Visual Fidelity Effects on Expressive Self-avatar in Virtual Reality: First Impressions Matter

Fang Ma*
Goldsmiths, University of London

Xueni Pan†
Goldsmiths, University of London

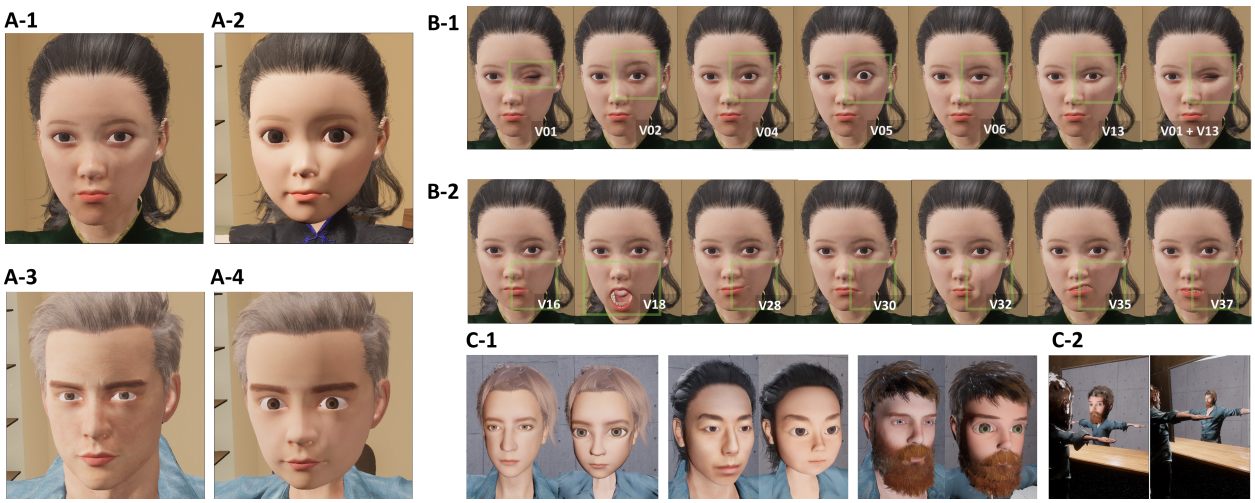

Figure 1: Study 1 used bespoke self-portrait avatars (C-1), while study 2 used standard avatars (A1-4) for a female and a male identity. The behaviour of two expressive avatars with realistic (A-1, A-3) and cartoonish (A-2, A-4) appearances were observed when they face a mirror and doing a presentation (C-2). The expressions driven by eye-tracking (B-1) and lip-tracking (B-2) were recreated based on the facial action units provided by VIVE Eye and Facial Tracking SDK (Sranipal) face blendShape document. Selected examples (V01-V37) are shown from 47 blendShapes in Study 2, the squeeze (V01+V13) is blended by two units.

ABSTRACT

Owning a virtual body inside Virtual Reality (VR) offers a unique experience where, typically, users are able to control their self-avatar's body via tracked VR controllers. However, controlling a self-avatar's facial movements is harder due to the HMD being in the way for tracking. In this work we present (1) the technical pipeline of creating and rigging self-alike avatars, whose facial expressions can be then controlled by users wearing the VIVE Pro Eye and VIVE Facial Tracker, and (2) based on this setting, two within-group studies on the psychological impact of the appearance realism of self-avatars, both the level of photorealism and self-likeness. Participants were told to practise their presentation, in front of a mirror, in the body of a realistic looking avatar and a cartoon like one, both animated with body and facial mocap data. In study 1 we made two bespoke self-alike avatars for each participant and we found that although participants found the cartoon-like character more attractive, they reported higher Body Ownership with whichever the avatar they had in the first trial. In study 2 we used generic avatars with higher fidelity facial animation, and found a similar "first trial effect" where they reported the avatar from their first trial being less creepy. Our results also suggested participants found the facial expressions easier to control with the cartoon-like character. Further, our eye-tracking data suggested that although participants were mainly facing their avatar during their presentation, their eye-gaze were focused elsewhere half of the time.

*e-mail: f.ma@gold.ac.uk
†e-mail: x.pan@gold.ac.uk

1 INTRODUCTION

The past decade has witnessed Virtual Reality (VR) becoming one of the most engaging media with increasing accessibility, enabled by a variety of Head-mounted Displays (HMDs) targeting at different markets: from the high-end Vajro XR-3, to the mid-range HTC VIVE Pro 2, to the consumer level standalone Oculus Quest 2. What makes this convenient access to virtual experiences even more compelling is the ability to express ourselves through our self-avatar, something that could be tailored to reflect our real-life personality and style, or in some cases, one which we desire. However, research on the relationship between self-avatar's level of realism and its psychological impact on the users has been inconclusive [4, 17, 28]. One of the challenges is that the level of avatar realism can be decomposed into two seemingly orthogonal yet intertwined factors: visual fidelity (appearance realism) and motion fidelity (behaviour realism) [5, 26].

In terms of visual fidelity, our ability to represent photo-realistic virtual characters in VR is currently limited only by the HMD's display resolution and the real-time rendering power (see Unreal's MetaHuman). More interestingly, most of the existing character creation platforms allow users to easily customise their avatars either via a parameterised system, or by uploading their own photos to generate look-alike avatars automatically. There is no doubt that in the near future, it would be almost effortless for us to create our "digital twins" in the 3D world which would look identical to ourselves in real-life.

When it comes to embodiment experiences in VR, having a look-alike avatar indeed would contribute towards the top-down process of *acceptance* of the virtual-self [36]. Other more "bottom-up" components in VR embodiment, such as *agency and control*, is generally recognised to be better sustained by motion-fidelity, or in this context often referred to as visual-motor synchrony [32]. Typically, in terms of body movement (posture, gesture), visual-motor synchrony can be achieved with real-time motion tracking system, or often an approximation using real-time inverse kinematics with data feed from the three 6DoF trackers (head and two controllers) available from most HMDs. Real-time facial tracking, on the other hand, has always been considered trickier, in particular when half of the user's face is hidden behind the HMD.

We have observed, however, the trend for new generations of HMDs to come with the option for real-time facial tracking. This typically involves eye-tracking enabled by build-in eye trackers inside the HMD, and lip-tracking by external infrared cameras attached to the lower part of the HMD. With this setup, participants could control not only their body movements but also their facial expressions in real-time, enabling both facial and body visual-motor synchrony in self-embodiment.

In this work, we use the HTC VIVE Pro Eye and the VIVE Facial Tracker (tracking lip movement) which came to market in 2021 for real-time facial tracking for embodiment. We present the pipeline of making self-alike avatars for participants to use it in VR, where their self-avatar not only follow their body movements but also their facial expressions in real-time. Moreover, we also explore the concept of visual-fidelity in the context of self-avatars, as here it could be interpreted with two different meanings: to which extent does the avatar look *photo-realistic*, and to which extent does the avatar look *like the user*. In particular, we examine this concept in relation to the Uncanny Valley theory [23], which proposes that, although human-likeness in general increases familiarity, highly photo-realistic humanoid avatars or robots could trigger eerie responses. Last but not least, we are interested in the relationship between visual and motion fidelity, and how different combinations of these two could impact users psychological feeling of Uncanny Valley, Embodiment, as well as their behaviour in VR in terms of hand and gaze behaviour.

