one of ten animated virtual humans with either an abstract or realistic appearance (see Figure 1). Our dependent variables assessed participants' ratings of the respective virtual human regarding its plausibility, affective appraisal, and social judgments.

3.2 Participants

A total of 106 undergraduate students were recruited through the University of Würzburg's participant management system. They received course credit for participation. Post-survey exclusion criteria ruled out the data from participants who had not spoken German for at least ten years ($n = 2$), had not corrected visual impairments ($n = 1$), experienced technical issues ($n = 3$), or knew at least one of the virtual human models in person ($n = 9$). This resulted in 91 valid data sets. For the 65 valid data sets in the low immersive condition, ages ranged from 18 to 53 years ($M = 21.52$, $SD = 4.29$), comprising 33 females, 31 males, and one non-binary participant. For the 26 valid data sets in the high immersive condition, ages ranged from 19 to 27 years ($M = 21.73$, $SD = 1.97$), comprising 20 females and six males. The study was conducted according to the Declaration of Helsinki and approved by the ethics committee of the Human-Computer-Media Institute of the University of Würzburg.

3.3 Measures

We assessed the virtual humans' plausibility (VHP) using the *Virtual Human Plausibility Questionnaire* (VHPQ) [38]. It comprises the virtual human's (1) appearance and behavior plausibility (ABP) and (2) match to the virtual environment (MVE). The 11 questions were rated on a 7-point Likert scale (*7 = highest VHP*). Further, we assessed affective appraisal in terms of (1) humanness and (2) eeriness with the *Uncanny Valley Index* (UVI) [21]. Pairs of statements were rated using semantic differentials ranging from -3 to 3. The results were mapped to a scale of 1 to 7 (*7 = highest humanness and eeriness*). Finally, we assessed social judgments towards the virtual humans in (1) trust and (2) sympathy with concise questions taken from Roth and Wienrich [53]. Responses were rated on a 7-point Likert scale (*7 = highest social judgment*).

3.4 Apparatus

3.4.1 Virtual Humans. Six realistic virtual humans striving for a lifelike appearance were created by applying a 3D-reconstruction photogrammetry pipeline [1]. Therefore, we scanned three male and three female volunteers wearing casual clothes. Their ages ranged from 21 to 25 ($M = 23.33$, $SD = 1.86$), they gave written consent, and they were not compensated for participation. Further, four unique and distinguishable variants of abstract virtual humans showing generic anthropomorphic features were created by coloring a wooden mannequin's upper body in green, yellow, blue, or brown. All virtual humans are depicted in the supplemental material of this work. The virtual humans utilized consistent pre-generated animations. Body motion was recorded using an OptiTrack Flex 3 motion capture system. Subtle and non-intrusive facial expressions, including randomized eye blinks, were pre-modeled. Due to fixed eyes and the absence of fingers of abstract virtual humans, only the realistic virtual humans moved their gaze and fingers slightly in a randomized manner. The resulting animation sequence comprised an initial idle motion (11 s) succeeded by a brief waving motion (7 s).

Table 1: The descriptive values of the dependent measures for each experimental condition and p-values of the multilevel linear models' main and interaction effects. Single-asterisks indicate significant and double-asterisks highly significant *p*-values.

	Low Immersion		High Immersion		Main Effect	Main Effect	Interaction Effect
	Abstract	Realistic	Abstract	Realistic	Immersion	Realism	
	M (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*p*	*p*	*p*
Appearance and Behavior (ABP)	4.53 (1.24)	4.68 (1.09)	4.39 (1.21)	5.28 (0.93)	.072	<.001**	<.001**
Match with VE (MVE)	4.60 (1.35)	4.54 (1.20)	4.49 (1.45)	4.58 (1.38)	.946	.815	.341
Humanness	2.49 (0.94)	4.17 (1.32)	2.41 (0.82)	5.02 (1.19)	.014*	<.001**	<.001**
Eeriness	3.07 (0.77)	3.37 (0.75)	3.28 (0.97)	3.19 (0.61)	.826	<.001**	<.001**
Sympathy	4.11 (1.47)	4.13 (1.44)	3.74 (1.51)	4.71 (1.45)	.820	<.001**	<.001**
Trust	3.63 (1.38)	3.78 (1.30)	3.40 (1.36)	4.36 (1.34)	.245	<.001**	<.001**

3.4.2 System Description. We implemented an application supporting both immersion conditions using Unity in version2019.4.8f1 [63]. A neutral, light-colored virtual room was utilized as the virtual environment. For the low immersive condition, we recorded ten dedicated videos of 18 s (30 fps, 1280 × 720 pixels), each featuring one of the virtual humans. The videos were streamed in an online study, implemented with LimeSurvey 4.5 [32]. For the high immersive condition, participants wore an HTC Vive Pro HMD (110° FOV, 90 Hz, 1440×1600 px per eye) [22], integrated using SteamVR [64] and the related Unity plug-in in version 2.6.1. The setup ran on a VR-capable workstation. The virtual environment was calibrated along the X and Z axes, aligning the participant's head position and rotation with the virtual camera's viewpoint of the video recordings. The distance between the virtual humans and the participants, or the virtual camera, respectively, was set to 2.25 m.

