A Large-Scale Study of Proxemics and Gaze in Groups

Mark Roman Miller*
Computer Science, Stanford University

Cyan DeVeaux†
Communication, Stanford University

Eugy Han‡
Communication, Stanford University

Nilam Ram§
Psychology, Communication, Stanford University

Jeremy N. Bailenson¶
Communication, Stanford University

2023 IEEE Conference Virtual Reality and 3D User Interfaces (VR) | 979-8-3503-4815-6/23/$31.00 ©2023 IEEE | DOI: 10.1109/VR55154.2023.00056

ABSTRACT

Scholars who study nonverbal behavior have focused an incredible amount of work on proxemics, how close people stand to one another, and mutual gaze, whether or not they are looking at one another. Moreover, many studies have demonstrated a correlation between gaze and distance, and so-called equilibrium theory posits that people modulate gaze and distance to maintain proper levels of nonverbal intimacy. Virtual reality scholars have also focused on these two constructs, both for theoretical reasons, as distance and gaze are often used as proxies for psychological constructs such as social presence, and for methodological reasons, as head orientation and body position are automatically produced by most VR tracking systems. However, to date, the studies of distance and gaze in VR have largely been conducted in laboratory settings, observing behavior of a small number of participants for short periods of time. In this experimental field study, we analyze the proxemics and gaze of 232 participants over two experimental studies who each contributed up to about 240 minutes of tracking data during eight weekly 30-minute social virtual reality sessions. Participants' non-verbal behaviors changed in conjunction with context manipulations and over time. Interpersonal distance increased with the size of the virtual room; and both mutual gaze and interpersonal distance increased over time. Overall, participants oriented their heads toward the center of walls rather than to corners of rectangularly-aligned environments. Finally, statistical models demonstrated that individual differences matter, with pairs and groups maintaining more consistent differences over time than would be predicted by chance. Implications for theory and practice are discussed.

Index Terms: Human-centered computing—Human computer interaction (HCI—Interaction paradigms—Virtual reality; Human-centered computing—Collaborative and social computing—Empirical studies in collaborative and social computing

1 INTRODUCTION

Virtual Reality (VR) captures an unprecedented richness of data from its users. In the common case, a device tracks head and hands position and rotation dozens of times per second. There are many questions regarding social behavior that this data can answer. Two of the most prominent are where people stand and were people look. These two constructs - proxemics and gaze - are easily available through this data and can be analyzed in high spatial and temporal fidelity in a paradigm known as behavioral tracing [34]. Understanding proxemics is important because it relates to several constructs of interest including liking, communication, and warmth [6]. Violations of personal space can be confusing and stressful, both when one is too far from another, or when one is too close. Gaze is similarly important, as it can signal attention and intimacy [23]. The study of proxemics and gaze is important to social virtual reality.

However, virtual reality is still a new medium, and effects of novelty and learning can influence findings. To this end, we studied the behavior of 232 participants who participated in social virtual reality eight times over eight weeks for about thirty minutes per session. We found several effects, including influences of facets of place, time, and dyad on these behaviors. For example, dyads increased their personal space over time, but also looked at each other more often over time. There was also a relationship between personal space and directness of gaze that corroborates what is known as equilibrium theory [1]. These results continue to show that social virtual reality carries over many of the patterns that we know from social interaction in real life. They also encourage future work studying the effects of virtual reality on behavior over time and within varying contexts.

2 RELATED WORK

This work builds on several threads of previous research: longitudinal studies of social virtual reality, proxemics, and gaze. We review each of these threads and then bring them as context to our research questions.

2.1 Social VR Over Time

Social VR has increased in popularity in recent years, spurred on by the availability of consumer VR devices. Software that enables these experiences vary from entertainment platforms like VRChat, Rec Room, and AltspaceVR to professional platforms like Mozilla Hubs, Meta Horizon Workrooms, and ENGAGE.

The earliest studies of social VR took place around 2000. For example, works led by Slater [28] and Garau [9] explore communication differences between virtual reality and face-to-face communication. Despite this early start, Han and colleagues [12] report only 37 social VR studies in their 2022 paper. There have been even fewer studies that have 3 or more participants sharing a virtual space in immersive virtual reality. Mütterlein and collaborators [19] studied groups of two to four, varying several facets of VR and observing their influence on intention to collaborate. Moustafa and Steed [18] performed an exploratory in-the-wild study of collaboration over headset-only VR with several groups of two to four participants. Roth and collaborators [26] studied groups of five participants in an augmented a virtual museum experience with visualized social signals like joint attention and eye contact. Finally, while not a study of immersive VR but rather a desktop-based virtual environment, Williamson and colleagues [33] studied the proxemics of 26 participants in a virtual workshop. Considering how many studies of virtual reality have been done, work on groups in VR has been difficult to come by.

It has been even rarer to find studies of social VR over time. Longitudinal studies are often difficult to coordinate, but are able to show adaptation to a given medium and study behavior of users who are well-acclimated to a system. Bailenson and Yee [4] studied three groups of three participants in 15 sessions over seven weeks and found substantial changes and adaptations over time in several

---

*e-mail: mrmillr@stanford.edu
†e-mail: cyanjd@stanford.edu
‡e-mail: eugyoung@stanford.edu
§e-mail: nilamram@stanford.edu
¶e-mail: bailenso@stanford.edu

2642-5254/23/$31.00 ©2023 IEEE
DOI 10.1109/VR55154.2023.00056

Authorized licensed use limited to: Stanford University. Downloaded on November 16,2023 at 05:39:35 UTC from IEEE Xplore. Restrictions apply.

variables. Roth and collaborators [27], Moustafa and Steed [18], and Khojasteh and Stevenson Won [15] all found several adaptations to the medium of virtual reality when studying participants behaviors over time. While individual adaptations depended on the affordances and frustrations of the VR hardware and software used in the study, it is clear that much of the adaptation was participants communicating more through the available social signals. This adaptation is a well-known result in computer-mediated communication [32]. To give an example, participants in the study by Moustafa and Steed [18] used a VR system with a headset and no hand controllers. Instead of a wave for a greeting or farewell, which would normally involve hand tracking, participants 'waved' using their heads, tilting their head left and right. Han and colleagues [12] found several effects of time, including greater presence, enjoyment, entitativity, and realism over time.

2.2 Proxemics and Gaze

Proxemics is the study of person-to-person proximity and its relations with affect, behavior, and cognition. Hall's work [10] on proxemics on middle-class American adults defined four levels of proximity: intimate (< 0.45m), personal (0.45m-1.2m), social (1.2m-3.6m) and public (> 3.6m). These thresholds are not universal but rather vary depending on large-scale, cultural variables like the prevalence of contact [5] and the relative importance of individualism versus collectivism [16].