Two within-subjects studies were conducted with a total of 18 participants. In study 1 (N=7), we made two bespoke self-alike avatars for each participant: one photo-realistic, the other cartoon-like. Both avatars were animated with real-time body and face tracking data using the HTC VIVE Pro 2 and the VIVE lip tracker. Our first hypothesis is that *participants would find the cartoon-like character less eerie* (**H1**). Our second hypothesis is that *there is no differences in the embodiment measurement other than the score being relatively high* as we are using self-alike avatars with both face and body visual-motor synchrony (**H2**). In study 2 (N=11), we investigate the same hypotheses with improved motion-fidelity by using additional tailor-made blendShapes to our virtual characters and at the same time changed to gender-matched generic avatars, again either cartoon-like or photo-realistic. As a result, study 2 has a reduced visual-fidelity (not self-alike) and an increased motion-fidelity. Additionally, we recorded participants behavioural data in order to better interpret our results (**H3**, more details see Section 5).

In summary, we created a complete pipeline for making self-avatars which resemble the user in terms of both their appearance and movements, incorporating both real-time body and facial animation. Through two user studies, this pipeline allowed us to investigate the psychological effects of different combinations of appearance and behavioral realism, in terms of the uncanny valley effect and the level of embodiment. We are not aware of any other work able to systematically evaluate the effect of self-alike avatars in VR, animated with real-time body animation and facial expression captured directly from the user inside an HMD.

In Section 2 we present related work followed by Section 3 our pipeline for avatar creation and real-time motor-synchrony implementation. We present the two studies in Section 4 and Section 5, and finally conclude our work with Section 6.

2 Related Work

2.1 Embodiment

Embodiment, or body ownership illusion, was first demonstrated by Botvinick and Cohen [2] in their 1998 study where participants were able to experience owning a body part that is not their own, triggered by visual-tactile synchrony. The same illusion is then found to be easily achievable in VR, where participants in HMDs would look down and discover a virtual body where they expect to see their real one [32]. Over the past few decades, there is plenty of evidence that "being someone else" in VR goes far beyond an experience of mere entertainment, but rather opportunities for us to better relate to the "virtual body" we are given in VR (e.g., reducing implicit racial bias [27] or ageism [24]), or change our behaviour because of characteristics of our avatars (e.g., "Proteus Effect", talking more confidently when assigned a taller self-avatar [37]).

VR Embodiment is often supported by the visual-proprioception synchrony (the virtual body being spatially where one expect their real body to be) as well as the visual-motor synchrony (your virtual body moves in syn with your real body) [34]. Typically full-body motion capture equipment is used where participants' body movements are fully-tracked and applied to their VR self-avatar in real-time. To maximise the effect, normally a virtual mirror is used where participant can see their virtual self via the mirror. The illusion could also be recreated by a simpler technical setup with only the three 6 DoF tracking points (Head, and two controllers) provided by most HMDs. Inverse kinematics are used to calculate the position of the rest of the body. This setup works well for tasks where participants stand relatively still and move mainly just their arms.

Although facial tracking is technically plausible [25], few research work has examined its psychological impact in embodiment. Kruzic et al. [16] highlighted that facial expression as a nonverbal cue contributes more than body language in virtual conversations. When it comes to embodiment, using pre-baked facial animation, Gonzales-Franco [12] found that participants had an increased level of self-identification than those without facial animation.

2.2 Uncanny Valley

Masahiro Mori first proposed the Uncanny Valley theory in 1970 [23]. Initially used for human robot interaction, it became a popular topics in virtual character [20, 31]. Mori believes that when the robot's appearance and movements are close to those of a human, the observer will generate positive, emotional feedback. However, when the similarity reaches a critical point, the observer's positive feelings will rapidly decline and turn into strong negative emotions. When the similarity continues to rise and approaches an average, healthy human, the affinity feelings will rise again and turn back into positive emotional responses. Researchers acknowledge the uncanny valley is related to both avatar's appearance and behaviour. Several studies attempted to understand the Uncanny Valley effect in VR [30]. However, the perception and sensitivity of this effect in the immersive situation still remains unclear.

2.3 Visual and Motion Fidelity

The appearance of virtual avatars could portray various images for audiences, for example cartoon-like emoticons and hyper-realistic humanoids [29]. The computer-generated characters have become highly realistic in recent projects (e.g., UE4 2021[1]).

However, it is not always the best idea to use avatars with the highest visual-fidelity. Some early pioneer studies demonstrate that expectations are often excessive when a virtual human is made to look like a human, such that one would expect the virtual human to behave exactly like a real one [9, 33]. Indeed, Garau's [9] research demonstrates that a mismatch between appearance and behaviour realism would decrease the co-presence level. Matching a highly

[1] https://www.unrealengine.com/en-US/digital-humans

simulated appearance to their vivid behaviour is still challenging for virtual humans in VR when designing virtual humans. Zibrek's studies on the appearance of the expressive virtual human indicated a significant positive correlation between render style on social presence, places illusion and empathetic concern, while there is no significant effect on embodiment and proximity [38–40]. However, only simple eye contact interactions were used in Zibrek's experiment and did not deal with more complete expression feedback. In a 2015 study, Kokkinara & McDonnel investigated the effect of real-time facial expression on self-avatar and found higher-level of animation being perceived as more appealing [15]. However, their work was conducted on a flat screen display.

Currently, few studies have addressed how fidelity matching will affect the user experience in VR. Drawing on insight from the experimental methods from a different perspective like appearance, behaviour animation and fidelity effects in VR [1, 12, 21, 39], in this study we will compare the psychological effect of self-avatars with different level of realism both animated with real-time mocap data for the body and the face.

3 Implementation

Figure 2 shows our technical pipeline for creating expressive self-avatar for both studies, separated by appearance and behaviour, and divided into three major steps: avatar creation and rigging, set-up for Unity real-time embodiment avatar animation (both body and face), and finally running the experiment with data collection for behavioural analysis.

3.1 Avatar Creation and Rigging

The modelling software Character Creator 3 [2] (CC3) and Maya were used. The self-avatar model part included creating the 3D character model with skeleton rigging and blendshapes. CC3 was used to create two bespoke avatars for each of the 7 participants in study 1. We asked each participant to send over a self-portrait photo at least 24 hours ahead of the experiment which is then used to generate a look-alike realistic 3D head for self-avatars using CC3 AI generator plugin Headshot [3] with some manually editing including adjusting the body shape, hair (including facial hair) and textures for face and body (wrinkles, blemishes, clothing) in CC3, as well as creating a cartoon-like version for each participant based on their realistic one by editing morph parameters in CC3 and wipe off skin texture's details in Photoshop. In study 2, we created 4 default models (see Figure 1(A1-4)) used the same approach.

We then exported the FBX file with auto-generated blendShapes and skeletal rigging into Maya. In study 1, we selected a set of blendShapes provided by CC3 in order to map 23 blendShapes provided by VIVE. In order to achieve a higher motion fidelity, in study 2, we reconstructed a set of 47 Vive blendShapes based on the Vive eye and lip documentation (52 blendShapes in total) using Maya, which is a combination (with different weights) of CC3's existing ExpressionPlus[4] (Based on ARKit 52 Standard Blendshape) blendShape and custom made ones using bone rigs.

3.2 Building Unity VR Project

The experiment project was created using Unity3D. We built an experiment process controller and wrote a UI editor script to facilitate the experiment process. In terms of visual fidelity, High Definition Render Pipeline (HDRP) and post-processing setup were used in Unity to achieve realistic rendering effects. To make the appearance of the realistic avatar look more high-fidelity, we installed the CC3 Unity plugin to automatically link PBR shader, subsurface scattering and humanoid rigging.

Full-body tracking was carried out using the FinalIK plug-in. The mirror implementation (symmetrical avatar script) is based on inverse kinematic bone rigs. To implement real-time face tracking, SRanipal SDK[5] for VIVE Pro Eye and Facial Tracker were used. We modified the eye and lip scripts in SRanipal SDK to solve the automatic assignment issue because of the avatar's blendshapes names could not be matched to SRanipal preset name after Maya editing.