3.5 Procedure

Upon registration, participants were either provided a link to the online study or invited to the local laboratory. As the study commenced, participants were informed about the procedure, provided explicit consent, reported demographic data, and received a briefing about the upcoming exposition phase. During each exposition, participants viewed one of the ten virtual humans performing the recorded animation for 18 seconds, either played as a video (low immersion) or presented as 3D representations in VR (high immersion). Subsequently, participants answered questions about the presented virtual human, including all dependent measures and whether they knew the virtual human model in person. In the immersive condition, participants removed the HMD and used a desktop computer to answer questions. The exposition's procedure was repeated for each stimulus in a randomized order, resulting in ten expositions per participant. In the study, virtual humans were referred to as virtual characters, preventing implicit expectations for resembling human characteristics. There was no explicit framing of whether the virtual humans were human or computer-controlled. The experimental procedure took an average of 36 min (desktop) and 64 min (VR). A depiction can be found in the supplemental material of this work.

3.6 Statistical Analysis

Statistical analyses were performed in R (version 4.3.0) [49] and conducted at a significance level of $\alpha = .05$. First, we conducted a confirmatory factor analysis (CFA) [52] and calculated Cronbach's α [66] to confirm the factor structure and reliability of the VHPQ for a total of 260 appraisals in VR. We tested for multivariate normality (Henze-Zirkler) [27] and reported recommended fit measures [6, 23] including RMSEA (cut-off ≤ .06), CFI (cut-off ≥ .95), and TLI (cut-off ≥ .95) as well as standardized loading estimates. Second, we calculated multilevel linear models with maximum likelihood estimation [12, 48] to examine the effects of immersion and virtual humans on the dependent measures. Descriptive statistics for each condition and p-values for the multilevel linear models' main and interaction effects are reported in Table 1. A planned contrast was set to compare realistic and abstract virtual humans. In case main and interaction effects became significant, we interpreted the highest order effect only, i.e., the significant interaction [12]. For significant interaction effects, we performed post hoc comparisons leveraging mixed-effects models to compare virtual humans across immersion groups. We applied the planned contrast, accounted for repeated measures, and adjusted for multiple comparisons (Bonferroni) within each between-group [31].

4 RESULTS

4.1 Confirmatory Factor Analysis (VHPQ)

We performed a robust maximum likelihood estimation with Yuan-Bentler corrected test statistics ($factor = 1.701$, $HZ = 2.51$, $p < .001$) [68]. The 11 items had a satisfactory fit within the structure of ABP and MVE ($\chi^2(43) = 188.11$, $RMSEA = .061$, $CFI = .964$, $TLI = .955$), with only RMSEA marginally exceeding the cut-off value. High internal reliability was found for both ABP ($\alpha = .90$) and MVE ($\alpha = .95$). Standardized loading estimates are reported in this work's supplemental material.

4.2 Virtual Human Plausibility

The interaction between the degree of immersion and virtual humans, $\chi^2(1) = 30.25$, $p < .001$, significantly affected ABP. Pairwise comparisons and the interaction plot (Figure 2) indicate that realistic virtual humans were rated significantly higher in ABP than abstract virtual humans for low immersion, $t(817) = 2.03$, $p = .042$, $r = .07$, and high immersion, $t(817) = 7.84$, $p < .001$, $r = .26$. Further, the degree of immersion, $\chi^2(1) = .01$, $p = .95$, virtual humans, $\chi^2(1) = .06$, $p = .82$, and their interaction, $\chi^2(1) = .91$, $p = .34$, had no significant effect on MVE.