Virtual reality is amenable to studying proxemics due to the built-in capacity to track a user's position. This position data can then be leveraged as a continuous measure, oftentimes with high spatial and temporal fidelity [34]. Proxemics has been used to inform both independent variables and dependent variables. Bailenson and colleagues [3] leveraged proxemics to demonstrate virtual characters receive more personal space and seem more real when more behaviorally realistic. Bönsch and collaborators [6] show that angry characters receive more personal space than happy characters. Choudhary and collaborators [8] varied two affordances of social VR, volume and head size, to investigate their effect on distance estimation. Head size affected distance estimation, but volume did not. Takahashi and collaborators [29] varied the speaking volume of a character and noted that in a walking task, participant gave more distance to the character when the character spoke louder.

Gaze is another nonverbal form of communication that can intention, attention, and intimacy [7]. The value of this communication has even motivated technical developments in various stereo-like displays [21, 22]. Overall, the use of gaze in mixed reality has been substantial and has recently been reviewed [23]. Vertegaal and collaborators [31] showed that for a majority of time, people look at the speaker or the target in multiparty conversation. Gaze also signals turn-taking. When a virtual conversation was instrumented with automated random gaze, participants spoke in a greater number of turns and shorter duration turns compared to no changes in gaze, but teams using this random gaze model did not complete a task as quickly as teams using realistic gaze [30]. Particular kinds of gaze, like eye contact, do not necessarily signal effective or ineffective communication on their own, but can do so if other criteria are met [24]. One of the changes over time that Bailenson and Yee [4] found was that participants looked at others less over time. They explain this effect due to the weight and discomfort of the headset and the lack of any facial cues on the virtual avatars.

Finally, proximity and gaze can relate to each other through equilibrium theory [1]. Because both proximity and gaze are signals of intimacy, extrinsic changes to one variable (e.g., stepping into a small space like an elevator) lead to a change in the other variable (less mutual gaze) so as to maintain an appropriate level of intimacy. This has been demonstrated in several contexts, including virtual reality [2, 35].

2.3 Research Questions

RQ1: How does interpersonal distance adapt over time in virtual reality? As far as we are aware, there are no studies that investigate personal space over time in virtual reality. Considering there have been other adaptations over time, whether and how people's personal space develops is an important question.

RQ2: How can the virtual reality environment affect spacing? There has been some evidence of the influence of environment on proxemics in general and interpersonal distance in particular [20], and sparse work using VR [13]. However, there is still much to explore, as there are quite many variables that define a space.

RQ3: What is the relative size of inter-dyad difference factors? It is known there are cultural differences in interpersonal distance, and some inter-dyad differences have been studied, like the gender composition of a pair [35]. However, the intersection of social VR studies that also collect data over time is small.

RQ4: How does gaze change over time? Given that the reasons gaze behavior changed in the work by Bailenson and collaborators [4] were discomfort due to headset weight and lack of nonverbal cues, what happens now that the headset is lighter and nonverbal cues like hand tracking are commonplace?

3 METHODS

This work reports on data collected as part of the Stanford Longitudinal VR Classroom Dataset over the course of two studies of classroom immersive VR. Students met in small groups ranging from four to 12, and consented to have their verbal, nonverbal, and performance continually tracked during each course, typically for about eight weekly sessions which lasted about 30 minutes per session. In addition, each student provided self report about their experience after each session (see [11] for a detailed description). The current paper utilizes previously unreported data from the dataset, and focuses on gaze and proxemics. These studies from which we report data were each run using a social VR platform called ENGAGE. Each study consisted of its own participant pool and conditions. Participant consent went through an extremely rigorous process, approved by two separate organizations within Stanford. Moreover, there was a 3rd party arbitrar who oversaw data collection during the course, and students had an interactive, hour-long discussion of the study procedures and data collection before deciding to consent. Data recorded included position and rotation of each participant's headset and hand controllers.

3.1 Apparatus

In both datasets, participants used the Meta Oculus Quest 2 headsets (503g) and two hand controllers (126g) in their own personal environments. The combined field-of-view of the headset is 104.00° horizontal FOV, 98.00° FOV. In the avatar study dataset, two participants opted to participate with owned personal headsets (both PC-based Valve Index). The headsets did not perform eye tracking, and in order to complete this experiment at the scale necessary, we opted not to include add-on eye tracking devices. Note that here, head orientation is used as a proxy for gaze.

The software in use was the ENGAGE virtual communications platform, versions 1.7 through 2.0.1, produced by ENGAGE PLC. The virtual environments in which the participants met w. In dataset 1, all participants met in the same "Engineering Workshop" room. In dataset 2, participants met in one of 192 uniquely-built environments each week. These environments differed in size of moving area, height, and whether it was indoors or outdoors. Figure 1 shows screen captures of anonymized virtual students in discussion.

3.2 Participants

There were a total of 232 participants in the study across the two subject populations ($n_1 = 86$), ($n_2 = 146$). Participants were university students enrolled in one of two 10-week courses about VR.

Authorized licensed use limited to: Stanford University. Downloaded on November 16,2023 at 05:39:35 UTC from IEEE Xplore. Restrictions apply.

Figure 1: Participants performing discussion activities in the VR environment.

While all students who were part of the course took part in all the VR activities, only those who consented to participate in the study had their data included in the study. Of the 101 students in Study 1 and 171 in Study 2, 93 and 158 consented to participate in the study, respectively.

In Study 1 (Female = 30, Male = 47, Other = 2, declined or did not answer = 7), participants were between 18 and 58 years old (M = 22.3, SD = 5.2; $n_{18-23} = 68$, $n_{24-29} = 7$, $n_{30-34} = 3$, $n_{35-39} = 1$, $n_{55-59} = 1$, $n_{declined} = 6$ and identified as African American or Black ($n = 11$), Asian or Asian American ($n = 30$), Hispanic or Latinx ($n = 9$), Middle Eastern ($n = 1$), White ($n = 21$), more than one race ($n = 5$), or declined to or did not respond ($n = 9$). Participants had varying levels of experience with VR, with 41 (51.2%) having never used VR before. Prior to the course, 38 participants were not familiar with anyone in their discussion group, and others reported knowing one ($n_1 = 13$) or more members ($n_2 = 12$, $n_3 = 1$, $n_4 = 2$, $n_5 = 2$).

In Study 2 (Female = 59, Male = 79, declined or did not respond = 4), participants were between 18 and 49 years old (M = 20.9, SD = 2.8; $n_{18-23} = 133$, $n_{24-29} = 4$, $n_{45-49} = 1$, $n_{declined} = 4$ and identified as African American or Black ($n = 12$), Asian or Asian American ($n = 47$), Hispanic or Latinx ($n = 8$), Indigenous/Native American, Alaska Native, First Nations ($n = 2$), Middle Eastern ($n = 1$), Native Hawaiian or other Pacific Islander ($n = 5$), White ($n = 41$), more than one race ($n = 19$), a racial group not listed ($n = 1$), or declined to or did not respond ($n = 2$). Participants had varying levels of experience with VR, with 50 (36.2%) having never used VR before. Prior to the course, 67 participants were not familiar with anyone in their discussion group, and others reported knowing one ($n_1 = 36$) or more members ($n_2 = 10$, $n_3 = 4$, $n_4 = 5$, $n_5 = 1$, $n_7 = 1$).