3.3 Realtime Mocap and Behavioural Data Collection

The real-time facial motion capture hardware relied on the HTC Vive Pro Eye and Facial Tracker. We installed Vive SRanipal drive (SR_Runtime in Step B3) as the system link between hardware (real-time motion capture data sender) and Unity (motion capture data receiver). The body is driven by the headset and controllers which, via Steam VR to drive the avatar's heads, hands and body in Unity. We also collected behavioural data for analysis. In study 2, we used a VR experiment tool called UXF (Unity Experiment Framework) [3] for behaviour data collection.

4 Study 1: Look-alike Self-avatar

We conducted a within-subject study where each participant had two different self-avatars: realistic, or high visual fidelity (see Figure 1 (A1, A3)) and cartoon-like, or low visual fidelity (see Figure 1 (A2, A4)). Each participant has both sessions in a randomised order: either the realistic looking avatar first followed by the cartoon-like one (condition 1), or the other way (condition 2). We measure the level of uncanny valley and embodiment after each session.

4.1 Procedure

All participants (N = 7, 5 female) had some experience with VR. The procedure was approved by the Ethics Committee of Goldsmiths Computing. Participants read the Participant Information Sheet and signed the Consent Form before they started the experiment.

There were two almost identical sessions during the experiment, during each session the participant was told to practise their presentation in front of a virtual mirror in VR, where they can see their look-alike self-avatar. The only difference between the sessions is the visual fidelity of the avatar: one photorealistic, the other one cartoon-like. Both avatars are self-portraits generated from their photo in advance (see Section 3 for more detail).

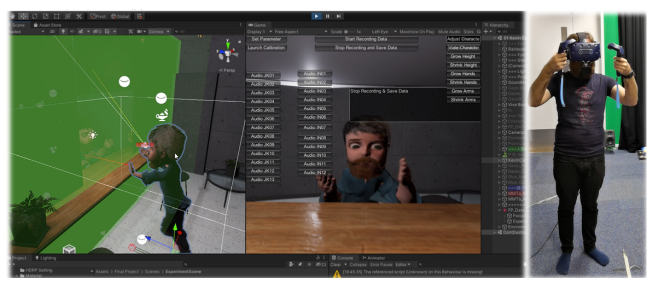

Figure 3: The interface when a participant did presentation practice during the experiment.

After putting on the VR headset, the participant could see their self-avatar in front of a mirror. They had two minutes to adapt to the virtual world, during which they were guided by a pre-recorded video to complete an eye calibration process and rescale their avatar height. The IPD (Inter-Pupillary Distance) was adjusted in this process which also helps in reducing simulation sickness [35]. Then, the main presentation practice session started. Similar to the study

[2]https://www.reallusion.com/

[3]https://www.reallusion.com/character-creator/headshot/

[4]https://manual.reallusion.com/character_creator_3/enu/3/content/character_creator_3/3.4/08_animation/using-arkit.htm

[5]/https://developer-express.vive.com/resources/vive-sense/eye-and-facial-tracking-sdk/

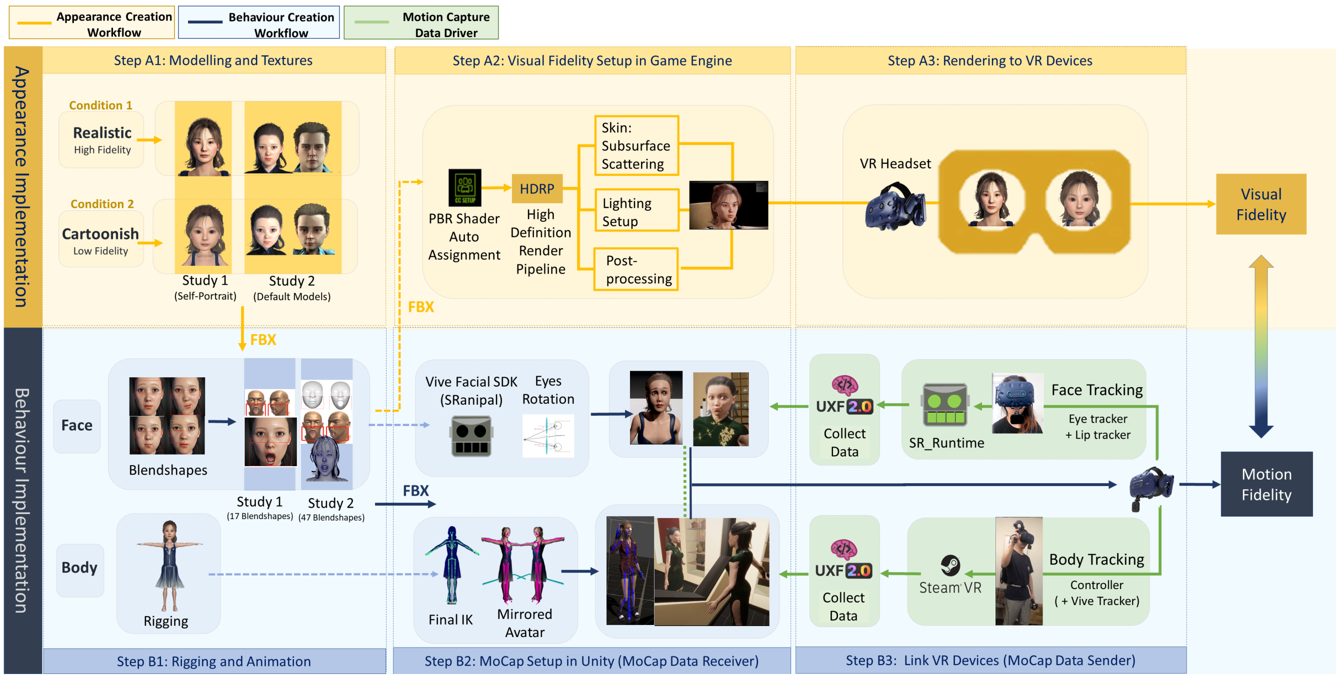

Figure 2: Implementation Process

by Gonzalez-Franco et al. [12], 13 audio clips from J.K. Rowling's Harvard Commencement speech was played. After each audio clip, the participants repeated as if they were practicing their own presentation. After each session, participants were told to take off the HMD and fill in an experience questionnaire.

4.2 Measurement and Data Analysis

We were interested in if there is a different uncanny valley impact and embodiment illusion degree at higher or lower of avatar appearance realism. Both were measured by questionnaires. After each session, participants were asked to answer 18 questions from Ho and MacDorman's Uncanny Valley Questionnaire [14]. Four dimensions are used in the uncanny valley measurement for data analysis including: one perceptual-cognitive dimension *Perceived Humanness* (PH: 6 questions) and three affective dimensions *Eerie* (ER: 3 questions), *Spine-tingling* (ST: 5 questions) and *Attractiveness* (AT: 4 questions). Each dimension has a 7-Likert scale from -3 to 3, for example ugly (-3) to beautiful (3) in the attractive dimension.

Embodiment was measured by a selection of 7 questions from Gonzalez-Franco [12], as shown in Table 1. The questions are further classified and calculated by *Body Ownership* (BO), *Agency and motor Control* (AC), *Location of the Body* (LB) and *External Appearance* (Ext) for analysis [11]. In order for the scale of the experiment to be consistent, the questionnaire used the same 7-Likert scale for each question as all questionnaire scale, in which participants stated their agreement levels with embodiment from -3 (completely disagree) to 3 (completely agree). See Table 1 for the questions used and scoring.

