

# Augmenting Immersive Telepresence Experience with a Virtual Body

Nikunj Arora, Markku Suomalainen\*, Matti Pouke, Evan G. Center, Katherine J. Mimnaugh, Alexis P. Chambers, Sakaria Pouke and Steven M. LaValle



**Fig. 1:** Head-mounted display telepresence augmented with a virtual body; the participant (side images) is sitting in one place wearing a head-mounted display, but is telepresent in another location through a live streaming camera having a discussion with an experimenter (center image) and can see robot arms in one condition that match their real arms' motions.

**Abstract**—We propose augmenting immersive telepresence by adding a virtual body, representing the user's own arm motions, as realized through a head-mounted display and a 360-degree camera. Previous research has shown the effectiveness of having a virtual body in simulated environments; however, research on whether seeing one's own virtual arms increases presence or preference for the user in an immersive telepresence setup is limited. We conducted a study where a host introduced a research lab while participants wore a head-mounted display which allowed them to be telepresent at the host's physical location via a 360-degree camera, either with or without a virtual body. We first conducted a pilot study of 20 participants, followed by a pre-registered 62 participant confirmatory study. Whereas the pilot study showed greater presence and preference when the virtual body was present, the confirmatory study failed to replicate these results, with only behavioral measures suggesting an increase in presence. After analyzing the qualitative data and modeling interactions, we suspect that the quality and style of the virtual arms, and the contrast between animation and video, led to individual differences in reactions to the virtual body which subsequently moderated feelings of presence.

**Index Terms**—Telepresence, 360-degree live streaming, Presence, Place Illusion.

## 1 INTRODUCTION

Marvin Minsky coined the word 'telepresence' in his influential 1980s essay, suggesting a new paradigm of remote work where one could carry out complex physical tasks in remote locations while simultaneously receiving rich sensory feedback [31]. In his view, the biggest challenge of telepresence was achieving the feeling of *presence*, the sensation of 'being there'. This remains a challenge in current commercial telepresence robots using regular 2-D screens due to the screens' insufficient immersion. Though robot mobility increases presence over stationary video calls [37], users still do not really feel as 'being there', causing issues such as remote participants speaking less in group work tasks, finding the tasks more difficult [52], or having discussions less interactive [57] than in-person discussions. Modern Head-Mounted Displays (HMDs) and 360-degree video cameras can supply the immersion capable of remedying this deficiency and provide users a more similar experience as being physically at the camera's location; adding a panoramic camera on a mobile robot could further increase this feel-

ing. There are a wide variety of use cases for this setup; for example, people with mobility issues or living too far can join family events, and business people can join events requiring mobility, such as facility tours or workshops, while feeling as if they were really present at the event.

Following telepresence research, the term 'presence' was adopted by the Virtual Reality (VR) research community to describe one of the fundamental illusions elicited by VR systems. While an exact definition of presence is still lacking, most VR scholars agree on presence roughly referring to the sensation of 'being there,' similarly to Minsky's original essay (for further discussion, see [40]). One of the most influential definitions regarding this concept is Mel Slater's framework of immersion and presence, in which the former is defined as the physical system's capability of providing immersive experiences, whereas the latter refers to the subjective feeling of being in a virtual location [50]. Later, to reduce confusion amongst terminology, Slater proposed the terms Place Illusion (PI), to specifically refer to the sensation of being in another location [42], and Plausibility Illusion (PSI) to refer to the illusion of the virtual scenario actually happening to the user and the general sensation of realism [42]. Skarbez proposed a framework where the entirety of 'presence' would consist of PI, PSI as well as co- and social presence illusions, the latter referring to VR-generated illusions of being in the company of other people or interacting with others, respectively [5, 40].<sup>1</sup>

<sup>1</sup> Similarly, in this article, we also use 'presence' to address multiple components, while referring to the subcomponents as PI, PSI and co-presence.

- All authors are with University of Oulu, Finland: [firstname.lastname@oulu.fi](mailto:firstname.lastname@oulu.fi)
- \* Corresponding author ( [markku.suomalainen@oulu.fi](mailto:markku.suomalainen@oulu.fi) )

Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: [reprints@ieee.org](mailto:reprints@ieee.org). Digital Object Identifier: [xx.xxxx/TVCG.201x.xxxxxx](https://doi.org/10.1109/TVCG.201x.xxxxxx)

PI and PSI make VR interesting because they enable *realistic responses* in users. For example, a VR user can feel genuine fear of a virtual pit despite having certain knowledge that the pit is not actually there. It has been argued that these illusions and their capability to produce realistic responses are what make VR useful in a number of applications, ranging from therapy to rehabilitation as well as social interaction [42]. The same property can help achieve a real breakthrough in telepresence robotics. Though Minsky’s original work discussed issues with presence, there are limited studies about presence in telepresence, even though the concept is acknowledged in many ways in robotic telepresence research (for example, [37, 52, 54, 55, 57]). Thus, research on increasing presence in telepresence is timely and needed.

The role of proprioception in presence was identified early in VR research; the better the system can accommodate users’ physical actions (like render matching multimodal output when the user is moving around), the more immersive the experience, and therefore the greater the capability of eliciting PI [38, 42, 50]. The aforementioned requirements include a visible body that matches the user’s own proprioception; HMD-based VR systems achieve this by a virtual body coupled with a tracking system. PSI, on the other hand, is suggested to be dependent on *coherence*, which can be roughly defined as the general credibility of the virtual scenario and its ability to meet expectations [39]. There are studies, however, that indicate the benefits of having a visible body for PSI as well [41, 45].

Although the role of a virtual body in presence has been investigated in abundance in the field of VR, its role in 360-video, or robotic telepresence, is yet underexplored. We, however, argue that even a virtual nonphysical body augmented on top of a video feed has the potential to enhance the experience of a remote participant using a 360-video based telepresence system. In this paper we present the results of our study concerning the augmentation of a 360-video based telepresence experience with a virtual body and its effects on presence.

This paper contributes results from a rigorous user study on the effect of having a virtual body in HMD telepresence. After an initial 20-participant pilot, we hypothesized that having a virtual body would increase PI, co-presence, and preference of the user, measured by both questionnaires and behavioral metrics. After estimating effect sizes using the pilot study, we performed a 62-person confirmatory study. However, the confirmatory study did not confirm any of our questionnaire-based hypotheses, and from behavioral metrics we obtained mixed results depending on the exact metric used. We looked into open-ended questions to characterize preferences that could explain the unexpected results, and discovered that individual sentiment towards the robot body, often arising from the individual’s perception of the naturalness of the robot body, moderated participants’ sense of presence to a large degree. We believe this study, although not confirming our initial hypotheses, will be an important springboard for further research in understanding how individual differences can contribute to the adoption of HMD telepresence technology going forward.

## 2 RELATED WORK

Telepresence has different or complementary use cases compared to the common multi-user Collaborative Virtual Environments (CVEs), the earliest of which are already several decades old [33]. Contemporary CVEs, such as Rec Room, Glue, or Mozilla Hubs are available for the general public and allow remote participants to communicate in computer-generated virtual space, either through HMDs or traditional displays. The drawback of CVEs is that they require everyone to participate virtually; a mix of physical and virtual participants is not possible or very complex [26]. In contrast, there are often cases where one or several participants would like to join a physical event, such as a factory tour or grandchild’s birthday, remotely. For these types of events, it is important that the remote person also feels present, which is where telepresence robots are useful.

Whereas there is research on using 360-degree cameras as streaming devices during the COVID-19 pandemic to stream, for example, a funeral [58], there is a limited amount of research on using an HMD for face-to-face communication between telepresent and local people. Much work regarding an HMD in communication instead focuses

on shared experiences, such as an HMD user “riding along” with another [21, 56]. While most publications regarding telepresence robots consider 2-D telepresence robots which stream to a regular screen, there is an increasing amount of recent work regarding how a mobile telepresence robot could stream to an HMD to increase the feeling of presence of the remote participant [4, 19, 30, 53, 63]. The work presented in this paper is geared towards telepresence robots, but can be applied to stationary cameras as well.

### 2.1 Presence and Virtual Body

Over the years, a multitude of studies have been conducted regarding various aspects and their effect on presence in VR. Some examples include locomotion techniques [49, 51], visual fidelity [43, 62, 65], passive and active haptic feedback [25, 29], and audio [2]. The role of proprioception and virtual body for presence in VR is well-known [38, 42]. The more users can utilize natural body movements to control the VR system, the better [44]. Also, having a visible virtual body corresponding to user movements is beneficial for PI [47]. This is unsurprising, since *sensorimotor contingencies* [32], the match between your own sensorimotor actions and VR output, or *immersion*, is known to be a prerequisite for PI [42]. Interestingly, however, even though PSI does not depend on sensorimotor contingencies, virtual bodies seem to affect not only PI, but PSI as well [42]. A study aimed at separating various PI and PSI related aspects found that having a virtual body was important for both PI and PSI [45]. Also, in Skarbez’s experiment examining the relative importance of various factors for eliciting PSI, having a virtual body was found to be the most important [41].