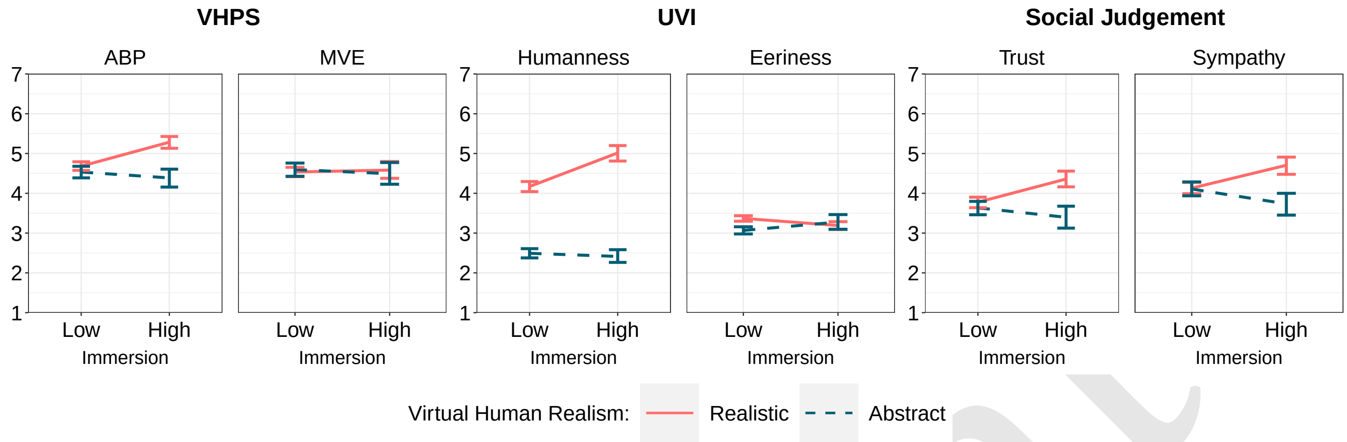

Figure 2: The interaction plots depict the degree of immersion (low or high) and the contrast between realistic and abstract virtual humans. Error bars represent 95 % confidence intervals estimated using bootstrapped standard deviations.

4.3 Affective Appraisal

The interaction between the degree of immersion and virtual humans significantly affected humanness, $\chi^2(1) = 51.90$, $p < .001$. Pairwise comparisons and the interaction plot (Figure 2) indicate that realistic virtual humans were rated significantly higher in humanness compared to abstract virtual humans for low, $t(817) = 24.813$, $p < .001$, $r = .66$, and high immersion, $t(817) = 24.335$, $p < .001$, $r = .65$. Further, the interaction between immersion and virtual humans significantly affected eeriness, $\chi^2(1) = 16.38$, $p < .001$. Pairwise comparisons and the interaction plot (Figure 2) indicate that realistic virtual humans were rated as significantly more eerie compared to abstract virtual humans for low immersion $t(817) = 5.90$, $p < .001$, $r = .20$, but not for high immersion, $t(817) = -1.07$, $p < .285$, $r = .04$.

4.4 Social Judgment

The interaction between the degree of immersion and virtual humans significantly affected sympathy, $\chi^2(1) = 28.42$, $p < .001$, and trust, $\chi^2(1) = 29.55$, $p < .001$. Pairwise comparisons and the interaction plot (Figure 2) indicate no significant differences between the types of virtual humans for low immersion in sympathy, $t(817) = 0.23$, $p = .818$, $r = .01$, and trust, $t(817) = 1.77$, $p = .077$, $r = .06$. However, for high immersion, realistic virtual humans were rated more sympathetic, $t(817) = 6.49$, $p < .001$, $r = .22$ and more trustworthy, $t(817) = 7.59$, $p < .001$, $r = 0.26$, than abstract ones.

5 DISCUSSION

5.1 Virtual Human Plausibility Questionnaire

We confirmed the factor structure and reliability of the VHPQ items using a total of 260 evaluations in VR. Overall, our CFA demonstrated a satisfactory goodness-of-fit for the structure of ABP and MVE with high internal reliability for both factors. Furthermore, our results highlight the sensitivity of the ABP scale in detecting (in)congruencies related to anthropomorphic and realism cues in animated virtual humans. Moving forward, it is crucial to further assess the questionnaire's validity [41], focusing on factor-specific congruence manipulations of virtual humans.

5.2 Virtual Humans Realism and Immersion

5.2.1 Virtual Human Plausibility. In line with H1.1, the appearance and behavior of realistic virtual humans were perceived as more plausible compared to abstract virtual humans. We attribute higher ratings of ABP for realistic virtual humans to congruencies between the virtual human's appearance and the physical world, but also their internal visual and behavioral features. Rendering realistic behavioral cues, i.e., motion-captured body animations and facial expressions, was eventually perceived as more congruent with the life and human-like appearance of realistic virtual humans. This interpretation goes along with related work naming the interplay between visual and behavioral realism an important aspect in virtual humans [2, 13, 40, 70]. Furthermore, in line with H1.2, a notable shift in ABP emerged as the effect increased from small, or even marginal, to intermediate [14] when transitioning from low to high immersion. The higher degree of immersion might have contributed to a sense of "being there" in the VE [18], allowing participants to perceive the virtual humans more naturally [59]. In that sense, when encountering a (virtually) life-sized virtual human in VR, individuals are likely to allocate greater attention to anthropomorphic and realistic cues and their congruencies. This tendency is also evident in the heightened attribution of humanness to realistic virtual humans in VR. Besides, comparable ratings of MVE across all conditions indicate that the rather neutral VE did not affect the plausibility of the virtual humans.

5.2.2 Affective Appraisal. In the low immersive condition, realistic virtual humans were rated more human and eerie than abstract virtual humans, indicating a possible uncanny valley effect. However, contradicting the marginal indications from related work [19, 29], our results do not imply an (intensified) uncanny valley effect in the immersive condition. A noticeable increase in humanness for realistic virtual humans between low and high immersion might reflect a recovery from the uncanny valley in VR. However, we must remain open-ended regarding whether the realistic virtual humans have either not yet entered or sufficiently recovered from the uncanny valley, as we only have two types of virtual humans as reference points. Future work will have to explore a wider range of animated virtual humans and their level of humanness concerning the uncanny valley effect in immersive VR.

