

**Social Virtual Reality Elicits Natural Interaction Behavior with Personalized and Generic
Avatars**

Gayoung Son and Marius Rubo

Cognitive Psychology, Perception and Research Methods, University of Bern, Switzerland

Author Note

Correspondence concerning this article should be addressed to Gayoung Son, University of Bern,
Department of Psychology, Fabrikstrasse 8, 3012 Bern. E-mail: gayoung.son@unibe.ch

Abstract

Social Virtual Reality (VR) allows to interact in shared virtual environments while embodying computerized avatars which display behavior in real-time. The technique mimics real social interactions in its preservation of the spatial relatedness of social gaze and other facets of non-verbal behavior, but the extent to which people behave naturally in such artificial situations remains largely unknown. Here we show in 128 participants who interacted in dyads that the coordination of gaze and speech behavior closely follows patterns known from face-to-face interactions: eye gaze to a partner's eye region was relatively enhanced while listening compared to while speaking and at the end of a speaking turn compared to the beginning of a turn. Gaze, speech and smiling behavior were sensibly adapted to differing conversation topics (small talk, personal talk, talk about conflicting opinions). In contrast to written communication on the internet, anonymization – here realized using generic as opposed to personalized avatars – was not associated with behavioral disinhibition or any differences in subjective experience, possibly due to a closeness-generating effect of direct eye contact despite the concealment of one's own and the interaction partner's identity. Our results indicate that social VR elicits natural interaction behavior and may be used to implement anonymized face-to-face interactions without the negative side-effects often associated with anonymization.

Keywords: Social Virtual Reality, Avatar Appearance, Nonverbal Communication, Gaze Behavior, Social Interaction

Social Virtual Reality Elicits Natural Interaction Behavior with Personalized and Generic Avatars

Social virtual reality (Social VR) is a form of digital communication which aims to closely mimic real-life social interactions (Hennig-Thurau et al., 2023; Maloney et al., 2020; McVeigh-Schultz et al., 2019; Y. Pan & Steed, 2017; Rogers et al., 2022; Smith & Neff, 2018; Wei et al., 2022): inhabiting a shared virtual environment as virtual avatars with others, people may express their non-verbal communication behavior in its natural integration in social space, including deictic gestures and gaze communication. While social VR may approximate real-life social situations in its replication of a shared environment, it might lead to a disruption of the expression of the natural interaction behavior. After all, social VR represents an artificial situation where individuals encounter each other as computerized avatars which, despite advances in visual quality (Weidner et al., 2023), cannot be mistaken for reality. Furthermore, the ability to customize avatars introduces a layer of anonymity that may affect how the interaction unfolds, as users can conceal their real identities or conversely present themselves as they appear in real-life. Although it is commonly argued that VR offers more ecologically valid settings in which to study natural behavior (Parsons, 2015), the extent to which interactions in VR are truly naturalistic remains relatively unclear.

Over the past years, several studies aimed to empirically assess the extent to which social interaction in social VR can be seen as naturalistic. Previous studies in the field often focused on subjective experiences to assess the quality of interaction in VR, with fewer studies assessing objective behavior in participants interacting in VR environments (Xenakis et al., 2022; Yaremych & Persky, 2019). Miller et al. (2023) found that participants in social VR kept greater interpersonal distances in larger rooms, a behavior known from real-life situations. Despite only having gesture and audio cues, participants embodying a robotic avatar in social VR displayed similar speech patterns to face-to-face interactions, reflected in higher rates of interruptions compared to interactions with no visible avatars (Smith & Neff, 2018). Attentional processes were sometimes assessed based on head orientation (e.g., Han et al., 2022; Ichino et al., 2023; Miller

et al., 2023; Smith & Neff, 2018), with only few studies assessing eye gaze behavior which is known to more accurately represent attention allocation (Abdullah et al., 2021; Ma & Pan, 2022).

