Nice is Different than Good: Longitudinal Communicative Effects of Realistic and Cartoon Avatars in Real Mixed Reality Work Meetings

Georgiana Cristina Dobre
Goldsmiths, University of London
London, United Kingdom
c.dobre@gold.ac.uk

Marta Wilczkowiak
Microsoft Research
Cambridge, United Kingdom
mawilczk@microsoft.com

Marco Gillies
Goldsmiths, University of London
London, United Kingdom
m.gillies@gold.ac.uk

Xueni Pan
Goldsmiths, University of London
London, United Kingdom
x.pan@gold.ac.uk

Sean Rintel
Microsoft Research
Cambridge, United Kingdom
serintel@microsoft.com

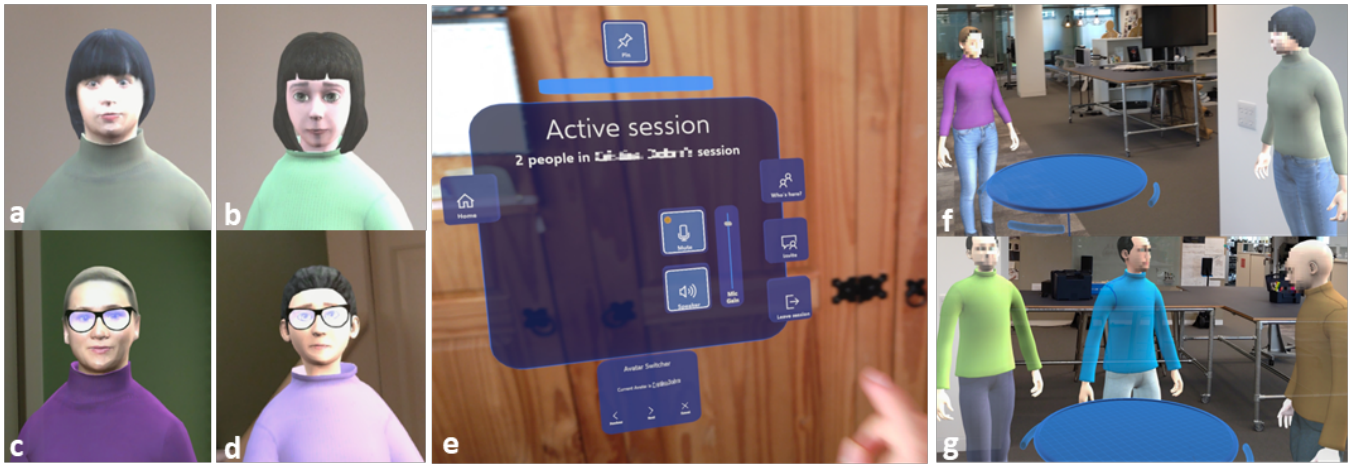

Figure 1: Example of realistic (a,c) and cartoon (b,d) avatar upper bodies. The main menu controlling the session (e). A meeting between two participants in realistic avatars (f) and three in cartoon avatars (g) with the adjustable blue table marking the centre.

ABSTRACT

We report a within-subjects study of the effect of realistic and cartoon avatars on communication, task satisfaction, and perceived sense of presence in mixed reality meetings. For 2 – 3 weeks, six groups of co-workers (14 people) held a recurring real work meeting using Microsoft HoloLens2 devices. Each person embodied a personalised full-body avatar with a realistic face and another with a cartoon face. Half the groups started in the realistic condition and the other half started in the cartoon condition; all groups switched conditions half-way. Initial results show that, overall, participants found the realistic avatars' nonverbal behaviour more appropriate for the interaction and more useful for understanding their colleagues compared to the cartoon one. Regarding the results over time, we identify different insights for cartoon and realistic avatars based on the type of avatar was embodied first. We discuss the implications of these results for mixed and virtual reality meetings.

KEYWORDS

avatar style, mixed reality, meetings, longitudinal, real-world

CHI '22 Extended Abstracts, April 29-May 5, 2022, New Orleans, LA, USA

ACM ISBN 978-1-4503-9156-6/22/04.
https://doi.org/10.1145/3491101.3519628

ACM Reference Format:
Georgiana Cristina Dobre, Marta Wilczkowiak, Marco Gillies, Xueni Pan, and Sean Rintel. 2022. Nice is Different than Good: Longitudinal Communicative Effects of Realistic and Cartoon Avatars in Real Mixed Reality Work Meetings. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts (CHI '22 Extended Abstracts), April 29-May 5, 2022, New Orleans, LA, USA.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3491101.3519628

1 INTRODUCTION

As virtual 'metaverses' expand from games to embrace potentially all digital engagement, the level of avatar realism could have a great impact on work meetings. The value of mixed and virtual reality (MR/VR) meetings is the preservation of spatial relationships [20] and social behaviours such as proximity or gaze [2]. Avatars represent people's identity, position, interest and activity [5]. These may range from spheres with hands that float and move based on the user's movement, to full or partial bodies with generic cartoon styling to highly customised realistic styling.