5.2.3 Social Judgments. Our findings align with previous work indicating increased sympathy and trust for realistic virtual humans when transitioning from low to high immersion [53]. Notably, this increase is not observed in abstract virtual humans. An intensified perception of (in)congruencies in VR may have led to acceptance or even the suspension of disbelief [18, 30] in virtual humans, potentially influencing social judgments. Interestingly, the interaction plot unveils a parallel pattern between ABP and social judgments, indicating a relationship between the plausibility of virtual humans' appearance and behavior and how we perceive them as social actors. This aspect warrants further analysis within ecologically valid scenarios with (social) interaction and validated measurements [33].

5.3 Limitations and Future Work

First, our study did not provide interaction possibilities, limiting our results to non-interactive animated virtual humans. However, our approach ensured controlled and comparable stimuli presentation in both conditions. When replicating our work in an interactive setting, one should be aware that interactions might elicit incongruencies and be prone to fall into an uncanny valley [42]. Second, an imbalance in sample sizes between low ($N = 65$) and high immersion ($N = 26$), and an imbalance in the number of stimuli concerning the virtual human's realism (six realistic and four abstract), may affect the findings' generalizability and warrant cautious application to broader populations. Third, for the low immersion condition, we relied on an online study. With this approach, we overcame COVID-19 restrictions during data collection, engaging a broader participant pool. However, it limited overall control, and our conditions diverged as VR participants had to come to the local laboratory, spending more time on the study. Lastly, to provide a comprehensive understanding of virtual human plausibility aiming towards a cross-platform Metaverse [9], future work should broaden the scope of XR and handheld devices. This expansion would enhance the generalizability of findings and offer insights into user experiences with virtual humans across different technological contexts.

5.4 Conclusion

We investigated the differences between low and high immersion on the plausibility of two types of virtual humans (realistic/abstract), distinct in realism and anthropomorphism. We first confirmed the factor structure and internal reliability of a preliminary virtual human plausibility questionnaire in VR to take a step further toward a validated scale. Moreover, our results revealed that the appearance and behavior of realistic virtual humans were perceived as more plausible compared to their abstract counterparts. Although these distinctions were evident in both conditions, a notable shift occurred when transitioning from 2D screens to VR, intensifying the effect with higher immersion. A comparable pattern emerged in social judgments, with discernible differences in trust and sympathy towards the virtual humans becoming evident exclusively with high immersion. Additionally, our findings indicate an uncanny valley effect for low but not for high immersion. The noticeable increase in perceived humanness for realistic virtual humans in VR might indicate a potential recovery from the uncanny valley; however, this interpretation should be approached with caution and requires further investigation. In summary, we indicate the potential of life-like virtual humans to contribute to a plausible immersive experience as the basis for creating credible systems in beneficial areas utilizing virtual humans [10]. Furthermore, our study affirms related work highlighting the potential of high immersion to elicit distinct and intensified effects in virtual humans.

ACKNOWLEDGMENTS

This research has been funded by the German Federal Ministry of Education and Research (BMBF) in the projects ViTraS (Grants 16SV8219 and 16SV8225), ViLeArn More (Grant 16DHB2214) and HiAvA (Grant 16SV8781), and by the Bavarian State Ministry For Digital Affairs in the project XR Hub (Grant A5-3822-2-16). Erik Wolf gratefully acknowledges a Meta Research Ph.D. Fellowship. We thank Andrea Bartl and Philipp Coblenz, who are with the HCI Group of the University of Würzburg, and Florian Heinrich, who is with the University of Magdeburg, for their valuable contributions.

REFERENCES

[1] Jascha Achenbach, Thomas Waltemate, Marc Erich Latoschik, and Mario Botsch. 2017. Fast generation of realistic virtual humans. In *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology* (Gothenburg, Sweden) *(VRST '17)*. Association for Computing Machinery, New York, NY, USA, Article 12, 10 pages. https://doi.org/10.1145/3139131.3139154

[2] Jeremy N Bailenson, Nick Yee, Dan Merget, and Ralph Schroeder. 2006. The Effect of Behavioral Realism and Form Realism of Real-Time Avatar Faces on Verbal Disclosure, Nonverbal Disclosure, Emotion Recognition, and Copresence in Dyadic Interaction. *Presence: Teleoperators and Virtual Environments* 15, 4 (08 2006), 359–372. https://doi.org/10.1162/pres.15.4.359

[3] Miguel Barreda-Ángeles and Tilo Hartmann. 2022. Psychological benefits of using social virtual reality platforms during the covid-19 pandemic: The role of social and spatial presence. *Computers in Human Behavior* 127 (2022), 107047. https://doi.org/10.1016/j.chb.2021.107047