In the context of VR and telepresence, inhabiting a foreign, or virtual, body is often referred as the *sense of embodiment* (SOE) [23]. SOE has been defined to consist of the following three subcomponents: *agency* (I control the body), *body ownership* (this is my own body), and *sense of self-location* (my location in relation to the body, distinct from PI) [23]. There are multiple studies showing that a personalized avatar (or a personalized virtual limb) can increase body ownership [1, 16, 20]. However, Lugrin et al. [27] found that there is a risk of uncanny valley with human-like avatars, and that cartoon-like and robot-like avatars provided better body ownership. Among many other things, SOE has been utilized to induce gender-swapping and race-swapping illusions [3, 6], and even out-of-body experiences [46].

### 2.2 Co-presence

Co-presence and social presence are presence-related constructs referring to the VR illusions of being together with company. While the exact terminology somewhat differs throughout the literature, Biocca [5] and Skarbez [40] suggest that the co-presence illusion refers to the illusion of other humans being present, whereas the social presence illusion refers to the illusion of interacting with someone. Co-presence has been the focus of various studies concerning both reactions to virtual characters [14], as well as human-controlled avatars in CVEs [7, 26].

The role of virtual bodies in co-presence and social presence illusions has been studied in the context of both CVEs and robotic telepresence. In a purely virtual setting, a realistic hand created the best social presence when communicating with sign language [61]. Tanaka et al. [55] found embodying a physical robot enhanced social telepresence in remote face-to-face communication. There exist some studies where video footage has been combined with virtual elements; in a study similar to ours, there was no difference in social presence between realistic and cartoon-like avatars in a collaboration between AR and VR users [60]. In a study focusing mostly on user experience, Danieau et al. [10] did not find significant differences for PI when augmenting pre-recorded 360 video with various virtual bodies; however, Chen et al. [9] did notice an increase in presence when users could see their hands and a table in front of them. In this study, the difference to the previous ones is that the discussion was live, not pre-recorded, forcing the participants to engage in conversation as well.

To conclude, there is interest and need for our study, namely in understanding the importance of seeing your “own” virtual body in VR-based telepresence. Moreover, we envision that future telepresence robots might have physical robotic arms for manipulation tasks and to

facilitate communication. Since we were not interested in researching individual virtual bodies, we decided to display the user's body to themselves as robot arms and a robot torso instead of showing realistic looking human arms or any visualization of legs or the lower body. This also afforded us the benefit of avoiding any uncanny valley effects.

### 3 HYPOTHESES

We performed an initial pilot study to further specify our hypotheses and analysis methods. Based on related literature, we predicted that adding a body would make the study participants feel greater presence, and as well that they would prefer having a body. We measured PI both with questionnaires and behavioral metrics, whereas co-presence and preference are measured with questionnaires only. Based on the results of the pilot study, we tested the following hypotheses:

- H1: Having a representation of an artificial robot body (as opposed to no virtual body at all) in a telepresence situation increases the feeling of PI
- H2: Having a representation of an artificial robot body (as opposed to no virtual body at all) in a telepresence situation increases the feeling of co-presence
- H3: The users prefer having the virtual body

## 4 METHODOLOGY

### 4.1 Participants

Participants gave written informed consent in accordance with the local ERB prior to participating in the study. Our required sample size was derived from our pilot study of 20 participants where the smallest effect size of interest was the Slater-Usch-Steed (SUS) questionnaire score difference between body and without body conditions ( $d_z=0.35$ ; see Section 4.7). We performed *a priori* power analysis using G\*Power software and determined that 62 participants would be needed to achieve 80% power to detect an effect of this magnitude or greater while still obtaining an equal number of participants per group for the between-participants hand motion comparison. Our final 62 participant sample had a mean age of 30.1 years old ( $sd=6.33$ ) and consisted of 20 females, 41 males, and one participant who preferred to self-describe their gender.

When replacing participants after initial exclusions, by error, replacement participants were not distributed into conditions in the same manner as the excluded participants they replaced, resulting in 33 participants in the "body first" condition order and 29 participants in the "without body first" condition order. This created an imbalanced ratio of 33 and 29 participants per group in the between-participants analyses (hand motion) and 62 participants per condition in the within-participants analyses (all others).

### 4.2 Exclusion Criteria

Preregistered exclusion criteria stated that participants who did not complete the entire experiment's protocol, or for whom the onset of the rat throw event (see Section 4.4) was not visible in the recording frame, would be excluded from the experiment. Protocol completion and recordings were verified after data from 62 participants had been collected, and 12 new participants were recruited to replace participants excluded due to a camera recording malfunction.

### 4.3 Study Setup

To test the effects of having a virtual body in telepresence, we presented participants with a scenario where they are introduced to equipment in a research laboratory. Participants were seated and stationary while one of the authors assumed the role of the laboratory introduction host and followed a predetermined script to interact with every participant. This researcher shall henceforth be referred to as the host.

The script is broken into two parts, where one session is experienced with a virtual body and the other without a visible body, in alternate orders counterbalanced among participants to make sure each session was experienced both with and without a body. Incoming participants were further counterbalanced by gender such that there would be an

equal number of males and females experiencing each condition order to preclude any potential effects, though we did not expect to see them. The script was designed in such a way that the host did not need to rely on the participants' responses to progress, though responses were solicited to make the participants more engaged. Throughout the interaction, the participant is asked to indicate the location of various objects in the lab so as to make the conversation interactive and to give the participant opportunities to use their hands (even if participants could not actually 'touch' anything they see). We note that we did not tell the participants whether the host could see their hands and they were never explicitly asked to use their hands; this was a deliberate decision, as we focused on how the body affected the participants themselves and wanted to observe the emergent participant behavior. Unbeknownst to the participant, the curtain separating the participant and host was slightly parted immediately prior to the beginning of the interaction so that the host could actually see the participant's real hands regardless of the participant having the virtual body or not. However, the host was at all times blind in regards to whether the participant was experiencing the with or without body condition. The curtain was lowered again before the participant took off the HMD to make sure they did not realize the interaction occurred so close physically.

The participant's arms were shown as robot arms (seen in Fig. 2) for two reasons; first, we wanted to investigate through simulation whether a future telepresence robot having real, physical arms could increase the experience also for the remote user. Second, even though personalized avatars have shown stronger presence and agency [20], other results indicate that cartoon or robot-like arms can be a good substitute [27]. We wanted to avoid ending up in the uncanny valley with a human-like hand [27], and personalizing arms to make them look like the user's own arms was not a motivation for this study. Inverse kinematics tracking for arms and simple finger gestures were provided by Oculus Rift S controllers. No other body parts were tracked; foot tracking has been shown to have no significant effect on the overall result [34] unless they analyzed the participants' movement behaviour; as the participants didn't locomote in our study due to the camera (in this study Ricoh Theta Z1 using full HD resolution) also being stationary, the body's legs were hidden and only the torso and arms were visible. The robot arms were based on the Space Robot Kyle from the Unity Asset Store<sup>2</sup> scaled up 1.3 times on all axes. The arm model had similar joints as a human, and the total length of each arm was 0.8 meters.

Keskinen et al. [22] observed that the user's own height with respect to the camera height had no effect on the viewing experience and concluded that a height of 150 cm was suitable for the camera mount, irrespective of the user's actual height or posture. However, during the initial tests of the system, users felt too high at 150 cm while sitting. Ultimately for both the pilot and main experiment, the camera height was set to be equal to the eye height of the host when seated (117 cm).

All connections were wired to minimize latency. Thus, the camera was chosen to be placed in close proximity to the participant's actual location, both to reduce latency in the video feed and provide directional audio (shown to be important for creating realistic feelings in a remote environment [2]) to the participant without relying on a 3D audio technique. The camera was connected via USB cable to the same laptop that the VR headset was connected to. The 360-degree camera footage was accessed as standard webcam footage in Windows from Unity, where the arms were added and the combination then streamed to the HMD. In practice, the participant was sitting only several meters

<sup>2</sup><https://assetstore.unity.com/packages/3d/characters/robots/space-robot-kyle-4696>

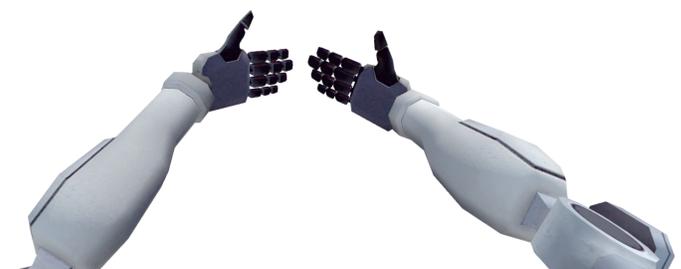


Fig. 2: The user's view of the robot arms with the background removed.

from the camera and the host, but the participant could not see the host setting due to a separating curtain. The host always entered and exited through a different door than the one used by the participant to avoid the participant realizing they were telepresent only behind a curtain. Possible lip synchronization issues, due to small delay in the video but no delay in the audio, were prevented by the host wearing a mask due to the COVID-19 protocols.

#### 4.4 Procedure

Upon arrival, participants were greeted by another researcher (the experimenter) and asked to sit in a chair that was fixed at a predetermined location. They were then asked for consent, after which the experimenter read out the instructions, presented the controls, and told the participant how to put on the VR headset. When they were ready to begin, the experimenter covertly signalled the host to begin the interaction. After the host entered the room, the curtains which separated the participant from the location of the camera were moved slightly without the participant's knowledge to allow the host to see the participant's gestures, both with and without the virtual body.