A common concern for realistic avatars is that they may trigger a mismatch between high expectations and delivery of nonverbal behaviour (i.e., movement, gesticulation, facial expressions), leading to the decreased user affinity and feelings of unease [22]. Cartoonish styling, whether generic or customized, may also lead users to be anxious about the appropriateness of non-realistic representation in a work context [1]. The majority of the research on avatars is focused on presence, workload or trust [9, 11, 13, 15, 24, 26], with mixed results [13, 27] (see Section 2). These topics are common because participants do not know each other before the study. Moreover, study participants often look only at short animations or still images of avatars [16, 22] and/or have one-off interactions with others [9, 10, 15, 24, 26, 28], making the findings prone to novelty effects [12, 21]. However, real-life collaborative work in immersive environments involves users who know each other and interact regularly, trying to get real work done. The *communicative functionality* of avatars, then, is essential. Since the spatial audio common to most immersive environments provides a highly naturalistic vocal representation, it is *nonverbal* communicative functionality that is primarily at issue, such as the ability to identify one another and then recognize facial expressions and gestures [6], negotiate proxemics [7], and, when presented virtually, to trust that these are authentic representations [19]. In sum, most of what we know about avatar appearance in meeting-style settings come from one-off lab studies in virtual reality environments. We know less about whether these findings apply to mixed reality, still less about effects in real-world contexts, and very little about the longitudinal effects on avatar acceptance. To our knowledge, there is a gap in virtual reality literature regarding this combination of aspects.

In this paper we address this gap by investigating how people feel about using full body avatars in real meetings, comparing two different styles of facial realism. For two to three weeks, six groups of co-workers (14 people) from a global technology company held a recurring real work meeting using Microsoft HoloLens2 devices. Each person embodied a personalised full-body avatar with a realistic face and another with a cartoon face. Half the groups started in the realistic condition and the other half started in the cartoon condition, and all groups switched conditions half-way through their study period. Ultimately, we were interested in whether various acceptance ratings for both realistic and cartoon conditions would change over time as the novelty effect wore off. Hence we have the following research questions (RQ): How does the avatar representations interact with:

RQ1: **the functional communicative value** based on (a) the identification of the other person (people); (b) the perceived authenticity of communications; (c) the perceived usefulness of expression and movement.

RQ2: **the task satisfaction** based on: (a) level of tasks impact, (b) comfort and (c) engagement.

RQ3: **the concept of presence** based on: (a) co-presence and (b) social presence.

In the following, we introduce the related work in Section 2 and our methodology in Section 3. We then present our results in Section 4. Next, we discuss these results in Section 5 and conclude in Section 6.

2 RELATED WORK

We first review related work in avatar appearance in remote communication. Then, we cover the importance of temporality in the field of human computer interaction and previous longitudinal studies performed in VR.

2.1 Avatars appearance in remote communication and collaboration in VR/MR

The Uncanny Valley theory, which describes users' reaction to virtual characters [17, 25], suggests that when the avatar gets more realistic, the users' affinity increases until a certain point, after which affinity drops quickly and the avatar makes users feel uneasy; this then picks up again as the avatar realism further increases. In the work of Yuan et al., participants watched in VR or on a flat screen a conversation between a photo-realistic and a cartoon character [27]. They did not participate in the conversation. Participants rated the photo-realistic avatars as more trustworthy and gave them higher affinity scores. These scores were also higher when participants used VR compared to seeing the conversation on a flat screen. Similarly, Jo et al. found that the realistic characters received higher trust levels even though they were not knowledgeable in the task [10], and that there was a positive association between the level of realism and the level of trust. They also compared a VR environment to an AR one with a real video background. They found that participants reported higher co-presence with the cartoon avatars compared to the realistic ones, but only in the AR environment.

On the other hand, Yoon et al. compared cartoon and photo-realistic avatars in AR under different three levels of body part visibility: head & hands, upper-body, and full-body [26]. Two studies were conducted in which participants were invited to solve a crossword puzzle and arrange furniture in a room to stimulate face-to-face discussion. They found a significant difference in social presence between different levels of body part visibility, with the full-body having the highest score, but no difference between the two types of avatars.