3.3 Procedure

Students opted in to the experiment with a consent form approved by the university institutional review board (IRB) and the university student's oversight committee. This IRB process required that researchers and course staff did not know which students opted in as participants in the experiment until after the course finished, so that there would be no plausible appearance of coercion to participate in the study. This also implied that all students were recorded in this study, and only data associated with consenting participants was used. Participants were reminded of the recording through a visual notification of recording at the beginning of a recording session or upon joining a session currently being recorded.

The activities that participants performed in a week included discussion on readings about virtual reality and development of room-scale virtual dioramas in groups of varying sizes. In all virtual environments, participants were able to move using physical motion such as walking or leaning and user-interface-based motion such as point-and-press-to-teleport and joystick movement. Participants were also able to create 3D drawings, write on personal whiteboards and 'paper' notes of varying sizes, add immersive effects and 3D objects, and display media content. There was a library of about one thousand virtual objects available for participants to create, move, organize, and delete in the virtual spaces. The platform accommodated use of 3D audio, which allowed for splitting off into smaller groups without audio overlap. The majority of time during the sessions did not use this feature. Sessions took place eight times over the course of eight to nine weeks, and the duration was about thirty minutes per session.

3.4 Conditions

3.4.1 Avatar Study

The avatar study consisted of two conditions with two levels each, counterbalanced across sessions in a Latin square. As there were eight weeks, this assignment was repeated for the second four weeks, and as there were eight groups, the assignment was identical for the second four groups.

Avatar: self vs. uniform. One condition that varied by group and week was the embodiment of either a *self-avatar*, in which a participant was told to create an avatar that 'looks and feels like you', or a predefined *uniform avatar*, determined through pre-testing to be the most gender- and racially-ambiguous among the options available.

Synchrony manipulation: present vs. absent. At the beginning of each session, participants performed either a synchronized motion activity, raising and lowering arms in unison with the rest of the group, or an individual drawing activity that matches the amount of motion in the synchrony activity condition, but not the synchronous nature of the activity.

3.4.2 Context Study

In the study on virtual context, there were three conditions with two levels each, counterbalanced across sessions with a Latin square. As there were twenty-four groups, three groups were assigned to each of the eight sequences in the Latin square. A total of 192 environments were created in a stimulus sampling paradigm [25], 48 per combination of the two environmental variables (view and setting).

View: panoramic vs. constrained. The environments varied in terms of the amount of space visible in horizontal and vertical directions. In panoramic environments, much more space was visible and accessible than in constrained environments.

Setting: indoors vs. outdoors. In order to tease apart the natural correlation between outdoor, panoramic spaces, and indoor, constrained spaces, we varied these two dimensions separately.

Motion: active vs. passive. Finally, some groups in some weeks were asked to minimize their UI-based motion, like teleporting and smooth movement. A manipulation check revealed this was only partially effective, i.e., there were differences in movement between the two conditions, but there was still substantial teleporting and smooth movement in the passive condition.

3.5 Data

To illustrate the scale and the structure of the data, Figure 2 shows the weekly session, session duration, group sizes, and participants this work. The highest level of data organization was the *study*, which was either the *avatar study*, collected in summer 2021, or *context study*, collected in fall 2021. These are represented in the figure as two separate plots. The next levels of organization are the week and the section. The *week* indicates which week of eight the data were obtained, and is laid out in columns. The *section* was the group and time participants met for discussion, and is laid out in rows. In the avatar study, there were eight sections, and in the context study, there were 24 sections. Each participant took part in only one section per week. Usually, a participant attended the same section week to week, but there were some exceptions. A *session* is one participant's data for one week. Each session lies entirely

Authorized licensed use limited to: Stanford University. Downloaded on November 16,2023 at 05:39:35 UTC from IEEE Xplore. Restrictions apply.

within one and only one section. In total, we obtained data on 1745 sessions that, on average, lasted 31.19 minutes (SD = 7.86 min).

Visual features in Figure 2 highlight several aspects of the data. Empty rectangles (e.g., Avatar, Week 2, Group 4) indicate both primary and backup recordings failed and data was lost, which happened occasionally but not often. Single-line recordings (e.g., Context, Week 3, Group 10) occurred when the recording failed but one group member attended a different section that week. Roughness on the left side of the plot (e.g., Group 12 Week 4 Inset) indicated variation in when participants arrived in the virtual world. A premature and sharp cutoff on the right side (e.g., Avatar Study, Week 2, Group 3) indicated a system crash, but a fuzzier break (e.g., Group 12 Week 4 Inset) indicated a normal group dismissal.

Several analyses use a *pair* as a unit of analysis. This is a complete pairing of participants within the same session, meaning in group of e.g., 6 participants, there will be $\binom{6}{2} = 15$ pairs for symmetric relations such as the distance between participant A and participant B and $6 \times 5 = 30$ pairs for non-symmetric relations such as the percentage of time participant A is in participant B's field of view. On average, there were 21.9 pairs per session, with more per session in the avatar study (34.4 pairs per session) than the context study (17.9 pairs per session) due to the difference in group sizes (avatar, 9.25 people per session; context, 6.43 people per session).

During a session, the data was collected at 30Hz and consisted of four tracked points with six degrees of freedom each. Three of the points were the headset, left hand controller, and right hand controller, and the fourth is the 'root', the transformation between the participant's physical space and the virtual space. The root changed when a participant translated or rotated their position with a UI control (e.g., teleporting by pointing and clicking, rotating 15 degrees left by tapping the controller joystick).

The coordinate system of the data follows the conventions used by the Unity game engine; namely, a left-handed coordinate system with Y upwards, Z forwards, and X rightwards, and intrinsic rotations in the order of yaw (Y), pitch (X), and roll (Z), where positive values indicate a left-handed rotation relative to the positive direction along the axis.

4 RESULTS

The results section is organized with respect to the variables of interest. First, we discuss the effect of experimental manipulations and time upon interpersonal distance. Then, we discuss the distributions of and effects on gaze, proxied by the forward direction of the headset. Finally, we note the relationship between the two described in equilibrium theory [1].

The statistical analyses in this work used mixed-effect models using the 'lmer' and 'lmerTest' packages in the R programming language. In addition to linear models that have an output variable and an input variable, mixed effect models allow the specification of grouping factors for correlated random effects. For example, one random effect is the individual differences due to participant, which avoids both collapsing across observations as with an average and inflating significance with correlated errors.