[4] Alejandro Beacco, Ramon Oliva, Carlos Cabreira, Jaime Gallego, and Mel Slater. 2021. Disturbance and Plausibility in a Virtual Rock Concert: A Pilot Study. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*. 538–545. https://doi.org/10.1109/VR50410.2021.00078

[5] Gary Bente, Ralf Schmälzle, Nolan T. Jahn, and Andrea Schaaf. 2023. Measuring the effects of co-location on emotion perception in shared virtual environments: An ecological perspective. *Frontiers in Virtual Reality* 4 (2023). https://doi.org/10.3389/frvir.2023.1032510

[6] P. M. Bentler. 1990. Comparative fit indexes in structural models. *Psychological Bulletin* 107 (1990), 238–246. https://doi.org/10.1037/0033-2909.107.2.238

[7] Ilias Bergström, Sérgio Azevedo, Panos Papiotis, Nuno Saldanha, and Mel Slater. 2017. The Plausibility of a String Quartet Performance in Virtual Reality. *IEEE Transactions on Visualization and Computer Graphics* 23, 4 (2017), 1352–1359. https://doi.org/10.1109/TVCG.2017.2657138

[8] Henrique Galvan Debarba, Sylvain Chagué, and Caecilia Charbonnier. 2022. On the Plausibility of Virtual Body Animation Features in Virtual Reality. *IEEE Transactions on Visualization and Computer Graphics* 28, 4, 1880–1893. https://doi.org/10.1109/TVCG.2020.3025175

[9] Yogesh K. Dwivedi, Laurie Hughes, Abdullah M. Baabdullah, Samuel Ribeiro-Navarrete, Mihalis Giannakis, Mutaz M. Al-Debei, ..., and Samuel Fosso Wamba. 2022. Metaverse beyond the hype: Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *International Journal of Information Management* 66 (2022), 102542. https://doi.org/10.1016/j.ijinfomgt.2022.102542

[10] Nina Döllinger, Carolin Wienrich, Erik Wolf, Mario Botsch, and Marc Erich Latoschik. 2019. ViTraS - Virtual Reality Therapy by Stimulation of Modulated Body Image - Project Outline. *Mensch und Computer 2019 - Workshopband*. https://doi.org/10.18420/muc2019-ws-633

[11] Nina Döllinger, Erik Wolf, David Mal, Stephan Wenninger, Mario Botsch, Marc Erich Latoschik, and Carolin Wienrich. 2022. Resize Me! Exploring the

User Experience of Embodied Realistic Modulatable Avatars for Body Image Intervention in Virtual Reality. *Frontiers in Virtual Reality* (2022). https://doi.org/10.3389/frvir.2022.935449

[12] Andy P. Field, Jeremy Miles, and Zoë Field. 2012. *Discovering statistics using R.* Sage.

[13] Maia Garau, Mel Slater, Vinoba Vinayagamoorthy, Andrea Brogni, Anthony Steed, and M. Angela Sasse. 2003. The impact of avatar realism and eye gaze control on perceived quality of communication in a shared immersive virtual environment. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Ft. Lauderdale, Florida, USA) *(CHI '03)*. Association for Computing Machinery, New York, NY, USA, 529–536. https://doi.org/10.1145/642611.642703

[14] Gilles E. Gignac and Eva T. Szodorai. 2016. Effect size guidelines for individual differences researchers. *Personality and Individual Differences* 102 (2016), 74–78. https://doi.org/10.1016/j.paid.2016.06.069

[15] Guilherme Gonçalves, Pedro Monteiro, Hugo Coelho, Miguel Melo, and Maximino Bessa. 2021. Systematic Review on Realism Research Methodologies on Immersive Virtual, Augmented and Mixed Realities. *IEEE Access* 9 (2021), 89150–89161. https://doi.org/10.1109/ACCESS.2021.3089946

[16] Manuel Guimarães, Rui Prada, Pedro A. Santos, João Dias, Arnav Jhala, and Samuel Mascarenhas. 2020. The Impact of Virtual Reality in the Social Presence of a Virtual Agent. In *Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents* (Virtual Event, Scotland, UK) *(IVA '20)*. Association for Computing Machinery, New York, NY, USA, Article 23, 8 pages. https://doi.org/10.1145/3383652.3423879

[17] Negin Hamzeheinejad, Daniel Roth, Samantha Monty, Julian Breuer, Anuschka Rodenberg, and Marc Erich Latoschik. 2021. The Impact of Implicit and Explicit Feedback on Performance and Experience during VR-Supported Motor Rehabilitation. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*. 382–391. https://doi.org/10.1109/VR50410.2021.00061

[18] Carrie Heeter. 1992. Being There: The Subjective Experience of Presence. *Presence: Teleoperators and Virtual Environments* 1, 2 (05 1992), 262–271. https://doi.org/10.1162/pres.1992.1.2.262