2.2 Temporality in VR/AR communication

Longitudinal studies in immersive environments (VR/MR) are less common, as they are more resource and time consuming. However, they are more ecologically valid, and could bring insights in how users' behaviour changes over time as they become more familiar with the system. For instance in Bailenson & Lee's work, over time, participants looked less at each other, teams became more connected, experienced less simulation sickness and there

was no significant change in the level of presence and co-presence [4]. In the work of Moustafa & Steed, groups of family/friends met $1-2$ times per week for 4 weeks, having an average session of 50 minutes in GearVR [18]. They performed activities such as watching videos/VR experiences, playing games or chatting, where they could embody a generic head & shoulders avatar and adjust it based on pre-defined features. They found that participants changed their avatars over time to better resemble themselves, and that they did so because their group found it otherwise uncomfortable to communicate naturally. As nonverbal behaviours and facial expressions were missing, initially participants reported difficulty interpreting social cues. Over time however, they learnt to adapt and to rely more on other cues, such as voice tone. Researchers also argued that the social group dynamics are transferable to VR interactions. In another longitudinal study by Khojasteh et al., participants met for 5 sessions in different groups of two [11], performing a different collaborative task each time in VR using Facebook Spaces. They found that participants felt more comfortable using the controllers and the app over time and they used voice tone and choice of words to determine the emotional expression in the lack of facial expressions. There was no significance differences in the level of presence. Although some people reported in the interview that they got better at completing tasks in VR, there was no significant change on workload over time. However, some participants mentioned in the interview that they better connected with their partners after being more comfortable using the VR equipment.

In these longitudinal studies in VR, participant's avatar was adapted by the user to better resemble them but keeping it in the same style [11, 18]. To our knowledge, there are no studies comparing different avatar styles over time in immersive environments. We propose investigating the effect on avatar appearance over time in mixed reality (MR) in groups of $2-3$ co-workers that know each other, comparing cartoon and realistic avatars. Based on Yoon et al.'s findings, we decide to use full-body avatars [26]. To maintain high ecological validity, we ask participants to perform their usual work meetings. Hence we are interested in how the avatar appearance interacts with the way participants communicate with each other, the perceived task satisfaction and with the participants perceived sense of presence.

3 METHODOLOGY

Device and application. The study run in Mixed Reality using the HoloLens 2 (HL2) device (microsoft.com/en-gb/hololens). We built a networked application using Unity3D game engine (unity.com) where users could see a hologram of a blue table (Figures 1f and 1g) and a control menu (Figure 1e). The table was adjustable and represented the centre of the meeting, all other participants in the meeting were located in space around the table. The control menu allowed the participants to go to the *'Home'* menu and to create a new meeting, to see who is in the current meeting, to join a meeting, to mute themselves, to adjust their microphone gain, to switch their avatar, to leave the meeting and to quit the application.

Avatars. Participants used two full-body avatars, one in a cartoon style and one in a realistic style. We created them to look like them using a picture from the shoulders up. With the local version of Avatar SDK (avatarsdk.com) we created the head for both avatars (cartoon: version 1.2.4; realistic version 2.0.5;) and 4 bodies (two male and two female, one of each with a cartoon and realistic appearance; see Figure 1f,g or the supplementary material). We also changed the avatars' clothing colours (i.e., Figures 1f and 1g). We animated the avatars in real-time using inverse kinematics. The input to avatar animation was the HoloLens hand and head tracking signal. Facial animation used a simple lip-flapping script based on voice amplitude and a blinking animation without gaze animation.

Participants. We sent participant recruitment emails for groups of two or three participants from the same company. The requirements were that they must know each other, work together, be part of daily work meetings, and be willing to have one of their regular daily meetings in mixed reality using HL2 for two-three weeks (ten meetings). We offered them a charity donation of £75 per person on their behalf. A total of 32 participants in 13 groups volunteered to take part, but seven groups (18 participants) could not due to time and logistical constrains. Thus, a total of 14 participants (7 female, 6 male, 1 non-binary; 21-45 of age) completed the study, forming 6 groups: 4 dyads and 2 triads. Out of these 6 groups, 4 were same-gender groups (2 male-only, 2 female-only), and 2 mixed-gender groups. One of the two groups with 3 participants was a mixed-gender group, and the other one same-gender. The members in each group remained the same for the duration of the study and no participant missed a meeting. Participants either had the HL2 device at home (8 participants), or we supplied a device to each of them (6 participants) for the duration of the study. Even though some participants used the HL2 before, none of them worked on remote MR meetings. We installed the application in each HL2. To maintain a high level of ecological validity, we did not ask the participants to perform a specific task. We allowed them to carry out their meeting as usual for at least $10-15$ minutes. Often, the meeting was in a format of a daily stand-up, a status report or a daily team catch-up.