Each participant was instructed to hold the controllers regardless of condition to avoid creating a bias due to not holding the controllers even when no virtual body was displayed. Recordings of participants were used to monitor their reactions to events in the script. Aside from video data, the position coordinates and Euler angles of the HMD and both controllers in Unity's world coordinate system were also recorded for every frame. During the pilot study, we had an induction phase where the participants were told to practice different finger gestures using the VR controllers (similarly as in, for example, [12]). However, this caused the participants to focus on memorizing the gestures and led to the expectation that they would need to use the optional gestures; thus, the induction phase did not evoke the emergent behavior that we wanted to see. Hence, we modified the instructions to reduce emphasis on practicing the gestures for the subsequent confirmation study.

In an attempt to elicit movement from the participant and gauge an implicit sense of presence, one key event took place during each session. In the first session, a plush rat doll was thrown towards the camera. After the participant was asked if they could see the rat, the host picked it up to show them and then shouted that the rat bit her, throwing the rat at the camera in the process. In the second session, the host walked close to the camera to pick up the thrown item, which could be interpreted by participants as an invasion of personal space. We chose hand motion as our dependent variable in the rat throw event and head motion as our dependent variable in the approach event based on the types of motion elicited in pre-pilot tests. After the host left, the curtains were covertly drawn closed again before asking the participant to remove the headset. We chose to focus only on hand motion in response to the rat throw event after reviewing pilot data (see Section 5.1). Depending on condition order, some participants had a body during the rat throw event while others did not, requiring a between-participants analysis of hand motion within the larger within-participants design (see Section 4.7).

After each session, participants were asked to fill out a questionnaire regarding their experience. After both sessions were completed, they were presented with a post-experiment questionnaire that queried demographic information and overall opinions. Finally, the participants received 20€ Amazon vouchers for participating in the study.

#### 4.5 Measures

Three sets of questionnaires were presented to the participant; two after finishing each of the sessions, and one post-experiment questionnaire after both sessions are over and their respective questionnaires have been submitted. The Google forms service was used to design and present the questionnaires to the participants. The quantitative questions were designed using a Likert scale [28] with a range of 1 to 7.

The questionnaires after the sessions were divided into three or four sections, depending on the virtual body condition for that session. The first section determines the participant's sense of PI. For this component, we used the extended version of the SUS questionnaire [48, 59], using "remote environment" and "sitting with the host" in place of virtual

environment (see Appendix 1). We believed that the SUS would be the most appropriate presence questionnaire for this part of the study because the SUS focuses solely on PI.

The second section of the questionnaire was used for measuring the co-presence with the host. The questionnaire was adapted and modified based on the co-presence questionnaire presented by Casanueva et al. [8] and the co-presence subsection of the social presence questionnaire by Biocca et al. [5]. The exact questions were chosen and modified to fit the scope of this study (see Appendix 1).

The third section of the questionnaire collected qualitative data regarding Breaks In Presence (BIP) [13]. The final section was exclusive to the condition in which the participant had a visible body. This questionnaire (a subset of questions taken from Gonzalez-Franco and Peck embodiment questionnaire [15]) measures participant's sense of embodiment to the virtual avatar.

After all these questionnaires were filled in both sessions, the participants were presented with a post-experiment questionnaire; forced-choice questions on which condition they preferred and which condition induced more co-presence, and their demographic information such as gender, age, video gaming history and history with VR. This final, more overt set of questions was reserved until all conditions had been completed in order to avoid the influence of demand characteristics on the previous, more covert sets of questionnaires.

#### 4.6 Pilot Study

We conducted a separate pilot study of 20 participants (13 male; mean age of 28.5,  $sd=4.52$ ) to fine tune protocol and estimate effect sizes for each analysis. The procedure and analyses to be used in the main study were then preregistered at <https://osf.io/daks8>. Protocol between the pilot and confirmatory study differed in that the host asked and repeated back the participant's name at the beginning of the interaction in the main study, rather than beginning the interaction without introductions as in the pilot, in order to communicate that the experience was interactive rather than a passive recording. Also, participants were not required to practice gestures with the controller before the study, though it was suggested that they try, and one request to "point" to an item was replaced with asking "do you see" the item.

#### 4.7 Analyses

Confirmatory analyses were selected based on pilot data. All other analyses are exploratory and are described under Section 5.3. Confirmatory analyses were preregistered as one-tailed tests, when applicable, while exploratory analyses were always two-sided tests, when applicable. We tested five preregistered hypotheses, applying the Holm correction to correct for multiple comparisons. Applying the Holm correction for multiple comparisons gave the following critical  $p$ -values for our five analyses: .01, .0125, .017, .025, and .05. After obtaining observed  $p$ -values for each confirmatory analysis, these thresholds applied to tests of hand motion, preference for body condition, explicit evaluation of PI, co-presence scores, and SUS scores, respectively.

We sought to measure PI implicitly, via behavior, explicitly, via asking participants directly in which session they felt more present, and somewhere between these two extremes, via the SUS questionnaire. Co-presence and preference were measured only by questionnaire or by asking participants directly whether they felt having a body improved the overall experience, respectively. Given that directly asking participants about their sense of PI, or for their preference, could induce demand characteristics, we place less weight on the results obtained through this method but report them nonetheless for completeness.

SUS score was calculated as the number of SUS questionnaire items given a 6 or 7 rating by a participant, resulting in a score that could range from 0 to 6 for each participant [48]. We performed a Wilcoxon signed-rank test (one-sided) to test for greater SUS scores in the body condition than in the without body condition.

Hand motion in response to the rat throw event was analyzed in the following way: an experimenter identified the frame (framerate: 60 fps) at which the rat left the host's hand and marked this frame as the onset of the event for each participant. The position of the participant's hands, as marked by Oculus Rift S controllers, at the time of this onset frame

served as the initial starting position for motion quantification. Motion was quantified by integrating the total distance in meters traveled from this initial starting position, in three dimensions, for each hand, over the next 200 frames. The integrations for left and right hands were summed to obtain a single measure of motion for each participant. As robustness tests, we also calculated motion as the average distance over the response window of summed left and right hands from a 10 frame pre-window motion baseline, as well as the maximum distance traveled over the response window of summed left and right hands from the same 10 frame pre-window motion baseline. We performed a Shapiro-Wilk test and found that the motion data was non-normally distributed,  $W=.76$ ,  $p<.001$ . We thus tested for greater hand motion in the body group than in the without body group using a non-parametric Wilcoxon rank-sum test (one-sided).

Explicit evaluation of PI (indication of greater PI in the body condition than in the without body condition) was analyzed using an exact binomial test (one-sided).

Co-presence scores were calculated similarly to SUS scores as the number of items in the co-presence questionnaire given a 6 or 7 rating by a participant. Because our co-presence questionnaire only had four items in comparison to the six of the SUS questionnaire, co-presence scores could range from 0 to 4 for each participant. We performed a Wilcoxon signed-rank test (one-sided) to test for greater co-presence scores in the body condition than in the without body condition.

Preference for body condition (preference for the body condition over the without body condition) was analyzed using an exact binomial test (one-sided).

## 5 RESULTS

### 5.1 Pilot Results

The pilot study consisted of 20 participants using a within-participants design. All analyses were run within-participants except head and hand motions, which were between-participants and thus had only 10 people per group). Pilot results indicated benefits of having a virtual body in regard to PI, co-presence, and preference. We observed a strong influence of having a virtual body each in terms of resulting hand motions in response to the rat throw event, co-presence scores, explicit presence, and explicit preference, while the effects in terms of PI as measured by SUS scores were moderate. In contrast, the virtual body manipulation produced negligible effects on resulting head motions in the approach event.

SUS scores in the body condition ( $m=3.80$ ) were not statistically greater than SUS scores in the without body condition ( $m=3.05$ ) as indicated by a Wilcoxon signed-rank test (two-sided),  $Z=1.48$ ,  $p=.14$ ,  $r=0.33$ , suggesting there was no substantial increase in PI when participants had a virtual body in this sample, as measured by the SUS survey.

The distribution of hand motion was non-normal as indicated by a Shapiro-Wilk normality test,  $W=.85$ ,  $p=.006$ , thus a non-parametric test was used. During the rat throw event, motion was significantly greater in the body group ( $m=1.30$ ) than in the without body group ( $m=0.41$ ) as indicated by a Wilcoxon rank-sum test (two-sided),  $Z=2.53$ ,  $p=.012$ ,  $r=0.57$ , suggesting that the inclusion of a virtual body increased participants' implicit sense of PI during the rat throw event.

The distribution of head motion was non-normal as indicated by a Shapiro-Wilk normality test,  $W=.89$ ,  $p=.025$ , thus a non-parametric test was used. During approach event, motion was not significantly greater in the body group ( $m=0.35$ ) than in the without body group ( $m=0.25$ ) as indicated by a Wilcoxon rank-sum test (two-sided),  $Z=0.04$ ,  $p=.97$ ,  $r=0.01$ , suggesting that the inclusion of a virtual body had little effect on participants' implicit sense of PI when rendered during the approach event.

Due to the small effect size related to the approach event, and the resulting onerous number of participants that would be required to reliably detect such an effect given it truly exists, the analysis of head motion was not included in our subsequent confirmatory experiment. We initially considered the rat throw and approach events to be equivalent in terms of events that could elicit motion from participants. After reviewing the pilot results, however, we reasoned that the difference

between the effects might derive from differences in the time scales and emotional aspects of the events. In particular, the rat throw is a sudden onset, physically threatening event, whereas the approach is more of a medium onset, socially discomforting event. It is also perhaps not surprising that including visible arms would have a larger effect on hand motion than on head motion.