4.1 Proxemics

We define our variable of interest, interpersonal distance, to be a function of the distance between participants' heads. For any given session, this is in fact a distribution of values, so there must be a summary function to collapse this distribution into one value [17]. In some previous work [3, 29], the minimum has been used. However, several concerns led us to select a different summary function. In contrast to face-to-face interaction, virtual reality allows spatial overlap between people. A participant can accidentally teleport into a position that is arbitrarily close to another participant. In addition to this, the sheer length of observation time (31 minutes average) increased the risk for this or other outliers in distance. Therefore,

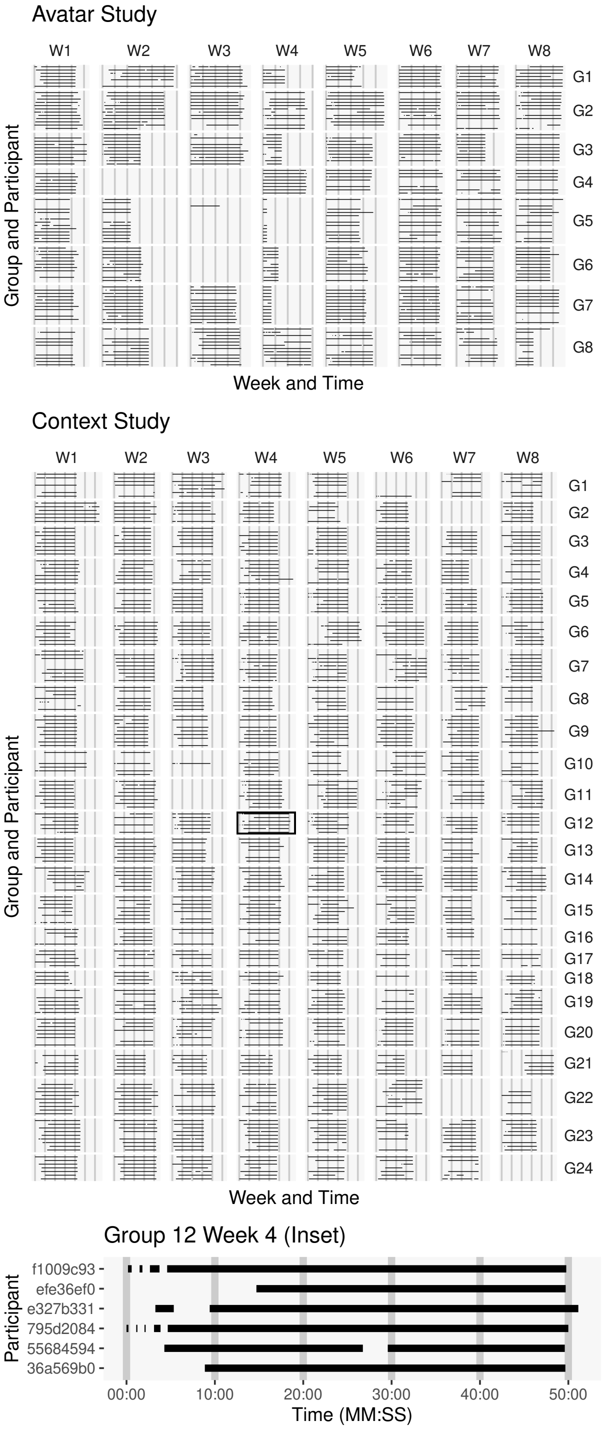

Figure 2: Figures showing the weekly sessions, duration, participants, and group size of the two studies. Panels 1 and 2 (Avatar Study and Context Study) consists of many facets. Each facet represents a week, given by its horizontal ordering, and a group, given by its vertical ordering. Within each facet, and in the final panel that shows a close-up of group 12 in week 4 from the context study, each participant receives a horizontal row on which a line is drawn if data is collected at that time. Vertical lines within each facet demarcate 10-minute intervals of time.

Authorized licensed use limited to: Stanford University. Downloaded on November 16,2023 at 05:39:35 UTC from IEEE Xplore. Restrictions apply.

instead of computing the minimum value (first smallest) of the set, we compute the n-th smallest value. We selected $n = 150$ so that the values of five seconds worth of samples are ignored. We judged five seconds to be appropriate for this buffer because it is long enough for participants to react and move away if one participant accidentally moved too close to another. The results we report were robust to variations in this parameter.

Interpersonal distance values were highly right-skewed, and were thus log-transformed. All values are reported given original units (meters) rather than the model term, log-meters. Consequently, differences between values, such as standard deviations and unstandardized effect sizes, become multipliers, which are written in this work as percentages. We also included the gender composition of the pairs as a covariate, as previous work [14, 35] found effects of pair gender composition on interpersonal distances.

4.1.1 Combined Datasets

The prototypical pair began week 1 at a distance of 1.44m and increased in distance by 7.0% per week over the eight weeks to 2.31m ($t(205.74) = 6.197$, $p < 0.001$). Although pairs in the Avatar study were 8.4% farther apart than pairs in the Context study (1.50m vs. 1.38m), this difference was not significant ($t(29.35) = 0.893$, $p = 0.379$). However, distance did differ across gender pairings (M-M = 1.33m, F-F = 1.42m, other pairs = 1.53m, $\chi^2(2, N = 5341) = 27.65$, $p < 0.001$). This result indicates interpersonal distance increases over time (**RQ1**).

The variance in distances was attributable to section, session, and pair differences. Variation uniquely due to section (the group) had a standard deviation of 18.4% ($\chi^2(1, N = 5341) = 11.92$, $p < 0.001$, M+SD = 1.70m, M-SD = 1.22m). Variation uniquely due to session had a standard deviation of 41.6% ($\chi^2(1, N = 5341) = 565.05$, $p < 0.001$, M+SD = 2.04m, M-SD = 1.02m). Finally, variation uniquely due to pair had a standard deviation of 24.8% ($\chi^2(1, N = 5341) = 83.33$, $p < 0.001$, M+SD = 1.80m, M-SD = 1.15m).

4.1.2 Avatar Study

In the Avatar study, the prototypical pair began week 1 at a distance of 1.61m and increased in distance by 5.4% per week over the eight weeks to 2.33m ($t(51.7) = 3.288$, $p = 0.002$). Neither condition of customized avatar nor shared task beforehand showed significant effects. The prototypical pair in the customized avatar condition was 8.0% closer than the prototypical pair in the uniform avatar condition pair (1.55m vs. 1.67m), which is not larger than would be expected by chance ($t(51.92) = -1.108$, $p = 0.273$). The prototypical pair in the synchrony activity condition was 2.7% closer than the prototypical pair in the condition with no synchrony activity (1.59m, 1.61m), which is not larger than would be expected by chance ($t(50.96) = -0.364$, $p = 0.718$). Distance did differ across gender pairings (M-M = 1.43m, F-F = 1.51m, other pairs = 1.77m, $\chi^2(2, N = 2028) = 18.38$, $p < 0.001$).

