# Comparing Mixed Reality Agent Representations: Studies in the Lab and in the Wild

Ben J. Congdon
ben.congdon.11@ucl.ac.uk
University College London
United Kingdom

Gun Woo (Warren) Park
warren@dgp.toronto.edu
University of Toronto
Canada

Jingyi Zhang
jy.zhang@ucl.ac.uk
University College London
United Kingdom

Anthony Steed
a.steed@ucl.ac.uk
University College London
United Kingdom


**Figure 1: The experiment environment in the head-and-shoulders-video vs 3D-model condition. Participants were placed in this environment in virtual reality and had to answer trivia questions. The two agents are advisors with pre-recorded scripts.**

## ABSTRACT

Mixed-reality systems provide a number of different ways of representing users to each other in collaborative scenarios. There is an obvious tension between using media such as video for remote users compared to representations as avatars. This paper includes two experiments (total n = 80) on user trust when exposed to two of three different user representations in an immersive virtual reality environment that also acts as a simulation of typical augmented reality simulations: full body video, head and shoulder video and an animated 3D model. These representations acted as advisors in a trivia quiz. By evaluating trust through advisor selection and self-report, we found only minor differences between representations, but a strong effect of perceived advisor expertise. Unlike prior work, we did not find the 3D model scored poorly on trust, perhaps as a result of greater congruence within an immersive context.

## CCS CONCEPTS

• **Human-centered computing** → **Virtual reality**; **Mixed / augmented reality**; **Computer supported cooperative work**.

## KEYWORDS

Virtual reality, mixed reality, avatars, collaboration, trust

**ACM Reference Format:**
Ben J. Congdon, Gun Woo (Warren) Park, Jingyi Zhang, and Anthony Steed. 2023. Comparing Mixed Reality Agent Representations: Studies in the Lab and in the Wild. In *29th ACM Symposium on Virtual Reality Software and Technology (VRST 2023), October 09–11, 2023, Christchurch, New Zealand.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3611659.3615719



*VRST 2023, October 09–11, 2023, Christchurch, New Zealand*

ACM ISBN 979-8-4007-0328-7/23/10.
https://doi.org/10.1145/3611659.3615719

## 1 INTRODUCTION

The formation of trust is an important part of interpersonal interactions [24]. It has been demonstrated that the media over which communication takes place affects that communication, and trust formation in particular [10, 26, 31]. As we increasingly rely on 'mediated' communication, it becomes more important to understand the effects of the media we use.

While there have long been critiques of the effectiveness of video for remote communication [11], the recent pandemic has raised awareness of its shortcomings [2]. In mixed-reality communications we have the potential to spatially arrange participants to alleviate some of the burden of attention to a single screen (see Section 2.1). We also have the ability to represent the remote users using video or avatar representations. These options come with different advantages and disadvantages including ease of capture, faithfulness to current appearance and expressive power with respect to the shared environment.

In this paper we are particularly interested in scenarios where a remote person has to give specific advice to a user. We are motivated by the need driven by systems such as Hololens[1] where a user might bring in a remote person for consultation. The representation of the remote person might significantly affect the user's ability to build trust with and collaborate with them.

We performed two experiments to investigate how a user develops trust with an agent representing a remote expert: a remote (Online) experiment and an in-the-lab (In-Person) experiment. In both experiments participants had to answer hard trivia questions, and they could ask advice from one of two potential advisors who were represented in different ways. The Online experiment was designed for in-the-wild data gathering [37], without direct contact with the experimenters. There are difficulties in performing augmented reality studies with this approach. Augmented reality head-mounted displays are less common than virtual reality displays, and provide less consistent results; we cannot be sure of the participant's room layout, environment lighting and so on, and variability in these environmental factors can heavily impact how the experiment components are displayed to participants [14]. We instead ran both experiments on virtual reality devices and aimed to keep the virtual environment believable, so that along with virtual reality, the results may also be applicable to pass-through augmented reality. The main difference between the trials is the Online study was run un-monitored with a short protocol, and thus was not too long for participants, whereas the In-Person experiment used a refined and longer protocol (see Section 3).

For experiment design we adapt prior work which looks at trust through the lens of user-advisor relationships [21, 29, 30, 32]. We compare three representations for our advisors: a head-and-shoulders video, as is typical when video-conferencing; a full-body video, as used in some augmented reality applications[2], additionally standing-in for a high-fidelity generated model of a user as in some virtual reality applications[3]; and a lower-fidelity 3D model. The design is kept intentionally similar to prior work [29, 32], as we aim to see how well these earlier studies can be replicated in immersive virtual reality, and whether we observe any differences due to the change in medium.

We measure trust both behaviourally through the rate of advice sought from a given advisor, and through participants' self-report in post-questionnaires. Based on prior work, we hypothesize:

[1]https://microsoft.com/hololens
[2]https://beem.me
[3]https://spatial.io

- **H1** Full-body video will prompt more trust than head-and-shoulders video [45]
- **H2** Video will prompt more trust than the 3D model [29, 32]
- **H3** Participants will gradually come to trust the expert advisor regardless of representation [29, 32]

We found some evidence in support of H1, and strong evidence to support H3. However, we found very little evidence to support H2, perhaps as a 3D model appears more congruent in an immersive virtual environment than in the on-screen windows used in previous studies.

## 2 RELATED WORK

### 2.1 Collaborative Mixed Reality

The area of collaborative mixed reality dates back to early work in simulators, teleoperators and virtual reality systems from 1990s and earlier [35, 38]. The ability to represent users to each other, and thus have a notion of shared embodiment in a space has thus long attracted attention within the field of human-computer interaction [5]. By representing a user as an avatar in a virtual environment, other viewers of the space can very rapidly understand that user's situation and intentions as they can exploit a lifetime of learning how to interpret human movement and expression in everyday situations. This then leads to a shared sense of co-presence between users [9] and this has significant impacts on short and long-term behaviours of users [33].

One key determinant of the impact of avatars is their visual appearance and behaviour complexity [3]. While it is possible to create very high quality humans with complex behaviours [22], a lot of recent systems have used low realism characters that might be cartoony or caricature in nature [13, 19]. There is a tension between making characters realistic but avoiding uncanniness of representation [20]. Thus, as we discuss in Section 2.2, one option that avoids uncanniness is to use video representations when available.

While the previous systems have mostly focused on desktop or immersive virtual reality, collaborative augmented reality or mixed reality has also been explored at great length, but mostly in a laboratory setting given the nature of the devices [7, 8]. A number of recent surveys outline the different ways of considering how to categorise different types of mixed-reality where different users might have different interfaces and/or be represented to each other in different ways [12, 23, 34].

Recently researchers have started looking directly at the effect avatar representation has on users of collaborative augmented reality [45]. In that paper Yoon et al. compare different virtual avatar representations: head and hands, upper body and full-body avatars. They found that the full body avatar was perceived as best. Our study will follow this finding for the 3D avatar cases.