When asked explicitly which condition lent greater PI, 18 out of 20 participants selected the body condition, and this preference towards the body condition was significant in an exact binomial test (two-sided),  $p<.001$ , suggesting that the inclusion of a virtual body increased participants' explicit sense of PI.

Co-presence scores in the body condition ( $m=3.00$ ) were significantly greater than co-presence scores in the without body condition ( $m=2.05$ ) as indicated by a Wilcoxon sign-rank test (two-sided),  $Z=2.45$ ,  $p=.014$ ,  $r=0.55$ , suggesting there was an increase in co-presence when participants had a virtual body, as measured by the co-presence survey.

When asked explicitly about their preference for virtual body, 16 out of 20 participants selected the body condition, and this preference towards the body condition was significant in an exact binomial test (two-sided),  $p=.012$ , suggesting that participants preferred having a virtual body to not having any body representation at all.

### 5.2 Confirmatory Results

The pattern of results in the main study differed from the pattern observed in the pilot. We found some support for an increase in implicit presence provided by the virtual body, but little to no support for benefits in regard to PI, co-presence, explicit presence, or explicit preference upon initial analysis.

**Having the body did not increase presence as measured by the SUS or when asked directly.** SUS scores in the body condition ( $m=2.69$ ) did not differ from SUS scores in the without body condition ( $m=2.74$ ) as indicated by a Wilcoxon sign-rank test (one-sided),  $Z=-0.48$ ,  $p=.63$ ,  $r=-0.06$ , providing no evidence of an increase in PI when participants had a virtual body, as measured by the SUS survey. When asked explicitly which condition created greater PI, 36 out of 62 participants selected the body condition, though this bias towards the body condition was not significant in an exact binomial test (one-sided),  $p=.13$ , providing no evidence that the inclusion of a virtual body increased participants' explicit sense of PI.

**There is partial evidence that having the body increases presence through behavioral metrics.** The distribution of integrated hand motion was non-normal as indicated by a Shapiro-Wilk normality test,  $W=.75$ ,  $p<.001$ , thus a non-parametric test was used. During the rat throw event, motion was significantly greater in the body group ( $m=0.636$ ) than in the without body group ( $m=0.420$ ) as indicated by a Wilcoxon rank-sum exact test (one-sided),  $Z=2.53$ ,  $p=.011$ ,  $r=0.32$ , although this difference did not survive the Holm-corrected alpha threshold ( $\alpha = .01$ ). Using the motion average robustness metric rather than motion integration (see Section 4.7), motion was significantly greater in the body group ( $m=0.102$ ) than in the without body group ( $m=0.045$ ) as indicated by a Wilcoxon rank-sum exact test (two-sided),  $Z=2.92$ ,  $p=.003$ ,  $r=0.37$ . Additionally, the maximum vector robustness metric (see Section 4.7) also revealed significantly greater motion in the body group ( $m=0.232$ ) than in the without body group ( $m=0.115$ ) as indicated by a Wilcoxon rank-sum exact test (two-sided),  $Z=2.93$ ,  $p=.003$ ,  $r=0.37$ . Despite the failure to survive correction for multiple comparisons for the confirmatory study's integrated motion effect, we take the combination of the pilot motion effect, the confirmatory motion effect, and both robustness checks in the confirmatory experiment as some evidence suggesting that having a virtual body offered participants an increased sense of implicit PI on an immediate time scale, given that each effect is statistically significant, or borderline statistically significant, in the predicted direction (see Fig. 3).

**Having the body did not increase co-presence.** Co-presence scores in the body condition ( $m=2.10$ ) were not significantly greater than co-presence scores in the without body condition ( $m=2.02$ ) as indicated by a Wilcoxon signed-rank test (one-sided),  $Z=0.85$ ,  $p=.40$ ,  $r=0.11$ , providing no evidence of an increase in co-presence when participants had a virtual body, as measured by the co-presence survey.

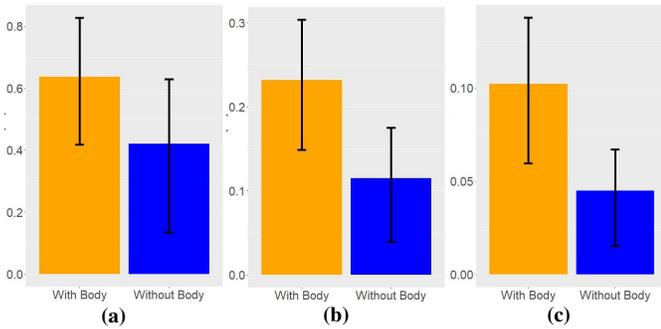


Fig. 3: Hand motion data from the rat throw session, measured with (a) integration, (b) max vector, and (c) mean vector. Motion, in meters, is on the y axis. Error bars represent 95% bootstrapped confidence intervals.

**There is, at best, weak evidence that the body was preferred.** When asked explicitly about their preference for virtual body, 39 out of 62 participants selected the body condition, and this bias towards the body condition was significant in an exact binomial test (one-sided),  $p=.028$ , however this difference did not surpass the Holm-corrected alpha threshold ( $\alpha = .0125$ ), providing at best weak evidence in combination with the pilot results that participants preferred having a virtual body to not having any body representation at all.

### 5.3 Exploratory Results

#### 5.3.1 Confound Analysis

**We observed no confounds related to the order of sessions and stimuli.** Here we describe tests for the influence of unintended confounding variables; namely, order effects potentially arising from experiencing the body condition in the first or second session, and session effects potentially arising from differences between the first session containing the “rat throw” event and the second session containing the “approach” event. We must note that we are unable to distinguish between an effect of session independent of mere time spent in the HMD, as script constraints dictated that the session containing the rat throw event always preceded the second session containing the approach event. We observed no order effects related to whether the body condition came in the participant’s first or second session, as indicated by Wilcoxon rank-sum tests (two-sided) for SUS scores,  $Z=0.42$ ,  $p=.68$ ,  $r=0.05$ , or co-presence scores,  $Z=1.49$ ,  $p=.14$ ,  $r=0.19$ . Likewise, we observed no session/HMD accommodation effects, as indicated by Wilcoxon signed-rank tests (two-sided) for SUS scores,  $Z=0.79$ ,  $p=.42$ ,  $r=0.10$ , or co-presence scores,  $Z=1.91$ ,  $p=.056$ ,  $r=0.24$ . We also observed no order effects, as indicated by two-sample proportion tests (two-sided) in terms of forced choice PI,  $\chi^2(1, N=62)=0.90$ ,  $p=.34$ , or forced choice preference  $\chi^2(1, N=62)=0.43$ ,  $p=.51$ . Confirmatory motion data was only analyzed from the first session as a between-participants test and thus the confounds described would not apply.

#### 5.3.2 Qualitative Data

We collected open-ended data to gather insights into why participants did or did not prefer having a virtual body. The data we collected included BIPs using open-ended responses, and typed open-ended responses that were justifications for forced-choice questions regarding PI and preference. The open-ended data was analyzed using the thematic analysis method with inductive approach [35]. In the first stage of analysis, two researchers independently identified codes from the response data and mutually agreed on the codings in the second phase.

**The stimulus events caused the most breaks in presence, with the visual proportions and the virtual body gathering the next most mentions.** The BIP related data contained a relatively small number of remarks regarding the virtual body (4 occurrences), or the lack thereof (6 occurrences). According to questionnaire data, the most common causes for BIPs in both conditions were the *rat throw* (8 and 11 occurrences in no-body and body conditions, respectively) and *approach* (11 and 9 occurrences respectively), the same events we used to collect behavioral responses. Aside from these events, *visual proportions* (7 occurrences without body, 4 occurrences with body)

were a common cause for BIPs (“when looking down I felt I am too tall” and “When I looked around and realized my proportion was out of place. Host also looked like a giant when they came near me.”). The remaining responses were mostly concerning unrelated disturbances such as external noises, comfort, video quality and HMD related issues.

**Participants who had a better sense of PI with the body thought the ability to interact was the biggest reason why, followed by answers related to plausibility and agency.** Thirty six out of 62 participants considered the virtual body to give a better sense of being next to the guide (PI). For these participants, the three most popular responses concerned *interaction* (16 occurrences), *plausibility* (7 occurrences), and *agency* (6 occurrences). *Interaction* related responses concerned the body being beneficial for either interacting with the guide (“because i could use my hands to interact with the host”), or the environment, (“It gave me a sense of feeling of being able to interact with the environment”), even though the remote environment did not contain interaction affordances. Responses coded as *Plausibility* contained aspects that could be attributed under PSI or coherence [39]. These responses concerned either general sensation of credibility or realism (“It felt normal to see your hands in front of you”), or matching expectations (“less disconnect with what I expect to see. Expecting to see hands but the robot has none causes confusion”). *Agency* referred to being able to control the virtual body without any specific interaction related context (“Because there was a sense of physical autonomy”). All codes and their frequencies for participants who considered the virtual body enhancing PI can be seen in Fig. 4.