The variance among session and among pair was significant, but variation among section was not significant. Variation uniquely due to section had a standard deviation of 11.8% ($\chi^2(1, N = 2028) = 2.15$, $p = 0.142$, M+SD = 1.80m, M-SD = 1.44m). Variation uniquely due to session had a standard deviation of 27.9% ($\chi^2(1, N = 2028) = 88.02$, $p < 0.001$, M+SD = 2.06m, M-SD = 1.26m). Finally, variation uniquely due to pair had a standard deviation of 34.5% ($\chi^2(1, N = 2028) = 54.9$, $p < 0.001$, M+SD = 2.17m, M-SD = 1.20m). In order to compare the fixed effects to individual differences for **RQ3**, we compare the difference in means of each to the standard deviation uniquely due to the pair. The effect due to avatar was 0.35 times and synchrony activity was 0.09 times the standard deviation of distance due to pair. Both of these indicate that distance due to pair is much more than distance due to either independent variable.

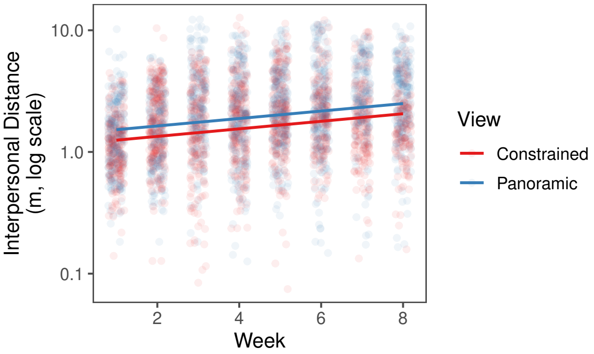

Figure 3: Plot of interpersonal distance as a function of week and view within the context study. Each pair is represented by a dot. Lines represent the prototypical pair in the constrained or panoramic condition, as indicated by the line's color.

4.1.3 Context Study

In the Context study, the prototypical pair began week 1 at a distance of 1.38m and increased in distance by 7.4% per week over the eight weeks to 2.27m ($t(162.51) = 5.744$, $p < 0.001$). The prototypical pair in the outdoor environment condition was 5.4% closer than the prototypical pair in the indoor environment condition (1.34m vs. 1.42m), which is not larger than would be expected by chance ($t(158.55) = -0.978$, $p = 0.329$). The prototypical pair in the active motion condition was 29.8% closer than the prototypical pair in the passive motion condition (1.21m vs. 1.57m), which is larger than would be expected by chance ($t(158.95) = 4.611$, $p < 0.001$). The prototypical pair in the panoramic view condition was 23.4% farther than the prototypical pair in the constrained view condition (1.53m vs. 1.24m), which is larger than would be expected by chance ($t(158.74) = 3.719$, $p < 0.001$). Distance did differ across gender pairings (M-M = 1.29m, F-F = 1.39m, other pairs = 1.42m, $\chi^2(2, N = 3313) = 9.06$, $p = 0.011$). Figure 3 shows the distribution of distances as well as the effect of week and view on distance. In regards to **RQ2**, we find evidence that panoramic views lead to larger interpersonal distance than constrained views, but no evidence that indoor or outdoor setting influences interpersonal distance.

The variance among each of section, session, and pair was significant. Variation uniquely due to section had a standard deviation of 22.4% ($\chi^2(1, N = 3313) = 16.54$, $p < 0.001$, M+SD = 1.69, M-SD = 1.13). Variation uniquely due to session had a standard deviation of 40.8% ($\chi^2(1, N = 3313) = 440.58$, $p < 0.001$, M+SD = 1.94m, M-SD = 0.98m). Finally, variation uniquely due to pair had a standard deviation of 16.1% ($\chi^2(1, N = 3313) = 19.94$, $p < 0.001$, M+SD = 1.60m, M-SD = 1.19m). In order to compare the fixed effects to individual differences for **RQ3**, we compare the difference in means of each to the standard deviation uniquely due to the pair. The effect due to environment condition was 0.35 times, motion was 1.75 times, and view was 1.41 times the standard deviation of distance due to pair.

4.2 Gaze through Head Orientation

Gaze, a useful measure of attention, can be inferred from headset direction. This is parameterized here in terms of yaw, pitch, and roll as shown in Figure 4 panel D.

4.2.1 Yaw, Pitch, and Roll

The distribution of yaw, shown in panel A of Figure 4 was relatively uniform. This reflects the fact that how one parameterizes the horizontal plane does not dramatically affect human behavior: we can rotate 30 degrees around the vertical axis and continue on

Authorized licensed use limited to: Stanford University. Downloaded on November 16,2023 at 05:39:35 UTC from IEEE Xplore. Restrictions apply.

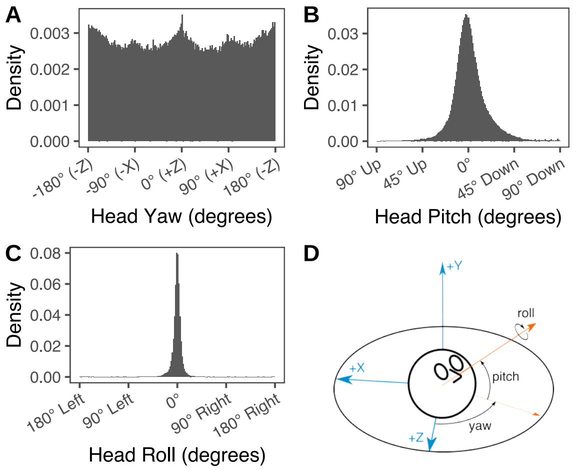

Figure 4: Panels showing histograms of Tait-Bryan angles of participants' headsets. Panel A displays yaw, Panel B displays pitch, and Panel C displays roll. Panel D displays a schematic of these angles relative to a participant's headset. Note that yaw and pitch are both in the negative direction in this schematic.

a conversation. Within this uniformity, there were apparent peaks at 90-degree intervals. These effects are discussed further in the following section.

The distribution of pitch is shown in panel B of Figure 4. Two characteristics of the distribution are noteworthy. First, the bulk of time was spent looking nearly horizontally. Second, participants exhibited a trend that looking downward was more common than looking upward for a given angular displacement from horizontal.

The distribution of roll is shown in panel C of Figure 4. This distribution was highly concentrated, with 95% of samples falling between -17 and 17 degrees.

4.2.2 Head Orientation Rectilinearity

The distribution of yaw over time shows peaks when yaw is at 0, -90, 90, or 180 degrees. These are directions aligned with the X and Z axis of the underlying coordinate system. There are no obvious indicators of these axes in the virtual space, and in principle every virtual object in the scene can be rotated all together by an arbitrary amount without any perceptible change on the part of the user.

