Visual Indicators Representing Avatars' Authenticity in Social Virtual Reality and Their Impacts on Perceived Trustworthiness

Jinghuai Lin, Johrine Cronjé, Carolin Wienrich, Paul Pauli, and Marc Erich Latoschik

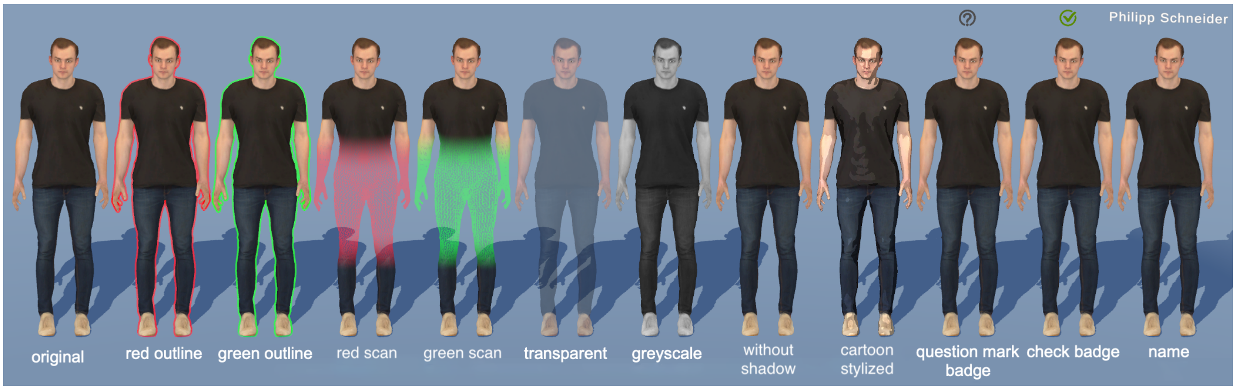

Fig. 1: Evaluated visual indicators of avatars' authenticity. From left to right: the original avatar, "red outline", "green outline", "red scan", "green scan", "transparent", "greyscale", "without shadow", "cartoon stylized", "question mark badge", "check badge", "name". Note that the "scan" effects are dynamic as the "scanned area" moving from head to toes.

Abstract—Photorealistic avatars show great potential in social VR and VR collaboration. However, identity and privacy issues are threatening avatars' authenticity in social VR. In addition to the necessary authentication and protection, effective solutions are needed to convey avatars' authenticity status to users and thereby enhance the overall trustworthiness. We designed several visual indicators (VIs) using static or dynamic visual effects on photorealistic avatars and evaluated their effectiveness in visualizing avatars' authenticity status. In this study we explored suitable attributes and designs for conveying the authenticity of photorealistic avatars and influencing their perceived trustworthiness. Furthermore, we investigated how different interactivity levels influence their effectiveness (the avatar was either presented in a static image, an animated video clip, or an immersive virtual environment). Our findings showed that using a full name can increase trust, while most other VIs could decrease users' trust. We also found that interactivity levels significantly impacted users' trust and the effectiveness of VIs. Based on our results, we developed design guidelines for visual indicators as effective tools to convey authenticity, as a first step towards the improvement of trustworthiness in social VR with identity management.

Index Terms—social VR, avatar, identity management, visual indicator, design guidelines, authenticity, trust

1 Introduction and Related Work

With the emergence and popularity of affordable VR devices, the number of VR applications and their users has been rapidly growing. Among these applications, social virtual reality (social VR or SVR), "*a growing set of multiuser applications that enable people to interact with one another in virtual spaces through VR head-mounted displays* [37]", is gaining considerable attention for its variety of applications [3, 15, 21, 51, 60]. Meanwhile, social VR is also evolving into a more integrated platform instead of a single application. As the concepts of virtual reality and social networking continue to merge, social VR has the potential to become an alternative realm for human socio-cultural activities that extends beyond real life [13, 45]. This trend also contributes to the shift of users' self-representations towards personalized (photo-)realistic avatars that resemble their likeness or represent their true identities [45].

However, current social VR platforms, especially those featuring photorealistic avatars (e.g., *Spatio*[1]), have barely addressed the potential challenges to users' identity and authenticity that such a trend could bring: how does one know that the photorealistic avatar they meet in the virtual world looks like the person controlling it? How can one confirm the identity of the person they are interacting with as they claim it to be? Design guidelines for social VR identity management are desired to enhance the trust and authenticity among users and increase the overall user experience in social VR.

Focusing on the impact of identity management on users' trust and acceptance toward photorealistic avatars and the interplay between privacy protections and social behavior, we initiated the first investigation into an identity management mechanism using visual indicators (VIs) that communicate the authenticity of an photorealistic avatar to users. We aim to answer the following research questions:

RQ1: What visual indicator designs can efficiently communicate the authenticity of avatars' identities?

- *Jinghuai Lin is with Human-Computer Interaction Group at University of Würzburg. E-mail: jinghuai.lin@uni-wuerzburg.de.*
- *Johrine Cronjé is with Department of Psychology I, Biological Psychology, Clinical Psychology and Psychotherapy at University of Würzburg. E-mail: johrine.cronje@uni-wuerzburg.de.*
- *Carolin Wienrich is with Psychology of Intelligent Interactive Systems Group at University of Würzburg. E-mail: carolin.wienrich@uni-wuerzburg.de.*
- *Paul Pauli is with Department of Psychology I, Biological Psychology, Clinical Psychology and Psychotherapy at University of Würzburg, and Center of Mental Health, Medical Faculty at University of Würzburg. E-mail: pauli@psychologie.uni-wuerzburg.de.*
- *Marc Erich Latoschik is with Human-Computer Interaction Group at University of Würzburg. E-mail: marc.latoschik@uni-wuerzburg.de.*

Manuscript received 25 March 2023; revised 17 June 2023; accepted 7 July 2023. Date of publication 3 October 2023; date of current version 31 October 2023. Digital Object Identifier no. 10.1109/TVCG.2023.3320234

[1] https://www.spatial.io/

RQ2: How do visual indicators affect the perceived trustworthiness of the avatars?

RQ3: Will the levels of interactivity influence the effect of visual indicators?

In our study, we first created eleven VIs and aimed to determine the design metaphors that can effectively impact users' trust towards an photorealistic avatar. We then further investigated whether the level of interactivity could influence the effectiveness of such an indicating system (the avatar was either presented in a static image, an animated video clip, or an immersive virtual environment to participants). We reported the analysis, results, and discussion, as the first contribution to the design guidelines for social VR that focuses on enhancing authenticity and trustworthiness.

2 Related Works

2.1 Identity infringement targeting digital bodies

An avatar is a graphical self-representation of a user that is driven by the user's movements in the virtual world [6]. Lin and Latoschik [32] have summarized three types of self-representations in social VR: uploaded avatars, customizable avatars, and personalized (photo-)realistic avatars. In the expected scenario of social VR as an extension of real-world socio-cultural activities, users will be encouraged to use an avatar that looks like themselves and uniquely represents their identity. This unique self-representation serves as the proxy of their physical body and can also be regarded as their digital bodies in the virtual world [32]. Recent advancements in 3D reconstruction technology have made it simple and affordable to create highly realistic, lifelike avatars with merely photos or videos [1, 61].

Identity infringement has long been a big concern for social media. For example, many fake accounts on Facebook impersonate others' identities to engage in fraud [49]. Similar infringement has been and will continue to take place in social VR. Falchuk et al. [14] pointed out the dangers of "social engineering" in social VR. This deceptive practice involves individuals being tricked by imposters posing as their friends, resulting in victims being defrauded of personal information and money. Lake [28] described the risks of identity theft in VR and pointed out that identity theft or infringement could happen when cybercriminals steal the digital bodies of others. Such infringement would be "far more intimate than the theft of a still image or a static fake Facebook profile [28]". As the trend of using photorealistic avatars to represent users' true identities in virtual reality grows, identity theft targeting these digital bodies could have profound impacts on users' trust and acceptance in social VR.

2.2 Identity management to ensure authenticity

Huang and Jung [22] defined the perceived authenticity of virtual humans as "*whether the virtual characters' claimed identities are perceived as authentic.*" For social VR citizens, authenticity is whether an avatar's designed social or individual identities are consistent with the claimed identity of the person controlling it.

We looked for measures that ensure authenticity in social VR. Promising progress has been made in user identification and authentication methods and algorithms utilizing VR sensors and data. For instance, Liebers et al. [30] utilized users' body motions in two task-driven scenarios to identify users with an accuracy of up to 90%, and used body normalization to increase the identification rate. Miller et al. [40] achieved 95% of identification rate using body motions with no specially designed identifying task. Schell et al. [53] compared three data representations (scene-relative, body-relative, and body-relative velocity data) of users' motions and different machine learning techniques for user authentication, and achieved 100% mean accuracy within 150s of motion recording with the combination of body-relative data and long short-term memory (LSTM) technique. Meanwhile, other modalities such as eye movement [54] and iris [26] have also been explored and utilized for authentication. With the recent release of the *Apple Vision Pro* Headset that uses iris scanning authentication (*Optic ID*)[2], we are optimistic that more and more manufacturers will provide hardware and software support for identity management.