Overall, findings on interaction behavior in social VR remain sparse and the field has not yet fully connected with broader accounts on interaction psychology, which, mostly by analyzing videotapes of real face-to-face interactions, have compiled a rich and detailed body of knowledge on face-to-face interactions throughout the last decades (Aron et al., 1997; J. A. Hall & Knapp, 2013; J. A. Hall et al., 2019). A comparison of the patterns of interaction behavior in social VR with patterns previously observed real-life situations may help to assess the extent to which social VR fosters realistic interactions.

A robust behavioral pattern observed in real-life interactions is an increased time spent gazing at a partner's eyes while listening to the partner compared to while speaking (e.g., Degutyte & Astell, 2021), which was also observed in social VR, albeit in a triadic interaction (Abdullah et al., 2021). Additionally, people in real-life interactions tend to gaze more frequently towards their partner's eyes at the end of speaking turns compared to the beginning of speaking turns (Hessels et al., 2019; Ho et al., 2015; Kendon, 1967) in order to manage turn-taking. This mechanism was replicated in principle in an interaction with a computer agent shown on a computer screen where the avatar's gaze avoidance at the beginning of turns was shown to influence turn-taking (Andrist et al., 2013), but little is known about the presence of this pattern in people interacting with each other through social VR.

People may additionally vary their interaction behavior depending on the type of conversation in which they are involved. For example, Rutter et al. (1978) found that participants gazed more at the end of a turn during a competitive conversation (in which participants were instructed to persuade the partner) compared to a cooperative conversation (in which participants discussed shared opinions). On the other hand, participants tend to divert their gaze when responding to personal (Exline et al., 1965) or challenging questions (Glenberg et al., 1998). Further, people may modulate the extent of their speaking to express dominance over a partner (J. A. Hall et al., 2005) and may smile to express positive affect, interest or trust (DeWall et al.,

2009; Krumhuber et al., 2007). While it was observed that computer characters which showed a smiling expression were similarly perceived as more positive (Oh et al., 2016), little is known about the use of smiling behavior in social VR.

Social VR setups additionally allow manipulation of social interaction properties in ways that have no direct equivalent in real-life interactions. A manipulation which attracted attention in VR research is the design of virtual avatars, which may visually resemble participants' physical appearance (described as personalized) but may likewise be given any other physical appearance (described as generic). Studies overall observed positive effects of avatar personalization on self-reported presence, social presence or embodiment (Waltemate et al., 2018; Weidner et al., 2023), although not all studies observed such an effect (e.g., Latoschik et al., 2017; Ma & Pan, 2022; Salagean et al., 2023; Suk & Laine, 2023).

While the mentioned studies on effects of avatar personalization were conducted on participants who entered a virtual environment in isolation, one previous study investigating group behavior in social VR (Han et al., 2022) observed increased self-reported presence but decreased enjoyment in participants using personalized as compared to generic avatars in interactions with others. To the best of our knowledge, no studies have yet directly investigated effects of avatar personalization on fine-grained gaze and speech behavior in social VR. The embodiment of generic avatars, which conceal a person's identity relative to personalized avatars, may be expected to exert a disinhibiting influence on social interaction behavior, similar to disinhibition effects observed in anonymous written communication on the internet (Suler, 2004). Such effects are thought to originate from a detachment of social information processing from regulatory processes as they naturally unfold in the presence of others. While disinhibition in written communication may be expressed in a more extreme choice of words, behavior in face-to-face interactions can be more subtly regulated in the balancing of inhibition and disinhibition. For instance, social inhibition was found to be associated with an aversion of eye contact (Chen & Clarke, 2017), whereas disinhibition was linked to louder speaking (J. A. Hall et al., 2005) and more interrupting behavior (Keltner et al., 2003).