**Among participants for whom the virtual body did not enhance their PI, the most popular reasons why were naturalness (13 occurrences), plausibility (6 occurrences) and other (3) occurrences.** Responses coded as *naturalness* contained remarks about the virtual body seeming too artificial or unnatural (“The robot body wasn’t very natural”). Examples of *plausibility* related responses include “Somehow seeing the robot hands broke the illusion of actually being somewhere else, and made it feel more like a video game. I immediately did not feel as immersed as in the first time.” and “Robot body does not felt [sic] like I am living in reality”. The category *other* contained responses that did not give a particular reason, were difficult to interpret or did not fit other existing categories (for example, “i am not sure. i felt the first one was more clear to visualize”). All codes and their frequencies for participants who did not consider the virtual body enhancing PI can be seen in Fig. 5.

**People who preferred the virtual body named interaction and sense of place as the biggest factors.** Thirty nine out of 62 participants considered the virtual body to improve the overall experience. Again, *interaction* (16 occurrences) related responses were most popular (e.g., “Because I could interact more with the host”). The second most popular category was *sense of place* (12 occurrences) referring to PI (“Seeing the robot move in the virtual environment made me feel like I was there. Other way around I felt like I was just watching the room via a camera.” and “It gave me the feel that I am in some other environment”). Again, *agency* (7 occurrences) was the third most popular category (“I have feedback on my senses eg Hand movement”). In Fig. 6 are codes and their frequencies for participants considering the virtual body to improve their experience.

**For participants who disliked the virtual body in terms of overall experience, the most popular responses why were related to naturalness (10 occurrences) and pointlessness (4 occurrences)** (for example, “Because it had no physical function”). The remaining codes *sense of place*, *plausibility*, and *visual proportions* were each mentioned once. All codes and their frequencies for participants who did not consider the virtual body improving their experience can be seen in Fig. 7.

#### 5.3.3 Modeling Interaction Effects

The contrast in open-ended answers between users who liked or disliked the body, together with the lack of main effects in confirmatory analyses (in contrast to the moderate to large effects observed in the pilot study) prompted us to dig deeper into the data to look for potential crossover interactions that could be masking main effects. To model

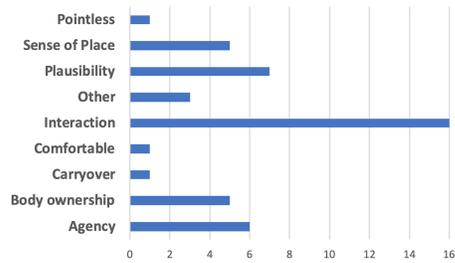


Fig. 4: Qualitative codes for participants who considered virtual body beneficial for their PI.

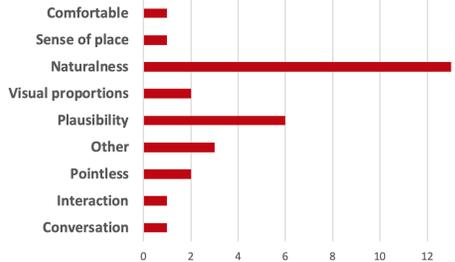


Fig. 5: Qualitative codes for participants who did not consider the virtual body as beneficial for their PI.

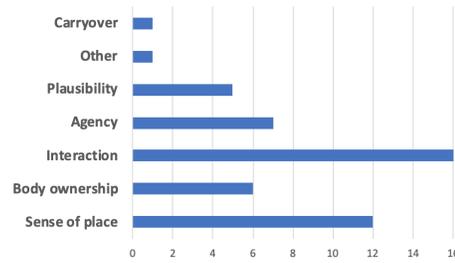


Fig. 6: Qualitative codes for participants who considered the virtual body as generally improving their experience.

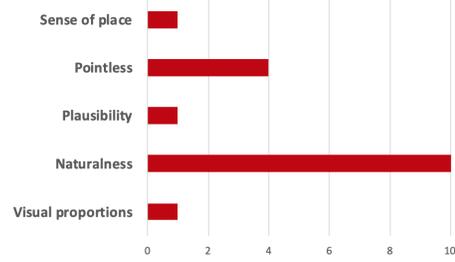


Fig. 7: Qualitative codes for participants who did not consider the virtual body as generally improving their experience.

SUS and co-presence scores we used cumulative link mixed models from the R package 'Ordinal' ([18]) designed to handle repeated measure predictors and ordinal response variables. In each of these models, between-participant variance was specified as a random intercept term.

**Presence was well explained by saturated interaction models using condition, session, and sentiment as predictors.** We took inspiration from the qualitative results; as described in Section 5.3.2, there was a strong divide concerning whether having a virtual body improved the overall telepresence experience, with some participants describing how they appreciated the virtual body while others described how it bothered them. Rather than treat their sentiment towards the virtual body as an outcome variable, we instead tried using it as a predictor to investigate whether sentiment modulated presence effects. Participants' "yes/no" responses to the question asking whether the virtual body improved the overall experience was inserted into the model as a proxy for their sentiment towards the virtual body. The three-way interaction models predicting SUS and co-presence via condition (SUS:  $\beta = 6.32$ ,  $CI = 3.09, 9.54$ ; co-presence:  $\beta = 4.82$ ,  $CI = 2.07, 7.56$ ), session (SUS:  $\beta = 3.57$ ,  $CI = 0.60, 6.54$ ; co-presence:  $\beta = 4.48$ ,  $CI = 1.79, 7.17$ ), and sentiment (SUS:  $\beta = 3.34$ ,  $CI = 0.58, 6.10$ ; co-presence:  $\beta = 3.60$ ,  $CI = 1.19, 6.02$ ) and their interaction terms were successful (Appendix 2, Interaction Models). Thus, having a virtual body increased senses of PI and co-presence but only when participants held a positive sentiment towards the virtual body, and in a manner that was also sensitive to the participant's current session, with more pronounced differences due to sentiment arising from the approach session than the rat throw session. These saturated interaction models outperformed the null models and additive models (Appendix 2), and the remaining possible unsaturated model configurations.

**Sentiment towards the virtual body is key to predicting presence across conditions and sessions.** Figures 8 and 9 display two way interactions between condition and sentiment, session and sentiment, and condition and session, for SUS and co-presence scores. Figure 8a depicts the interaction between condition and sentiment for SUS scores, revealing that scores were similar when participants had a virtual body regardless of whether they liked the virtual body; however, when the virtual body was absent, those who disliked the virtual body recorded *higher* SUS scores than those who liked the virtual body. Figure 8c portrays a similar story for co-presence scores, yet here the larger difference due to sentiment comes during the with body condition rather than in the without body condition. Figure 8b depicts the interaction between session and sentiment for SUS scores, revealing that sentiment towards the virtual body had little effect during the rat throw session, yet during the approach session, those who liked the virtual body recorded

higher scores than those who disliked it. This interaction played out differently for co-presence scores (Figure 8d), where generally higher scores were recorded during the rat throw event than in the approach event, and in particular for those who liked the virtual body. Figures 9a and 9b depict the interaction between condition and session for SUS scores and co-presence scores, respectively. Scores tended to be higher in the rat throw condition, and differences due to sentiment, particularly for co-presence scores, were larger in the body condition than in the without body condition.

**Sentiment additionally predicts the magnitude of motion in response to the rat throw event.** Although there was some evidence of a main effect of motion per confirmatory analyses, we were curious whether the motion effect would also interact with sentiment. If not, it would suggest that seeing a representation of our hands causes us to move them in response to a threatening stimulus, regardless of whether we like our virtual hands or find their virtual representation to be believable. If instead the motion effect did interact with sentiment, this would suggest that even our implicit, automatic reactions can be shaped by higher level cognition regarding whether we like our virtual hands or find them to be believable. We find evidence in support of the latter. We used the R package 'fitdistrplus' [11] to graphically compare fits for our motion distribution, and found it to reconcile best with a gamma distribution. We then fit a gamma regression model using condition, sentiment, and their interaction term to predict integrated hand motion. The interaction term was a good predictor of motion ( $\beta = 4.60$ ,  $CI = 1.76, 8.15$ ), while the main effects of condition ( $\beta = -0.73$ ,  $CI = 0.72, -2.47$ ) and sentiment ( $\beta = -0.62$ ,  $CI = 0.63, -2.35$ ) were not. Still, this interaction model significantly outperformed the null model, the condition predictor only model, and the additive model (Appendix 2). Figure 10 depicts this hand motion interaction. For those who did not like the virtual body, movement was largely equivalent whether or not they had a virtual body, while for those who did like the virtual body, they tended to move more when they had it and barely move at all without it.

### 5.3.4 Demographic Analyses and Embodiment

**Demographics failed to predict presence.** No demographic information such as age, gender, time spent playing video games, or time spent using VR predicted whether a participant was more likely to respond that having a virtual body improved their overall experience in binomial regression models, nor did they predict differences in SUS or co-presence scores between conditions in ordinal regression models (all model coefficient 95% bootstrapped CIs include 0).