However, to our knowledge, no existing social VR identity management system is integrated with such authentication, and there has been little research explicitly looking into the protection of digital bodies. In summary, there is a severe lack of relevant research, neither focusing on technical implementation nor providing design guidelines.

2.3 Measuring trust in virtual humans

Measuring trust in virtual agents and avatars is crucial in the evaluation and building of trust in the virtual worlds. While self-report questionnaires are commonly used in trust measurement, many validated questionnaires (such as the "Interpersonal Trust scale" (ITS) [52] and the "KUSIV3" [7]) either primarily assess participants' propensity to trust other (generalized trust) rather than trust in specific trustees (specific trust) [31], nor not suitable for virtual humans and virtual environments. In the literature of avatars, Bente et al. [10, 11] developed a 20-item questionnaires to measure interpersonal trust in avatar-mediated net-based collaboration. Surprenant [56] investigated avatar-mediated self-disclosure behaviors using a combination of several scales including the ITS and the Self-Disclosure Index [39].

Apart from subjective measures, behavioral measurements are increasingly gaining attention in the study of trust in virtual humans. Experimental methods such as a trust game [9, 20, 31] and the ask-endorse paradigm [20, 31, 48] were employed. Additionally, collaborative behavior [48] and mutual gaze during conversations [4] were also used as behavioral clues to investigate trust.

3 Methods

We aim to answer our **RQ1** with the VI creation and the pre-study, as described in **Section 3.1** and **Section 3.2**. Then, with the main study described in **Section 3.3** we explored **RQ2** and **RQ3**.

3.1 Visual indicator creation

As a first step, several VIs have been created as candidates for the evaluations (see Figure 1). The design of VIs went in two directions. On the one hand, VIs with explicit information can easily indicate the identity or status of a user in traditional social media. For example, *Twitter* used to use a blue "verified" checkmark next to the username to indicate an "active, notable, and authentic account of public interest."[3] This motivated us to use either a "grey question mark" or a "green checkbadge" as VIs. Besides, information that provides or indicates the identity of users, such as a full name, can also be a good sign of authenticity. On the other hand, avatars' appearance could intuitively influence users, with which they subconsciously apply their prior knowledge [24], without the information being explicitly communicated, potentially impacting trust-building [47]. For instance, human behavior and emotions can be influenced by hue (e.g., green, blue, red) and saturation (e.g., greyscale, dull color, rich color) [55]. Adopting the idea of image schemas and image-schematic metaphors [23–25], we used the combinations of colors, rendering style, and visual effects applied to the avatar features as metaphors to represent the authenticity of users, thus creating our implicit VIs. We further explain the metaphors of selected rendering styles/visual effects below.

[2]https://www.apple.com/newsroom/2023/06/introducing-apple-vision-pro/

[3]https://help.twitter.com/en/managing-your-account/legacy-verification-policy. Note: starting April 1, 2023, *Twitter* only provides the "blue checkmark" to accounts with an active subscription service: https://help.twitter.com/en/managing-your-account/about-twitter-verified-accounts

Color coding. Different colors have various or even opposite symbolic meanings in different cultures [50]. However, with globalization, the general public has gradually accepted some standard color coding concepts in visual design (Figure 2) [5]. Despite its positive symbolism in many cultures, such as of passion, energy, and excitement [42,50], the red color could physiologically increases blood pressure and respiratory rate [18], and is often associated with "danger" and "forbidden" in signage and GUI design. In addition, research has shown that the color red is often associated with distrust [55]. As the most restful color to the eye [42], green is often used as a symbol of safety and viability [42]. Such associations with green and red have been found across different occupational and cultural groups [46], as well as in GUI design of computer applications. Moreover, red and green are usually used in pairs for binary operations (e.g., red for "decline" and green for "accept" an incoming call. Figure 2A) [5]. Therefore, we chose red to warn users about suspicious identities and green as the opponent to symbolize that the avatar's identity is safe and trustworthy.

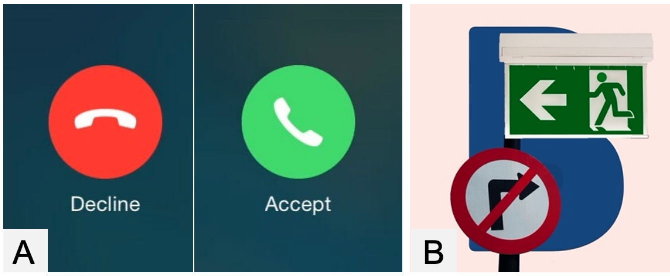

Fig. 2: Red and green are usually used as opponents in GUI and signage design. A) red for "decline" and green for "accept" an incoming call [5]. B) green is used on "safety exit" sign and red is used on "turing prohibited" sign. Image source: *www.beaverswood.co.uk/what-are-the-4-types-of-safety-signs/*

1) Outline. In visual design, the outline effect has an attention-grabbing, emphatic effect and is often used in text and icons. The outline effect is also used in gaming UI to highlight specific attributes of an object or a character, and the outline's color often indicates its status (e.g., teammates or enemies). We combined the red-green color coding with the outline effect to highlight the negative and positive symbolisms of the colors.

2) Scanning. The scanning effect is a dynamic visual effect that "reveals" the polygonal mesh underneath the avatar's texture from head to toe, reminiscent of the "scanning" process of a person commonly seen in movies or games. Similar to the outline effect, the scanning effect is combined with the red-green color coding to highlight the color symbolisms. Furthermore, the association of "scanning" with "revealing" in daily life (e.g., passengers going through security scanning at the airport) is the conceptual metaphor for "revealing" the true identity of an avatar. Besides, the metaphor of "scanning" combined with color coding has been widely used in computer games (e.g., *Assassin's creed*, *Cyberpunk 2077*), as an effective indication to highlight objects and enemies.

3) Opacity. The opacity image schema is based on the physical property of translucency, where an object or material allows light to pass through it to varying degrees. In western culture, a character with low opacity (i.e., a semi-transparent avatar) can often be associated with a "ghost" or a "spirit" [38]. Such association can be considered the instance of a "transparent is not solid" conceptual metaphor, which can further semantically dissolve the notion of "realness" and "authenticity". Visually, low opacity also decreases the visibility of information, which could be a sign of dishonesty. Thus, we adjusted the opacity of the avatar to 60% as another indicator of low authenticity.

Table 1: List of authenticity visual indicator designs.

	Negative metaphor	Positive metaphor
1) Outline	red outline	green outline
2) Scanning	red scanning	green scanning
3) Opacity	60% opacity	–
4) Saturation	0% saturation	–
5) Shadow	without shadow	–
6) Stylized	cartoonish	–
7) Badge/sign	"grey question mark"	"green check"
8) Name	–	full name

4) Saturation. The saturation image schema is based on the physical property of color intensity. In many cultural contexts, a greyscale image may be associated with a lack of vitality and energy. For instance, greyscale photos are often used to commemorate the deceased in some Asian cultures. A greyscale avatar might be perceived as "less lively" or "less engaging", stimulating subconscious distrust. Color also contains information and is often associated with emotions and attitudes [50]. Decreasing the saturation of an avatar can create a sense of ambiguity or uncertainty, which can be further associated with a lack of authenticity.

5) Shadow. Removing the shadow of an avatar is a subtle way of symbolizing that the person is "abnormal". In some cultures, the absence of shadow is interpreted as a symbol of supernature and a lack of humanity. Similar to a transparent avatar, the "without shadow" metaphor may evoke "spiritual" associations and can semantically indicate a low level of "realness" and "authenticity".

6) Stylized. The cartoon stylized design is characterized by simplified features, bright and contrasting colors, and unrealistic lighting. Compared to the original photorealistic avatar, the stylized avatar has low "fidelity" and could affect emotional reactions during social interaction [59]. In existing social VR applications, cartoon stylized designs are often used in playful contexts (e.g., *Rec Room*, *AltspaceVR*), where real-life identity and personal information are less involved.

Figure 1 demonstrates how the eleven visual indicators look on an photorealistic avatar. In Table 1, we list all VI designs and divide them into two columns based on their hypothesized impact on the perceived trustworthiness of the avatar: those with a positive impact (indicating the avatar's identity is authentic) and those with a negative impact (indicating the avatar's identity is false/suspicious).

3.2 Pre-study

To explore **RQ1**, we conducted the pre-study as an online survey in which participants helped us with the following tasks:

Avatar selection: Select one male and one female photorealistic avatar with medium attractiveness and high realism rating for the main study. Avatars' appearances might influence participants' subjective judgments, especially concerning perceived trustworthiness. It has been studied and confirmed [64] that attractiveness strongly influences the perceived trustworthiness of people. Thus, in the pre-study, participants rated the perceived attractiveness and realism of the candidates with likert-like scales. From the result, we then selected one male and one female avatar with a medium attractiveness rating and high realism rating.