All data analysis was performed with IBM SPSS Statistics v27. We first conducted a Repeated Measure One-Way ANOVA test regardless of data distribution normality due to ANOVAs being considered as fairly robust to deviations from normality [19], even for extremely small sample sizes [6]. Whenever a significant difference was identified by the ANOVA test, we conducted a test of normality. In cases where normality was rejected, we conducted the appropriate non-parametric test to validate our result. A similar method is also used in [4]. As for study 1 our sample size is considered extremely

Table 1: Embodiment Measurement and Scoring

Embodiment Subset	Questions
BO = Q1-Q2	1. I felt as if the virtual body I saw when I looked down was my body. 2. It seemed as if I might have more than one body.
AC = Q3	3. It felt like I could control the virtual body as if it was my own body.
LB = Q4 - Q5 + Q6	4. I felt as if my body was located where I saw the virtual body. 5. I felt out of my body. 6. I felt as if my (real) body were drifting towards the virtual body or as if the virtual body were drifting towards my (real) body.
Ext = Q7	7. I felt like I was wearing different clothes from when I came to the laboratory.

small ($N < 10$) where the normality test itself can be less robust, even when normality was not rejected, we took extra steps to validate our data using additional statistical parameters suggested in [6].

4.3 Result

4.3.1 Uncanny Valley

Overall, as expected, mean value on all four categories indicated that participants reacted towards the Cartoon-like avatar (C) more positively than the Realistic one (R) (see Figure 4). They found the cartoon-like avatar more Attractive (Mean±Std C: 0.95 ± 0.76; R: 0.33 ± 0.54), more Human-like (C: 0.10 ± 0.92, R: -0.45 ± 0.82), less Eerie (C: 0.29 ± 0.78, R: 0.38 ± 0.93), and generates less feeling of Spine-tingling (C: 0.23 ± 0.44, R: 0.51 ± 0.38). A Repeated Measure One-Way ANOVA confirmed that participants found the Cartoon avatar significantly more **attractive** than the Realistic one ($F(1,6) = 6.3, \mathbf{p = 0.045}, \eta^2 = 0.51$). No other significant differences were found (Humanness: $F(1,6) = 2.1, p = 0.19, \eta^2 = 0.26$; Eerie: $F(1,6) = 1.0, p = 0.77, \eta^2 = 0.02$; Spine-tingling: $F(1,6) =$

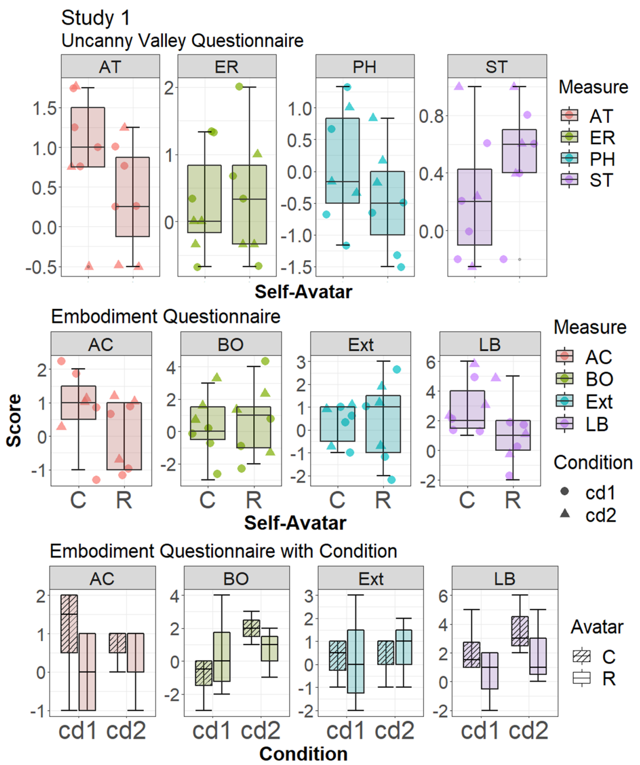

Figure 4: Questionnaire results from Study 1. Top: Uncanny Valley result of Attractiveness (AT), Eerie (ER), Perceived Humanness (PH), and Spine-tingling (ST). Middle: Embodiment Questionnaire of Agency and Motor Control (AC), Body Ownership (BO), External Appearance (Ext) and Location of the Body (LB). Bottom: Embodiment questionnaire by condition where cd1 (condition 1) is when participants had the realistic avatar first, and cd2 the cartoon-like avatar first

$1.8, p = 0.23, \eta^2 = 0.23$). Normality was not rejected for *attractiveness* by the Shapiro-Wilk test (C: $p = 0.72$; R: $p = 0.57$). Although Shaprio-Wilk test is recommended for small sample sizes, it might still be less robust for extremely small ones (i.e., $N < 10$) [10]. However, paired t-test (i.e., repeated ANOVA) is proven to be fairly robust for extremely small sample sizes ($N > 3$), given the within-pair correlation coefficient being high [6]. Our within-pair correlation is 0.54, thus we are fairly confident with our result.

4.3.2 Embodiment

When it comes to embodiment, the results are more mixed: on average, participants had lower body ownership (BO) and external appearance illusion (Ext) with the cartoon-like avatar as compared to the realistic one (Mean±Std: BO - C: 0.29 ± 2.0, R: 0.57 ± 0.21; Ext - C: 0.29 ± 0.95, R: 0.43 ± 1.81), but higher sense of agency & control (AC) and location of body (LB) with the cartoon-like one (AC - C: 0.86 ± 1.07, R: 0.14 ± 1.07; LB - C: 2.86 ± 1.95 R: 1.14 ± 2.19). However, a Repeated Measure One-Way ANOVA found no statistical significance for any of the measures (BO: $F(1,6) = 0.1, p = 0.72., \eta^2 = 0.024$; AC: $F(1,6) = 1.1, p = 0.33, \eta^2 = 0.155$; LB: $F(1,6) = 3.0, p = 0.14., \eta^2 = 0.0.33$; Ext: $F(1,6) = 0.032, p = 0.86, \eta^2 = 0.005$).

Interestingly, when we use "conditions" (i.e., the order in which participants went through the sessions) as a between-group variable, we found an interaction effect of **BO x Condition** ($F(1,5) = 7.1$, $\mathbf{p = 0.044}$, $\eta^2 = 0.59$). As shown in Figure 4, it seems that participants who had the Cartoon-like avatar in their first trial tended to experience a higher level of BO with the cartoon-like avatar than the realistic one, but when they had the realistic one first they reported a higher BO with the realistic one. In other words, there seems to be a "first trial effect" where participants tended to have reported a higher BO with the first avatar they interacted with, regardless of its level of realism. Normality for BO was not rejected by the Shapiro-Wilk test (C: cd1 $p = 0.16$ cd2 $p = 1.0$; R: cd1 $p = 0.69$ cd2 $p = 0.64$). Again, we took extra steps to further validate our results due to normality test might not be robust for extremely small samples. According to [6], independent sample t-tests are fairly robust when the effect size is large. We confirmed our result with an independent t-test comparing the differences between the two conditions, giving us the same results (two-tailed $p = 0.044$). We then calculated the effect size Cohen's $d = 2.17$. Given our effect size being very large here, we are fairly confident with our result.