[19] Daniel Hepperle, Christian Felix Purps, Jonas Deuchler, and Matthias Wölfel. 2022. Aspects of visual avatar appearance: Self-representation, display type, and uncanny valley. *The Visual Computer* 38, 4 (2022), 1227–1244. https://doi.org/10.1007/s00371-021-02151-0

[20] Daniel Hepperle, Hannah Ödell, and Matthias Wölfel. 2020. Differences in the Uncanny Valley between Head-Mounted Displays and Monitors. In *2020 International Conference on Cyberworlds (CW)*. 41–48. https://doi.org/10.1109/CW49994.2020.00014

[21] Chin-Chang Ho and Karl F. MacDorman. 2017. Measuring the Uncanny Valley Effect. *International Journal of Social Robotics* 9, 1 (2017), 129–139. https://doi.org/10.1007/s12369-016-0380-9

[22] HTC Corporation. 2018. Vive Pro. https://www.vive.com/eu/product/vive-pro/.

[23] Dennis L. Jackson, J. Arthur Gillaspy, and Rebecca Purc-Stephenson. 2009. Reporting practices in confirmatory factor analysis: An overview and some recommendations. *Psychological Methods* 14, 1 (2009), 6–23. https://doi.org/10.1037/a0014694

[24] Do Yuon Kim, Ha Kyung Lee, and Kyunghwa Chung. 2023. Avatar-mediated experience in the metaverse: The impact of avatar realism on user-avatar relationship. *Journal of Retailing and Consumer Services* 73 (2023), 103382. https://doi.org/10.1016/j.jretconser.2023.103382

[25] Kangsoo Kim, Divine Maloney, Gerd Bruder, Jeremy N. Bailenson, and Gregory F. Welch. 2017. The effects of virtual human's spatial and behavioral coherence with physical objects on social presence in AR. *Computer Animation and Virtual Worlds* 28, 3-4 (2017), e1771. https://doi.org/10.1002/cav.1771

[26] Silvia Erika Kober, Jürgen Kurzmann, and Christa Neuper. 2012. Cortical correlate of spatial presence in 2D and 3D interactive virtual reality: An EEG study. *International Journal of Psychophysiology* 83, 3 (2012), 365–374. https://doi.org/10.1016/j.ijpsycho.2011.12.003

[27] Selcuk Korkmaz, Dincer Goksuluk, and Gokmen Zararsiz. 2014. MVN: An R Package for Assessing Multivariate Normality. *The R Journal* 6, 2 (2014), 151–162. https://journal.r-project.org/archive/2014-2/korkmaz-goksuluk-zararsiz.pdf

[28] Peter Kullmann, Timo Menzel, Mario Botsch, and Marc Erich Latoschik. 2023. An Evaluation of Other-Avatar Facial Animation Methods for Social VR. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems* (, Hamburg, Germany,) *(CHI EA '23)*. Association for Computing Machinery, New York, NY, USA, Article 33, 7 pages. https://doi.org/10.1145/3544549.3585617

[29] Marc Erich Latoschik, Daniel Roth, Dominik Gall, Jascha Achenbach, Thomas Waltemate, and Mario Botsch. 2017. The effect of avatar realism in immersive social virtual realities. In *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology* (Gothenburg, Sweden) *(VRST '17)*. Association for Computing Machinery, New York, NY, USA, Article 39, 10 pages. https://doi.org/10.1145/3139131.3139156

[30] Marc Erich Latoschik and Carolin Wienrich. 2022. Congruence and Plausibility, Not Presence: Pivotal Conditions for XR Experiences and Effects, a Novel Approach. *Frontiers in Virtual Reality* 3 (2022). https://doi.org/10.3389/frvir.2022.694433

[31] Russell V. Lenth. 2023. *emmeans: Estimated Marginal Means, aka Least-Squares Means.* https://CRAN.R-project.org/package=emmeans R package version 1.8.7.

[32] LimeSurvey GmbH. 2020. LimeSurvey: An open source survey tool. https://www.limesurvey.org.

[33] Jinghuai Lin, Johrine Cronjé, Ivo Käthner, Paul Pauli, and Marc Erich Latoschik. 2023. Measuring Interpersonal Trust towards Virtual Humans with a Virtual Maze Paradigm. *IEEE Transactions on Visualization and Computer Graphics* 29, 5 (2023), 2401–2411. https://doi.org/10.1109/TVCG.2023.3247095

[34] Jinghuai Lin and Marc Erich Latoschik. 2022. Digital body, identity and privacy in social virtual reality: A systematic review. *Frontiers in Virtual Reality* 3 (2022). https://doi.org/10.3389/frvir.2022.974652

[35] Jean-Luc Lugrin, Marc Erich Latoschik, Michael Habel, Daniel Roth, Christian Seufert, and Silke Grafe. 2016. Breaking Bad Behaviors: A New Tool for Learning Classroom Management Using Virtual Reality. *Frontiers in ICT* 3 (2016). https://doi.org/10.3389/fict.2016.00026