VI design pre-screening: Determine which VI designs are more effective at conveying authenticity, as well as other potential emotional and metaphorical responses. In this task, participants were given background information about social VR, avatars, and potential identity theft targeting digital bodies. All the VI designs described in **Section 3.1** were presented to them, and they rated which VI designs are more effective at conveying authenticity, as well as other potential emotional and metaphorical responses. Different VI variances with the same design metaphor (e.g., green outline and red outline) were presented side by side to participants.

3.2.1 Materials and procedure

Avatar. The personalized photorealistic avatars were generated with the method by Achenbach et al. [1]. Seven male and seven female avatar candidates were randomly selected from the anonymous avatar dataset.

Participants. 50 (12 males and 38 females) participants were recruited via the university participant recruitment system, with an average age of 21.02 (*SD* = 2.42).

Procedure. Participants completed the task in an online survey implemented with Limesurvey 4.5.0[4]. After basic demographics questions, participants were provided background information on social VR, photorealistic avatars, identity theft targeting digital bodies, and the significace of authenticity and trustworthiness under such context. Then, on each page they watched a short video clip showcasing one avatar candidate and rated how much they think the presented avatar is attractive and looks like a real human on a 7-point Likert's scale, respectively. In the second task, we further explained the purpose of the visual indicators. Participants rated VI designs ("name" was not included in the pre-study but added afterward) on whether it is easy to notice the indicators and how much they can associate the design with the property of authenticity and other properties, such as trustworthiness and friendliness. The experiments of the pre-study was approved by the university's ethics committee. Each participant was compensated either with 5 euro or experiment attendance credit worth of 0.5 hour.

3.2.2 Pre-study results

Avatar selection. A female avatar (see Figure 3A) (attractiveness: *M* = 3.86 out of 7, *SD* = 1.58; realism: *M* = 4.80 out of 7, *SD* = 1.31) and a male avatar (see Figure 3B) (attractiveness: *M* = 3.72 out of 7, *SD* = 1.37; realism: *M* = 5.04 out of 7, *SD* = 1.23) were selected.

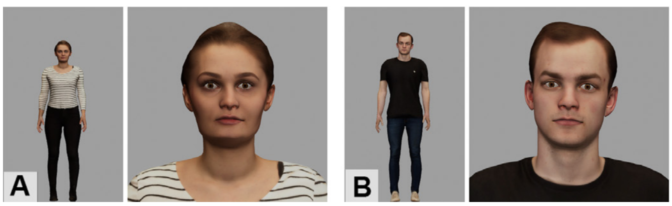

Fig. 3: A) selected female avatar with medium attractiveness and high realism. B) selected male avatar with medium attractiveness and high realism.

VI design pre-screening. To limit the conditions of the main study, we primarily chose VIs according to the scores associated with authenticity (see subplot *"Authenticity"* in Figure 4), thus "outline", "(with/without) shadow", and "cartoon stylized" were initially excluded. Meanwhile, with the aggregated measure of all the properties as shown in Figure 4, "outline" shows a significantly higher association with "trustworthiness", "friendliness", "helpfulness" and "sympathy"and was finally included for the main study.

[4]www.limesurvey.org

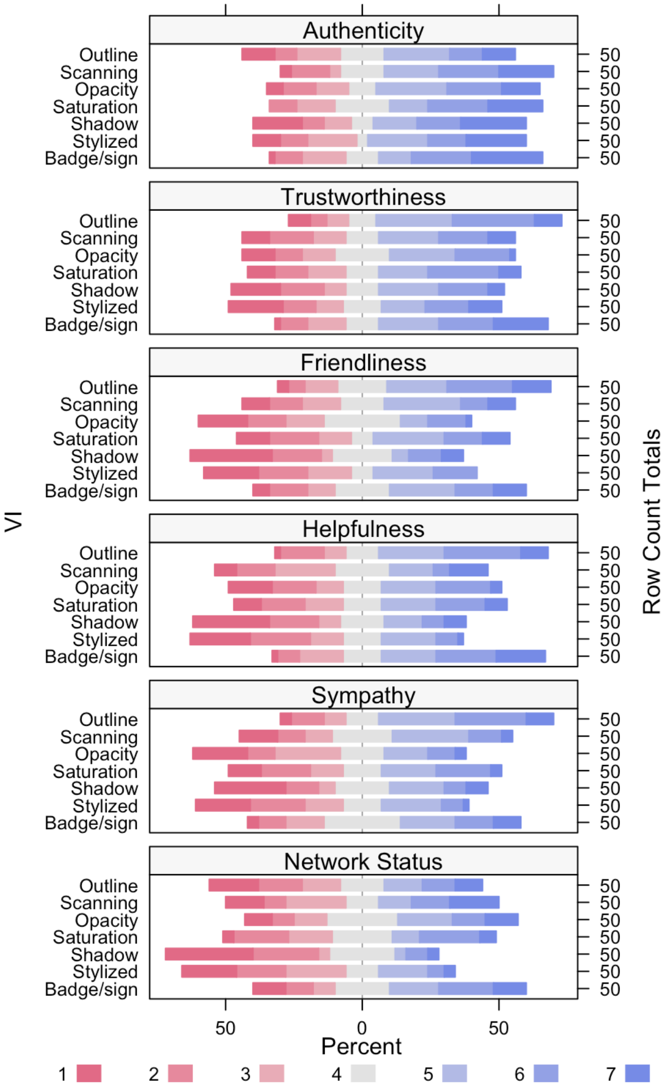

Fig. 4: Stacked bar chart for the signaling effectiveness rating (7-point Likert scale) associated with different properties of the VI designs. 1 - Strongly disagree, 7 - Strongly agree.

3.3 Main study: impacts of VIs on perceived trustworthiness

With the preliminary results from the pre-study, we applied the nine most effective visual indicators on the selected avatars. In the main study, we aim to answer **RQ2** and therefore constructed the following hypotheses:

H0: The VIs do not affect the perceived trustworthiness of avatars.

H1: Positive VIs will increase the perceived trustworthiness of avatars, and inversely for negative VIs.

Suppose **H1** is proven to be true, that certain VIs could affect the perceived trustworthiness of avatars, we are also interested in whether different interactivity levels of avatars affect the effectiveness of these VIs (**RQ3**). High interactivity is an essential feature that distinguishes social VR from traditional social networks and communication tools. As Lake [28] has pointed out, identity infringement in social VR would be far more intimate and harmful than the theft of a static profile picture. Literature shows that higher socio-spatial interactivity could positively affect social presence [16, 36, 58] and potentially lead to a more profound impact on factors with strong social attributes such as interpersonal trust towards avatars [11]. Therefore, we design three levels of interactivity, and make the following hypothesis:

H2: The perceived trustworthiness of the avatar will be different between different interactivity groups.

H3: With the level of interactivity and immersion increase, the effects of visual indicators will increase.

In the three levels of interactivity, the avatar will be presented either in a static image, resembling the condition of traditional social network profile image; in a animated video clip displayed on screen, resembling the condition of 3D online social platform (e.g., *Second Life*[5], MMORPGs); or in an immersive VR experience, representing the condition of modern social VR.

3.3.1 Experimental design.

The main study was conducted as a 3 x 10 mixed design, with three interactivity conditions (low interactivity, medium interactivity, and high interactivity) as the between-subject variable, and ten different VI conditions (the original avatar + nine different VIs) as the within-subjects variable.

Within-subjects: each participant went through ten different VI conditions, including the 1) original avatar without any VI, and the same avatar with VIs: 2) greyscale (0% saturation), 3) transparent (60% opacity), 4) green outline, 5) red outline, 6) green scanning, 7) red scanning, 8) "green check" badge, 9) "grey question mark" badge, and 10) name. Participants were asked to imagine that they were exploring a social VR application and to rate the avatars' perceived trustworthiness one by one.

Considering the number of VI designs used as within-subject conditions, it becomes necessary to address the potential order of presentation, which was explicitly considered in this study. We assumed that a carryover effect might not be presented in all between conditions in evaluating trust in avatars for several reasons. First, detailed background information and context have been provided to participants in advance. Both the avatar and the virtual environment were presented in a simple and intuitive way. It was unlikely that participants acquired extra knowledge through experience in previous conditions. Second, no theory or evidence suggests that learning, fatigue, habituation, sensitization, or adaption would impact how trustworthy participants rate the avatars. Lastly, the cognitive processes involved in evaluating trust are complex and multifaceted [8]. Participants may interpret different VIs from different perspectives based on various factors such as context and personal experience. These factors could lead to unique trust evaluations in each condition, further reducing the likelihood of carryover effects.

Nevertheless, we cannot rule out the possibility of a carryover effect. Research has shown evidence that a high level of immersion increases arousal responses to stimuli [17]. The emotional responses to the avatar in the immersive virtual environment might be enhanced, and carryover effects might appear in the immersive condition (Group C). Thus, we present the VI conditions either in ascending or descending order to exploratively investigate if such carryover effects exist.

Between-subjects: participants were divided into three groups with different interactivity levels. In Group A, "low interactivity", participants saw the avatar as a static image (see Figure 5a) embedded in the online survey. In Group B, "medium interactivity", participants saw a short video in which the avatar was animated. Both Groups A and B participated online with desktop display. The resolution of the image and the video clip were 2560 × 1440 pixels and participants could freely zoom in or out on their monitors. In Group C, "high interactivity", participants interacted with the avatar in an immersive VR experience while wearing an *Oculus Quest 2* HMD (1832 × 1920 pixels per eye).