Dataset. We collected questionnaire data, telemetry data from daily meetings and audio-video data from one focus group from each group. Apart from the consent forms, participants filled in the following questionnaires: demographic, on-boarding (covering their expectation of having meetings in MR) and a daily questionnaires that they completed after each meeting. We also collected head and hands movement data and whether and when participants are speaking. We did not collect speech data to keep the meetings confidential. In this paper we present the initial results of our analysis on the daily questionnaire data, the rest of the data being left out for future work.

Each group had half of their meetings (for approx. one week) using one avatar type, and the second half using the other. In total there were 54 meetings, generating a total of 124 daily questionnaire responses (one dyad's questionnaire was missing from their last meeting with cartoon avatars due to a technical error). Hence, there are 63 questionnaire responses from meetings where they used realistic avatars and 61 from cartoon avatars (2 missing). Half of the groups used the cartoon avatars first and the other half used first the realistic ones. Due to limitations out of our control, 2 groups were not able to run all ten sessions: one team had 8 sessions (4 for each type of avatar, realistic first) and the other group 6 sessions (3 for each type of avatar, cartoon first). We balanced the avatars order

#	RQ	Questionnaire Item	Cartoon		Realistic	
			m	sd	m	sd
1	2c	I felt engaged in the meeting.	5.46	0.87	5.63	0.92
2	2c	I felt that my colleagues were engaged in the meeting.	5.41	0.95	5.55	1.04
3	1b	The avatars communicated like my colleagues.	3.6	1.57	3.9	1.41
4	2a	The appearance of the avatars affected the meeting tasks.	3.86	1.56	3.77	1.19
5*	2b	The appearance of the avatars affected how comfortable I felt in the meeting.	4.05	1.6	3.86	1.55
6*	3b	The appearance of the avatars mattered to me.	4.73	1.88	4.66	1.7
7	3a	I felt that I was in the presence of my colleagues.	4.67	1.49	5.18	1.6
8*	1a	I could identify my colleagues.	5.12	1.54	5.78	0.98
9	3b	I perceive my colleagues' avatars as being only computerized images, not real people.	6.17	1.11	5.78	1.2
10*	3b	There were obvious unnatural nonverbal behaviours from my colleagues' avatars.	5.34	1.27	5.48	1.23
11*	1b,c	The avatars' nonverbal behaviour was appropriate for the context.	3.08	1.36	3.79	1.04
12*	1c	The avatars' nonverbal behaviour was useful for understanding my colleagues.	2.72	1.15	3.55	1.26

Table 1: The items in the daily questionnaire. Participants answered on a 1−7 Likert scale. The star(*) items showed significance. *RQ* stands for *research question*. *m* and *sd* stand for mean and standard deviation, showing the descriptive statistics for each question.

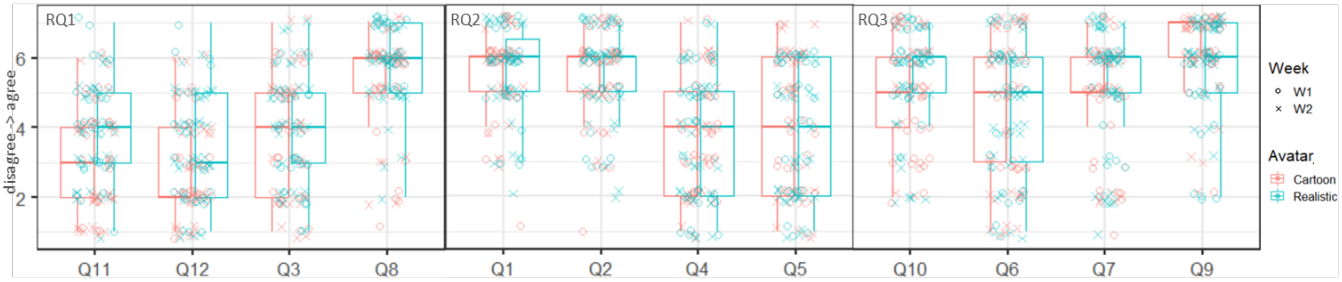

Figure 2: Boxplots for each question grouped by the research question (RQ) and separated by the type of avatar.The y-axis shows the rating (1 = *strongly disagree*; 7 = *strongly agree*).

for the triads too, one triad starting with the cartoon avatars and the other with the realistic ones. The daily questionnaire contained 12 items, 4 for each *RQ*, with responses in a 1 – 7 Likert scale, 1 meaning *strongly disagree*, and 7 *strongly agree*. A full list is in Table 1. We selected these questionnaire items from [3, 8, 14, 23] and adapted them to fit the study design and the RQs.