### 2.2 Video Avatars

Given the varied responses to avatars, many systems have supported video representations of users within shared virtual environment. An early demonstration was in the DIVE system [5] or in the real-time 3D reconstruction system of Ohya et al. [28].

One of the challenges with 2D video avatar representations is that eye gaze directions are not faithfully preserved so that multiple viewers might infer subtly different gaze behaviour [26]. This then might reduce the types of cue that viewers can rely on to interpret their collaborator's non-verbal behaviour [39]. Moving into AR or VR for a single viewer does allow the system to alleviate some of the effects, by, for example, facing the video avatar towards the user. In our experimental system described later, video avatar will directly face the experiment participant, so as to alleviate this problem. This also means that we do not, in this work, deal with the need for volumetric or other 3D representations.

## 2.3 Trust in Telecommunications

Developing trust is an important part of social interaction [24]. In face-to-face interactions, nonverbal cues, encompassing both visual cues like eye contact and head nods, and audio cues such as pitch, hold significant influence over how trustworthiness is perceived. These cues not only provide insights into an individual's background, including their education and origins, but also reveal intrinsic qualities like sincerity and confidence [43]. Several theories, including social presence theory [9, 27], cues-filtered-out theory [42], media richness theory [18], and social information processing theory [41], have explored the role of communication bandwidth and nonverbal cues. There is a general consensus that the loss or degradation of nonverbal channels in media affects users' perception of trustworthiness.

With video avatars, the most obvious cause of loss in nonverbal channels is the quality of the video. Horn et al. found a nonlinear relationship between quality of video and user performance at lie detection performance [16]. While some video degradation impaired performance, performance improved with low quality video suggesting that users were paying more attention to the audio. Bekkering & Shim investigated the effect of eye contact in video-mediated communication on trust [4]. Videos that did not support eye contact resulted in lower perceived trust than videos that enabled eye contact. Nguyen & Canny showed that in group conferencing situations, the spatial arrangement of the conferencing environment affected trust in videoconferencing [26]. They reported that gaze support and awareness were the main influencing factors on trust.

One source of potential bias in telecommunications systems is the relative scale and height of representations of users. In the second of a seminal pair of studies about the Proteus Effect, Yee & Bailensen showed that avatars represented as taller acted more confidently [44]. Huang et al. found that if a video conferencing system makes a person look relatively taller, that person has more social impact [17]. In our own experiments we will make sure that apparent height is the same by matching agent eye-lines.

A number of studies have compared the response to video or 3D modelled avatars with mixed results. Bente et al. found that video and avatars were treated similarly, but this was in a task where the users were in pairs and just facing each other in a fairly constrained situation [6]. Abdullah et al. compared multi-way video conferencing in a manner similar to typical Zoom or Teams-style arrangement of windows, against embodied avatar representations in virtual reality [1]. They note that users with embodied avatars had different patterns of communication than video conferencing, in particular more social connection maintenance was seen in video.

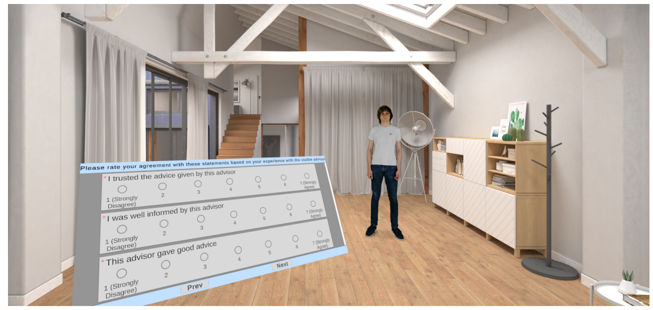


**Figure 2: The experiment environment at the questionnaire stage. The participant is being asked to evaluate a full-body-video advisor**

Similarly, in a study comparing avatars, video and robotic embodiment, Pan & Steed found that avatars were perceived as worse than video or robot embodiment [29]. Appearance of agents may not be the only factor; Torre et al. found that altering outward cues of emotions had only a small impact on trust when compared with differences in the agent's behaviour [40].

A number of papers have explored trust by examining how users follow advice given [21, 29, 30, 32]. We will build upon this work in our experiment. But the general idea is that rather than questioning users about their attitudes towards or beliefs about other users, we measure a user's objective responses to them by asking them to follow advice.

# 3 METHOD

We performed two experiments on the topic of trust in agent representation. The experimental design was influenced by prior work [21, 29, 30, 32]. Our experiments were designed in part to determine whether the findings from these earlier works could be replicated with current technology in an immersive context. The two experiments (Online and In-Person) have slightly different designs that build on the core described in this section. See Section 4 for these differences.

## 3.1 Experiment Design

*3.1.1 Protocol.* Participants were placed in a virtual environment and asked a series of difficult trivia questions (13 for the Online experiment, 32 for the In-Person experiment). For each question, participants could ask one of two advisors for help (optionally one advisor in the Online experiment, exactly one in the In-Person experiment). Participants then selected an answer and were presented with the next question. Participants could select advisors and answers by pointing their controller and pulling the trigger. The environment, position of advisors and question panel can be seen in Figures 1 and 2. Advisors were placed 2m away at a 15 degree angle from the participant.

After the final question, a screen popped up to inform the participant that the experiment was now over. Participants were then asked to fill out three questionnaires in VR. Participants answered the questionnaire in the same environment as the trivia quiz, with the same interaction mechanism. All questions were on a seven point Likert scale. First, the leftmost advisor was shown, and a questionnaire panel appeared with 6 questions on trust (e.g., "I was well informed by this adviser"). Second, both advisors were hidden, and the participant answered three questions from the Slater-Usoh-Steed presence questionnaire [36]. Finally, the rightmost advisor was shown, and the participant answered the same questions again for this advisor. The questionnaire can be seen in Figure 2.

The experiment was then concluded, and participants were given a small financial compensation. Note that exact reward structure varied for the Online and In-Person studies. See Section 4 for more details. Both protocols were approved by the University College London Research Ethics Committee.

*3.1.2 Conditions.* We aimed to compare 3 representations for advisors:

- **FB** Full-body video, as if the advisor was present in the virtual room (on the right in Figure 2)
- **HS** Head-and-shoulders video, on a television in the virtual room (on the left in Figure 1)
- **3D** A 3D model, as if the advisor was using the model as an avatar (on the right in Figure 1)

Each advisor would have one of three representations, and participants would always have two advisors with different representations. One advisor spoke an 'expert' script. The other spoke the 'non-expert' script. This gave us six sets of data (expert in bold): **FB** vs HS , FB vs **HS**, **FB** vs 3D, FB vs **3D**, **HS** vs 3D, and 3D vs **HS**. To avoid small cell counts we group conditions based on the expert, giving us three conditions: FB-Expert, HS-Expert and 3D-Expert.