Fig. 5: The virtual environment and the avatar. A) Participants in Group A saw static images of the avatar on desktop display. B) In Group B and C the avatar was animated, walked towards and waved at the participants. The avatar in Group B was presented in video clips with the same viewing angle and resolution as the images in Group A, on desktop display; the avatar in Group C was presented in an immersive virtual environment with an HMD. C) The questionnaire appeared in the virtual environment in Group C.

Avatar. Males and females may differ in trust socialization and evaluation [8]. To minimize potential gender influences in our experiment, each participant was exposed to an avatar representing the same gender as themselves. This approach aligns with established practices in trust research [43,62]. The avatar's animation was recorded with the OptiTrack[6] mocap system: the avatar walks slowly towards the participant and waves at them to greet them, with a slight smile and its head and eyes following the participants naturally.

Virtual environment. The virtual environment was implemented using Unity Engine[7] 2020.3.14f1. A modern office asset that provided a neutral emotional environment with a similar level of realism to the avatars was chosen (Figure 5). Additionally, in Group C the questionnaire appeared in the virtual environment next to the avatar (Figure 5C). Participants could use their VR controllers to select and complete the questionnaire.

Trust measurement questionnaire. We created a 9-item 7-point Likert-like rating scale (see Table 2) to subjectively measure participants' trust in the avatars. One item was created to assess the trust explicitly at an overall level (Q9). As the construct of trust is complex and can be subject to multiple interpretations [8], we developed items to further break down trust into distinct dimensions. Drawing from the widely recognized ABI model of trust [35], we considered the dimensions of perceived ability (competence), benevolence (good intention), and integrity. Moreover, in the realm of net-based communication where privacy concerns tend to arise, self-disclosure is closely associated with trust [27, 44, 57] and has frequently been measured as either a dimension or an product of trust [11,56]. Thus, in addition to the explicit trust rating we created two items for each of the dimension (intergrity, self-disclosure, competence, and intention) that contribute to trust.

Pre-questionnaire and post-questionnaire. The pre-questionnaire included basic demographics, internet and social network usage, the Trust Short Scale (KUSIV3) [7], and the Same-Sex Stranger questions from the Self-Disclosure Index (SDI) [39]. The post-questionnaires included the Virtual Human Plausibility Questionnaire [34] and two open questions for qualitative feedback on the VIs.

Participants. For Groups A and B, participants were recruited via the university participant recruitment system and Amazon Mechanical

[5]https://secondlife.com/

[6]https://optitrack.com/

[7]unity.com

Table 2: Scale for rating trust towards an avatar in social VR

Please rate the following statements from 1 (strongly disagree) to 7 (strongly agree) according to your intuition.	
(1) I believe the appearance of the avatar matches the person controlling it.	integrity
(2) I am willing to tell the avatar (the person controlling it) my real name.	self-disclosure
(3) I am willing to tell the avatar (the person controlling it) what I have done today.	self-disclosure
(4) I believe the avatar (the person controlling it) would keep a secret that is harmful to my reputation.	integrity
(5) If I need some advice, I would feel comfortable asking the avatar (the person controlling it) for help.	competence
(6) If we were performing a task together, I feel I could follow the guidance from the avatar (the person controlling it).	competence
(7) I would feel comfortable being with the avatar (the person controlling it) in a room.	intention
(8) In general, I think the avatar (the person controlling it) has good intentions.	intention
(9) In general, I find this avatar trustworthy.	explicit

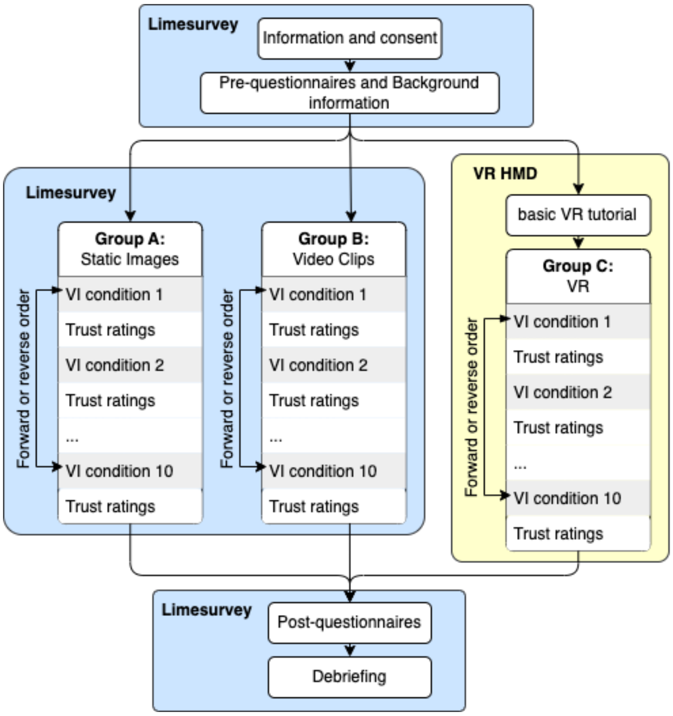

Fig. 6: Illustration of the experiment procedure. Groups A and B participated online and completed the experiment on Limesurvey. Group C participated on-site. They completed the VI rating in virtual environment with a VR HMD, and the rest of the experiment on Limesurvey with a laptop provided

Turk[8], and participated online. Group A comprises 29 participants (13 males and 16 females) with an average age of 29.41 years ($SD = 9.36$). Group B comprises 31 participants (12 males and 19 females) with an average age of 30.61 ($SD = 11.44$). For Group C, 67 participants were recruited via the university participant recruitment system and participated onsite. Due to the limited availability of male participants to attend on-site, we only recruited female participants in Group C with an average age of 22.95 ($SD = 3.07$). In the exploratory analyses of Group A and B (see **Section 4** Exploratory measures), no significant difference in the trust rating between male and female participants was observed. Thus we considered it appropriate to only include female participants in Group C.

Procedure. For Groups A and B, after completing the pre-questionnaire, they were provided background information about social VR, photorealistic avatars, and identity theft targeting digital bodies. Thereafter, participants proceeded further to the next pages, where the same avatar in different VI conditions was presented one by one in static images or animated video clips, and they were asked to imagine that they were exploring a social VR application while this avatar was standing in front of them. They needed to rate the avatar with the trust measurement questionnaire under all the conditions. Afterward, they completed the post-questions followed by the debriefing about the main purpose of the study. The experiment was implemented as an online survey using Limersurvey 4.5.0. For Group C, participants still needed to complete the pre-questions and post-questions on a laptop, but they viewed and rated the avatar and the VIs in immersive VR with an HMD. In all groups, there was no time limit for the exposure to the stimuli or the rating of the avatars and VIs. The experiments of the main study was approved by the university's ethics committee. Each participants were compensated either with 5 euro or experiment attendance credit worth of 0.5 hour. Figure 6 illustrated the procedure of the main study.

4 Results

4.1 Main results

We used the score of the explicit trust item (see Q9 in Table 2) as the main measurement of perceived trustworthiness, along with the scores for the four dimensions of trust: integrity, self-disclosure, competence, and intention (average of the two related items). Regarding the unequal sample sizes between groups, it is widely accepted that an ANOVA does not require equal sample sizes as a precondition [19, 63]. Although it may lead to a reduction in statistical power, an ANOVA is generally robust to unequal sample sizes in terms of Type I error [12]. Nevertheless, initial analysis indicated violations of normality and variance homogeneity. To address these violations, we employed the robust version of the two-way mixed ANOVAs [33]. This robust approach enhances the method's resilience to unequal sample sizes and violations of the assumptions, thus increasing the validity of the results. The robust two-way mixed ANOVAs were performed on the 20% trimmed means using the bwtrim function of the WRS2 package in R [33], with interactivity level as the between-subject variable and visual indicators as the within-subject variable. A robust version of Cohen's d [2] was computed as effect size for the main effects using the akp.effect function of WRS2. We also performed post-hoc tests with 20% trimmed means using the rmmcp function of WRS2 when applicable. All reported means (M) and standard deviations (SD) are trimmed values.

There was a significant main effect of the VI condition on the explicit rating, $F(9, 36.80) = 10.89$, $p < .001$, robust Cohen's $d = -0.19$, and the four trust-dimension ratings (Integrity: $F(9, 36.07) = 20.29$, $p < .001$, robust Cohen's $d = -0.06$; Self-disclosure: $F(9, 36.32) = 13.99$, $p < .001$, robust Cohen's $d = -0.31$; Competence: $F(9, 34.80) = 9.94$, $p < .001$, robust Cohen's $d = -0.26$; Intention: $F(9, 36.37) = 14.76$, $p < .001$, robust Cohen's $d = -0.05$). We found no significant main effect of interactivity level on explicit rating and all trust dimensions. However, there was a tendency ($p = .0568$) that interactivity level can impact self-disclosure. There was no significant interaction between the VI condition and interactivity on the explicit rating and competence; there were significant interactions between the VI condition and interactivity level on integrity ($F(18, 32.25) = 3.47$, $p < .001$), self-disclosure ($F(18, 32.05) = 2.97$, $p = .003$), and intention ($F(18, 35.06) = 4.23$, $p < .001$).