Procedure. After consenting, participants filled in the demographic and on-boarding questionnaires and then sent a head and shoulders picture of themselves. We used this to create their cartoon and realistic avatars. Next, we installed the application on their HL2 and set-up credentials for each participant to access the application. After this, the participating group and the researcher had a test meeting in MR to introduce the application's functionality and to perform a walk-through. For each daily scheduled session, the researcher was available to troubleshoot. Participants had the following procedure for each session: open the application from the HL2 application menu; sign in with their credentials; adjust the blue table so that there is enough local space around it (the rest of the group appears around that table). One group member creates a meeting and adds all the others. The rest of the group joins the meeting as participants are invited to; they change their avatar to the corresponding one for that week (cartoon or realistic avatars); they have their meeting as usual. At the end of it, they leave the meeting and then quit the HL2 application. After the meeting, the researcher reminds them to fill in the questionnaires for that session. They repeate this until the last session. The study was approved by the organisation's IRB (ID: 10112).

4 RESULTS

We first treated the experiment like a standard within-group study, where we compare the averaged scores for each participant using cartoon-like (C) and realistic (R) avatars. We then investigate the longitude effect by computing regression models for each dependent variable, taking into account the temporal feature.

4.1 Overview Effect of Realism

For each participant and for each question, we calculated two averages, one over all sessions (i.e., up to five) with C avatar, the other one with the R avatar. A Repeated Measure ANOVA was conducted to test the effect of realism. See Table 1 for the descriptive statistics and Figure 2 for the boxplot representation of each questions.

RQ1 Function Communicative Value: on average participants reported higher scores for all four questions (Q3,8,11,12). A Repeated Measure One-Way ANOVA found a significant difference

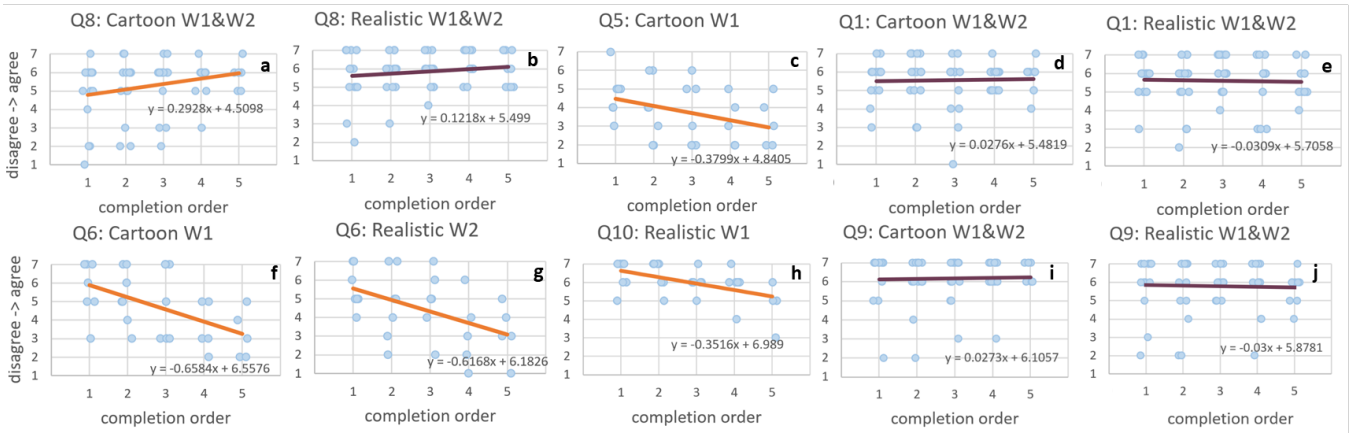

Figure 3: Scatter plots showing the questionnaire ratings over time. The y-axis in all graphs shows the rating (1 = *strongly disagree*; 7 = *strongly agree*). The x-axis represents the meetings in a chronological order. *W1*: Week 1; *W2*: *week 2. Q*: *question.* The line is the linear regression, coloured orange for the statistically significant and purple for the non-significant.

for Q11 ($F(1,13) = 7.14$, **p=.019**, $\eta^2 = .355$) and Q12 ($F(1,13) = 5.5$, **p=.036**, $\eta^2 = .296$), but not for Q8 ($F(1,13) = 3.53, p = .08, \eta^2 = .217$) or Q3 ($F(1,13) = .718, p = .41, \eta^2 = .52$). This result indicated that participants found their colleagues' nonverbal behaviour being appropriate for the context ($Q11$) and useful for understanding the colleagues ($Q12$) with the realistic avatar than the cartoon-like one.