*3.1.3 Variants.* Though we only had three conditions, we wanted to control for a variety of other factors may have had a confounding effect on our results. These factors were: opposing representation (e.g., FB vs HS, FB vs 3D), side of the expert (left or right), and actor portraying the expert (see Section 3.2.2). This leaves us with 24 variants. To preserve statistical power, we assigned variants to participants sequentially so that each condition had an equal balance of each, but did not otherwise include variants in analysis.

*3.1.4 Questions.* To compare our results with the prior works, we used a list of questions from the literature [29]. In prior work the first two questions were considered practice questions and not included in the analysis. We instead include all questions for analysis as we were particularly interested to find out whether representations would have an impact on participants' initial advisor selection, as they had not yet had a chance to interact. Answers were 4-option multiple choice. The ordering of the options was randomised between participants. To minimise the risk of participants already knowing the answer to the question, we have specifically chosen questions that are regarded as difficult, as demonstrated in the prior work.

*3.1.5 Expert vs Non-Expert Scripts.* For every participant, one advisor would read the expert script and one the non-expert. This variation in responses was to give the participant a reason to try both advisors. In our analysis, we can also use this to weigh the impact of behaviour against representation for advisors. The expert script answered 81% of questions correctly, while the non-expert answered 40% correctly. These are the same counts as in [32], but with the practice questions included. As in [32], the expert was correct much more frequently than the non-expert, but was not flawless. This was to avoid the expert seeming artificially perfect, and to keep the participant trying both advisors.

When correct, actors delivered the text confidently, and the text itself was kept short and simple (e.g., "The term was coined by Gadamer"). When incorrect, actors would sound and act unsure, and would give uncertain advice ("Not at all sure, but my guess is Frank L Wright?"). Note that this was true regardless of expertise; for the six incorrect answers given by the expert, for example, the performance and text was uncertain.

## 3.2 Experiment Setup

*3.2.1 Application.* The study application was made with the Unity game engine and was compatible with Meta Quest 1 and 2 headsets. Care was taken to ensure the application was able to reach the maximum frame rate of each device, and was free from stutters or hitches. The application communicated with a server to distribute conditions and to store results. Upon contacting the server, the application was issued a five digit unique user identifier (UUID) and a variant. This variant was temporarily claimed on the server such that it would no longer be issued to connecting applications. If the study was not completed after 24 hours, the claim would expire, and the server would re-issue the condition. This system was designed to allow multiple participants to complete the study at the same time, without duplicating variants.

*3.2.2 Representation Setup.* Performances for the representations were captured by recording video of two actors against a green screen. We used two actors so that in variants where both actors were visible, (e.g., FB vs HS), participants did not see two identical advisors, which may have seemed uncanny. Actors were intentionally selected to be similar in appearance. Both actors were male, and both actors were in their 20s and approximately 180cm tall. The first actor can be seen on the left in Figure 1, and the second on the right in Figure 2. Each actors was given both the 'expert' and 'non-expert' script, and therefore recorded two responses for each question. For both advisors, when they are playing an expert advisor role, they spoke with confidence with a decisive language. On the other hand, when they were playing a non-expert advisor role, they have said in a hesitant tone, often with more delays between phrases.

Each actor recorded two responses for each question, one for the 'expert' script, one for 'non-expert', giving us 4 videos total. The videos were recorded at 4K resolution and 30fps with a Xiaomi Redmi 4X. The camera was positioned approximately 2m away to match the distance from the participant to the representation in the virtual environment. The green screen background was removed with Adobe Premiere. The timestamps for the response to each question were stored in the study application, so the correct timestamp could be jumped to on-demand. Basic 'idle' behaviour was approximated by looping the 4-5 seconds of video captured between the actor's responses to provide blinks and small movements. To support the impression of an unbroken video stream, transitions were masked by writing the video frame before jumping to texture, and the post-jump video was blended with the pre-jump texture over the course of 300ms.

FB and HS used the audio and video performances of the actors. For HS, the video was cropped to the head and shoulders. 3D only used the audio performance. The FB and 3D representations were turned to face the participant in the transverse plane. This was to give the sense that they were looking at and talking directly to the participant. The model for the 3D condition was sourced from the Rocketbox library [15]. We selected the model as it was a good match for the actors, appearing male and approximately the same age and height. It did, however, differ from the actors in hair colour and clothing. We applied a looping 'idle' animation to the model to prevent it appearing unnaturally static. Unlike the other representations, this movement was not linked to the performance or current question. To give the impression of speech from the model, we used a simple lipsync library to generate lip motions from just the audio performances.

## 3.3 Outcome Measures

*3.3.1 EASR.* The primary outcome measure is the participant's Expert-Advice-Seeking-Rate (EASR). This is calculated by dividing the number of times the participant asked for advice from the expert by the number of opportunities the participant had to do so. For example, if a participant asked the expert for advice four times out of five questions, they would have an EASR of 0.8 over that window.

*3.3.2 Self-Reported Trust.* These are the responses to the post-study questionnaire on trust, collected per-advisor. The trust score is calculated by averaging the responses to the 6 questions participants were asked about the advisor. As responses were recorded on a seven point Likert scale, the trust score is between 1 and 7.

*3.3.3 Slater-Usoh-Steed Presence Score.* These are the responses to the Slater-Usoh-Steed (SUS) post-study questionnaire on presence [36]. For this paper, we calculate a SUS score by averaging the responses to the 3 questions participants were asked about the experience as a whole. As responses were recorded on a 7 point Likert scale, the SUS score is between 1 and 7.

*3.3.4 First Question.* This is the EASR for the first question. As this occurs before participants have yet had a chance to interact with advisors, we are interested to see if it reveals bias towards a particular representation.

# 4 EXPERIMENTS

This section describes the protocol for the two experiments where this differs from the core protocol described in Section 3.

## 4.1 Online

*4.1.1 Participants.* We recruited participants by making the application available on online applications portals SideQuest[4] and XRDRN[5]. Of 118 downloads, 51 participants went on to complete the study. However, 15 did not select an advisor at any point in the study, and two only selected the same advisor for the entire study. These results were therefore entirely excluded from analysis and participant totals throughout this paper. The participant count was therefore 34 for the Online experiment. Due to the nature of an in-the-wild experiment, demographics information was not recorded, beyond requiring participants to be over 18. To ensure that there is no existing trust with the actors shown in the user study, eligibility for the user study also included a restriction on prior connection with the actors.

*4.1.2 Questions.* As Section 3.1.4, but rather than the full set of 32 questions, we selected 13. For these questions, the expert was always correct, and the non-expert always incorrect, which is closer to the ratios used in [29]. This change was on the basis that for an in-the-wild experiment, a shorter duration and more straightforward study would help maintain participant engagement.