[8]mturk.com

Post-hoc tests for within-subject main effects. For the explicit trust rating, the avatar with the VI "name" was rated significantly more trustworthy (M = 5.08, SD = 1.26) than the original avatar (M = 4.83, SD = 1.19), p <.001. While there was no significant difference between the VI "check badge" and the original avatar, all the other VI conditions were rated significantly less trustworthiness than the original, including "positive" VIs. VIs with red color and "greyscale" were the most effective indicators to reduce trust. Besides, avatars with VIs "red outline", "red scanning", and "question mark badge" were rated as significantly less trustworthy than their "positive" variants, i.e., "green outline", "green scanning", "check badge". Other trust-dimension ratings show similar results, except that the VI "check badge" (M = 5.23, SD = 1.69) also significantly decreased self-disclosure rating, compared to the original avatar (M = 5.29, SD = 1.68), p = 0.003. From the above analysis, the null hypothesis **H0** is rejected, and **H1** can be partially accepted. While most negative VIs did significantly decrease the perceived trustworthiness of the avatar, except for the VI "name", positive VIs did not show the positive effects as we expected. Figure 7 shows the stacked bar chart for the explicit trust rating of different VI conditions; Figure 8 shows the bar chart for trust-dimension rating of different VI conditions.

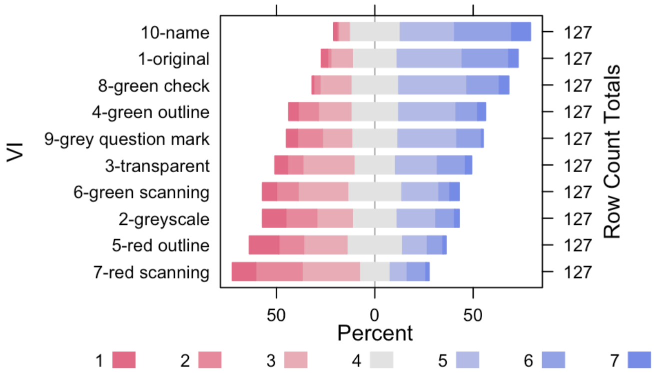

Fig. 7: Stacked Bar chart for the explicit trust rating of different VI conditions.

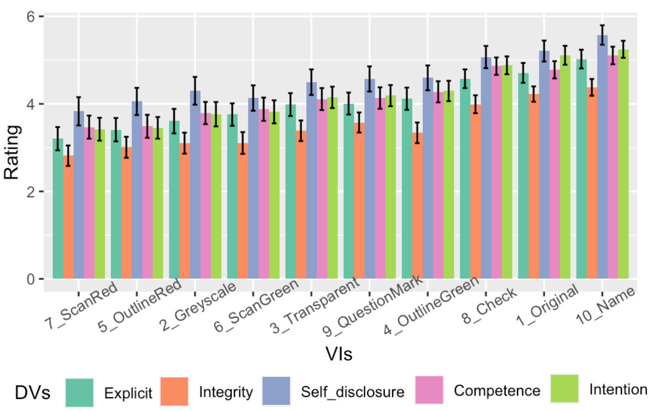

Fig. 8: Bar chart for trust-dimension ratings of different VI conditions.

Post-hoc tests for interactions. As there were significant interactions between VI condition and interactivity on integrity, self-disclosure, and intention, we tested the simple main effects of interactivity level at VIs and performed pairwise comparisons among group levels with Bonferroni correction. For integrity, ratings on both VI "green scanning" and "red scanning" significantly decreased while interactivity increased. For self-disclosure, ratings on all VIs except "greyscale", "green scanning", and "red scanning" significantly increased while interactivity increased. For intention, while ratings on VI "name" and "question mark badge" increased along with interactivity level, ratings on VI "green scanning" show an opposite tendency. Lastly, for the explicit rating, although the robust mixed-ANOVA detected no significant interaction, we discovered a tendency of increasing ratings for VIs "transparent" and "green outline" with interactivity increased, and an opposite tendency for "red scanning". These results are illustrated in Figure 9. Therefore, **H2** can be accepted. **H3** only holds for certain VIs within certain trust dimensions, while opposite conclusions emerge in other dimensions.

4.2 Exploratory measures

We performed robust mixed-ANOVAs with gender as a between-subject variable on data collected from Groups A and B. No significant main effect of gender nor interaction effect for all dependent variables was found. We also performed Bayesian two-sample t-tests on both Groups using the ttestBF function of the BayesFactor package in R [41]. Results showed anecdotal or moderate evidences for the null hypotheses (that the means of ratings from both genders are equal) over alternative hypotheses (Bayes Factor = 0.45 for Group A explicit ratings; Bayes factor = 0.13 for Group B explicit ratings). Therefore, gender was not considered a confounding variable in this study.

We also performed robust mixed-ANOVAs with the orders of VIs presentation (ascending v.s. descending) as the between-subject variable. For Groups A and B, there was no significant main effect of orders nor interaction. For Group C, there was no significant main effect of orders but significant interactions for all dependent variables except for self-disclosure. Using the explicit trust rating as the main measurement, we found that "green scanning" tends to be rated more trustworthy in descending order (M = 4.09, SD = 1.42) than in ascending order (M = 3.38, SD = 1.34), p = 0.039. Similarly, both "scanning" effects received higher scores in the integrity, competence and intention dimensions in descending order. In contrast, "greyscale" received higher scores in the integrity and competence dimensions in ascending order. Results indicated no carryover effect under low level of immersion (Groups A and B); under high level of immersion (Group C) carryover effects presented for certain VI conditions.

4.3 Qualitative feedback

Responses to the open questions also qualitatively revealed how attributes in the VI designs influence participants' trust. From the total of 127 responses, a considerable amount of (n = 25) participants explicitly mentioned that they trusted avatars that appeared "more realistic", "look more human", or "are more similar to a real (natural) person". Mainly referring to the "scanning", "greyscale" and "transparent" VIs, 22 participants mentioned that visual add-ons to the avatars made them look "less plausible", "creepy", "scary", "unnatural", "weird" and "disturbing", thus decreasing their trust. Many (n = 26) pointed out that they trust avatars with red color less and (or) trust more with green color. Some (n = 19) mentioned that they trusted avatars with the name or the "check badge", and 4 of them directly pointed out the similarity of how such explicit information influences their trust to that on traditional social media, such as Twitter. Besides, participants also judged the trustworthiness of the avatars on other attributes, such as whether the avatars were making eye contact and how plausible they moved.

5 Discussion

5.1 Visual indicators

The presented study showed that VIs influenced users' trust towards photorealistic avatars, even when they are not informed of the purpose of these indicators. The only VI that increased the avatar's perceived trustworthiness is "name"; "check badge" received the second highest score. This suggests that incorporating explicit information to reflect users' authenticity would help gain trust.

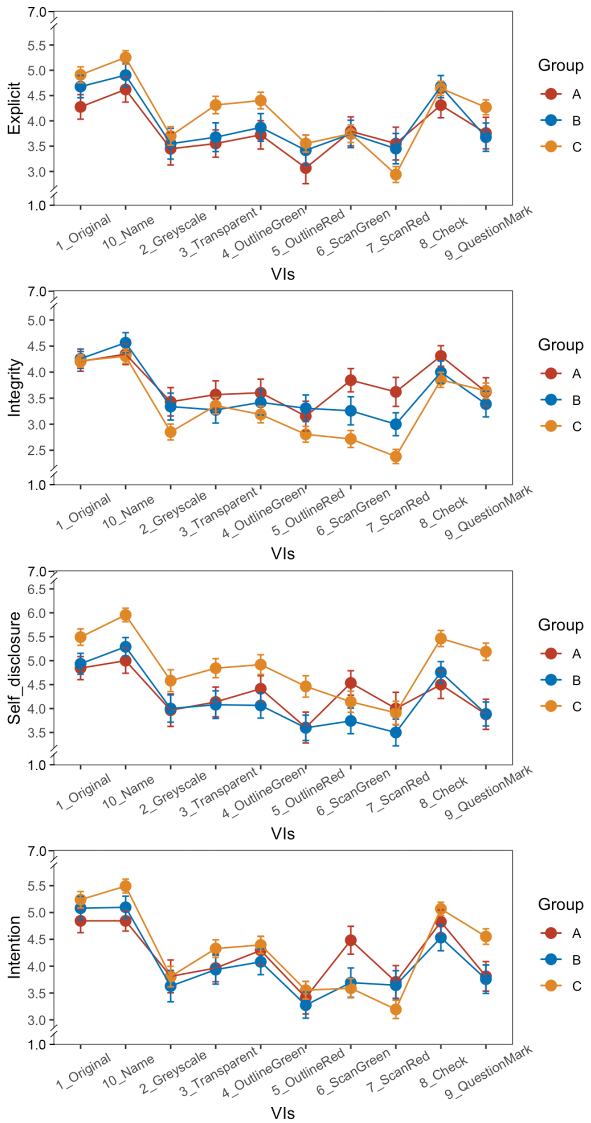

Fig. 9: Line charts of trust-dimension ratings of different VIs from three interactivity groups

As expected, participants can associate negative metaphors with untrustworthy attributes. Contrary to our expectations, "positive" VIs except for "name" and "check badge" also led to a decrease in trust, although to a lesser extent than their "negative" variants. It could be that the changes of color and visual add-ons on the avatar's appearance intuitively weakened the coherence of the avatar to the virtual environment and the (implicit) sensory impressions of a virtual human and thus its plausibility [29, 34], especially that the avatar and the virtual environment have a high level of realism. Such an inference is also in line with the qualitative feedback, that the visual add-ons made them look "less plausible" and "creepy".