The Present Study

The present study investigates whether established behavioral patterns known from face-to-face interactions are likewise traceable in a social VR environment, indicating the extent to which interactions in social VR may be seen as natural or lifelike, and whether the use of generic avatars leads to disinhibited social behavior. We implemented a comprehensive display and recording of behavioral information (with visible head and eye movements, hand and finger movements, facial expressions and lip synchronization) as recommended by (X. Pan & Hamilton, 2018; Wei et al., 2022; Weidner et al., 2023), and focused on an in-depth analysis of interaction behavior. A key variable of interest was eye gaze behavior which was described as a particularly important communication channel in navigating dyadic interactions (Argyle & Cook, 1976; Degutyte & Astell, 2021), but we also assessed speaking and smiling behavior.

Participants engaged in dyadic interactions on different conversational topics, either embodying personalized or generic avatars. We hypothesized that participants would show behavioral patterns known from face-to-face interactions (more gaze at a partner's eye region while listening compared to while speaking and at the end of a turn compared to the beginning of a turn), and that participants embodying generic as compared to personalized avatars would show signs of disinhibition (more gaze to the eye region, louder speaking, more interrupting) and report lower presence, social presence and body ownership. Analyses testing these hypotheses were complemented with a presentation of several social behavioral measurements within and across conversation topics to allow for a comprehensive characterization of behavior in social VR.

Materials and Methods

Participants

128 participants (105 female, 22 male, 1 diverse, aged 18 to 51 years, $M = 22.60$, $SD = 3.57$) interacted in dyads in social VR. The sample size was predefined to allow for the detection of between-group effects of $d = 0.5$ with a probability of 80% with alpha set to .05. The majority (122 students, 95.31%) were psychology students enrolled at the University of Bern, at the Bachelor level. Exclusion criteria included neurological illnesses, use of centrally acting agents

and strong eye sight problems which require glasses that do not fit under a VR equipment. The study was approved by the ethics committee of the University of Bern and was preregistered (<https://osf.io/p53fu>). Data and scripts are available at (<https://osf.io/fyr2q/>). Before the preregistered sample size was reached, we excluded one dyad which reported that they spoke English in the interaction, another dyad in which the computers lost connection during the interaction, and seven dyads in which the system failed to transmit facial expressions. Participants reported various degrees of prior experience with VR. 55 participants had no experience with VR (42.97%), 43 had one prior experience (33.59%), 23 reported two to three experiences (17.97%) and 7 participants disclosed 3 or more experiences with VR (5.47%). No participants reported regular VR use.

Each dyad was randomly assigned to either a condition where both participants embodied either personalized or generic avatars. All dyads conversed on three topics in a randomized order (small talk, personal talk and opinion talk; see section Tasks). The entire procedure and data pre-processing were pre-registered. Pre-registered hypotheses were complemented with novel hypotheses (e.g., a more in-depth comparison of gaze to the eye region at the beginning and end of a turn) to provide a more comprehensive picture of participant behavior while also controlling for multiple comparisons.

Implementation

Apparatus and Software

In each testing, two desktop laptops (Windows 11, NVIDIA GeForce RTX 4070 GPU, Intel Core i9-13900HX CPU, 32 GB RAM) were used to run a custom social VR software developed in Unity 3D (<https://unity.com>). The computers were connected via Ethernet and communicated in a shared local area network (LAN). Network communication was developed with *Fish-Net networking* (<https://github.com/FirstGearGames/FishNet/>). Network tick rate was set to 120 Hz. We monitored round-trip-times (RTT) for signals throughout all sessions. RTT was defined as the total time elapsed from when a signal was processed in a client's FixedUpdate-Loop, sent to the server, returned to and received by the client. Across all sessions,

the average RTT was 47.85 ms, with 95% of measured RTT values falling between 27.50 ms and 71.00 ms. To visualize the VR environment, Meta Quest Pro (Model: DK94EC; <https://www.meta.com>) head-mounted displays (HMDs) were used. These HMDs feature sensors to record eye movements, facial expressions, hand gestures and voice data, which was collected at 100 Hz and rendered onto the participant's avatar concurrently. Audio communication was realized using *Dissonance Voice Chat* (<https://placeholder-software.co.uk/>).