RQ2 Task Satisfaction: there is no difference between the two avatars for the level of engagement for themselves (Q1: $F(1,13) = .51, p = .49, \eta^2 = .04$), the perceived level of engagement of their colleagues (Q2: $F(1,13) = .44, p = .52, \eta^2 = .03$), the impact of the appearance over the task (Q4: $F(1,13) = .08, p = .79, \eta^2 = .01$) or the reported level of comfort (Q5: $F(1,13) = .50, p = .50, \eta^2 = .04$).

RQ3 Presence: again we found no differences in participants' report of the extent to which the avatar matters to them (Q6: $F(1,13) = .07, p = .80, \eta^2 = .01$), the level of co-presence they felt (Q7: $F(1,13) = 2.1, p = .17, \eta^2 = .14$), and whether they perceived their colleagues' avatar as being more as digital images (Q9: $F(1,13) = 2.1, p = 0.17, \eta^2 = .14$) or unnatural (Q10: $F(1,13) = .44, p = .52, \eta^2 = .03$).

4.2 Temporal scatter plots

To account for the temporal aspect, we calculated the regression statistics for each dependant variable, in respect to which avatar type the participants embodied. Hence, we computed this for each avatar type, accounting for whether the participants embodied that type in their first (W1) or second week (W2).

RQ1: The functional communicative value. We found significance positive correlation over time for being able to recognise their colleagues ($Q8$ regarding $RQ1a$), when participants were embodying in C ($p = 0.04$ Fig 3a), but not in R (Fig 3b). The rest of the $RQ1$ questions ($Q3, 11, 12$) did not show significance.

RQ2: The task satisfaction. In week 1, participants' responses on $Q5$ ($RQ2b$) were significant with C ($p = 0.05$, Fig 3c), meaning that with the cartoon avatar their level of comfort increased over time. No other significant effect was found for $Q5$ or other questions for $RQ2$ ($Q1, 2, 4$).

RQ3: The concept of presence. Participants reported that the avatar's appearance mattered less over time ($Q6$, $RQ3b$) when they embodied C avatars in $W1$ ($p = 0.002$ Fig 3f) and for R avatars in $W2$ ($p = 0.005$ Fig 3g). There is no significance for when they used the R avatars in $W1$ and the C in $W2$. Regarding obvious unnatural nonverbal behaviours in the avatars, ($Q10$, $RQ3b$), participants reported less of these over time for R avatars in $W1$ ($p = 0.003$, Fig 3h). No other significance was found on $Q10$ or regarding $RQ3b$, $Q9$.

5 LIMITATIONS AND DISCUSSION

Participants in this study had different prior experience with using MR devices. Almost half of them (6 out of 14) never used a MR device, whereas the rest used it at different frequencies in the past 6 months (two participants- more than two times a week, two participants- once a week, four participants- 1 – 3 times a month). Due to the small number of participants we could not be test whether the prior MR experience of some participants influenced their responses. All participants were employees in a large tech company. Hence, there might be a chance that they are more accepting of innovations in virtual environments. We consider these limitations of our study and propose further work to address them.

In the remaining of this section we discuss the implications on the results reported in the previous section, in terms of the difference between C and R avatars, followed by the trends over time.

5.1 Cartoon vs Realistic overall

Realistic avatars' nonverbal behaviour was reported more useful for understanding others and more appropriate for the context compared to the cartoon avatars. Although there was no difference in the technical implementation of the nonverbal behaviour (see chapter 4.1), the participants reported significantly higher ratings of usefulness and appropriateness for the realistic avatar. This indicates that the realistic avatars showed a higher functional communicative value based on the perceived authenticity of communication ($RQ1b$)

and the perceived usefulness of expression and movement ($RQ1c$) compared to the cartoon avatars. None of the other questions were significant on this measure.

5.2 Cartoon vs Realistic over time

RQ1. Participants reported that when using the cartoon avatars ($W1$ and $W2$), they identified their colleagues better over time (Figure 3a) and reported a constant high values overall with realistic avatars (Figure 3b). Despite the cartoon appearance, participants learnt to identify their colleagues (who they knew prior to the study) over time. None of the other questions regarding $RQ1$ ($Q3, 11, 12$) showed significance.

RQ2. Participants reported that the impact of the avatar appearance on their comfort reduced over time, but only in the case that they used the cartoon avatars before the realistic avatars. This could imply that seeing their colleagues in R avatars affected their comfort, long term, even after they changed to a C avatar. Regarding $RQ2c$, participants reported that overall they were engaged in the meeting when using either avatars (Figures 3d, e). There was no significant trend related to the impact on meeting tasks.