If this is the case, how did the direction of these axes influence participant's behavior? We believe these dimensions are visible through other rectangularly aligned objects, like the walls of a room or the orientation of a bench. We suggest that an environment designer finds it easier to align rectangular world elements with an underlying rectangular grid, and the global XZ grid provides that grid for the designer.

The relation of yaw to the environment is made visually apparent by considering the data from the two studies separately, as in Figure 5. In the avatar study there was one environment, a rectangular-shaped high-ceiling room. Instructions for the week's task were posted at the far ends of the room, at +Z and -Z directions. Of these, participants looked more often in the -Z direction, as it was closer. In the Context study, there were 192 separate environments participants saw, and so there was much larger potential variation, as well as three variables that may influence head orientation.

To statistically investigate the alignment of head orientation with the different axes, we define *head orientation rectilinearity*. Given a density function $f(\theta)$, $\theta \in (-\pi, \pi]$ representing the distribution of head yaw, head orientation rectilinearity is:

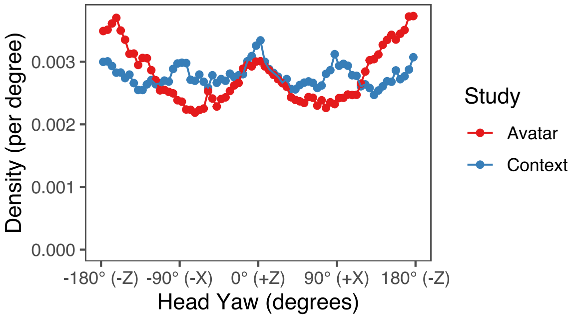

Figure 5: Distribution of yaw by study (avatar or context). Context study shows four peaks, avatar study shows two.

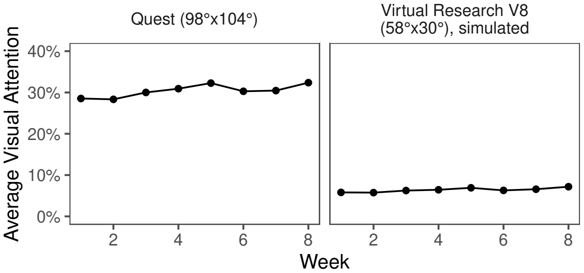

Figure 6: Proportion of time one participant was within field-of-view of the other, defined for both the headset in use (Oculus Quest 2) and the range for comparative previous work (Virtual Research V8).

$$\int_{-\pi}^{\pi} f(\theta)cos(4\theta)d\theta$$

For intuition, consider that the $cos(4\theta)$ term weights angles around 90-degree intervals positively, and angles away from those interval points negatively.

A mixed-effect model was fit to the rectilinearity data with fixed effects of week, motion ability, environment, and visible space and random effects of individual, section, and session within section. All fixed effects were tested for significance. The only statistically significant result was the intercept was different from zero, i.e., that rectilinearity was on average positive ($t(20.59) = 4.801, p < 0.001$), all other terms $p > 0.102$). Participants were more likely to be looking in directions that aligned with the horizontal axes of the global environment's coordinate system than at angles diagonal to the coordinate system.

4.2.3 Head Orientation towards Others

In previous work that investigated gaze over time, it was found that attention to others decreased over the course of the study. Considering the rarity of datasets that collect gaze over a long span of time, we investigated this same effect in the present dataset. We define visual attention for two people to be the percentage of time one person's head is within another person's field of view. Note that this is subtly different from Bailenson and Yee [4] as they use the percentage of time a person has at least one person in their field of view. Averaging this visual attention across all participants across all sessions within a week, we find that the amount of visual attention increases over time, beginning at 28.28% in week one and increasing 0.47% per week ($t(6) = 3.003$, $p = 0.024$). These values are in Figure 6 in the left panel.

Authorized licensed use limited to: Stanford University. Downloaded on November 16,2023 at 05:39:35 UTC from IEEE Xplore. Restrictions apply.

We also perform an auxiliary analysis on the same data and process, save that the size of the field-of-view is not the Quest's but rather a smaller region, namely the headset in the original work (Virtual Research V8). We do this in the case that the effect in previous work [4] is not due to the field-of-view per se, but rather to the angular dimensions that merely happened to be the field of view in that previous work. The effect was also significant and in the same direction. The average visual attention in the Virtual Research V8's field-of-view began at week one with 5.63% and increased 0.17% per week ($t(6) = 3.852$, $p = 0.008$). These values are in Figure 6 in the right panel. In regards to **RQ4**, both analyses indicate that participants looked at each other more over time.

4.3 Distance-Gaze Equilibrium

Equilibrium theory [1] posits that two people maintain a constant level of intimacy by balancing two cues, interpersonal distance and gaze directness. In previous work in naturalistic settings [35], balancing these two cues manifested as a negative correlation between indirectness of gaze and distance between people. In contrast to this previous work, we have data very rich in time. However, the non-independence of sample-level data adds significant complexity to a statistical model, and the equilibrium effect is in our case so strong that we do not attempt to summarize the distribution. Instead, we first consider only the moments in time for which the distance between participants is 3.66m or less. This threshold is the same as in [35] and stems from work by Hall [10]. From there, we randomly select a single moment in time from this thresholded set and use this moment in time as a pair-level data point.

To measure the indirectness of gaze, we follow Yee and collaborators [35] by calculating the *gaze sum*. The gaze sum is the the sum of two angles based upon two participant's head positions and orientations at a given moment. The first addend is the angle between participant A's forward vector and participant B's head with the vertex of the angle at participant A's head, and the second addend is analogous for participant B: the angle around participant B's head from participant B's forward direction to participant A's head. To measure interpersonal distance, we followed the same procedure as in Sect. 4.1 but did not log-transform the data, as the thresholding step disrupted its log-normal distribution.

The analysis was performed with a mixed-effect model with fixed effects of week, gaze sum, and dataset, and random effects of section and session within section. The prototypical participant pair at a gaze-sum of 0 degrees, that is, directly looking at each other, had an distance of 2.66m in week 1 which decreased by a non-significant amount of 0.007m per week ($t(160.1) = -1.114$, $p = 0.267$). The effect of gaze sum on the prototypical pair was -0.00113m per degree ($t(4877) = -5.57$. $p < 0.001$), meaning the prototypical pair facing the same direction (180 degree gaze sum) was 0.20m closer than the prototypical pair facing each other directly. There was also a significant effect of study on interpersonal distance, such that the prototypical pair in the Context study were 0.14m farther apart than the prototypical pair in the Avatar study ($t(19.2) = 4.511$, $p < 0.001$).