4.4 Participants' Verbal Feedback and Discussion

First of all, our results confirmed our first hypothesis **H1** that participants would react towards the Cartoon-like character more positively, in that all four categories indicated the same trend, with attractiveness rated significantly higher for the Cartoon-like avatar than the realistic one. One of the participants specifically commented on their preference for the cartoon-like avatar: "I prefer the cartoonish avatar than the realistic one because I do not expect the cartoonish avatar to look exactly like me...when I look at the cartoonist avatar, I'm thinking it doesn't necessarily have to look like me so I do kind of in a weird way acknowledge myself. And I feel like it (cartoonist avatar) represents me more."

Despite the clear preference for the cartoon-like avatar, and some participants even thought the cartoon-like represent them better, there is no significant results when it comes to embodiment. This is inline with our second hypothesis **H2**. Interestingly, we were able to identify a "first trial effect" where participants tended to report a higher Body Ownership Illusion with whichever the first avatar they had. Due to the low number of participants in study 1, we decided to run more participants to investigate this effect further.

In terms of the task, most participants reported that they enjoyed the presentation scenario. Some even expressed that they found it useful: "I think if I had to talk to myself in real life to be practising with a mirror, it would be more awkward than in virtual reality. I feel that the avatar was making me feel like it is me, but I felt somehow more confident speaking there than in real life."

Further, some participants commented about their own behaviours in the study ("I find that cartoon one makes me want to move more because they are fun."). We therefore decided to include behavioural analysis in study 2.

Finally, when asking about the fidelity of facial mocap, one participants commented "I felt like I wasn't a character when I was moving my mouth and it (the virtual mouth) wasn't moving like mine (real mouth)." Therefore in study 2 we improved our implementation pipeline on lip tracking.

5 Study 2: Generic Self-avatar

A second study was conducted with higher motion-fidelity for the self-avatars. A total of 24 blendShapes were added to the original blendShapes in study 1, 12 of which are lip morph targets. The lip synchronisation thus became more realistic when participant spoke which made the self-avatar more expressive. However, due to time constraints, the need for an increased number of participants, and added blendShape for each avatar, we used two generic (cartoon-like or realistic, see Figure 1A) gender-matched avatars for each participant.

Our first two hypotheses remain the same (H1: Uncanny Valley; H2: Embodiment). We also recorded behaviour data and we hypothesise that *participants would move more with the Cartoon-like avatar* (**H3**), as this is something highlighted by participants in the interview in study 1 (see section 4.4). Finally, we also included participants' gaze data. In particular, we will compare two different measures for gaze: the actual Eye-tracking data, and Head ray, which is often used as an approximation for gaze [12].

5.1 Procedure

Participants (N = 11, 3 female) attended study 2. Similar to study 1, they did two presentation sessions, one with each avatar, in a randomised order. After each session, participants were asked to fill a similar questionnaire as study 1, with an additional enfacement questionnaire. During the two sessions, again they were played 13 audio clips, manually triggered by an experimenter who was in the same room. They were told to repeat the same sentence after each audio clip as if they were practising their presentation in front of the mirror. Their behavioural data were recorded during their presentation, by a script through UXF which is converted to CSV files to analyse.

Before the experiment, similarly to study 1, participants went through first the avatar rescale and eye-tracking calibration process. After finishing the calibration process, each participant was given a 30 second period to familiarise themselves with the environment and their self-avatar. We also recorded their behavioural data during these 30 seconds.

5.2 Measurement

Similar to study 1, we used the Uncanny valley and Embodiment questionnaires. In addition, we have also included four questions on Effacement and included behavioural data.

5.2.1 Enfacement

The four additional enfacement questions in Table 2 are from Estudillo and Bindemann's questionnaire [8]. We use this to measure the perception of the virtual self-avatar's face, as the animation of the facial expression has been enhanced in study 2.

Table 2: Additional Enfacement Questions.

Enfacement	Questions
Ownface	1. I felt like the virtual face was my face.
Belonged	2. I felt like the virtual face belonged to me.
Mirror	3. I felt like I was looking at my own face reflected in a mirror.
(out of) Control	4. I felt like my own face was out of my control.

5.2.2 Hand Movement

Both the left and right hand controllers' positions were recorded per frame (approximately 30 frames per second). We then use this data to calculate the total distance of hand movements. The distance in the virtual environment is measured in metres. For each participant, we calculated the average hand movements (metre per second) for both the initial 30 seconds of preparation time, and the accumulated periods of the 13 blocks of recorded presentation.

5.2.3 Gaze Data

We used the Vive eye-tracking to provide Eye-tracking data, we then used the ray-cast function and colliders in Unity to record the object the eye-tracking ray hits for each frame. In this study, it is mostly on different body/face parts of the mirrored characters. Moreover, using the same method, we also recorded the head-ray data (the forward ray from the HMD device), which is commonly used in previous experiments [12] as an indicator for gaze data when eye-tracking is not available. We are interested in whether the head-ray is indeed a good indicator for gaze.

5.3 Result

Similarly to study 1, we perform a normality check for all significant results from our Repeated ANOVA test. When normality was not

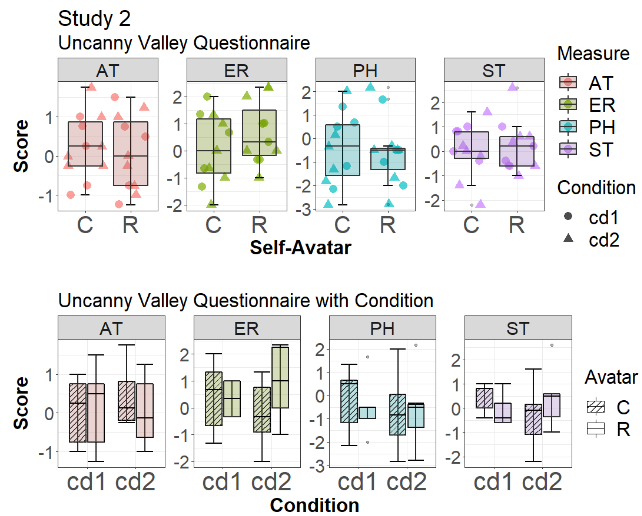

Figure 5: Study 2 Top: Uncanny Valley result of Attractiveness (AT), Eerie (ER), Perceived Humanness (PH), and Spine-tingling (ST). Bottom: Uncanny Valley by condition where cd1 (condition 1) is when participants had the realistic avatar first, and cd2 the cartoon-like avatar first

rejected, as our sample size is over 10 for this study, we did not perform any additional checks. In cases where normality was rejected, we conducted suitable non-parametric tests.

5.3.1 Uncanny Valley

The mean values of all four categories for the Uncanny Valley questionnaire are very similar for both Cartoon-like and Realistic avatars. A Repeated Measure One-Way ANOVA found no significant difference for any of the four categories (Attractiveness: $F(1,10) = 0.32, p = 0.58, \eta^2 = 0.03$; Erie: $F(1,10) = 1.75, p = 0.22, \eta^2 = 0.15$; Spine-tingling: $F(1,10) = 0.54, p = 0.48, \eta^2 = 0.05$). However, when we use "condition" (the order in which they experienced the two avatars) as a between-group variable, there is a strong interaction effect of Spine-tingling x Condition ($F(1,9) = 18.13$, **p =0.002**, $\eta^2 = 0.668$). Similar to the "first trial effect" previously identified in study 1, participants seem to find the avatar they had in their first trial less creepy, regardless of it being a cartoon-like or realistic one. The Shapiro-Wilk test did not reject the normality of our dataset (C: cd1 $p = 0.22$ cd2 $p = 86$; R: cd1 $p = 0.16$ cd2 $p = 0.40$).

5.3.2 Embodiment

In terms of embodiment, we found no statistical differences between the two avatars in any of the categories, nor have we found any interaction effect with condition (i.e., order effect). In Figure 6 we plotted study 2 embodiment against the result of study 1, and despite not having an avatar that actually looked like them in study 1, participants in study 2 reported similar levels of embodiment across all four categories.