Color as an attribute of the VIs has an impact on users' trust. Green was seen as an implication of trustworthiness, and red was significantly associated with distrust. Although the addition of the color outline and the scanning effect can be disturbing and reduce the perceived trustworthiness, colors may still be useful to indicate the authenticity in a more plausible and less intrusive way, e.g., a wristband with different colors, which is left to be explored in future research.

5.2 Interactivity

As interactivity increased, there was a consistent tendency for the original avatar to be perceived as more trustworthy. It might be that the appearance of the original avatar already created a relatively high level of trust and was further reinforced by increased interactivity and social presence. Future experiments could explore how the increase in interactivity influences the perceived trustworthiness of an avatar with an untrustworthy appearance.

With the addition of VIs, the impact of interactivity shows different manifestations. Notably, the dynamic scanning effect was significantly more effective in reducing trust at higher levels of interactivity, regardless of color. It could be that at high levels of interactivity and immersion, the scanning effect was more intrusive and incoherent to the virtual environment, thus triggering lower trust.

5.3 Carryover effects

In the exploratory measures, we discovered no carryover effect in Groups A and B, but in Group C with the trust rating of VI "green scanning". Figure 10 plots the explicit trust rating from the two reversed presentation orders. The increase of trust in VI "green scanning" may be because participants in descending orders viewed the explicit VI designs "name" and "badge" first and might have a clearer perception of the meanings of these VIs. When the VI "green scanning" was presented, they recognized the conceptual metaphor initially designed to evoke trust more easily. Such carryover effect was further enlarged by the high level of immersion in Group C, in which participants experienced higher presence and co-presence with the avatar. Similarly, increased arousal responses might explain the differences in ratings of VI "red scanning" and "greyscale" in other trust dimensions. Despite the difference in the trust rating of "green scanning", our deduced findings summarized in **Sections 5.1** and **5.2** still hold for both orders.

Our findings are the first insights into the effects of the chosen VIs on the perceived trustworthiness of avatars. Despite our assumption, our results give rise to a suspected carryover. Thus, our results provide a foundation for future studies and highlight the importance of the consideration of carryover effects on trust evaluation under similar scenarios.

5.4 Implications for design

Based on our previous observations and discussions, we provide a summary of the implications and suggestions for design guidelines in social VR. Specifically, we focus on conveying the authenticity of photorealistic avatars to enhance their overall trustworthiness.

Display explicit information. In many traditional social media, users use their own identity and real name; this makes users feel that they are communicating with a "person" who has a real identity and therefore trust them more. The same could be true for social VR if it aims to be the extension of real life and encourages users to enter the virtual world as themselves. Meanwhile, displaying more personal information under users' preferences, such as their region, may further increase the authenticity and trustworthiness of the avatars.

Reduce visual add-ons for higher trust. To achieve a higher level of trust, our evidence suggests to avoid visual add-ons such as outlines that would undermine avatars' realism and plausibility. When it is desired for verified users to signal their authenticity, it is wiser to embed such information around the avatar, e.g., a floating text or symbol over the head of the avatar, rather than modifying the avatar intrusively.

Use combination of color and explicit information. Colors are suitable indicators of authenticity and trust, according to both quantitaive and qualitative feedbacks. The combination of colors and

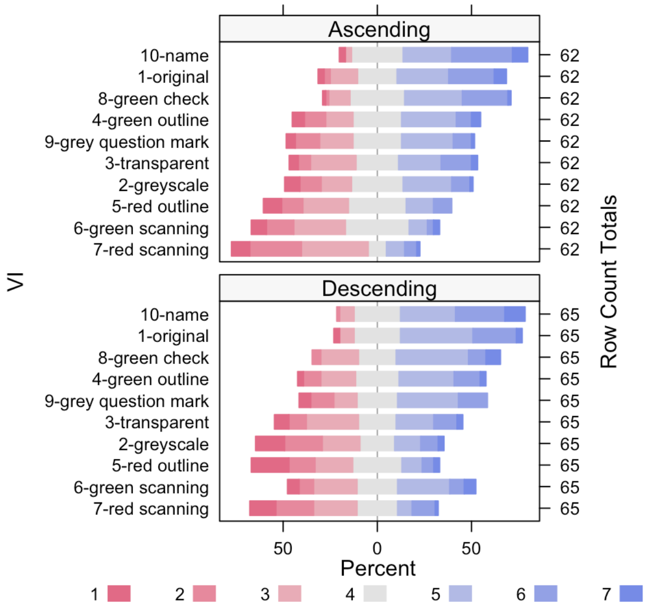

Fig. 10: Stacked bar chart of the explicit trust rating from two presentation orders

explicit information could indicate whether the information is genuine or not. For example, a name in red could indicate that it might be a fake name.

Use dynamic and intrusive effects for warning. In the case when users need to be warned about the suspicious identity or behavior of a specific avatar, our results indicate that VIs with more intrusive and dynamic effects such as the "scanning" and the "greyscale" would be helpful. These VIs could intuitively trigger lower trust and protect users from potential security and privacy issues such as social engineering. This could be especially true in immersive social VR with high interactivity.

Option to enable and disable the visual indicators. As the VIs could lead to the break of realism and plausibility, they do not necessarily need to be displayed continiously. For example, a social VR user could press a button on their controllers to "scan" all the avatars around, and different colors could mark the avatar's status.

5.5 Limitations

Our study is not without limitations. Firstly, participants were only asked to imagine being in a social VR application rather than really being exposed to one, and the "avatars" participants saw were only pre-scripted agents. Although we have provided background information on social VR and digital bodies, it was challenging to ensure that participants fully understood the contexts and placed themselves in the imaginary scenario. Secondly, the unequal sample sizes among groups and the inclusion of only female particiapants in Group C (despite no gender difference in Groups A and B was found) might hinder the generalization of conclusions. Thirdly, the lack of avatar diversity may create certain bias. Although not intentionally but through the criteria of medium attractiveness and high realism, both avatars selected in the pre-study were of Caucasian ethnicity. Furthermore, we discovered that while the suggested rendering styles have minimal influence on the recognizability of avatars, they do, to some extent, reduce the realism and coherence of the avatars within the virtual environment. These effects might have additional consequences for the avatars beyond their perceived authenticity and trustworthiness, which have yet to be investigated. Lastly, using an unvalidated subjective trust measurement might lead to low construct validity and reliability.

5.6 Future work

In future studies, it is preferred to have an actual social VR application combined with existing privacy protection mechanisms for experiments. In such a platform, users who have their own photorealistic avatars could interact with each other and have more comprehensive usages and understanding of the mechanisms. With such an platform, we could also explore the conditions with a higher level of interactivity: the photorealistic avatar (controlled by another person) actively respond to participants' actions, such as starting a conversation or shaking hands. Besides, recruiting regular social VR users with more experience from current social VR platforms (e.g., *VRChat*) could also provide more constructive input.

The suggested rendering styles offer insights into the manipulation of the perceived trustworthiness and authenticity of avatars. However, it is crucial to consider other potential factors (such as emotional afffect) of these rendering styles on users in future research. Moreover, investigating the influence of various appearance attributes such as body shape and hairstyles on trust could be another future direction. In addition to visual indications, future studies could consider combining multisensory signals (e.g., audio, haptics), as multimodal interactions are integral to social VR and will likely enhance the effects. Furthermore, to prevent bias among different user groups, future studies should take into account the diversity of avatars, encompassing not only different ethnicities and genders but also considering a broader range of characteristics.

In the evaluation of trust, validated subjective and objective measurements (e.g., behavioral measurements) designed for social VR scenarios [31] could be developed and used to evaluate trust sufficiently.