RQ3. When it comes to social presence ($RQ3b$), the avatars' appearance mattered less over time with C in $W1$ and R in $W2$ ($Q06$). Interestingly, this trend was not for the avatar type, but for a certain order. When participants used C avatars first, they reported that the appearance mattered less. This was not seen for C in $W2$ after using R avatars. Furthermore, they reported the same trend (the appearance matter less over time) for the R in $W2$ avatars. Similarly, this was not seen for R in $W1$. It seems that using cartoon avatars first, made them feel that their colleagues' appearance did not matter that much after a few sessions and this trend continued for R avatars in $W2$.

This trend was also seen in the free-text comments. When they embodied C in $W1$, participants wrote in the third (out of five) session: *"I do find that I notice the avatar less over time. I accept it is <PARTICIPANT NAME> now I and don't notice its appearance as much anymore."* And after the last session with the cartoon avatar: *"I feel like I am starting to care less about the avatar's appearance"*. When they embodied the realistic avatars in $W2$, a participant commented in session 1: *"the avatar's appearance was more top-of-mind again"*, and at end of the week: *"I don't notice the avatars' [appearance] as much anymore [...]"*. This enforces that over time, participants got used with a certain appearance and it did not get in the way of work tasks.

Participants also reported a decrease in the other's avatar obvious unnatural nonverbal behaviours ($RQ3b$, $Q10$) when using R in $W1$ (Figure 3h). This could imply that the appearance of the realistic avatars compensated for the unnatural nonverbal behaviours in the avatars, resulting in reporting less so of it. The trend did not appear for C ($W1$ or $W2$) nor for R in $W2$. These results showed that even though they might have found obvious unnatural nonverbal behaviours at the beginning with R avatars, they seemed to decrease over time, which might show a recovery from the uncanny valley. This demonstrates the value of longitudinal studies in overcoming novelty effects. This trend was only seen in for R in $W1$, not in any C condition or in R following a C condition. This was possibly due to a strong initial mismatch between realistic appearance and less realistic behaviour, that participants adjusted to over time. Participants starting with C would not have such a great mismatch and would be accustomed to the behaviour by the time they used the R avatars.

There was no significant trend in whether participants perceived their colleagues' avatars as more computerized rather than real or whether they felt they were in the presence of their colleagues. In the former case the scores were high throughout (the avatar felt computerized), perhaps because the avatars were transparent holograms.

6 CONCLUSION

We presented the initial results from a longitudinal study on avatars' appearance during work-related meetings between co-workers. We investigated how the avatar appearance interacts with the way participants communicate with each other, perceived task satisfaction and perceived sense of presence. In total 14 participants in 6 groups had their usual work meetings in MR while embodying a resembling cartoon-like or realistic avatar. Based on 124 questionnaires from 54 meetings we found that the realistic avatar's nonverbal behaviour was perceived more appropriate for the interaction and more useful for understanding the others compared to the cartoon avatar. When looking at these results over time, there were different insights for cartoon and realistic avatars based on which one was embodied first. As such, this study raises the possibility that developing lifelike avatars in real work meetings over time should not be the assumed only valuable endpoint, i.e. what people consider to be nice for an avatar does not equate to the communicative good that it provides. The implication is that designing for longitudinal avatar acceptance is different to designing acceptable avatars.

ACKNOWLEDGMENTS

This project was partly supported by grant $EP/L015846/1$ for the Centre for Doctoral Training in Intelligent Games and Game Intelligence (iggi.org.uk) from the Engineering and Physical Sciences Research Council (EPSRC).

REFERENCES

[1] Jeremy N. Bailenson and Andrew C. Beall. 2006. Transformed Social Interaction: Exploring the Digital Plasticity of Avatars. In *Avatars at Work and Play: Collaboration and Interaction in Shared Virtual Environments*, Ralph Schroeder and Ann-Sofie Axelsson (Eds.). Springer Netherlands, Dordrecht, 1–16. https://doi.org/10.1007/1-4020-3898-4_1

[2] Jeremy N Bailenson, Jim Blascovich, Andrew C Beall, and Jack M Loomis. 2001. Equilibrium theory revisited: Mutual gaze and personal space in virtual environments. *Presence: Teleoperators & Virtual Environments* 10, 6 (2001), 583–598.

[3] Jeremy N Bailenson, Jim Blascovich, Andrew C Beall, and Jack M Loomis. 2003. Interpersonal distance in immersive virtual environments. *Personality and social psychology bulletin* 29, 7 (2003), 819–833.

[4] Jeremy N Bailenson and Nick Yee. 2006. A longitudinal study of task performance, head movements, subjective report, simulator sickness, and transformed social interaction in collaborative virtual environments. *Presence: Teleoperators and Virtual Environments* 15, 6 (2006), 699–716.

[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*. 242–249.