5.3.3 Enfacement

A Repeated Measure One-Way ANOVA over the four Enfacement questions revealed a significant difference over (out of) **Control** ($F(1,10) = 6.26$, **p = 0.031**, $\eta^2 = 0.385$), suggesting that participants felt more in control with over the facial expression of the cartoon-like avatar (C) than the realistic one (R) (Out of Control $Mean \pm Std$ C: -1.00 ± 1.41, R: 0.09 ± 0.141). No other questions came out significant (Belonged: $F(1,10) = 0.000, p = 1.000, \eta^2 = 0.000$; Mirror: $F(1,10) = 0.000, p = 1.000, \eta^2 = 0.000$; Ownface:

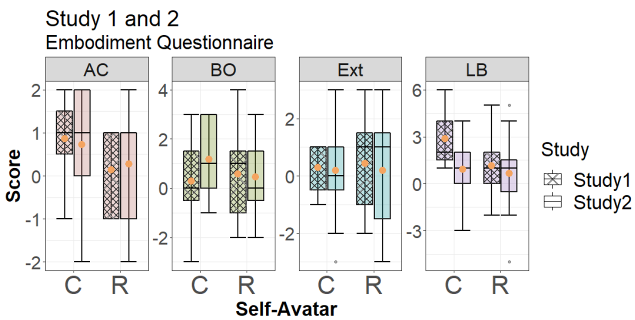

Figure 6: Scores of Embodiment compare study 1 and study 2.

$F(1,10) = 0.850, p = 0.602, \eta^2 = 0.028$). The assumption of normality was not rejected (C: $p = 0.07$; R: $p = 0.33$).

When taking "condition" as a between group variable we found an interaction effect of **Control x Condition** ($F(1,9) = 10.53$, **p = 0.010**., $\eta^2 = 0.530$). As shown in Figure 7, there seems to be an order effect where although in general participants felt the realistic avatar's facial expressions were more out of control, this is worse if they experience the realistic avatar first. As normality was rejected (C: cd1 $p = 0.15$ cd2 $p = 0.24$; R: cd1 $p = 0.31$ cd2 $p = 0.03$), we conducted a non-parametric test for independent-sample using Mann-Whitney U test, comparing the differences between the two trials for the two conditions and found a similar significant result ($U = 1.5$, **p = 0.009**), validating our result.

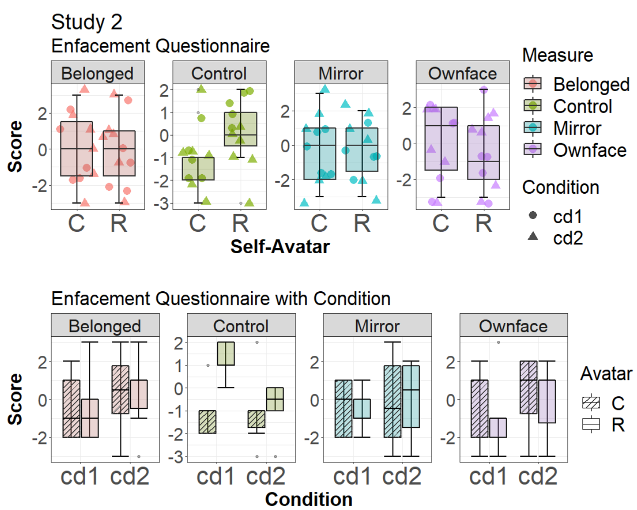

Figure 7: Scores of Enfacement Illusion in Study 2. Note Control means "out of control".

5.3.4 Hand Behaviour

One participant has been excluded from this analysis due to system failure, leaving a total number of 10 participants. A Repeated Measure One-Way ANOVA was conducted for both Hand30 (the initial 30 seconds of the preparation time) and HandP (the accumulated periods of the presentation). No significant differences were found (Hand30: $F(1,9) = 0.473 p = 0.509, \eta^2 = 0.050$; HandP:$F(1,9) = 2.047, p = 0.186, \eta^2 = 0.185$).

As we used generic avatars for this session, we hypothesise that the appearance of our avatars could have had an impact over participants behaviour. In particular, 4 out of the 10 participants in our session would be identified as from the same ethnic group as our avatars (i.e, Chinese). With **"ethnicity"** as a between-group condition (either Chinese or non-Chinese), we found an interactive effect of condition x Hand30 ($F(1,8) = 5.335$, **p = 0.050**, $\eta^2 = 0.400$). It seems that, during the initial 30 seconds of the adaptation period, Chinese participants gestured more with the realistic avatar and non-Chinese participants with the cartoon-like avatar. As the test of normality rejected the normality of our data (C: cn $p = 0.68$ non $p = 0.24$; R: cn $p = 0.21$ non $p = 0.004$), we conducted a non-parametric test for independent-sample using Mann-Whitney U test, comparing the differences between the two trials for the two conditions. However, our result was not significant ($U = 21.0$, **p = 0.067**). Therefore, we are not able to validate our result here.

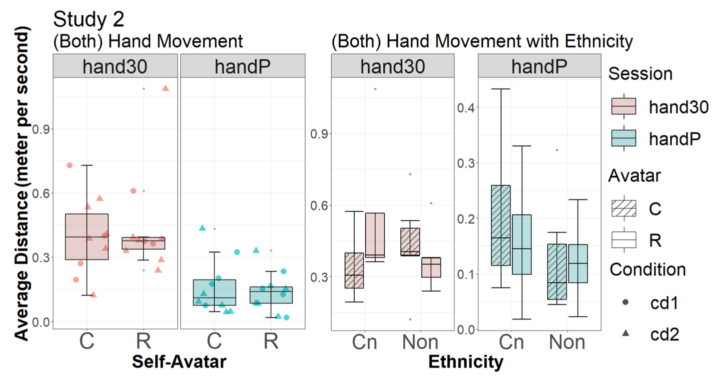

Figure 8: Hand Behaviour.

5.3.5 Gaze Behaviour

We recorded both Eye-tracking data and Head-ray for gaze analysis. Using the collider on the mirrored avatar's body, we calculate the gaze percentage for when the participants were looking at their avatar, and when they did, each body part. All 11 participants' data were recorded, however, due to a bug in the collider setting, 3 female participants's Head-ray data for the Cartoon-like avatar is missing, leaving 8 male participants. As shown in 9, surprisingly, the eye-tracking data and head-ray appear to be very different. It seems that the head-ray hits 100% the body of the avatar, meaning that during their presentation, all participants were facing towards the avatar. However, according to the eye-tracking data, only half of the time (51% for Cartoon-like avatar, and 56% for the realistic one) participants were indeed "looking" at their avatar.

5.4 Discussion

Contrary to our prediction, no significant differences were found with our Uncanny Valley measurements, and participants' reaction towards all four categories were very similar for both avatars. Our hypothesis **H1** is not supported. The reason why in study 1 **H1** was supported could be because participants found the cartoon-like character a better match with the motion fidelity, and the realistic with higher visual fidelity generated a mismatch. In this study, we have lowered the visual-fidelity for both avatars (no longer self-alike) and increased the motion fidelity (more blendShapes), making the realistic avatar more acceptable.

Interestingly, we found a similar "first trial effect" as identified in Study 1, where participants tended to react to the very first avatar (regardless of its level of realism) more favourably. In study 1, participants reported a higher Body Ownership level with their first trial avatar; in this study, they found their first avatar less creepy.