5 DISCUSSION

There are several variables that affect distance and affect gaze. In regards to distance, we found that participants increased their interpersonal distance over time **(RQ1)**. At first, this may seem counterintuitive, as participants should become closer by getting to know each other better during this experience. In contrast, we find the reverse because participants adapted to the medium. Hall, in defining proxemic spaces, gives the constraint that the larger end of conversational space is where one can hear another [10]. In these virtual environments, a majority of time was spent without 3D audio enabled. This led to no volume attenuation over distance, and so this restriction is lifted. This effectively extends conversational space much larger than exists in physical environments. Additionally, participants often needed larger ranges of spaces to complete some of the discussion activities that involved working with 3D models, which shifted the balance in favor of moving away. The opportunity to move away was important, as panoramic spaces also led to greater personal space than constrained spaces **(RQ2)**.

In regards to gaze, we found a pattern that participants tended to look in directions more aligned with the horizontal grid. This was not significantly related to any of the three variables in the context study. We hypothesize that this is either a carry-over from real-world behavior in following rectangularly-aligned seating in a room, or that focal points that draw attention are in or near the centers of walls, rather than corners. In regards to time and social attention, we found that the percentage of time one participant included the other in their field of view increased slightly over time **(RQ4)**. This was different from the results found in Bailenson and Yee [4] that found decreasing attention over time. The most likely explanation we give to this is that headsets are not as heavy as in 2006, and the visual information of looking at another avatar is more useful to the conversation than it was before, due to more fluid tracking and better inverse kinematics. It is worth noting that visual attention was not calculated in the same way (specifically, computing the percentage of participants withing view averaged over time versus the percentage of time at least one participant was in view), so one cannot make absolute comparisons even using the same field-of-view.

We also demonstrate effects due to the pair in each case **(RQ3)**. Variables such as familiarity and liking may have influenced these individual differences, and more follow-up work is necessary. One factor that did affect the distances is the gender composition. In each example, the closest avatars were male-male, followed by female-female, and the farthest pairs were all others. These results do not follow in line with previous VR research [35] and merit further investigation. Future work can also explore the cause of these per-pair effects. For example, it is possible the pair-specific differences in personal space are simply consistency, e.g. participants became comfortable in the relative positions they selected arbitrarily at the beginning of the quarter. It is also possible that certain participants knew each other beforehand and clustered together, certain participants grew to like each other, or consistent factors such as gender could make interpersonal distances predictable beforehand.

Finally, we corroborated previous work [2, 35] on personal distance and mutual gaze. Curiously, both of these values increased over time: participants got farther away while also looking at each other more. Did one cause the other? It is difficult to say. We still believe it is more likely that participants adapted to the medium in both its quirks and its novelty.

Limitations of this study include a lack of preregistration. All of these results are exploratory, as the procedure and analysis were not specified *a priori*. As a field study, there are sacrifices made to control for the sake of realism. Some of those in this dataset included the active/passive motion manipulation and consistent groups and group sizes. Additioanlly, the heterogeneity of the physical settings participants occupied while attending these sessions may have introduced unknown moderators.

While the value of week-scale time was used effectively in these analyses, the value of second- or minute-scale time was not leveraged. Future work ought to explore time-dependent ways to view proxemics, such as models for predicting dynamics or importance of distance in a moment. Future work can also more deeply investigate the inter-dyadic and inter-group effects for interpersonal distance.

6 CONCLUSION

In this work, we report findings regarding the proxemics and gaze of a large longitudinal study in social VR. Participants adapted their interpersonal distances based on affordances in the medium as well as the virtual environments in which they worked. There

Authorized licensed use limited to: Stanford University. Downloaded on November 16,2023 at 05:39:35 UTC from IEEE Xplore. Restrictions apply.

were also substantial interdyadic and intergroup differences, too. We also found that participants tended to look at each other more over time, contrary to previous longitudinal research. Taken together, these findings encourage future work in understanding adaptations to the medium of VR with more longitudinal studies as well as investigations into the inter-dyad and inter-group differences in these important aspects of human interaction. Finally, understanding human behavior in virtual reality may generalize to human behavior more broadly.

ACKNOWLEDGMENTS

The authors wish to thank Daniel Scott Akselrad, Tobin Asher, Brian Beams, Ryan Moore, Patricia Jeanne Yablonski, and Mark York for their support with the course. The authors also wish to thank Benjamin Liao, Casey Manning, Benjamin Martinez, Umar Patel, and Neha Vinjapuri for their help with developing the virtual environments for the study. This research was supported by the National Science Foundation grant (Award 1800922).

REFERENCES

[1] M. Argyle and J. Dean. Eye-Contact , Distance and Affiliation. *Sociometry*, 28(3):289–304, 1963.

[2] J. N. Bailenson, J. Blascovich, A. C. Beall, and J. M. Loomis. Equilibrium Theory Revisited: Mutual Gaze and Personal Space in Virtual Environments. *Presence: Teleoperators and Virtual Environments*, 10(6):583–598, 2001. doi: 10.1162/105474601753272844

[3] J. N. Bailenson, J. Blascovich, A. C. Beall, and J. M. Loomis. Interpersonal Distances in Virtual Environments. *Personality and Social Psychology Bulletin*, 29(7):819–833, 2003. doi: 10.1177/0146167203253270

[4] J. N. Bailenson and N. Yee. A Longitudinal Study of Task Performance, Head Movements, Subjective Report, Simulator Sickness, and Transformed Social Interaction in Collaborative Virtual Environments. 15(6):699–716, 2006. doi: 10.1162/pres.15.6.699

[5] C. M. Beaulieu. Intercultural study of personal space: A case study. *Journal of Applied Social Psychology*, 34(4):794–805, 2004. doi: 10.1111/j.1559-1816.2004.tb02571.x

[6] A. Bonsch, S. Radke, H. Overath, L. M. Asche, J. Wendt, T. Vierjahn, U. Habel, and T. W. Kuhlen. Social VR: How Personal Space is Affected by Virtual Agents' Emotions. *25th IEEE Conference on Virtual Reality and 3D User Interfaces, VR 2018 - Proceedings*, pp. 199–206, 2018. doi: 10.1109/VR.2018.8446480

[7] S. E. Brennan, J. E. Hanna, G. J. Zelinsky, and K. J. Savietta. Eye gaze cues for coordination in collaborative tasks. *DUET Workshop, CSCW'12, Seattle, Washington, USA*, pp. 11–15, 2012.

[8] Z. Choudhary, M. Gottsacker, K. Kim, R. Schubert, J. Stefanucci, G. Bruder, and G. F. Welch. Revisiting distance perception with scaled embodied cues in social virtual reality. *Proceedings - 2021 IEEE Conference on Virtual Reality and 3D User Interfaces, VR 2021*, pp. 788–797, 2021. doi: 10.1109/VR50410.2021.00106

[9] M. Garau, M. Slater, V. Vinayagamoorthy, A. Brogni, A. Steed, and M. A. Sasse. The impact of avatar realism and eye gaze control on perceived quality of communication in a shared immersive virtual environment. *Conference on Human Factors in Computing Systems - Proceedings*, (5):529–536, 2003. doi: 10.1145/642611.642703

[10] E. Hall. *The Hidden Dimension: Man's Use of Space in Public and Private*. Anchor Books, Hamburg, Germany, 1969.