**Greater embodiment was associated with improved experiences**

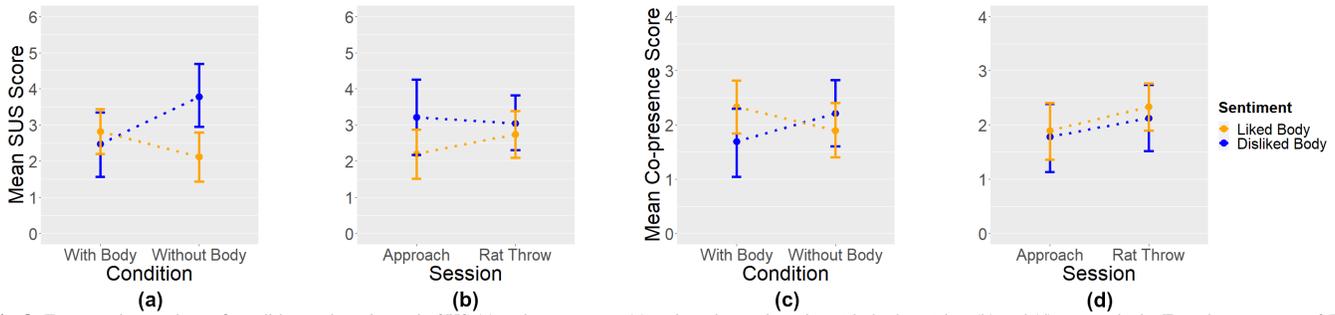


Fig. 8: Two-way interactions of condition and sentiment in SUS (a) and co-presence (c) and session and sentiment in both metrics, (b) and (d), respectively. Error bars represent 95% bootstrapped confidence intervals.

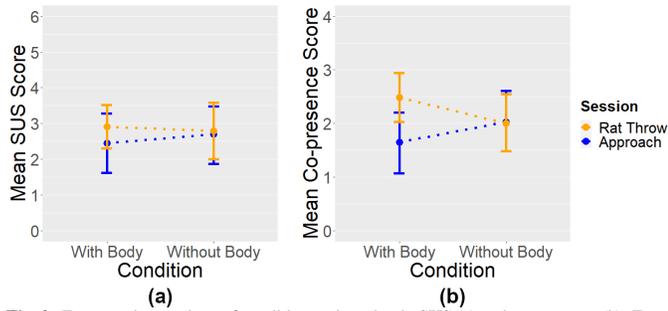


Fig. 9: Two-way interactions of condition and session in SUS (a) and co-presence (b). Error bars represent 95% bootstrapped confidence intervals.

**with the virtual body, but not with increased presence.** The embodiment scores did not predict SUS or co-presence scores in ordinal regression models; however, embodiment scores did moderately predict whether a participant felt that having a virtual body improved the overall experience in a binomial regression model (Appendix 2). Those who felt greater embodiment were more likely to report that the virtual body improved their experience ( $\beta = 0.88$ ;  $CI = 0.28, 1.58$ ). Despite this relationship, sentiment outperformed the embodiment predictor in interaction models, suggesting that while embodiment could influence sentiment, there remained contributions of variance from sources other than embodiment. Embodiment was thus not included in the final interaction models.

### 5.3.5 Confidence in Reciprocal Viewing

**Greater confidence that the host could see the participant’s arms was associated with greater presence.** After completing the experiment, participants were asked to rate their confidence that the host could see their hands, separately for each condition, using a 7-point Likert-scale. Participants were significantly more confident that the host could see their hands in the virtual body condition, as revealed by a Wilcoxon signed-rank test (two-sided),  $Z=4.48, p < .001, r=0.57$ . Furthermore, confidence that the host could see their hands predicted SUS scores ( $\beta = 0.27$ ;  $CI = 0.05, 0.49$ ) and co-presence scores ( $\beta = 0.24$ ;  $CI = 0.03, 0.46$ ) in ordinal regression models, although interactions with condition did not reach significance (SUS:  $\beta = 0.47$ ;  $CI = -0.02, 0.97$ ; co-presence:  $\beta = 0.48$ ;  $CI = -0.02, 0.98$ ). This overall pattern suggests that while participants were more likely to believe that the host could see their movements when they had a virtual body, a stronger belief that the host could see them led to stronger presence regardless of whether they had a virtual body or not.

## 6 DISCUSSION

This work suggests that individual differences may play a role in attempts to increase the strength of the PI and feelings of co-presence. The difference between pilot and confirmatory study results, not confirming the pre-registered hypotheses, was initially a surprise since the pilot study hinted towards moderate to large effect sizes for each result. The only consistent results between the pilot and confirmatory analyses were the extra behavioral metrics for PI. However, exploratory analysis on the confirmatory study revealed that those who preferred the body experienced greater presence when they had a body, while conversely those who disliked the body experienced greater presence when they

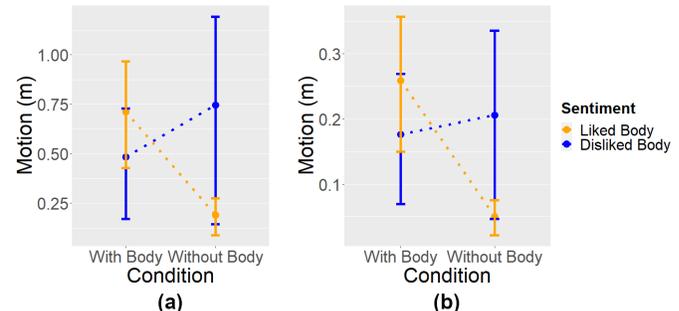


Fig. 10: Two-way interactions of condition and sentiment measured in hand motions, using either the integration (a) or max vector (b) metric. Error bars represent 95% bootstrapped confidence intervals.

did not have a body. Thus, the lopsided proportion of people liking the body in the pilot study (18/20) seems to have caused this discrepancy. It is also interesting to note that earlier work on whether seeing arms while watching 360 videos increases presence provided conflicting results, with one paper finding supporting evidence [9] while the other did not [10]. Though the use case here is different, with a truly interactive scenario, further research is merited.

To delve deeper into the reasons for dependency between sentiment towards the body and feelings of presence, we analyzed the answers to open-ended questions asking why the body did or did not increase the PI. Thirteen out of 26 participants who did not consider the virtual body increasing PI complained about the *naturalness* of the body. Similar responses came up also in the second forced choice question asking for general preferences (10 out of 23 mentioning *naturalness* as the reason for not preferring the body). This connection is supported by previous research: visual inconsistencies, and inconsistencies in general, seem to have a detrimental effect on PI [13, 39], and personalized limbs, on the other hand, increase body ownership (see Section 2.1). The choice of robot arms, based on the use case of real robot arms and avoiding the uncanny valley, seemed to have a stronger effect than we anticipated. Also, other use cases with virtual agents show that perceived likeness to virtual agents increases acceptance [64]. Thus, we conclude that there may be many people who benefit from seeing their own arms in immersive telepresence videos, if care is taken to make sure the arms look and feel real. It would be beneficial to repeat the experiment with more realistic appearing arms in the future to confirm this.

It is not clear why disliking the body in the approach session caused higher PI scores and having a body in the rat throw caused higher co-presence. Given the differences in the response-eliciting events between the sessions (see Section 5.1), we reason that this effect likely arises due to the timescale or the nature of event, or some combination of both factors. Perhaps another human slowly entering and leaving the space where they should have made physical contact with the participant led to a stronger PI break than a briefly thrown object, or led to a longer PI break which a participant could weigh more heavily than a brief one.

Although our main focus was PI and co-presence, the open-ended data contained many responses that were PSI related. For example, “it just in some way felt more real cause I could see my hands in that environment, even if they were robot hands” when stating a preference towards virtual body, or “because seeing my hands feels like a little bit that everything is unreal” when stating why PI was stronger without

body. Although we used the term *naturalness* when coding responses concerning unnatural virtual arms, these remarks can be also interpreted as concerning the *coherence* of the virtual body (for example, “*the body felt more being like a pat [sic] of something else than my own.*”). According to many of these remarks, the computer generated robot arms simply seemed out of place for many participants. Although PI and PSI are generally treated as separate illusions (one can have PSI without PI and vice versa), in our case, having a virtual body that broke PSI seemed to affect PI as well (participants reporting preference towards not having a virtual body reporting significantly lower PI scores). However, this is in line with previous research in the sense that virtual body seems to be an aspect that affects both PI and PSI [42, 45].

This work also demonstrates the importance of running sufficiently powered studies. If we had trusted the results of our initial 20 participant pilot without performing a confirmatory analysis in a larger sample, we could have falsely concluded that rendering a robot-styled virtual body has a strong positive impact on many facets of the telepresence experience, when the larger, and thus, more likely to be representative, sample indicates that it can have the *opposite* effect depending on how the specific user reacts to the virtual body. To better generalize the results, similar studies with different arms, both high-quality robot arms and personalized human-like arms, should be used; however, both the open-ended answers and Section 2.1 hint that better graphics and more individualized arms would increase the presence. That said, the results concerning the modulating effect of sentiment are experimental, and should be taken with caution until replicated in future experiments.

## 6.1 Limitations

The stimuli we chose, throwing the plush rat doll at the participant and violating their personal space, seemed to evoke reactions from the participants, but also to break their presence. The reasons were well documented in the open-ended questions; in the personal space violation, the approach, participants often mentioned that the host came so close it was unrealistic they did not bump into the participant (“*The host essentially passed my arm as she was getting Chewie*”, “*She came too close, would have touched me, but of course no feeling of it.*”). Similar comments, if not as many, were observed with the rat throwing (“*Being hit by the rat but not really feeling it, causing me to realize it was virtual*”, “*When i realised that the white mouse is not actually going to touch my skin*”). This risk was already evident in the pilot study, but we chose to keep the same stimuli in the confirmatory study because they did evoke visible reactions in the participants, and we wanted to have an objective measurement of PI besides the questionnaires. We feared that more typical events, such as dropping a pen, would not provoke strong enough reactions from the participants to be registered by our measures. We also note that there may have been individual differences across sessions on how close exactly the host went to the camera, due to inaccuracies in the throwing, but we took steps to make this as consistent as possible.