*4.1.3 Reward.* Due to the difficulty in distributing vouchers internationally and the potential for abuse of the study system over the internet, we distributed the reward to participants in the Online experiment as a prize draw. After the participant completed the experiment, they could send their UUID and their email address to the experimenter using an online form. The experimenter confirmed the server records indicated the UUID had belonged to a record that had completed the study, and the participant was added to a list of possible winners. 6 prizes were awarded, with a total value of £200 (GBP).

*4.1.4 Procedure.* Participants discovered the application on an online portal, then downloaded and installed ('sideloaded') the application on their Meta Quest 1 or 2 device. From here the experiment proceeded using the protocol described in Section 3.1.1. After the participant completed the experiment, they could submit their UUID and email address on an online form to be included in the prize draw. The total experiment duration was approximately 10 minutes.

## 4.2 In-Person

*4.2.1 Participants.* We gathered participants through internal email advertisement at University College London. 48 participants (22 male, 26 female) completed the experiment. Two of these participants selected the same advisor for the entire study. As in Section 4.1.1 these participants were fully excluded from our analysis and participant counts, giving us 46 participants total.

*4.2.2 Questions.* The In-Person experiment used the full question list as described in Section 3.1.4.

*4.2.3 Reward.* As participants were physically present, we gave a reward rather than the lottery system used in the Online study. We also reintroduced variable reward from [29]. The total reward available was between £6 and £12. Participants were given a base reward of £6, plus around £0.10 per question (summing to £3 should participants answer every question correctly). Another £3 was provided for the correct answer to a 'high stakes' final question, as in the prior studies [29, 32].

*4.2.4 Two-Alternative Forced Advisor Choice.* Motivated by the large number of participants who selected no advisor in the Online study (see Section 4.1.1), for the In-Person study we implemented advisor selection as a two-alternative-forced-choice. Participants

[4]https://sidequestvr.com
[5]https://xrdrn.org/

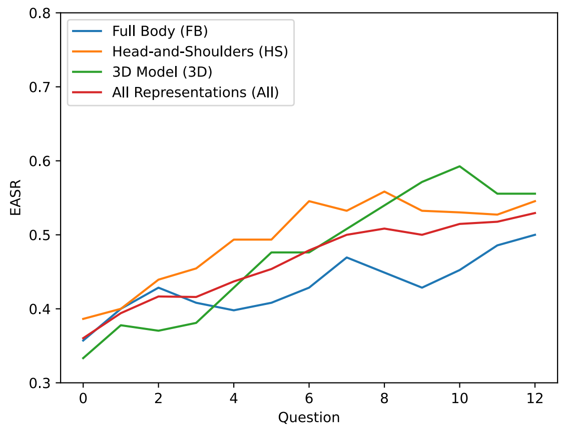


**Figure 3: Online expert-advice-seeking-rate (EASR) by question. An increasing EASR over time suggests participants were able to identify which of the advisors was the expert. Note these are smoothed with a window around the question (n=7). Individual responses were noisy as participants tended to alternate between advisors.**

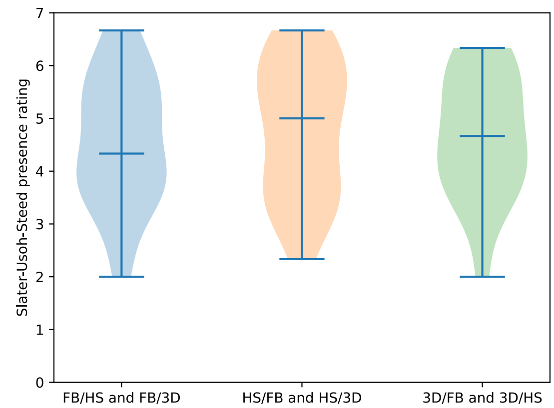


**Figure 4: Online responses to the Slater-Usoh-Steed presence questionnaire.**

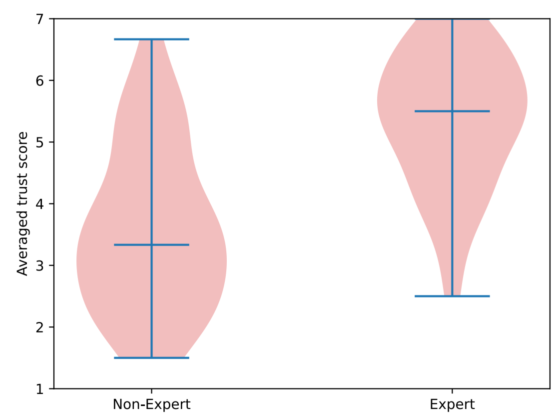


**Figure 5: Online self-reported trust scores, averaged across all representations.**

were thereby required to select an advisor before moving on from a question.

*4.2.5 Procedure.* Participants came to our lab and were equipped with a Meta Quest 2 head-mounted display. The experiment then proceeded following the main protocol described in Section 3.1.1, with the addition of the 2-alternative forced choice of advisor (see Section 4.2.4). At the end the study, after all questionnaires were complete, the experimenter calculated the total as described in Section 4.2.3. Finally the participant was informed of what they would receive, which was then distributed as an online voucher. The total experiment duration was approximately 25 minutes.

# 5 RESULTS

## 5.1 Online

*5.1.1 EASR over time.* As in Section 5.2, we assess the participant's EASR by combination of question and advisor. This measure can be seen in Figure 3. We follow the same process described above, grouping questions into sets of 6 and taking the average of each group. As we have 13 questions for the Online study, question 7 is discarded to avoid uneven group sizes. We then perform a mixed ANOVA with 1 between subjects factor (expert-representation: FB-Expert, HS-Expert or 3D-Expert) and 1 within subjects factor (question-group: 1to6, 8to13).

The results were normally distributed as assessed by Q-Q plot. There was homogeneity of variances and covariances, as assessed by Levene's and Box's test respectively ($p > .05$). The results were free of significant outliers.

There was no statistically significant 2-way interaction between the expert-representation and question-group on EASR, $F(2, 31) = 1.818$, $p = 0.179$, partial $\eta^2 = 0.105$. Expert-representation had no significant effect on EASR, $F(2,31) = 0.155$, $p = 0.857$. Question-group had a significant main effect on EASR between the 8to13 group (M = 0.515, SD = 0.291) and the 1to6 group (M = 0.417, SD = 0.198), $F(1, 31) = 6.974$, $p = 0.013$, partial $\eta^2 = 0.184$.

*5.1.2 Trust.* We evaluate the participant's self-reported trust score for each representation as described in Section 3.3.2. These are collected at the end of the experiment, so we should also consider advisor expertise. We perform a mixed ANOVA with 1 between subjects factor (expert-representation: FB-Expert, HS-Expert or 3D-Expert) and 1 within subjects factor (expertise: expert or non-expert).