[11] E. Han, M. R. Miller, C. DeVeaux, H. Jun, K. L. Nowak, J. T. Hancock, N. Ram, and J. N. Bailenson. People, places, and time: a large-scale, longitudinal study of transformed avatars and environmental context in group interaction in the metaverse. 28(2):zmac031. doi: 10.1093/jcmc/zmac031

[12] E. Han, M. R. Miller, N. Ram, K. L. Nowak, and J. N. Bailenson. Understanding group behavior in virtual reality: A large-scale, longitudinal study in the metaverse.

[13] M. Han, X. M. Wang, and S. G. Kuai. Social rather than physical crowding reduces the required interpersonal distance in virtual environments. *PsyCh Journal*, (February):1–10, 2022. doi: 10.1002/pchj.595

[14] L. A. Hayduk. Personal Space: Where We Now Stand. *Psychol. Bull.*, 94(2):293, 1983.

[15] N. Khojasteh and A. S. Won. Working Together on Diverse Tasks: A Longitudinal Study on Individual Workload, Presence and Emotional Recognition in Collaborative Virtual Environments. *Frontiers in Virtual Reality*, 2(June):1–24, 2021. doi: 10.3389/frvir.2021.643331

[16] B. McSweeney. Hofstede's model of national cultural differences and their consequences: A triumph of faith - A failure of analysis. *Human Relations*, 55(1):89–118, 2002. doi: 10.1177/0018726702551004

[17] M. R. Miller, H. Jun, and J. N. Bailenson. Motion and Meaning: Sample-Level Nonlinear Analyses of Virtual Reality Tracking Data. In *2021 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pp. 147–152. IEEE, Bari, Italy, 2021.

[18] F. Moustafa and A. Steed. A longitudinal study of small group interaction in social virtual reality. *Proceedings of the ACM Symposium on Virtual Reality Software and Technology, VRST*, 2018. doi: 10.1145/3281505.3281527

[19] J. Mütterlein, S. Jelsch, and T. Hess. Specifics of collaboration in virtual reality: How immersion drives the intention to collaborate. *Proceedings of the 22nd Pacific Asia Conference on Information Systems - Opportunities and Challenges for the Digitized Society: Are We Ready?, PACIS 2018*, 2018.

[20] A. Ozdemir. Shopping malls: Measuring interpersonal distance under changing conditions and across cultures. *Field Methods*, 20(3):226–248, 2008. doi: 10.1177/1525822X08316605

[21] Y. Pan and A. Steed. Effects of 3d perspective on head gaze estimation with a multiview autostereoscopic display. 86:138–148. doi: 10.1016/j.ijhcs.2015.10.004

[22] Y. Pan and A. Steed. A gaze-preserving situated multiview telepresence system. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 2173–2176. ACM. doi: 10.1145/2556288.2557320

[23] A. Plopski, T. Hirzle, N. Norouzi, L. Qian, G. Bruder, and T. Langlotz. The Eye in Extended Reality: A Survey on Gaze Interaction and Eye Tracking in Head-worn Extended Reality. *ACM Computing Surveys*, 55(3):1–39, 2023. doi: 10.1145/3491207

[24] E. Prytz, S. Nilsson, and A. Jönsson. The importance of eye-contact for collaboration in AR system. *9th IEEE International Symposium on Mixed and Augmented Reality 2010: Science and Technology, ISMAR 2010 - Proceedings*, pp. 119–126, 2010. doi: 10.1109/ISMAR.2010.5643559

[25] B. Reeves, L. Yeykelis, and J. J. Cummings. The Use of Media in Media Psychology. *Media Psychology*, 19(1):49–71, 2016. doi: 10.1080/15213269.2015.1030083

[26] D. Roth, C. Klelnbeck, T. Feigl, C. Mutschler, and M. E. Latoschik. Beyond Replication: Augmenting Social Behaviors in Multi-User Virtual Realities. *25th IEEE Conference on Virtual Reality and 3D User Interfaces, VR 2018 - Proceedings*, pp. 215–222, 2018. doi: 10.1109/VR.2018.8447550

[27] D. Roth, J. L. Lugrin, D. Galakhov, A. Hofmann, G. Bente, M. E. Latoschik, and A. Fuhrmann. Avatar realism and social interaction quality in virtual reality. *Proceedings - IEEE Virtual Reality*, 2016-July:277–278, 2016. doi: 10.1109/VR.2016.7504761

[28] M. Slater, A. Sadagic, M. Usoh, and R. Schroeder. Small-group behavior in a virtual and real environment: A comparative study. *Presence: Teleoperators and Virtual Environments*, 9(1):37–51, 2000. doi: 10.1162/105474600566600

[29] K. Takahashi, Y. Inoue, and M. Kitazaki. Interpersonal Distance to a Speaking Avatar: Loudness Matters Irrespective of Contents. *Proceedings - 2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops, VRW 2022*, pp. 774–775, 2022. doi: 10.1109/VRW55335.2022.00238

[30] R. Vertegaal and Y. Ding. Explaining effects of eye gaze on mediated group conversations: Amount or synchronization? *Proceedings of the ACM Conference on Computer Supported Cooperative Work*, pp. 41–48, 2002.

[31] R. Vertegaal, R. Slagter, G. Van Der Veer, and A. Nijholt. Eye gaze patterns in conversations: There is more to conversational agents than meets the eyes. *Conference on Human Factors in Computing Systems - Proceedings*, pp. 301–308, 2001.

Authorized licensed use limited to: Stanford University. Downloaded on November 16,2023 at 05:39:35 UTC from IEEE Xplore. Restrictions apply.

[32] J. B. Walther. Computer-Mediated Communication: Impersonal, Interpersonal, and Hyperpersonal Interaction. *Communication Research*, 23(1):3–43, 1996.

[33] J. R. Williamson, J. Li, and V. Vinayagamoorthy. Proxemics and social interactions in an instrumented virtual realityworkshop. *Conference on Human Factors in Computing Systems - Proceedings*, 2021. doi: 10.1145/3411764.3445729

[34] H. E. Yaremych and S. Persky. Tracing physical behavior in virtual reality: A narrative review of applications to social psychology. *Journal of Experimental Social Psychology*, 85(April):103845, 2019. doi: 10.1016/j.jesp.2019.103845

[35] N. Yee, J. N. Bailenson, M. Urbanek, F. Chang, and D. Merget. The unbearable likeness of being digital: The persistence of nonverbal social norms in online virtual environments. *Cyberpsychology and Behavior*, 10(1):115–121, 2007. doi: 10.1089/cpb.2006.9984

Authorized licensed use limited to: Stanford University. Downloaded on November 16,2023 at 05:39:35 UTC from IEEE Xplore. Restrictions apply.