Also, always having rat throw in the first session and approach in the second could have caused an HMD accommodation effect (which could also explain the effect condition had in the interactions in Figs. 8(b,d),9). However, we argue that this is unlikely; participants took long pauses between sessions to answer questionnaires where the PI could have had time to reset, and even if some effect carried over between sessions, we would expect that more time in the HMD, and thus the second session (approach), would receive higher scores, whereas our results clearly show a trend of higher scores in the first session (rat throw).

As discussed earlier, the appearance of the robot arms turned out to be a confounding factor. In addition, participants complained about incorrect visual proportions. These complaints came up especially in the BIP responses (7 occurrences and 4 occurrences without body and with body, respectively). The responses were either referring to participants’ sense of own height, the apparent size of the host (for example, “*The host body shape and height seems bit different, may be due to camera*”), or the relationship of the virtual body to the rest of the environment (for example, “*When I looked around and realized my proportion was out of place*”). The incorrect proportions were most likely due to the properties of the 360 camera making objects

in the video appear larger in comparison to the virtual body. This can be considered a confound causing unintended BIPs in our study. Finally, there was a risk that using real audio, due to the short physical distance between the host and the participants, could have confounded the study; however, only one participant reported a break in presence due to realizing that the host is in the same room (and two participants even thought they were watching a video), making us confident that real audio did not confound the study.

The main metric used for measuring hand motions during the stimuli (used as the objective measurement of presence), integration, did not provide statistically significant results, whereas both of the robustness measures, max vector and mean vector, did. A downside of using integration is its susceptibility to noise, as ‘jittery’ participants could potentially artificially inflate their motion values. An upside, however, is the lack of a requirement to choose a baseline, a process which itself could prove to be a source of error if chosen unwisely (though it is worth noting that the max and mean vectors are established metrics when measuring various biosignals and perhaps could be developed into standardized metrics here as well). Ultimately visual inspection of the pilot graphs did not reveal a significant amount of noise, and the motion data in the pilot study were very similar across metrics, which led us to choose integration as the primary metric. Given the consistency of the effect sizes in the same direction among the three, it is likely that they all would have come into agreement with a larger sample and thus more statistical power.

## 7 CONCLUSIONS AND FUTURE WORK

Here we reported findings concerning the augmentation of an immersive telepresence experience with a virtual body and its effect on presence and preference. We used questionnaire data and behavioral measures for confirming our hypotheses. We reported the results of our 20 participant pilot study, followed by a 62 participant confirmatory study. The confirmatory study did not replicate the results of the smaller pilot study regarding greater presence and preference with the virtual body, except for a subset of the behavioral metrics. Exploratory analysis indicated that individual differences could explain the difference between these results. Participants not liking the virtual body often brought up the unnaturalness of the virtual body. During exploratory analysis, we found preference sentiment predicting SUS and co-presence scores; those who liked the virtual body reported higher presence when embodied whereas those who did not reported higher presence scores when not.

There are multiple ways to improve the unnaturalness of the virtual body, a major confounding factor based on open-ended answers. First, our participants used controllers, even though contemporary hardware can directly track hands, and even fingers. However, the controllers did not come up frequently in the open-ended questions, so it seems this was not a major concern for the participants. Secondly, the visual fidelity of the virtual body could be improved by improving its consistency with the remote environment, for example, by adding environment mapping based reflections [17] and lighting that are based on the remote environment; this could potentially reduce the perceived unnaturalness of the body. Additionally, some modern HMDs have the video pass-through ability, which in theory allows the participants to see their own, actual hands, even though edge artifacts are likely to slightly decrease the quality. Even though one of our motivations was using a robot with real, physical arms, the video passthrough could, in theory, also overlay the participant’s real arms on top of the robot’s arms. Deficiencies in visual proportions can possibly be improved upon by transforming the virtual body size and virtual interpupillary distance (IPD) in software to compensate for the incorrect proportions caused by the camera since manipulating IPD in VR can be used to scale the perception of sizes and distances [24, 36]. Finally, we wish to see more research on mixing real and virtual worlds for live telepresence communication.

## ACKNOWLEDGMENTS

This work was supported by the Business Finland project HUMOR 3656/31/2019, Academy of Finland projects PERCEPT 322637, PIXIE 331822, and SRC project COMBAT 293389, and European Research Council project ILLUSIVE 101020977.