The results were approximately normally distributed as assessed by Q-Q plot. As some of the plots were irregular, a Shapiro-Wilk test of normality was also conducted and the results confirmed to


**Figure 6: In-Person expert-advice-seeking-rate by question. As Figure 3, an increasing EASR over time suggests participants identified the expert advisor. Note these are smoothed with a window around the question (n=7). Individual responses were noisy as participants tended to alternate between advisors.**

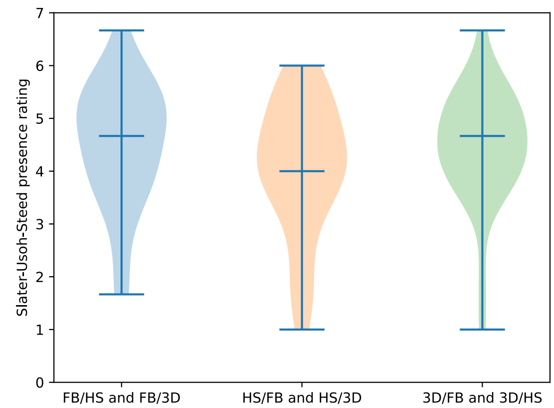


**Figure 7: In-Person responses to the Slater-Usoh-Steed presence questionnaire.**

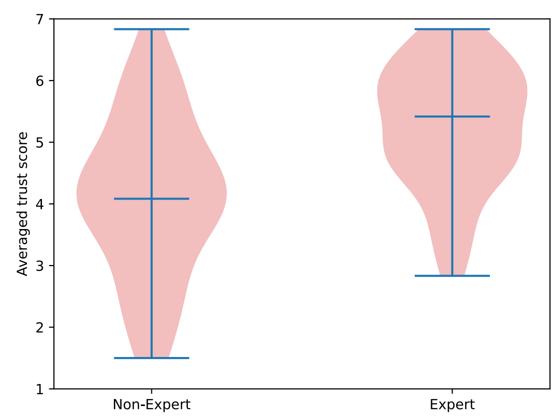


**Figure 8: In-Person self-reported trust scores, averaged across all representations**

be sufficiently normal ($p > 0.05$ for each cell). There was homogeneity of variances and covariances, as assessed by Levene's and Box's test respectively ($p > .05$). The results were free of significant outliers. Mauchly's test of sphericity indicated that the assumption of sphericity was violated for the two-way interaction, $p = 0.001$, so values below use Greenhouse-Geisser adjustment.

There was no statistically significant 2-way interaction between the expert-representation and expertise on trust score, $F(2, 31) = 1.252$, $p = 0.300$, partial $\eta^2 = 0.075$. Expert-representation had no significant main effect on trust score, $F(2,31) = 0.519$, $p = 0.600$. Expertise had a significant main effect on trust scores between the expert group ($M = 5.338$, $SD = 1.104$) and the non-expert group ($M = 3.593$, $SD = 1.335$), $F(1, 31) = 38.975$, $p < 0.001$, partial $\eta^2 = 0.557$ (see Figure 5).

*5.1.3 Presence.* Responses to the Slater-Usoh-Steed (SUS) presence questionnaire as described in Section 3.3.3. Figure 4 contains responses for the Online study.

*5.1.4 First Question.* We assess for the impact of the participant's first visual response to the advisors by comparing EASR for the first question alone. Of the 34 participants, the proportions that asked the expert were 4 of 9 (44.4%) in 3D-Expert, 5 of 14 (35.7%) in FB-Expert, and 3 of 11 (27.3%) in HS-Expert. For this measure, we tested for significance with Fisher's test rather than the typical Chi-Square. This is because in this case, 66% of the resulting 2 x c table had expected counts greater than or equal to 5, below the typical threshold of 80% used for Chi-Square tests. The result was a non-statistically significant difference in proportions as assessed by Fisher's exact test, $p = 0.823$.

## 5.2 In-Person

*5.2.1 EASR over time.* This measure can be seen in Figure 6. We perform our tests as described in Section 5.1.1, but as we have 32 questions, questions 13 and 20 are instead discarded. We then perform a mixed ANOVA with 1 between subjects factor (expert-representation: FB-Expert, HS-Expert or 3D-Expert) and 1 within subjects factor (question-group: 1to6, 7to12, 14to19, 21to26 or 27to32).

The results were approximately normally distributed as assessed by Q-Q plot. There was homogeneity of variances and covariances, as assessed by Levene's and Box's test respectively ($p > .05$). The results were largely free of significant outliers, with the exception of the 3D condition in the 1to6 group, which contained 2: a result of 1 and a result of 0. We leave these as-is, on the basis that they are present only in 1 of the 15 cells, and taken together will not affect skew. Mauchly's test of sphericity indicated that the assumption of sphericity was met for the two-way interaction, $\chi^2(9) = 14.99$, p = 0.091.

There was no statistically significant 2-way interaction between the expert-representation and question-group on EASR, F(8, 172) = 0.484, p = 0.867, partial $\eta^2 = 0.022$. Expert-representation had no significant main effect on EASR, F(2,43) = 0.273, p = 0.762. Question-group had a significant main effect on EASR, F(4, 172) = 3.144, p = 0.016, partial $\eta^2 = 0.068$. Looking further into pairwise interactions, 27to32 (M = 0.641, SD = 0.232) was significantly higher than 1to6 (M = 0.536, SD = 0.201), p = 0.044 (including Bonferroni adjustments for multiple comparisons).

*5.2.2 Trust.* As in Section 5.1.2, we perform a mixed ANOVA with 1 between subjects factor (expert-representation: FB-Expert, HS-Expert or 3D-Expert) and 1 within subjects factor (expertise: expert or non-expert).

The results were normally distributed as assessed by Q-Q plot. There was homogeneity of variances and covariances, as assessed by Levene's and Box's test respectively (p > .05). The results were free of significant outliers.

There was no statistically significant 2-way interaction between the expert-representation and expertise on trust score, F(2,43) = 1.08, p = 0.348, partial $\eta^2 = 0.048$. Expert-representation had a significant main effect on trust score, F(2,43) = 9.01, p < 0.001, partial $\eta^2$ = 0.295. Pairwise interactions reveal significant differences between HS-Expert and FB-Expert (p < 0.001) and between HS-Expert and 3D-Expert (p = 0.011). Expertise had a significant main effect on trust between the expert group (M = 5.25, SD = 1.01) and the non-expert group (M = 4.12, SD = 1.299), F(1,43) = 21.684, p < 0.001, partial $\eta^2 = 0.335$ (see Figure 8).

*5.2.3 Presence.* Responses to the SUS presence questionnaire as described in Section 3.3.3. Figure 7 contains responses for the In-Person study.

*5.2.4 First Question.* As in Section 5.1.4. Of the 46 participants, the proportions that asked the expert were 7 of 16 (43.8%) in 3D-Expert, 11 of 14 (78.6%) in FB-Expert, and 6 of 16 (37.5%) in HS-Expert. This is a suggestive but ultimately non-statistically significant difference in proportions as assessed by the Pearson Chi-Square test, p = 0.057.