Avatar Creation

Both personalized and generic avatars were created using Character Creator 4 (CC4; <https://www.reallusion.com/character-creator/>) and the Headshot plugin. This tool generates a realistic 3D model from a single portrait photograph. For each avatar, we manually adjusted facial details such as eye color and blendshapes as well as hair and body size. The avatars were dressed identically. Only the head, torso, and white silhouettes of the hand were visible in the virtual environment, which is a common configuration in social VR environments and is often preferred over full-body visualizations (Weidner et al., 2023). In the personalized condition, we used the participant photo to generate a look-alike avatar. In the generic condition, twelve different generic avatars (6 female, 6 male) were created based on artificially created portraits of non-existent individuals (Karras et al., 2018, <https://this-person-does-not-exist.com/en>) and were selected randomly for each participant with only gender matched. Avatars were exported as FBX (Filmbox) files incorporating wrinkle textures, the CC4 extended expression profile and 15 visemes (1:1 Direct system; <https://manual.reallusion.com/Character-Creator-4>). The CC4 extended expression profile includes 140 blendshapes to control nuanced facial expressions. See Figure 1 for examples of personalized avatars generated from portraits and generic avatars.

Head movements, eye movements, hand and finger movements as well as facial expressions were directly mapped onto avatars based on live streamed data from Meta's API (<https://developer.oculus.com/>). In addition, lip synchronization was realized based on audio streams from the microphone using Salsa Lip-Sync (<https://crazyminnowstudio.com>). Since delay of audio transmission often exceeds delay of transmitting individual values in digital networks

Figure 1

Close-up portrait photographs of two experimenters (first column), personalized avatars generated from these photographs (second column) and two generic avatars (third column).



and lip synchronization likewise has a delay of several hundred milliseconds, each computer performed lip synchronization computations for data from its own microphone and transmitted resulting phonemes to the other client for direct implementation.

Procedure

Participants received an information email beforehand asking them not to wear excessive make-up, which could interfere with the avatar creation process and were asked to bring contact lenses if they needed eyesight correction. The two participants of a dyad did not interact, see or hear each other physically at any point in the experiment but were located in different rooms while connected via social VR. Additionally, participants arrived and departed at staggered times to further reduce the possibility of encountering each other. In the current study, most participants

did not recognize each other: Only five out of 128 participants answered that they recognized their partner (3.91%). Two dyads, both in the personalized condition, answered they were previously acquainted with their respective partners. One participant, who was in the generic condition, did not specify their relationship.

On arrival to the lab, participants were informed about the experiment and signed the consent form. Participants in the personalized condition were additionally informed that they would be embodying an individualized avatar generated from their own photos. We then took a portrait photo of these participants and created a personalized avatar as the participants were filling out questionnaires. All participants were fitted with a heart rate sensor, answered a short demographic survey and completed a battery of questionnaires on individual traits which are not subject to the present analysis. After answering the questionnaires, participants were introduced to the VR headset. They were guided on how to fit the headset and adjust the interpupillary distance (IPD) setting. They then performed a 9-point eye-tracking calibration and checked the calibration outcomes. Once participants were comfortable with the headset, they were introduced to the virtual environment where they first adjusted their seating height and orientation so that a table in the virtual environment was aligned in orientation to the physical table in front of them. Located on the virtual table, a mirror in the virtual environment allowed participants to see their own avatar and its replication of their head, eye, hand and finger movements along with their facial expressions. Based on participant feedback, an experimenter adjusted their avatar's body size, head pitch, re-adjusted the eye tracking if necessary and gave them time to look around and get accustomed to the situation. Note that incorporating a wide range of behavioral modalities is essential for creating a realistic and engaging social VR experience (Xenakis et al., 2022) and that previous research reported on reduced social presence when facial expression displays were omitted in social VR (Kimmel et al., 2023; Tarnec et al., 2023).