Similarly to our study 1 and as expected, there is no evidence to reject our **H2** that participants experienced similar level of embodiment with both avatars. However, our enfacement questionnaire indicated that participants found the facial expressions more out of control with the realistic avatar as compared to the cartoon-like one, and that it is worse if they had the realistic avatar first. It seems that if they had the cartoon-like avatar first, they would still find the realistic one more difficult to control, but to a lesser extent.

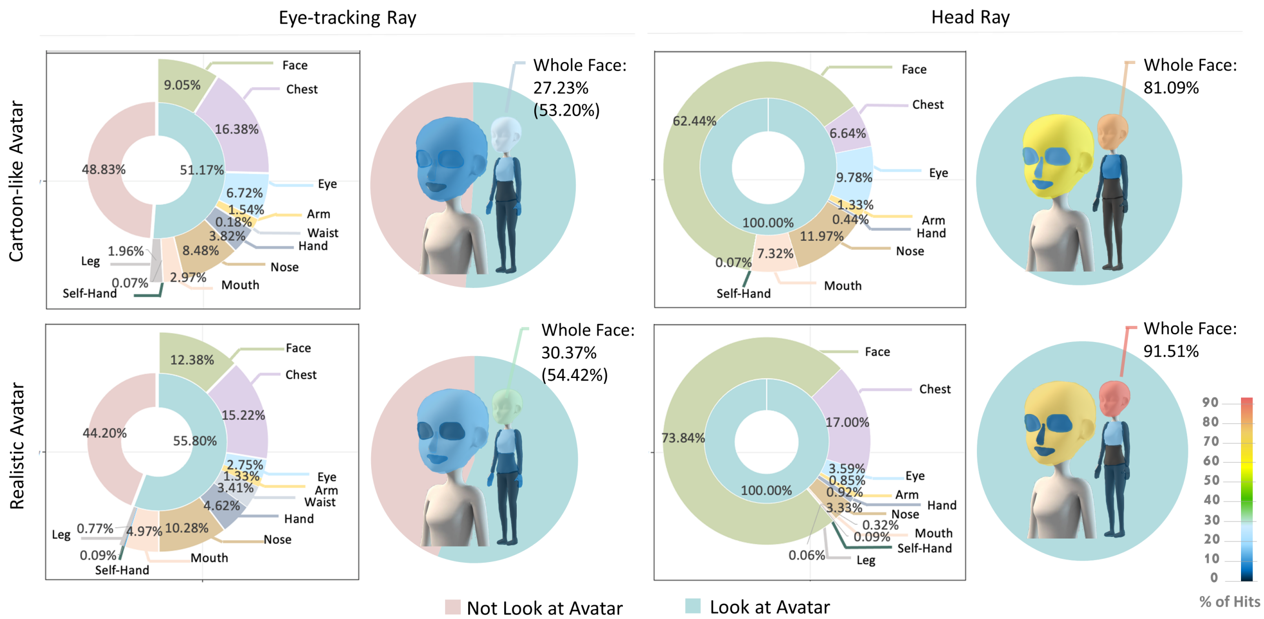

Figure 9: Gaze.

During our study 1 we observed that some participants used more gesturing with the cartoon-like avatar, and thus we included hand gesture as a measure in this study. Although no differences was found in our initial analysis, an ANOVA test indicated that non-Chinese participants moved their hands more in the initial adaptation period with the cartoon-like character, supporting our **H3**. It is not clear why Chinese participants did not follow the same pattern. It could be that the realistic version of the avatar looked more Chinese, so the Chinese participants were more interested in the realistic one rather than the cartoon-like, more western looking avatars. However, our data failed the normality test and a non-parametric test no longer finds the effect significant. More studies will have to be conducted to understand this effect better.

Last but not least, we compares two different methods of capturing gaze: eye-tracking data and head ray data. As eye-tracking is not yet a standard feature for HMDs, head-ray is commonly used as an approximation for user gaze. Our result, however, indicated that this method could be problematic: we found that, during the presentation stage, although participants were facing their avatars, almost 50% of they time their actual gaze was pointing elsewhere. This is inline with literature in social psychology and more recent multi-modal analysis studies [7], which states that the speaker would look at the listener's face about 60% of the time (in our case, the speaker was the participants, and the "listener" was their avatar in the mirror).

6 Conclusion

In this work, we presented a technical pipeline for generating body ownership illusion with self-alike avatars in VR. Users were able to drive their self-avatar not only with their body movements, but also their facial expressions. We then studied the psychological impact of self-avatar illusion with two experiments. In terms of Uncanny Valley, we found that participants preferred the cartoon-like character in study 1. However, as we decreased visual-fidelity in terms of self-alikeness and increased motion fidelity in study 2, participants found the realistic avatar equally acceptable as the cartoon-like one. We found no differences for both studies in terms of embodiment. We have, however, identified an order effect, or a "first trial effect" where participants seemed to have a preference for the very first avatar they had: they reported a higher body ownership illusion in study 1 and a lower level of creepiness in study 2 with their first avatar. Finally, we reported behaviour patterns of participants' gesture, measured by total hand movement, and gaze, measured by eye-tracking and hand ray. Our findings are inline with existing literature on social psychology, while revealing that using head-ray to approximate gaze in VR could be problematic.

We believe our interdisciplinary work contributed towards our understanding of embodiment, uncanny valley, and user behaviours in VR. Our work also provides useful insights for researchers in VR and psychology, and those interested in the potential of VR from the industry. For example, for studies in cognitive neuroscience where participant could see their self-avatar in the mirror, but perhaps in a different ethnicity which has a proven effect of reducing racial bias [18]; or an anorexia assessment/psychotherapy session where they can see a self-avatar but in a different body shape [22]. In future, we plan to use this pipeline to conduct experiments with a larger number of participants and a more balanced gender profile. We also plan to include additional measurements such as personality, which has been demonstrated to have an effect on level of embodiment [13].

Acknowledgments

Both Ma and Pan are funded by the AHRC Project "Immersive, Innovative, and Interactive Experience" AH/T011416/1.

References

[1] J. Bailenson, K. Swinth, C. Hoyt, S. Persky, A. Dimov, and J. Blascovich. The independent and interactive effects of embodied-agent appearance and behavior on self-report, cognitive, and behavioral markers of copresence in immersive virtual environments. *Presence*, 14:379–393, 08 2005. doi: 10.1162/105474605774785235

[2] M. Botvinick and J. Cohen. Rubber hands 'feel'touch that eyes see. *Nature*, 391(6669):756–756, 1998.

[3] J. Brookes, M. Warburton, M. Alghadier, M. Mon-Williams, and F. Mushtaq. Studying human behavior with virtual reality: The unity

experiment framework. *Behavior research methods*, 52(2):455–463, 2020.

[4] T. Collingwoode-Williams, Z. O'Shea, M. F. P. Gillies, and X. Pan. The impact of self-representation and consistency in collaborative virtual environments. *Frontiers in Virtual Reality*, 2:45, 2021.

[5] A. W. de Borst and B. de Gelder. Is it the real deal? perception of virtual characters versus humans: an affective cognitive neuroscience perspective. *Frontiers in psychology*, 6:576, 2015.

[6] J. C. De Winter. Using the student's t-test with extremely small sample sizes. *Practical Assessment, Research, and Evaluation*, 18(1):10, 2013.

[7] G. C. Dobre, M. Gillies, P. Falk, J. A. Ward, A. F. d. C. Hamilton, and X. Pan. Direct gaze triggers higher frequency of gaze change: An automatic analysis of dyads in unstructured conversation. In *Proceedings of the 2021 International Conference on Multimodal Interaction*, pp. 735–739, 2021.