*5.2.5 Final Question.* To assess the impact of the 'high-stakes' question in the In-Person study, we compare EASR for this final question alone. Of the 46 participants, the proportions that asked the expert were 8 of 16 (50%) in 3D-Expert, 11 of 14 (78.6%) in FB-Expert, and 10 of 16 (62.5%) in HS-Expert. This is a non-statistically significant difference in proportions as assessed by the Pearson Chi-Square test, p = 0.270.

# 6 DISCUSSION

## 6.1 Trust and Representation

Representation had a comparatively small effect on EASR throughout. In both the Online and In-Person experiments, we found no significant effect of representation on EASR at any stage of the study. Additionally, observing Figures 6 and 3, we see little difference in EASR between representation, either initially or over time. We expected that participants' advisor selection for the very first question would be a good indicator of implicit preference, as it takes place before the participant has had a chance to interact with advisors and thereby develop an understanding of expertise. However, this too produced no significant impact based on representation. For this first question, the results did suggest a slight but non-significant preference towards FB-Expert for the In-Person study. This was not replicated in the Online study.

Participants' self-reported trust score was significantly affected by representation in the In-Person experiment. In this case the results were clear against HS-Expert, with significant differences between this condition and the other two conditions. There were no other significant pairwise differences. Supporting this, SUS presence scores appear lower in conditions involving HS (see Figure 7). However, despite these differences in self-reported trust and presence in the In-Person experiment, no difference was observed between representations in the Online experiment. The In-Person experiment is likely a better indicator, as the participant count was higher and conditions could be more carefully controlled.

This result somewhat supports prior work, with some interesting differences. Riegelsberger et al. [32] compared rich-media advisors in the form of audio, video and a 3D avatar with a text-only advisor when interacted with via computer screen. They found a significant difference between rich-media advisors over the text advisor (p < 0.001, effect size not reported), but only non-significant differences (p < 0.062) between rich-media advisors. As our advisors are rich-media, this aligns with our result. Their results were, however, still somewhat suggestive of less trust awarded to the avatar condition. Similarly, Pan [29] performed a similar study comparing a physical robot, and a 3D model and video on a computer screen, and also found participants appeared to award significantly less trust to the 3D model initially, and were significantly slower to trust it than the other representations.

As we replicate the work of Pan and Riegelsberger [29, 32] in an immersive context, the discrepancy here is of interest. While they observe less trust being attributed to their 3D model condition, our 3D-Expert condition scores similarly to the others, and our HS-Expert condition appears less popular. One explanation for this discrepancy could be the different environments in which these 3 experiments took place. [32] was on a computer screen, while [29] placed digital advisors on a computer screen and the robot physically adjacent to the participant. Head-and-shoulders video is otherwise a very common teleconferencing approach, but our study took place within a 3D virtual environment. Social VR typically uses 3D avatars; in this context, the 3D model belongs, but HS is incongruous. It is not commonly used for this purpose, so is unfamiliar. It also has many visual differences to the virtual environment in which it is placed, such as being captured at a different resolution and frame rate, not receiving lighting from the environment. It's possible that participants are responding in part to the coherence of the representation within the virtual environment, and that representations which do not appear to belong are on some level jarring to the user. However, it is worth noting that EASR was no lower for the HS condition, either through analysis of EASR or Figure 6. This suggests that while participants did not respond as positively to the HS advisor, they implicitly still trusted this representation to have good answers.

## 6.2 Trust and Expertise

Expertise had a significant effect on EASR, consistent across both the In-Person and Online experiments. In Figures 6 and 3, initial results suggest random selection: in the In-Person experiment, EASR starts at approximately 0.5, while Online, initial EASR is around 0.35, as here the participants had three options (one advisor, the other advisor, or no advisor at all). The following upward trend in these figures indicates participants were then able to identify the expert advisor and seek advice from them preferentially. We also found a statistically significant effect of question number on EASR between the first and last groups in the In-Person and Online experiments, confirming this trend of increased EASR over time.

Self reported trust scores were significantly affected by expertise in both the Online and In-Person experiments. These can be seen in Figures 8 and 5. As these scores were collected at the end of the experiment, participants had already spent time with the advisors and come to conclusions about expertise, so this is perhaps not surprising. Combined with the effect on EASR, though, this does suggest participants were aware of their decisions about advisors rather than it being the result of implicit bias.

This ability to identify the expert is notable. Participants were not informed when they had answered correctly or incorrectly until the end of the experiment. Participants were therefore required to observe the performance of each representation and allocate trust based on these performances. It is possible that a participant would already know the correct answer to a question, and therefore decide on an advisors expertise based on their own knowledge. However, the questions posed were challenging and on a wide variety of subjects [29]. To further avoid this possibility, the expert advisor script also included incorrect and uncertain answers to a few questions, so trust is built only in aggregate. As we found no significant differences between EASR by representation (see Section 6.1), or any confounding 2-way interactions, we can conclude that all representations were sufficient to impart these performances and expertise information to participants at a similar rate.

This replicates the findings of [29] and [32]. In [29] we see a very similar trend by question. [32] does not break down EASR by question, but compares overall ASR and finds it significantly higher for the expert. Our findings are effectively the same here; regardless of representation, participants were able to effectively identify expertise by the end of the experiment, and preferentially sought advice from the more expert advisor.

## 6.3 Online and In-Person

In the Online study, 17 of 51 participants never selected an advisor and were fully excluded from all analysis, as described in Section 4.1.1. This was permitted under the Online experiment study design, but made the results hard to interpret; it's possible these participants did not ask for advice intentionally, but it's more likely that they were either disengaged with the study, did not understand the task, or associated asking for advice with some negative repercussion. These issues likely arise at least in part from attempting to replicate lab studies in an in-the-wild context [25, 37].

We made several changes to the In-Person design informed by the above issues, described in Section 4.2. Despite these changes, however, we find only small differences between the results of the two experiments overall. In both the Online and In-Person experiments, representation had no significant effect on EASR, overall, per-question group, or the first or final question. The high-stakes question in the In Person study therefore likely did not have a strong effect on advisor selection. Despite the increased incentive and greater question count, we did not observe a greater improvement of EASR in the In-Person experiment over the Online experiment; on the contrary, the effect size was slightly smaller. One theory would be that making advisor selection optional helped to filter out participants that were not fully engaged, as those that were engaged sought the advisor without incentive in the Online experiment.

## 6.4 Limitations

This experiments described in this paper are an attempt to compare representations of users in mixed reality. As mixed-reality hardware uptake remains low, gathering a suitable number of participants for in-the-wild experiments is challenging. We therefore simulate the above in a virtual reality context, with agents rather than avatars. Our results should be considered with this in mind. Additionally, while animations for the FB and HS conditions were based on actor movement and varied for each question, the 3D model used a simple idle animation for the duration of the experiment. Motion-captured animation may have been a fairer test. This condition did, however, still perform well in our experiments.