6 Conclusion

Our study proposed a visual indicator mechanism with several VI designs to represent photorealistic avatars' authenticity in social VR. With a pre-study and the main study, we investigated the effective attributes and design metaphors that convey avatars' authenticity status. Furthermore, the influence of their perceived trustworthiness, and how interactivity impacts the effectiveness of our designs were investigated. As a result, we found that using the full name of the owner of the avatar as an indicator can enhance trust, while most other visual indicators can significantly reduce users' trust. Tthe level of interactivity significantly impacts users' trust in photorealistic avatars and the effectiveness of the visual indicators. In the summary of our finding, we provided implications on how the visual indicators can be designed and applied. Our work is a first step in the direction of design guidelines for conveying avatars' authenticity and improving the overall trustworthiness in social VR featuring photorealistic avatars. Such guidelines play an essential role in identity management that protects users from the threats of identity infringement and social engineering. We also identified potential carryover effects under high level of immersion and provided insights for trust evaluation under similar scenarios. With the goal of establishing a secure and trustworthy social VR community that could serve as an alternative realm of human socio-cultural activities, future work could continue the exploration of design guidelines against different threats with various protection mechanisms to protect users' digital bodies.

Acknowledgments

This work is part of the Privacy Matters (PriMa) project. The PriMa project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 860315. Special thanks to Dr. Jean-Luc Lugrin for providing feedback and suggestions on the manuscript.

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, pp. 1–10. ACM, Gothenburg Sweden, Nov. 2017. doi: 10.1145/3139131.3139154 2, 4

[2] J. Algina, H. J. Keselman, and R. D. Penfield. An alternative to Cohen's standardized mean difference effect size: a robust parameter and confidence interval in the two independent groups case. *Psychological methods*, 10(3):317, 2005. Publisher: American Psychological Association. 6

[3] S. Arlati, V. Colombo, D. Spoladore, L. Greci, E. Pedroli, S. Serino, P. Cipresso, K. Goulene, M. Stramba-Badiale, G. Riva, A. Gaggioli, G. Ferrigno, and M. Sacco. A Social Virtual Reality-Based Application for the Physical and Cognitive Training of the Elderly at Home. *Sensors*, 19(2):261, Jan. 2019. Number: 2 Publisher: Multidisciplinary Digital Publishing Institute. doi: 10.3390/s19020261 1

[4] S. Aseeri and V. Interrante. The Influence of Avatar Representation on Interpersonal Communication in Virtual Social Environments. *IEEE Transactions on Visualization and Computer Graphics*, 27(5):2608–2617, May 2021. Conference Name: IEEE Transactions on Visualization and Computer Graphics. doi: 10.1109/TVCG.2021.3067783 2

[5] N. Babich. Using Red and Green in UI Design, Jan. 2019. https://uxplanet.org/using-red-and-green-in-ui-design-66b39e13de91 (accessed Mar. 16, 2023). 3

[6] J. N. Bailenson, A. C. Beall, J. Loomis, J. Blascovich, and M. Turk. Transformed Social Interaction: Decoupling Representation from Behavior and Form in Collaborative Virtual Environments. *Presence: Teleoperators and Virtual Environments*, 13(4):428–441, Aug. 2004. doi: 10.1162/1054746041944803 2

[7] C. Beierlein, C. J. Kemper, A. Kovaleva, and B. Rammstedt. Kurzskala zur Messung des zwischenmenschlichen Vertrauens: Die Kurzskala Interpersonales Vertrauen (KUSIV3). 2012. Publisher: DEU. 2, 5

[8] A. Ben-Ner and F. Halldorsson. Trusting and trustworthiness: What are they, how to measure them, and what affects them. *Journal of Economic Psychology*, 31(1):64–79, Feb. 2010. doi: 10.1016/j.joep.2009.10.001 5

[9] G. Bente, T. Dratsch, S. Rehbach, M. Reyl, and B. Lushaj. Do You Trust My Avatar? Effects of Photo-Realistic Seller Avatars and Reputation Scores on Trust in Online Transactions. In F. F.-H. Nah, ed., *HCI in Business*, Lecture Notes in Computer Science, pp. 461–470. Springer International Publishing, Cham, 2014. doi: 10.1007/978-3-319-07293-7_45 2

[10] G. Bente, S. Rüggenberg, and N. C. Krämer. Social presence and interpersonal trust in avatar-based, collaborative net-communications. In *PRESENCE 2004: Proceedings of the 7th Annual International Workshop on Presence*, pp. 54–61, 2004. 2

[11] G. Bente, S. Rüggenberg, N. C. Krämer, and F. Eschenburg. Avatar-Mediated Networking: Increasing Social Presence and Interpersonal Trust in Net-Based Collaborations. *Human Communication Research*, 34(2):287–318, Apr. 2008. doi: 10.1111/j.1468-2958.2008.00322.x 2, 5

[12] M. J. Blanca Mena, R. Alarcón Postigo, J. Arnau Gras, R. Bono Cabré, and R. Bendayan. Non-normal data: Is ANOVA still a valid option? 2017. Accepted: 2018-05-07T11:07:07Z Publisher: Facultad de Psicología de la Universidad de Oviedo y el Colegio Oficial de Psicólogos del Principado de Asturias. 6

[13] J. D. N. Dionisio, W. G. B. Iii, and R. Gilbert. 3D Virtual worlds and the metaverse: Current status and future possibilities. *ACM Computing Surveys*, 45(3):1–38, June 2013. doi: 10.1145/2480741.2480751 1

[14] B. Falchuk, S. Loeb, and R. Neff. The Social Metaverse: Battle for Privacy. *IEEE Technology and Society Magazine*, 37(2):52–61, June 2018. Conference Name: IEEE Technology and Society Magazine. doi: 10.1109/MTS.2018.2826060 2

[15] K. Foerster, R. Hein, S. Grafe, M. E. Latoschik, and C. Wienrich. Fostering Intercultural Competencies in Initial Teacher Education. Implementation of Educational Design Prototypes using a Social VR Environment. In *Innovate Learning Summit*, pp. 95–108. Association for the Advancement of Computing in Education (AACE), 2021. 1

[16] D. R. Fortin and R. R. Dholakia. Interactivity and vividness effects on social presence and involvement with a web-based advertisement. *Journal of Business Research*, 58(3):387–396, Mar. 2005. doi: 10.1016/S0148-2963(03)00106-1 4

[17] D. Gall and M. E. Latoschik. Visual angle modulates affective responses to audiovisual stimuli. *Computers in Human Behavior*, 109:106346, Aug. 2020. doi: 10.1016/j.chb.2020.106346 5

[18] R. M. Gerard. *Differential effects of colored lights on psychophysiological functions*. PhD Thesis, University of California, Los Angeles., 1958. 3

[19] K. Grace-Martin. When Unequal Sample Sizes Are and Are NOT a Problem in ANOVA, Dec. 2020. https://www.theanalysisfactor.com/when-unequal-sample-sizes-are-and-are-not-a-problem-in-anova/ (accessed June. 14, 2023). 6

[20] J. Hale, M. E. Payne, K. M. Taylor, D. Paoletti, and A. F. De C Hamilton. The virtual maze: A behavioural tool for measuring trust. *Quarterly Journal of Experimental Psychology*, 71(4):989–1008, Apr. 2018. Publisher: SAGE Publications. doi: 10.1080/17470218.2017.1307865 2

[21] A. Heath. Inside Facebook's metaverse for work, Aug. 2021. https://www.theverge.com/2021/8/19/22629942/facebook-workrooms-horizon-oculus-vr (accessed June. 14, 2023). 1

[22] J. Huang and Y. Jung. Perceived authenticity of virtual characters makes the difference. *Frontiers in Virtual Reality*, 3, 2022. 2

[23] J. Hurtienne. How cognitive linguistics inspires HCI: image schemas and image-schematic metaphors. *International Journal of Human–Computer Interaction*, 33(1):1–20, 2017. Publisher: Taylor & Francis. 2

[24] J. Hurtienne and L. Blessing. Design for Intuitive Use-Testing image schema theory for user interface design. In *DS 42: Proceedings of ICED 2007, the 16th International Conference on Engineering Design, Paris, France, 28.-31.07. 2007*, pp. 829–830, 2007. 2

[25] J. Hurtienne and J. H. Israel. Image schemas and their metaphorical extensions: intuitive patterns for tangible interaction. In *Proceedings of the 1st international conference on Tangible and embedded interaction*, pp. 127–134, 2007. 2

[26] B. John, S. Koppal, and E. Jain. EyeVEIL: degrading iris authentication in eye tracking headsets. In *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications*, pp. 1–5, 2019. 2

[27] A. Joinson, U.-D. Reips, T. Buchanan, and C. B. P. Schofield. Privacy, Trust, and Self-Disclosure Online. *Human-Computer Interaction*, 25(1):1–24, Jan. 2010. doi: 10.1080/07370020903586662 5

[28] J. Lake. Hey, You Stole My Avatar!: Virtual Reality and Its Risks to Identity Protection. *EMORY LAW JOURNAL*, 69:48, 2020. 2, 4

[29] M. E. Latoschik and C. Wienrich. Coherence and Plausibility, not Presence?! Pivotal Conditions for XR Experiences and Effects, a Novel Model. *arXiv:2104.04846 [cs]*, June 2021. arXiv: 2104.04846. 8

[30] J. Liebers, M. Abdelaziz, L. Mecke, A. Saad, J. Auda, U. Gruenefeld, F. Alt, and S. Schneegass. Understanding User Identification in Virtual Reality Through Behavioral Biometrics and the Effect of Body Normalization. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pp. 1–11. ACM, Yokohama Japan, May 2021. doi: 10.1145/3411764.3445528 2