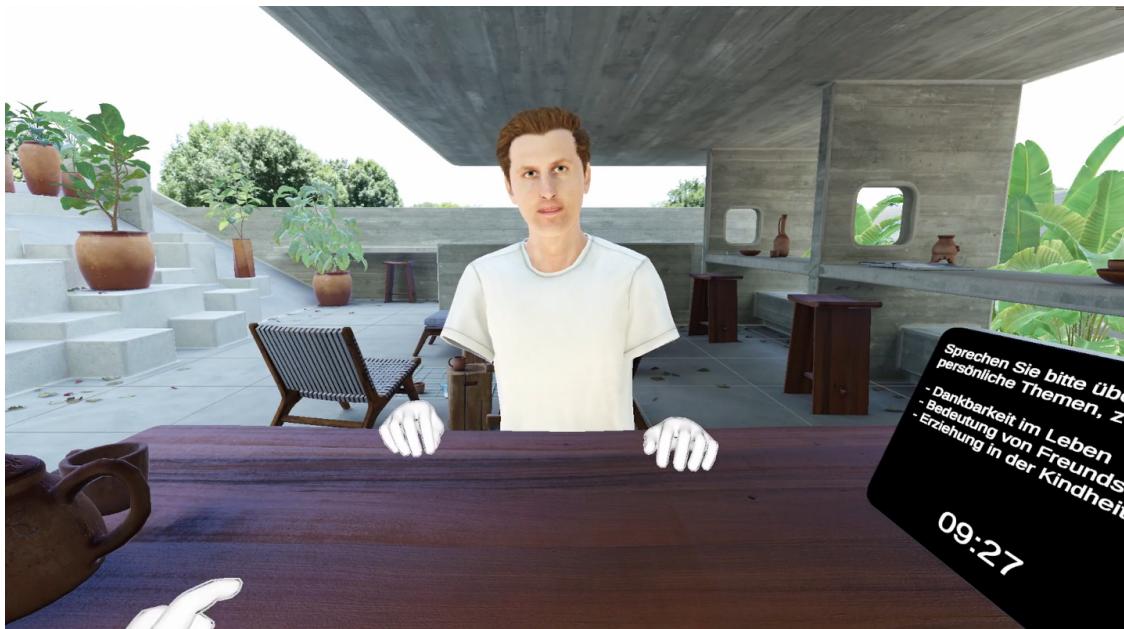
Before connecting the two participants, experimenters briefly outlined the various conversation topics and assured participants that no video or audio recordings would be obtained. The experimenters then initiated the connection with the other participant in the social VR

program and exited the room. After connection, the mirror on the table was moved to another side of the room and participants could see their partner's avatar located on the other side of the table. Participants were now able to interact with each another (see Figure 2 and Figure 3). Instructions were delivered through a computerized female voice. Following a brief introduction in which the participants were reminded not to reveal their own names to their partner, the dyads were instructed to talk about each topic for 10 minutes. The topics were (a) *small talk* involving topics on everyday subjects, (b) *personal talk* including closeness-generating topics and (c) *opinion talk* including a discussion on conflicting opinions. The rationale behind asking participants to engage in differing conversations was to assess modulations of behavior according to conversation types (e.g., speaking amplitude) but also to test the robustness of general behavioral markers such as the interaction of speech and gaze across contexts. For each conversation topic, a computerized voice presented the topic and recited example questions. A keyword for each question was displayed on a small screen on the virtual table which also showed the remaining time for each topic. For the small talk and personal talk topics, example questions were adapted from Aron et al. (1997). For opinion talk, participants were directed to choose an issue on which they held divergent views and to attempt to persuade their partner of their own perspective. We selected the three issues with the highest variance in response distributions from a pilot survey of university students to increase the probability that partners would hold divergent views to at least one of the proposed topics. These discussions were informed by research showing that opinion discrepancies among group members can induce discomfort and stress, as evidenced in studies where distress arises from perceived disagreement on controversial topics (Matz & Wood, 2005). The questions of each topic in English and in the experimental language are listed in Supplementary Table S1.