## REFERENCES

- [1] F. Argelaguet, L. Hoyet, M. Trico, and A. Lécuyer. The role of interaction in virtual embodiment: Effects of the virtual hand representation. In *2016 IEEE Virtual Reality (VR)*, pp. 3–10. IEEE, 2016.
- [2] J. J. Baldis. Effects of spatial audio on memory, comprehension, and preference during desktop conferences. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '01*, p. 166–173. Association for Computing Machinery, New York, NY, USA, 2001. doi: 10.1145/365024.365092
- [3] D. Banakou, P. D. Hanumanthu, and M. Slater. Virtual embodiment of white people in a black virtual body leads to a sustained reduction in their implicit racial bias. *Frontiers in human neuroscience*, 10:601, 2016.
- [4] I. Becerra, M. Suomalainen, E. Lozano, K. J. Minnaugh, R. Murrieta-Cid, and S. M. LaValle. Human perception-optimized planning for comfortable vr-based telepresence. *IEEE Robotics and Automation Letters*, 5(4):6489–6496, 2020.
- [5] F. Biocca, C. Harms, and J. Gregg. The networked minds measure of social presence: Pilot test of the factor structure and concurrent validity. In *4th annual international workshop on presence, Philadelphia, PA*, pp. 1–9, 2001.
- [6] E. Bolt, J. T. Ho, M. R. Lesur, A. Soutschek, P. N. Tobler, and B. Lenggenhager. Effects of a virtual gender swap on social and temporal decision-making. *Scientific Reports*, 11(1):1–15, 2021.
- [7] J. Casanueva and E. Blake. The effects of avatars on co-presence in a collaborative virtual environment. 2001.
- [8] J. Casanueva and E. Blake. The effects of avatars on co-presence in a collaborative virtual environment. In *Proc Ann Conf SA Inst of Computer Scientists and Information Technologists*, pp. 1–N, 01 2001.
- [9] J. Chen, G. Lee, M. Billinghurst, R. W. Lindeman, and C. Bartneck. The effect of user embodiment in av cinematic experience. 2017.
- [10] F. Danieau, T. Lopez, N. Mollet, B. Leroy, O. Dumas, and J.-F. Vial. Enabling embodiment and interaction in omnidirectional videos. In *2017 IEEE International Conference on Multimedia and Expo (ICME)*, pp. 697–702. IEEE, 2017.
- [11] M. L. Delignette-Muller, C. Dutang, et al. fitdistrplus: An r package for fitting distributions. *Journal of statistical software*, 64(4):1–34, 2015.
- [12] E. Ebrahimi, L. S. Hartman, A. Robb, C. C. Pagano, and S. V. Babu. Investigating the effects of anthropomorphic fidelity of self-avatars on near field depth perception in immersive virtual environments. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 1–8. IEEE, 2018.
- [13] M. Garau, D. Friedman, H. R. Widenfeld, A. Antley, A. Brogni, and M. Slater. Temporal and spatial variations in presence: Qualitative analysis of interviews from an experiment on breaks in presence. *Presence: Teleoperators and Virtual Environments*, 17(3):293–309, 2008.
- [14] M. Garau, M. Slater, D.-P. Pertaub, and S. Razaquae. The responses of people to virtual humans in an immersive virtual environment. *Presence: Teleoperators & Virtual Environments*, 14(1):104–116, 2005.
- [15] 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
- [16] G. Gorisse, O. Christmann, S. Houzangbe, and S. Richir. From robot to virtual doppelganger: Impact of visual fidelity of avatars controlled in third-person perspective on embodiment and behavior in immersive virtual environments. *Frontiers in Robotics and AI*, 6:8, 2019.
- [17] N. Greene. Environment mapping and other applications of world projections. *IEEE computer graphics and Applications*, 6(11):21–29, 1986.
- [18] R. Hirk, K. Hornik, and L. Vana. mvord: An r package for fitting multivariate ordinal regression models. *Journal of Statistical Software, Articles*, 93(4):1–41, 2020. doi: 10.18637/jss.v093.i04
- [19] B. Jones, Y. Zhang, P. N. Wong, and S. Rintel. Belonging there: Vrooming into the uncanny valley of xr telepresence. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW1):1–31, 2021.
- [20] S. Jung, C. Sandor, P. J. Wisniewski, and C. E. Hughes. Realme: The influence of body and hand representations on body ownership and presence. In *Proceedings of the 5th Symposium on Spatial User Interaction*, pp. 3–11, 2017.
- [21] S. Kasahara and J. Rekimoto. Jackin head: immersive visual telepresence system with omnidirectional wearable camera for remote collaboration. In *Proceedings of the 21st ACM symposium on virtual reality software and technology*, pp. 217–225, 2015.
- [22] T. Keskinen, V. Mäkelä, P. Kallioniemi, J. Hakulinen, J. Karhu, K. Ronkainen, J. Mäkelä, and M. Turunen. The effect of camera height, actor behavior, and viewer position on the user experience of 360° videos. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 423–430, 2019. doi: 10.1109/VR.2019.8797843
- [23] K. Kilteni, R. Groten, and M. Slater. The sense of embodiment in virtual reality. *Presence: Teleoperators and Virtual Environments*, 21(4):373–387, 2012.
- [24] J. Kim and V. Interrante. Dwarf or giant: the influence of interpupillary distance and eye height on size perception in virtual environments. In *27th International Conference on Artificial Reality and Telexistence, ICAT 2017 and the 22nd Eurographics Symposium on Virtual Environments, EGVE 2017*, pp. 153–160. Eurographics Association, 2017.
- [25] M. Kim, C. Jeon, and J. Kim. A study on immersion and presence of a portable hand haptic system for immersive virtual reality. *Sensors*, 17(5), 2017. doi: 10.3390/s17051141
- [26] T. Koskela, M. Mazouzi, P. Alavesas, M. Pakanen, I. Minyaev, E. Paavola, and J. Tuliniemi. Avatarex: Telexistence system based on virtual avatars. In *Proceedings of the 9th Augmented Human International Conference*, pp. 1–8, 2018.
- [27] J.-L. Lugin, J. Latt, and M. E. Latoschik. Avatar anthropomorphism and illusion of body ownership in vr. In *2015 IEEE Virtual Reality (VR)*, pp. 229–230, 2015. doi: 10.1109/VR.2015.7223379
- [28] S. A. McLeod. Likert scale. simply psychology., Aug 2019.
- [29] M. Meehan, B. Insko, M. Whitton, and F. P. Brooks Jr. Physiological measures of presence in stressful virtual environments. *Acm transactions on graphics (tog)*, 21(3):645–652, 2002.
- [30] K. J. Minnaugh, M. Suomalainen, I. Becerra, E. Lozano, R. Murrieta-Cid, and S. M. LaValle. Analysis of user preferences for robot motions in immersive telepresence. In *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2021.
- [31] M. Minsky. Telepresence. 1980.
- [32] A. Noë, A. Noë, et al. *Action in perception*. MIT press, 2004.
- [33] H. NOMA, Y. KITAMURA, T. MIYASATO, and F. KISHINO. Multi-point virtual space teleconferencing system. *IEICE transactions on communications*, 78(7):970–979, 1995.
- [34] Y. Pan and A. Steed. How foot tracking matters: The impact of an animated self-avatar on interaction, embodiment and presence in shared virtual environments. *Frontiers in Robotics and AI*, 6:104, 2019. doi: 10.3389/frobt.2019.00104
- [35] M. Q. Patton. Qualitative research. *Encyclopedia of statistics in behavioral science*, 2005.
- [36] M. Pouke, K. J. Minnaugh, A. P. Chambers, T. Ojala, and S. M. LaValle. The plausibility paradox for resized users in virtual environments. *Frontiers in Virtual Reality*, 2:48, 2021.
- [37] I. Rae, B. Mutlu, and L. Takayama. Bodies in motion: mobility, presence, and task awareness in telepresence. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 2153–2162, 2014.
- [38] M. V. Sanchez-Vives and M. Slater. From presence to consciousness through virtual reality. *Nature Reviews Neuroscience*, 6(4):332–339, 2005.
- [39] R. Skarbez, F. Brooks, and M. Whitton. Immersion and coherence: Research agenda and early results. *IEEE transactions on visualization and computer graphics*, 2020.
- [40] R. Skarbez, F. P. Brooks, Jr, and M. C. Whitton. A survey of presence and related concepts. *ACM Computing Surveys (CSUR)*, 50(6):1–39, 2017.
- [41] R. Skarbez, S. Neyret, F. P. Brooks, M. Slater, and M. C. Whitton. A psychophysical experiment regarding components of the plausibility illusion. *IEEE Transactions on Visualization and Computer Graphics*, 23(4):1369–1378, 2017. doi: 10.1109/TVCG.2017.2657158
- [42] M. Slater. Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1535):3549–3557, Dec. 2009. doi: 10.1098/rstb.2009.0138
- [43] M. Slater, P. Khanna, J. Mortensen, and I. Yu. Visual realism enhances realistic response in an immersive virtual environment. *IEEE computer graphics and applications*, 29(3):76–84, 2009.
- [44] M. Slater, J. McCarthy, and F. Maringelli. The influence of body movement on subjective presence in virtual environments. *Human factors*, 40(3):469–477, 1998.
- [45] M. Slater, B. Spanlang, and D. Corominas. Simulating virtual environments within virtual environments as the basis for a psychophysics of presence. *ACM Transactions on Graphics (TOG)*, 29(4):1–9, 2010.
- [46] M. Slater, B. Spanlang, M. V. Sanchez-Vives, and O. Blanke. First person

- experience of body transfer in virtual reality. *PLoS one*, 5(5):e10564, 2010.
- [47] M. Slater and M. Usoh. Representations systems, perceptual position, and presence in immersive virtual environments. *Presence: Teleoperators & Virtual Environments*, 2(3):221–233, 1993.
- [48] M. Slater, M. Usoh, and A. Steed. Depth of presence in virtual environments. *Presence: Teleoperators & Virtual Environments*, 3(2):130–144, 1994.
- [49] M. Slater, M. Usoh, and A. Steed. Taking steps: the influence of a walking technique on presence in virtual reality. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 2(3):201–219, 1995.
- [50] M. Slater and S. Wilbur. A framework for immersive virtual environments (five): Speculations on the role of presence in virtual environments. *Presence: Teleoperators & Virtual Environments*, 6(6):603–616, 1997.
- [51] A. Stavar, L. M. Dascalu, and D. Talaba. Design, test and experimental validation of a vr treadmill walking compensation device. In L. M. Camarinha-Matos, ed., *Technological Innovation for Sustainability*, pp. 402–409. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011.
- [52] B. Stoll, S. Reig, L. He, I. Kaplan, M. F. Jung, and S. R. Fussell. Wait, can you move the robot? examining telepresence robot use in collaborative teams. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, pp. 14–22, 2018.
- [53] M. Suomalainen, K. J. Mimnaugh, I. Becerra, E. Lozano, R. Murrieta-Cid, and S. M. LaValle. Comfort and sickness while virtually aboard an autonomous telepresence robot. In *EuroXR 2021: Virtual Reality and Mixed Reality*, pp. 3–24. Springer International Publishing, Cham, 2021.
- [54] M. Suomalainen, B. Sakcak, A. Widagdo, J. Kalliokoski, K. J. Mimnaugh, A. P. Chambers, T. Ojala, and S. M. LaValle. Unwinding rotations improves user comfort with immersive telepresence robots. In *Accepted for publication in the 2022 ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2022.
- [55] K. Tanaka, H. Nakanishi, and H. Ishiguro. Robot conferencing: physically embodied motions enhance social telepresence. In *CHI'14 Extended Abstracts on Human Factors in Computing Systems*, pp. 1591–1596. 2014.
- [56] A. Tang, O. Fakourfar, C. Neustaedter, and S. Bateman. Collaboration with 360° videochat: Challenges and opportunities. In *Proceedings of the 2017 Conference on Designing Interactive Systems*, pp. 1327–1339, 2017.
- [57] J. E. F. Tree, S. Whittaker, S. C. Herring, Y. Chowdhury, A. Nguyen, and L. Takayama. Psychological distance in mobile telepresence. *International Journal of Human-Computer Studies*, 151:102629, 2021.
- [58] D. Uriu, K. Toshima, M. Manabe, T. Yazaki, T. Funatsu, A. Izumihara, Z. Kashino, A. Hiyama, and M. Inami. Generating the presence of remote mourners: a case study of funeral webcasting in japan. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pp. 1–14, 2021.
- [59] M. Usoh, E. Catena, S. Arman, and M. Slater. Using Presence Questionnaires in Reality. *Presence: Teleoperators and Virtual Environments*, 9(5):497–503, 10 2000. doi: 10.1162/105474600566989
- [60] B. Yoon, H.-i. Kim, G. A. Lee, M. Billinghurst, and W. Woo. The effect of avatar appearance on social presence in an augmented reality remote collaboration. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 547–556. IEEE, 2019.
- [61] B. Yoon, H.-i. Kim, S. Y. Oh, and W. Woo. Evaluating remote virtual hands models on social presence in hand-based 3d remote collaboration. In *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 520–532. IEEE, 2020.
- [62] I. Yu, J. Mortensen, P. Khanna, B. Spanlang, and M. Slater. Visual realism enhances realistic response in an immersive virtual environment-part 2. *IEEE Computer Graphics and Applications*, 32(6):36–45, 2012.
- [63] G. Zhang and J. P. Hansen. People with motor disabilities using gaze to control telerobots. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1–9, 2020.
- [64] S. Zhou, T. Bickmore, M. Paasche-Orlow, and B. Jack. Agent-user concordance and satisfaction with a virtual hospital discharge nurse. In *International conference on intelligent virtual agents*, pp. 528–541. Springer, 2014.
- [65] P. Zimmons and A. Panter. The influence of rendering quality on presence and task performance in a virtual environment. In *IEEE Virtual Reality, 2003. Proceedings.*, pp. 293–294. IEEE, 2003.