# 7 CONCLUSION

In this paper we reported two experiments on trust in agent representations. Our results suggested a strong effect of perceived agent expertise on trust, across both experiments and all agent representations. We also found some evidence that in our In-Person experiment, full-body video prompted more trust than the head-and-shoulders video more typical in teleconferencing applications. Unlike in prior studies, the 3D model performed well on trust, equivalent to full-body video and significantly better than head-and-shoulders video. This could be a reflection of the immersive nature of the study.

While our intention was to simulate mixed reality teleconferencing with these experiments, our results do indicate that the context and environment surrounding the agent has a significant effect. An interesting avenue for future work would therefore be to conduct the same experiment in a head-mounted augmented reality display. Here the 3D model may once again be perceived as incongruent, and we may see different results. The lower trust apportioned to the head-and-shoulders view is also interesting as this is currently the most commonly available set-up for teleconferencing. Keeping the head-and-shoulders within a frame was a choice designed to emphasise familiarity, but it may be worth exploring other ways of presenting video.

# ACKNOWLEDGMENTS

This project has received funding from UCL and from the European Union's Horizon 2020 Research and Innovation program under grant agreement No 739578 (RISE).

# REFERENCES


[1] Ahsan Abdullah, Jan Kolkmeier, Vivian Lo, and Michael Neff. 2021. Videoconference and Embodied VR: Communication Patterns Across Task and Medium. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (Oct. 2021), 453:1–453:29. https://doi.org/10.1145/3479597

[2] Jeremy N. Bailenson. 2021. Nonverbal overload: A theoretical argument for the causes of Zoom fatigue. *Technology, Mind, and Behavior* 2, 1 (Feb. 2021). https://doi.org/10.1037/tmb0000030

[3] Jeremy N Bailenson, Nick Yee, Dan Merget, and Ralph Schroeder. 2006. The effect of behavioral realism and form realism of real-time avatar faces on verbal disclosure, nonverbal disclosure, emotion recognition, and copresence in dyadic interaction. *Presence: Teleoperators and Virtual Environments* 15, 4 (2006), 359–372. Publisher: MIT Press.

[4] Ernst Bekkering and JP Shim. 2006. Trust in videoconferencing. *Commun. ACM* 49, 7 (2006), 103–107.

[5] Steve Benford, John Bowers, Lennart E. Fahlén, Chris Greenhalgh, and Dave Snowdon. 1995. User embodiment in collaborative virtual environments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '95)*. ACM Press/Addison-Wesley Publishing Co., USA, 242–249. https://doi.org/10.1145/223904.223935

[6] Gary Bente, Sabine Rüggenberg, Nicole C. Krämer, and Felix Eschenburg. 2008. Avatar-Mediated Networking: Increasing Social Presence and Interpersonal Trust in Net-Based Collaborations. *Human Communication Research* 34, 2 (April 2008), 287–318. https://doi.org/10.1111/j.1468-2958.2008.00322.x

[7] Mark Billinghurst and Hirokazu Kato. 2002. Collaborative augmented reality. *Commun. ACM* 45, 7 (July 2002), 64–70. https://doi.org/10.1145/514236.514265

[8] M. Billinghurst, S. Weghorst, and T. Furness. 1998. Shared space: An augmented reality approach for computer supported collaborative work. *Virtual Reality* 3, 1 (March 1998), 25–36. https://doi.org/10.1007/BF01409795

[9] Frank Biocca, Chad Harms, and Judee K. Burgoon. 2003. Toward a More Robust Theory and Measure of Social Presence: Review and Suggested Criteria. *Presence: Teleoperators and Virtual Environments* 12, 5 (Oct. 2003), 456–480.

[10] Nathan Bos, Judy Olson, Darren Gergle, Gary Olson, and Zach Wright. 2002. Effects of four computer-mediated communications channels on trust development. In *SIGCHI*. ACM, 135–140.

[11] Carmen Egido. 1988. Video conferencing as a technology to support group work: a review of its failures. In *Proceedings of the 1988 ACM conference on Computer-supported cooperative work (CSCW '88)*. Association for Computing Machinery, New York, NY, USA, 13–24. https://doi.org/10.1145/62266.62268

[12] Barrett Ens, Joel Lanir, Anthony Tang, Scott Bateman, Gun Lee, Thammathip Piumsomboon, and Mark Billinghurst. 2019. Revisiting collaboration through mixed reality: The evolution of groupware. *International Journal of Human-Computer Studies* 131 (Nov. 2019), 81–98. https://doi.org/10.1016/j.ijhcs.2019.05.011

[13] Guo Freeman and Divine Maloney. 2021. Body, Avatar, and Me: The Presentation and Perception of Self in Social Virtual Reality. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW3 (Jan. 2021), 239:1–239:27. https://doi.org/10.1145/3432938

[14] Joseph L Gabbard, J Edward Swan, and Deborah Hix. 2006. The effects of text drawing styles, background textures, and natural lighting on text legibility in outdoor augmented reality. *Presence* 15, 1 (2006), 16–32.

[15] Mar Gonzalez-Franco, Eyal Ofek, Ye Pan, Angus Antley, Anthony Steed, Bernhard Spanlang, Antonella Maselli, Domna Banakou, Nuria Pelechano, Sergio Orts-Escolano, Veronica Orvalho, Laura Trutoiu, Markus Wojcik, Maria V. Sanchez-Vives, Jeremy Bailenson, Mel Slater, and Jaron Lanier. 2020. The Rocketbox Library and the Utility of Freely Available Rigged Avatars. *Frontiers in Virtual Reality* 0 (2020). https://doi.org/10.3389/frvir.2020.561558

[16] Daniel B Horn, Lana Karasik, and Judith S Olsen. 2002. The effects of spatial and temporal video distortion on lie detection performance. In *SIGCHI*. ACM, 714–715.

[17] Wei Huang, Judith S Olson, and Gary M Olson. 2002. Camera angle affects dominance in video-mediated communication. In *CHI'02 Extended Abstracts on Human Factors in Computing Systems*. ACM, 716–717.

[18] Kumi Ishii, Mary Madison Lyons, and Sabrina A. Carr. 2019. Revisiting media richness theory for today and future. *Human Behavior and Emerging Technologies* 1, 2 (2019), 124–131. https://doi.org/10.1002/hbe2.138 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/hbe2.138.