[6] Judee K Burgoon, Laura K Guerrero, and Kory Floyd. 2016. *Nonverbal communication*. Routledge.

[7] Edward T Hall, Ray L Birdwhistell, Bernhard Bock, Paul Bohannan, A Richard Diebold Jr, Marshall Durbin, Munro S Edmonson, JL Fischer, Dell Hymes, Solon T Kimball, et al. 1968. Proxemics [and comments and replies]. *Current anthropology* 9, 2/3 (1968), 83–108.

[8] Chad Harms and Frank Biocca. 2004. Internal consistency and reliability of the networked minds measure of social presence. (2004).

[9] Paul Heidicker, Eike Langbehn, and Frank Steinicke. 2017. Influence of avatar appearance on presence in social VR. In *2017 IEEE Symposium on 3D User Interfaces (3DUI)*. IEEE, 233–234.

[10] Dongsik Jo, Ki-Hong Kim, and Gerard Jounghyun Kim. 2017. Effects of avatar and background types on users' co-presence and trust for mixed reality-based teleconference systems. In *Proceedings the 30th Conference on Computer Animation and Social Agents*. 27–36.

[11] Negar Khojasteh and Andrea Stevenson Won. 2021. Working Together on Diverse Tasks: A Longitudinal Study on Individual Workload, Presence and Emotional Recognition in Collaborative Virtual Environments. *Frontiers in Virtual Reality* 2 (2021), 53.

[12] Michael Koch, Kai von Luck, Jan Schwarzer, and Susanne Draheim. 2018. The novelty effect in large display deployments–Experiences and lessons-learned for evaluating prototypes. In *Proceedings of 16th European Conference on Computer-Supported Cooperative Work-Exploratory Papers*. European Society for Socially Embedded Technologies (EUSSET).

[13] 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*. 1–10.

[14] Matthew Lombard, Theresa B Ditton, and Lisa Weinstein. 2009. Measuring presence: the temple presence inventory. In *Proceedings of the 12th annual international workshop on presence*. 1–15.

[15] Jean-Luc Lugrin, Johanna Latt, and Marc Erich Latoschik. 2015. Anthropomorphism and Illusion of Virtual Body Ownership.. In *ICAT-EGVE*. 1–8.

[16] Karl F MacDorman and Debaleena Chattopadhyay. 2016. Reducing consistency in human realism increases the uncanny valley effect; increasing category uncertainty does not. *Cognition* 146 (2016), 190–205.

[17] Masahiro Mori, Karl F MacDorman, and Norri Kageki. 2012. The uncanny valley [from the field]. *IEEE Robotics & Automation Magazine* 19, 2 (2012), 98–100.

[18] Fares Moustafa and Anthony Steed. 2018. A longitudinal study of small group interaction in social virtual reality. In *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology*. 1–10.

[19] 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

[20] Oliver Otto, Dave Roberts, and Robin Wolff. 2006. A review on effective closely-coupled collaboration using immersive CVE's. In *Proceedings of the 2006 ACM international conference on Virtual reality continuum and its applications*. 145–154.

[21] Dhaval Parmar. 2017. *Evaluating the effects of immersive embodied interaction on cognition in virtual reality*. Ph. D. Dissertation. Clemson University.

[22] Mincheol Shin, Se Jung Kim, and Frank Biocca. 2019. The uncanny valley: No need for any further judgments when an avatar looks eerie. *Computers in Human Behavior* 94 (2019), 100–109.

[23] Mel Slater. 1999. Measuring presence: A response to the Witmer and Singer presence questionnaire. *Presence* 8, 5 (1999), 560–565.

[24] Thomas Waltemate, Dominik Gall, Daniel Roth, Mario Botsch, and Marc Erich Latoschik. 2018. The impact of avatar personalization and immersion on virtual body ownership, presence, and emotional response. *IEEE transactions on visualization and computer graphics* 24, 4 (2018), 1643–1652.

[25] Shensheng Wang, Scott O Lilienfeld, and Philippe Rochat. 2015. The uncanny valley: Existence and explanations. *Review of General Psychology* 19, 4 (2015), 393–407.

[26] 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)*. IEEE, 547–556.

[27] Lingyao Yuan, Alan Dennis, Kai Riemer, et al. 2019. Crossing the uncanny valley? Understanding affinity, trustworthiness, and preference for more realistic virtual humans in immersive environments. In *Proceedings of the 52nd Hawaii International Conference on System Sciences*.

[28] Katja Zibrek, Elena Kokkinara, and Rachel McDonnell. 2018. The effect of realistic appearance of virtual characters in immersive environments-does the character's personality play a role? *IEEE transactions on visualization and computer graphics* 24, 4 (2018), 1681–1690.