[8] A. Estudillo and M. Bindemann. Can gaze-contingent mirror-feedback from unfamiliar faces alter self-recognition? *The Quarterly Journal of Experimental Psychology*, 70:1–39, 03 2016. doi: 10.1080/17470218.2016.1166253

[9] M. Garau. The impact of avatar fidelity on social interaction in virtual environments. 01 2003.

[10] A. Ghasemi and S. Zahediasl. Normality tests for statistical analysis: a guide for non-statisticians. *International journal of endocrinology and metabolism*, 10(2):486, 2012.

[11] M. Gonzalez-Franco and T. C. Peck. Avatar embodiment. towards a standardized questionnaire. *Frontiers in Robotics and AI*, 5:74, 2018. doi: 10.3389/frobt.2018.00074

[12] M. Gonzalez-Franco, A. Steed, S. Hoogendyk, and E. Ofek. Using facial animation to increase the enfacement illusion and avatar self-identification. *IEEE Transactions on Visualization and Computer Graphics*, 26(5):2023–2029, 2020. doi: 10.1109/TVCG.2020.2973075

[13] D. Higgins, R. Fribourg, and R. McDonnell. Remotely perceived: Investigating the influence of valence on self-perception and social experience for dyadic video-conferencing with personalized avatars. *Frontiers in Virtual Reality*, 2:60, 2021.

[14] C.-C. Ho and K. F. MacDorman. Revisiting the uncanny valley theory: Developing and validating an alternative to the godspeed indices. *Computers in Human Behavior*, 26(6):1508–1518, 2010. Online Interactivity: Role of Technology in Behavior Change. doi: 10.1016/j.chb.2010.05.015

[15] E. Kokkinara and R. McDonnell. Animation realism affects perceived character appeal of a self-virtual face. In *Proceedings of the 8th ACM SIGGRAPH Conference on Motion in Games*, pp. 221–226, 2015.

[16] C. O. Kruzic, D. Kruzic, F. Herrera, and J. N. Bailenson. Facial expressions contribute more than body movements to conversational outcomes in avatar-mediated virtual environments. *Scientific Reports*, 10, 2020.

[17] M. E. Latoschik, D. Roth, D. Gall, J. Achenbach, T. Waltemate, and M. Botsch. The effect of avatar realism in immersive social virtual realities. In *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology*, pp. 1–10, 2017.

[18] L. Maister, M. Slater, M. V. Sanchez-Vives, and M. Tsakiris. Changing bodies changes minds: owning another body affects social cognition. *Trends in cognitive sciences*, 19(1):6–12, 2015.

[19] S. E. Maxwell, H. D. Delaney, and K. Kelley. *Designing experiments and analyzing data: A model comparison perspective*. Routledge, 2017.

[20] R. McDonnell and M. Breidt. Face reality: Investigating the uncanny valley for virtual faces. In *ACM SIGGRAPH ASIA 2010 Sketches*, SA '10. Association for Computing Machinery, New York, NY, USA, 2010. doi: 10.1145/1899950.1899991

[21] R. McMahan, D. Bowman, D. Zielinski, and R. Brady. Evaluating display fidelity and interaction fidelity in a virtual reality game. *IEEE transactions on visualization and computer graphics*, 18:626–33, 04 2012. doi: 10.1109/TVCG.2012.43

[22] S. C. Mölbert, A. Thaler, B. J. Mohler, S. Streuber, J. Romero, M. J. Black, S. Zipfel, H.-O. Karnath, and K. E. Giel. Assessing body image in anorexia nervosa using biometric self-avatars in virtual reality: Attitudinal components rather than visual body size estimation are distorted. *Psychological medicine*, 48(4):642–653, 2018.

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

[24] S. Oh, J. Bailenson, E. Weisz, and J. Zaki. Virtually old: Embodied perspective taking and the reduction of ageism under threat. *Computers in Human Behavior*, 60:398–410, 07 2016. doi: 10.1016/j.chb.2016.02.007

[25] K. Olszewski, J. J. Lim, S. Saito, and H. Li. High-fidelity facial and speech animation for vr hmds. *ACM Trans. Graph.*, 35(6), Nov. 2016. doi: 10.1145/2980179.2980252

[26] X. Pan and A. F. d. C. Hamilton. Why and how to use virtual reality to study human social interaction: The challenges of exploring a new research landscape. *British Journal of Psychology*, 109(3):395–417, 2018.

[27] T. C. Peck, S. Seinfeld, S. M. Aglioti, and M. Slater. Putting yourself in the skin of a black avatar reduces implicit racial bias. *Consciousness and cognition*, 22(3):779–787, 2013.

[28] D. Roth, J.-L. Lugrin, D. Galakhov, A. Hofmann, G. Bente, M. E. Latoschik, and A. Fuhrmann. Avatar realism and social interaction quality in virtual reality. In *2016 IEEE Virtual Reality (VR)*, pp. 277–278. IEEE, 2016.

[29] R. Schwartz and W. Steptoe. The immersive vr self: performance, embodiment and presence in immersive virtual reality environments. In *A networked self and human augmentics, artificial intelligence, sentience*, pp. 108–116. Routledge, 2018.

[30] V. Schwind. Implications of the uncanny valley of avatars and virtual characters for human-computer interaction. 2018.

[31] V. Schwind, K. Wolf, and N. Henze. Avoiding the uncanny valley in virtual character design. *Interactions*, 25(5):45–49, Aug. 2018. doi: 10.1145/3236673

[32] M. Slater, D. Pérez Marcos, H. Ehrsson, and M. V. Sanchez-Vives. Inducing illusory ownership of a virtual body. *Frontiers in neuroscience*, 3:29, 2009.

[33] M. Slater and A. Steed. A virtual presence counter. *Presence*, 9:413–434, 10 2000. doi: 10.1162/105474600566925

[34] B. Spanlang, J.-M. Normand, D. Borland, K. Kilteni, E. Giannopoulos, A. Pomés, M. González-Franco, D. Perez-Marcos, J. Arroyo-Palacios, X. N. Muncunill, and M. Slater. How to build an embodiment lab: Achieving body representation illusions in virtual reality. *Frontiers in Robotics and AI*, 1:9, 2014. doi: 10.3389/frobt.2014.00009

[35] K. Stanney, C. Fidopiastis, and L. Foster. Virtual reality is sexist: But it does not have to be. *Frontiers in Robotics and AI*, 7:4, 2020. doi: 10.3389/frobt.2020.00004

[36] T. Waltemate, D. Gall, D. Roth, M. Botsch, and M. E. Latoschik. The impact of avatar personalization and immersion on virtual body ownership, presence, and emotional response. *IEEE transactions on visualization and computer graphics*, 24(4):1643–1652, 2018.

[37] N. Yee and J. Bailenson. The proteus effect: The effect of transformed self-representation on behavior. *Human Communication Research*, 33:271 – 290, 07 2007. doi: 10.1111/j.1468-2958.2007.00299.x

[38] K. Zibrek, E. Kokkinara, and R. Mcdonnell. The effect of realistic appearance of virtual characters in immersive environments - does the character's personality play a role? *IEEE Transactions on Visualization and Computer Graphics*, 24(4):1681–1690, 2018. doi: 10.1109/TVCG.2018.2794638

[39] K. Zibrek, S. Martin, and R. McDonnell. Is photorealism important for perception of expressive virtual humans in virtual reality? *ACM Trans. Appl. Percept.*, 16(3), Sept. 2019. doi: 10.1145/3349609

[40] K. Zibrek and R. McDonnell. Social presence and place illusion are affected by photorealism in embodied vr. In *Motion, Interaction and Games*, MIG '19. Association for Computing Machinery, New York, NY, USA, 2019. doi: 10.1145/3359566.3360064