[36] Jean-Luc Lugrin, Johanna Latt, and Marc Erich Latoschik. 2015. Avatar anthropomorphism and illusion of body ownership in VR. In *2015 IEEE Virtual Reality (VR)*. 229–230. https://doi.org/10.1109/VR.2015.7223379

[37] Elhassan Makled, Florian Weidner, and Wolfgang Broll. 2022. Investigating User Embodiment of Inverse-Kinematic Avatars in Smartphone Augmented Reality. In *2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 666–675. https://doi.org/10.1109/ISMAR55827.2022.00084

[38] David Mal, Erik Wolf, Nina Döllinger, Mario Botsch, Carolin Wienrich, and Marc Erich Latoschik. 2022. Virtual Human Coherence and Plausibility – Towards a Validated Scale. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. 788–789. https://doi.org/10.1109/VRW55335.2022.00245

[39] David Mal, Erik Wolf, Nina Döllinger, Carolin Wienrich, and Marc Erich Latoschik. 2023. The Impact of Avatar and Environment Congruence on Plausibility, Embodiment, Presence, and the Proteus Effect in Virtual Reality. *IEEE Transactions on Visualization and Computer Graphics* 29, 5 (2023), 2358–2368. https://doi.org/10.1109/TVCG.2023.3247089

[40] Samah Mansour, Mostafa El-Said, Carolyn Rude-Parkins, and Jagadeesh Nandigam. 2006. The interactive effect of avatar visual fidelity and behavioral fidelity in the collaborative virtual reality environment on the perception of social interaction. In *Proceedings of the 10th WSEAS International Conference on Communications (ICCOM'06)*. World Scientific and Engineering Academy and Society (WSEAS), Wisconsin, USA, 387–395.

[41] Tenko Raykov Marcoulides, George A. 2010. *Introduction to Psychometric Theory.* Routledge, New York. https://doi.org/10.4324/9780203841624

[42] Ryan P. McMahan, Chengyuan Lai, and Swaroop K. Pal. 2016. Interaction Fidelity: The Uncanny Valley of Virtual Reality Interactions. In *Virtual, Augmented and Mixed Reality*. Springer International Publishing, Cham, 59–70. https://doi.org/10.1007/978-3-319-39907-2_6

[43] Joshua McVeigh-Schultz, Anya Kolesnichenko, and Katherine Isbister. 2019. Shaping Pro-Social Interaction in VR: An Emerging Design Framework. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, Glasgow Scotland Uk, 1–12. https://doi.org/10.1145/3290605.3300794

[44] Masahiro Mori, Karl F. MacDorman, and Norri Kageki. 2012. The Uncanny Valley [From the Field]. *IEEE Robotics & Automation Magazine* 19, 2 (2012), 98–100. https://doi.org/10.1109/MRA.2012.2192811

[45] Kristine L. Nowak and Jesse Fox. 2018. Avatars and computer-mediated communication: a review of the definitions, uses, and effects of digital representations. *Review of Communication Research* 6 (2018), 30–53. https://doi.org/10.12840/issn.2255-4165.2018.06.01.015

[46] Federica Pallavicini, Alessandro Pepe, and Maria Eleonora Minissi. 2019. Gaming in Virtual Reality: What Changes in Terms of Usability, Emotional Response and Sense of Presence Compared to Non-Immersive Video Games? *Simulation & Gaming* 50, 2 (2019), 136–159. https://doi.org/10.1177/1046878119831420

[47] Tomislav Pejsa, Michael Gleicher, and Bilge Mutlu. 2017. Who, Me? How Virtual Agents Can Shape Conversational Footing in Virtual Reality. In *Intelligent Virtual Agents*. Springer International Publishing, Cham, 347–359.

[48] José Pinheiro, Douglas Bates, and R Core Team. 2023. *nlme: Linear and Nonlinear Mixed Effects Models.* https://CRAN.R-project.org/package=nlme R package version 3.1-162.

[49] R Core Team. 2023. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

[50] Imogen C. Rehm, Emily Foenander, Klaire Wallace, Jo-Anne M. Abbott, Michael Kyrios, and Neil Thomas. 2016. What Role Can Avatars Play in e-Mental Health Interventions? Exploring New Models of Client–Therapist Interaction. *Frontiers in Psychiatry* 7 (2016). https://doi.org/10.3389/fpsyt.2016.00186

[51] Gabriela Ripka, Silke Grafe, and Marc Erich Latoschik. 2020. Preservice Teachers' encounter with Social VR – Exploring Virtual Teaching and Learning Processes in Initial Teacher Education. In *Proceedings of SITE Interactive Conference 2020*. Association for the Advancement of Computing in Education (AACE), 549–562.