[19] Anya Kolesnichenko, Joshua McVeigh-Schultz, and Katherine Isbister. 2019. Understanding Emerging Design Practices for Avatar Systems in the Commercial Social VR Ecology. In *Proceedings of the 2019 on Designing Interactive Systems Conference (DIS '19)*. Association for Computing Machinery, New York, NY, USA, 241–252. https://doi.org/10.1145/3322276.3322352

[20] Marc Erich Latoschik, Daniel Roth, Dominik Gall, Jascha Achenbach, Thomas Waltemate, and Mario Botsch. 2017. The effect of avatar realism in immersive social virtual realities. In *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology (VRST '17)*. Association for Computing Machinery, New York, NY, USA, 1–10. https://doi.org/10.1145/3139131.3139156

[21] Ian Li, Jodi Forlizzi, Anind Dey, and Sara Kiesler. 2007. My agent as myself or another: effects on credibility and listening to advice. In *Proceedings of the 2007 conference on Designing pleasurable products and interfaces*. ACM, 194–208.

[22] Nadia Magnenat-Thalmann and Daniel Thalmann. 2005. *Handbook of Virtual Humans*. John Wiley & Sons. Google-Books-ID: HSfi4r2Gf9IC.

[23] Bernardo Marques, Samuel Silva, João Alves, Tiago Araújo, Paulo Dias, and Beatriz Sousa Santos. 2022. A Conceptual Model and Taxonomy for Collaborative Augmented Reality. *IEEE Transactions on Visualization and Computer Graphics* 28, 12 (Dec. 2022), 5113–5133. https://doi.org/10.1109/TVCG.2021.3101545 Conference Name: IEEE Transactions on Visualization and Computer Graphics.

[24] Roger C Mayer, James H Davis, and F David Schoorman. 1995. An integrative model of organizational trust. *Academy of management review* (1995), 709–734.

[25] Aske Mottelson and Kasper Hornbæk. 2017. Virtual reality studies outside the laboratory. In *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology (VRST '17)*. Association for Computing Machinery, New York, NY, USA, 1–10. https://doi.org/10.1145/3139131.3139141

[26] David T Nguyen and John Canny. 2007. Multiview: improving trust in group video conferencing through spatial faithfulness. In *SIGCHI*. ACM, 1465–1474.

[27] Catherine S. Oh, Jeremy N. Bailenson, and Gregory F. Welch. 2018. A Systematic Review of Social Presence: Definition, Antecedents, and Implications. *Frontiers in Robotics and AI* 5 (2018). https://doi.org/10.3389/frobt.2018.00114 Publisher: Frontiers.

[28] J. Ohya, Y. Kitamura, H. Takemura, F. Kishino, and N. Terashima. 1993. Real-time reproduction of 3D human images in virtual space teleconferencing. In *Proceedings of IEEE Virtual Reality Annual International Symposium*. 408–414. https://doi.org/10.1109/VRAIS.1993.380751

[29] Ye Pan and Anthony Steed. 2016. A Comparison of Avatar-, Video-, and Robot-Mediated Interaction on Users' Trust in Expertise. *Frontiers in Robotics and AI* 3 (2016). https://www.frontiersin.org/articles/10.3389/frobt.2016.00012

[30] Ye Pan, William Steptoe, and Anthony Steed. 2014. Comparing flat and spherical displays in a trust scenario in avatar-mediated interaction. In *SIGCHI*. ACM, 1397–1406.

[31] Irene Rae, Leila Takayama, and Bilge Mutlu. 2013. In-body Experiences: Embodiment, Control, and Trust in Robot-Mediated Communication. *interaction* 15, 28 (2013), 36.

[32] Jens Riegelsberger, M Angela Sasse, and John D McCarthy. 2006. Rich media, poor judgement? A study of media effects on users' trust in expertise. In *People and Computers XIX—The Bigger Picture*. Springer, 267–284.

[33] Ralph Schroeder. 2001. *The Social Life of Avatars: Presence and Interaction in Shared Virtual Environments*. Springer Science & Business Media.

[34] Alexander Schäfer, Gerd Reis, and Didier Stricker. 2022. A Survey on Synchronous Augmented, Virtual, andMixed Reality Remote Collaboration Systems. *Comput. Surveys* 55, 6 (Dec. 2022), 116:1–116:27. https://doi.org/10.1145/3533376

[35] Sandeep Singhal and Michael Zyda. 1999. *Networked virtual environments: design and implementation*. Addison-Wesley, Reading, MA.

[36] Mel Slater, Martin Usoh, and Anthony Steed. 1994. Depth of Presence in Virtual Environments. *Presence: Teleoperators and Virtual Environments* (1994). https://doi.org/10.1162/pres.1994.3.2.130

[37] Anthony Steed, Sebastian Frlston, Maria Murcia Lopez, Jason Drummond, Ye Pan, and David Swapp. 2016. An 'In the Wild' Experiment on Presence and Embodiment using Consumer Virtual Reality Equipment. *IEEE Transactions on Visualization and Computer Graphics* 22, 4 (April 2016), 1406–1414. https://doi.org/10.1109/TVCG.2016.2518135 Conference Name: IEEE Transactions on Visualization and Computer Graphics.

[38] Anthony Steed and Manuel Oliveira. 2009. *Networked Graphics: Building Networked Games and Virtual Environments*. Elsevier. Google-Books-ID: 76C_quJqVXcC.

[39] William Steptoe, Anthony Steed, Aitor Rovira, and John Rae. 2010. Lie tracking: social presence, truth and deception in avatar-mediated telecommunication. In *SIGCHI*. ACM, 1039–1048.

[40] Ilaria Torre, Emma Carrigan, Rachel McDonnell, Katarina Domijan, Killian McCabe, and Naomi Harte. 2019. The Effect of Multimodal Emotional Expression and Agent Appearance on Trust in Human-Agent Interaction. In *Proceedings of the 12th ACM SIGGRAPH Conference on Motion, Interaction and Games* (Newcastle upon Tyne, United Kingdom) *(MIG '19)*. Association for Computing Machinery, New York, NY, USA, Article 14, 6 pages. https://doi.org/10.1145/3359566.3360065

[41] Joseph B. Walther. 2015. Social Information Processing Theory (CMC). In *The International Encyclopedia of Interpersonal Communication*. John Wiley & Sons, Ltd, 1–13. https://doi.org/10.1002/9781118540190.wbeic192 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/9781118540190.wbeic192.

[42] Joseph B Walther and Malcolm R Parks. 2002. Cues filtered out, cues filtered in: Computer-mediated communication and relationships. *Handbook of interpersonal communication* 3 (2002), 529–563.

[43] Ederyn Williams. 1977. Experimental comparisons of face-to-face and mediated communication: A review. *Psychological Bulletin* 84, 5 (1977), 963.

[44] Nick Yee and Jeremy Bailenson. 2007. The Proteus Effect: The Effect of Transformed Self-Representation on Behavior. *Human Communication Research* 33, 3 (July 2007), 271–290. https://doi.org/10.1111/j.1468-2958.2007.00299.x

[45] Boram Yoon, Hyung-il Kim, Gun A. Lee, Mark Billinghurst, and Woontack Woo. 2019. 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)*. 547–556. https://doi.org/10.1109/VR.2019.8797719 ISSN: 2642-5254.