[31] J. Lin, J. Cronjé, I. Käthner, P. Pauli, and M. E. Latoschik. Measuring Interpersonal Trust towards Virtual Humans with a Virtual Maze Paradigm. *IEEE Transactions on Visualization and Computer Graphics*, pp. 1–11, 2023. Conference Name: IEEE Transactions on Visualization and Computer Graphics. doi: 10.1109/TVCG.2023.3247095 2, 9

[32] J. Lin and M. E. Latoschik. Digital body, identity and privacy in social virtual reality: A systematic review. *Frontiers in Virtual Reality*, 3, 2022. 2

[33] P. Mair and R. Wilcox. Robust statistical methods in R using the WRS2 package. *Behavior research methods*, 52(2):464–488, 2020. Publisher: Springer. 6

[34] 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, Mar. 2022. doi: 10.1109/VRW55335.2022.00245 5, 8

[35] R. C. Mayer, J. H. Davis, and F. D. Schoorman. An integrative model of organizational trust. *Academy of management review*, 20(3):709–734, 1995. Publisher: Academy of Management Briarcliff Manor, NY 10510. 5

[36] M. P. McCreery, D. B. Vallett, and C. Clark. Social interaction in a virtual environment: Examining socio-spatial interactivity and social presence using behavioral analytics. *Computers in Human Behavior*, 51:203–206, Oct. 2015. doi: 10.1016/j.chb.2015.04.044 4

[37] J. McVeigh-Schultz, A. Kolesnichenko, and K. Isbister. Shaping Pro-Social Interaction in VR: An Emerging Design Framework. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pp. 1–12. ACM, Glasgow Scotland Uk, May 2019. doi: 10.1145/3290605.3300794 1

[38] A. Meier. How Ghosts Became Transparent, and Other Spectral Evolutions, Dec. 2017. http://hyperallergic.com/410649/the-ghost-a-cultural-history/ (accessed Mar. 16, 2023). 3

[39] L. C. Miller, J. H. Berg, and R. L. Archer. Openers: Individuals who elicit intimate self-disclosure. *Journal of Personality and Social Psychology*,

44:1234–1244, 1983. Place: US Publisher: American Psychological Association. doi: 10.1037/0022-3514.44.6.1234 2, 5

[40] M. R. Miller, F. Herrera, H. Jun, J. A. Landay, and J. N. Bailenson. Personal identifiability of user tracking data during observation of 360-degree VR video. *Scientific Reports*, 10(1):17404, Dec. 2020. doi: 10.1038/s41598-020-74486-y 2

[41] R. D. Morey, J. N. Rouder, T. Jamil, S. Urbanek, K. Forner, and A. Ly. BayesFactor: Computation of Bayes Factors for Common Designs, July 2022. https://CRAN.R-project.org/package=BayesFactor (accessed Mar. 08, 2023). 7

[42] J. Morton. *A guide to color symbolism*, vol. 28. Colorcom, 1997. 3

[43] S. Mukherjee. What Drives Gender Differences in Trust and Trustworthiness? *Public Finance Review*, 48(6):778–805, Nov. 2020. Publisher: SAGE Publications Inc. doi: 10.1177/1091142120960801 5

[44] L. Nemec Zlatolas, T. Welzer, M. Hölbl, M. Heričko, and A. Kamišalić. A Model of Perception of Privacy, Trust, and Self-Disclosure on Online Social Networks. *Entropy*, 21(8):772, Aug. 2019. Number: 8 Publisher: Multidisciplinary Digital Publishing Institute. doi: 10.3390/e21080772 5

[45] F. O'Brolcháin, T. Jacquemard, D. Monaghan, N. O'Connor, P. Novitzky, and B. Gordijn. The Convergence of Virtual Reality and Social Networks: Threats to Privacy and Autonomy. *Science and Engineering Ethics*, 22(1):1–29, Feb. 2016. doi: 10.1007/s11948-014-9621-1 1

[46] C. K. L. Or and H. H. L. Wang. Color–concept associations: A cross-occupational and -cultural study and comparison. *Color Research & Application*, 39(6):630–635, 2014. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/col.21832. doi: 10.1002/col.21832 3

[47] M. Pakanen, P. Alavesa, N. van Berkel, T. Koskela, and T. Ojala. "Nice to see you virtually": Thoughtful design and evaluation of virtual avatar of the other user in AR and VR based telexistence systems. *Entertainment Computing*, 40:100457, Jan. 2022. Place: Oxford Publisher: Elsevier Sci Ltd WOS:000701918100001. doi: 10.1016/j.entcom.2021.100457 2

[48] Y. Pan and A. Steed. The impact of self-avatars on trust and collaboration in shared virtual environments. *PloS one*, 12(12):e0189078, 2017. ISBN: 1932-6203 Publisher: Public Library of Science San Francisco, CA USA. 2

[49] A. Picchi. A problem for Facebook users: Identity scams, Apr. 2018. https://www.cbsnews.com/news/a-growing-problem-for-facebook-users-identity-scams/ (accessed Mar. 08, 2023). 2

[50] B. Racoma. Color Symbolism – Psychology Across Cultures, July 2019. https://www.daytranslations.com/blog/color-psychology/ (accessed Mar. 16, 2023). 3

[51] G. Ripka, S. Grafe, and M. E. Latoschik. Preservice teachers' encounter with social VR–exploring virtual teaching and learning processes in initial teacher education. In *SITE Interactive Conference*, pp. 549–562. Association for the Advancement of Computing in Education (AACE), 2020. 1

[52] J. B. Rotter. A new scale for the measurement of interpersonal trust. *Journal of personality*, 1967. ISBN: 1467-6494 Publisher: Blackwell Publishing. 2

[53] C. Schell, A. Hotho, and M. E. Latoschik. Comparison of Data Representations and Machine Learning Architectures for User Identification on Arbitrary Motion Sequences, Nov. 2022. arXiv:2210.00527 [cs]. doi: 10.48550/arXiv.2210.00527 2

[54] I. Sluganovic, M. Roeschlin, K. B. Rasmussen, and I. Martinovic. Using Reflexive Eye Movements for Fast Challenge-Response Authentication. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, pp. 1056–1067. ACM, Vienna Austria, Oct. 2016. doi: 10.1145/2976749.2978311 2

[55] L. Su, A. P. Cui, and M. F. Walsh. Trustworthy Blue or Untrustworthy Red: The Influence of Colors on Trust. *Journal of Marketing Theory and Practice*, 27(3):269–281, July 2019. Publisher: Routledge _eprint: https://doi.org/10.1080/10696679.2019.1616560. doi: 10.1080/10696679.2019.1616560 2, 3

[56] A. M. Surprenant. Measuring trust in virtual worlds: Avatar-mediated self-disclosure. 2012. 2, 5

[57] S. Taddei and B. Contena. Privacy, trust and control: Which relationships with online self-disclosure? *Computers in Human Behavior*, 29(3):821–826, May 2013. doi: 10.1016/j.chb.2012.11.022 5

[58] C.-H. Tu and M. McIsaac. The Relationship of Social Presence and Interaction in Online Classes. *American Journal of Distance Education*, 16(3):131–150, Sept. 2002. Publisher: Routledge _eprint: https://doi.org/10.1207/S15389286AJDE1603_2. doi: 10.1207/S15389286AJDE1603_2 4

[59] M. Volonte, S. V. Babu, H. Chaturvedi, N. Newsome, E. Ebrahimi, T. Roy, S. B. Daily, and T. Fasolino. Effects of Virtual Human Appearance Fidelity on Emotion Contagion in Affective Inter-Personal Simulations. *IEEE Transactions on Visualization and Computer Graphics*, 22(4):1326–1335, Apr. 2016. doi: 10.1109/TVCG.2016.2518158 3

[60] M. Wang. Social VR : A New Form of Social Communication in the Future or a Beautiful Illusion? *Journal of Physics: Conference Series*, 1518:012032, Apr. 2020. doi: 10.1088/1742-6596/1518/1/012032 1

[61] S. Wenninger, J. Achenbach, A. Bartl, M. E. Latoschik, and M. Botsch. Realistic Virtual Humans from Smartphone Videos. In *26th ACM Symposium on Virtual Reality Software and Technology*, pp. 1–11. ACM, Virtual Event Canada, Nov. 2020. doi: 10.1145/3385956.3418940 2

[62] Y. Wu, A. S. M. Hall, S. Siehl, J. Grafman, and F. Krueger. Neural Signatures of Gender Differences in Interpersonal Trust. *Frontiers in Human Neuroscience*, 14, 2020. 5

[63] Zach. How to Perform an ANOVA with Unequal Sample Sizes, Sept. 2021. https://www.statology.org/anova-unequal-sample-size/ (accessed June. 14, 2023). 6

[64] N. Zhao, M. Zhou, Y. Shi, and J. Zhang. Face attractiveness in building trust: Evidence from measurement of implicit and explicit responses. *Social Behavior and Personality: an international journal*, 43(5):855–866, 2015. Publisher: Scientific Journal Publishers. 3