[52] Yves Rosseel. 2012. lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software* 48, 2 (2012), 1–36. https://doi.org/10.18637/jss.v048.

i02

[53] Daniel Roth and Carolin Wienrich. 2018. Effects of Media Immersiveness on the Perception of Virtual Characters. In *Proceedings of the 15th Workshop on Virtual and Augmented Reality of the GI Special Interest Group VR/AR.*

[54] Ryan Schulz. 2023. Welcome to the Metaverse: A Comprehensive List of Social VR/AR Platforms and Virtual Worlds. https://ryanschultz.com/list-of-social-vr-virtual-worlds/. Accessed June, 2023.

[55] Andreas Siess and Matthias Wölfel. 2019. User color temperature preferences in immersive virtual realities. *Computers & Graphics* 81 (2019), 20–31. https://doi.org/10.1016/j.cag.2019.03.018

[56] Richard Skarbez, Frederick P. Brooks, and Mary C. Whitton. 2021. Immersion and Coherence: Research Agenda and Early Results. *IEEE Transactions on Visualization and Computer Graphics* 27, 10 (2021), 3839–3850. https://doi.org/10.1109/TVCG.2020.2983701

[57] Richard Skarbez, Frederick P. Brooks, Jr., and Mary C. Whitton. 2017. A Survey of Presence and Related Concepts. *ACM Comput. Surv.* 50, 6, Article 96 (nov 2017), 39 pages. https://doi.org/10.1145/3134301

[58] Richard Skarbez, Solène Neyret, Frederick P. Brooks, Mel Slater, and Mary C. Whitton. 2017. A Psychophysical Experiment Regarding Components of the Plausibility Illusion. *IEEE Transactions on Visualization and Computer Graphics* 23, 4 (2017), 1369–1378. https://doi.org/10.1109/TVCG.2017.2657158

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

[60] Mel Slater. 2018. Immersion and the illusion of presence in virtual reality. *British Journal of Psychology* 109, 3 (2018), 431–433. https://doi.org/10.1111/bjop.12305

[61] Mel Slater, Domna Banakou, Alejandro Beacco, Jaime Gallego, Francisco Macia-Varela, and Ramon Oliva. 2022. A Separate Reality: An Update on Place Illusion and Plausibility in Virtual Reality. *Frontiers in Virtual Reality* 3 (2022). https://doi.org/10.3389/frvir.2022.914392

[62] Sahinya Susindar, Mahnoosh Sadeghi, Lea Huntington, Andrew Singer, and Thomas K. Ferris. 2019. The Feeling is Real: Emotion Elicitation in Virtual Reality. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 63, 1 (2019), 252–256. https://doi.org/10.1177/1071181319631509

[63] Unity Technologies. 2019. Unity. https://unity3d.com. version 2019.4.8f1.

[64] Valve Corporation. 2021. SteamVR. https://store.steampowered.com/app/250820/SteamVR.

[65] Florian Weidner, Gerd Boettcher, Stephanie Arevalo Arboleda, Chenyao Diao, Luljeta Sinani, Christian Kunert, Christoph Gerhardt, Wolfgang Broll, and Alexander Raake. 2023. A Systematic Review on the Visualization of Avatars and Agents in AR & VR displayed using Head-Mounted Displays. *IEEE Transactions on Visualization and Computer Graphics* 29, 5 (2023), 2596–2606. https://doi.org/10.1109/TVCG.2023.3247072

[66] William Revelle. 2023. *psych: Procedures for Psychological, Psychometric, and Personality Research.* Northwestern University, Evanston, Illinois. https://CRAN.R-project.org/package=psych R package version 2.3.3.

[67] Erik Wolf, David Mal, Viktor Frohnapfel, Nina Döllinger, Stephan Wenninger, Mario Botsch, Marc Erich Latoschik, and Carolin Wienrich. 2022. Plausibility and Perception of Personalized Virtual Humans between Virtual and Augmented Reality. In *2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR).* 489–498. https://doi.org/10.1109/ISMAR55827.2022.00065

[68] Ke-Hai Yuan and Peter M Bentler. 2000. 5Three likelihood-based methods for mean and covariance structure analysis with nonnormal missing data. *Sociological methodology* 30, 1 (2000), 165–200.

[69] Katja Zibrek, Elena Kokkinara, and Rachel McDonnell. 2017. Don't stand so close to me: investigating the effect of control on the appeal of virtual humans using immersion and a proximity-based behavioral task. In *Proceedings of the ACM Symposium on Applied Perception* (Cottbus, Germany) *(SAP '17).* Association for Computing Machinery, New York, NY, USA, Article 3, 11 pages. https://doi.org/10.1145/3119881.3119887

[70] Katja Zibrek, Sean Martin, and Rachel McDonnell. 2019. Is Photorealism Important for Perception of Expressive Virtual Humans in Virtual Reality? *ACM Trans. Appl. Percept.* 16, 3, Article 14 (sep 2019), 19 pages. https://doi.org/10.1145/3349609

[71] Cindy Ziker, Barbara Truman, and Heather Dodds. 2021. Cross Reality (XR): Challenges and Opportunities Across the Spectrum. *Innovative Learning Environments in STEM Higher Education Opportunities, Challenges, and Looking Forward* (2021), 55–77.