After all topics were discussed, participants were asked to say goodbye to each other and were disconnected in the social VR setup, resulting in the partner's avatar disappearing from the scene for each participant. They then viewed a set of images in the virtual scene for about three minutes which are not covered here, before removing the headset and completing a final battery of questionnaires (see section Self-report Measures below). The entire study lasted around 1.5

Figure 2

Interaction in social VR as seen from a participant's perspective.



hours for both conditions.

Behavioral Measures

Behavioral variables

During the VR interactions, we continuously recorded markers of verbal and non-verbal behavior as they were described in research on human interactions in face-to-face settings, i.e., with no use of VR technology (Degutyte & Astell, 2021; Hessels et al., 2019). Table 1 in sec. Results summarizes the variables of interest in this study.

Data Preprocessing

For gaze behavior, we were primarily interested in the proportion of time spent looking towards the partner's eye region, which was defined as an area of 2 degrees around the center of either eye. To categorize gaze as being directed towards the partner's eye region or elsewhere at each moment in time, we first removed data during blinks (which accounted for an average 3.85% ($SD = 2.81$) of the time across participants, averaged across both eyes for each participant) as well

Figure 3

Two individuals engaging in a conversation in social VR (top image) while being located in different rooms (bottom images). Photos are taken from a third-person perspective but individuals in social VR view the scene from their own first-person perspective as in real-life situations.



as 100ms before and after each blink. Data was interpolated using Piecewise Cubic Hermite Interpolating Polynomial (PCHIP) and a 2nd order Butterworth filter with a cutoff frequency of 10Hz was applied. Gaze data were transposed to represented angular deviations from the center of each of the partner's eyes. To ensure measurement sensitivity towards this relatively small area of interest for a wide range of participants including those with strong eye dominance or alignment issues (e.g., strabismus), we considered at each moment in time the gaze ray from either eye or a combined gaze ray with the lowest angular deviation towards either of the partner's eyes.

Categorization of time points to phases where a participant was speaking or not speaking was based on the sound amplitude received at the HMD's built-in microphone on a scale from 0 (minimum amplitude) to 100 (maximum amplitude). The signal was first smoothed using a moving average filter with a width of 100ms. Time points with an amplitude of 2 or above and time points within 100ms of these time points were provisionally labelled as speaking. Pauses between two speaking phases of up to 2 seconds were considered to represent the same speaking turn and were thus also labelled as speaking. Speech phases lasting shorter than 1 second were then considered to be backchanneling vocalizations (e.g., "hmm" or "uh-huh") and were labelled as not speaking. Moments when a person was not speaking but the partner was speaking were labelled as listening. Moments when both partners were speaking were considered interruptions, with the participant who was listening directly prior to the interruption being the person interrupting their partner.

Smiling was determined based on data acquired through the HMD's face tracking feature which uses action units based on the Facial Action Coding System (FACS: <https://developer.oculus.com/documentation/unity/move-face-tracking/>). Extent of smiling on both sides of the mouth was averaged to one value between 0 (no smiling expression) and 100 (maximum smiling expression).

Measurements were averaged within each conversation in each participant. Outliers beyond 3 standard deviations from the mean were set to 3 standard deviations from the mean. The percentage of corrected outliers in behavioral variables ranged from 0% to 2.08% (see

Supplementary Table S4 for details).

Self-report Measures

After the interaction, the participants answered questions on the quality of the interaction, avatar appearance and their VR experience. Additional questions on the perception of their partner's traits are described in a separate report. For questions related to the quality of the interaction and the avatar appearance, we developed a set of questions which are detailed in Supplementary Materials (see sec. Questionnaires).

Quality of Interaction

For each of the three topics, participants evaluated their success as a dyad in adhering to the specific topic, their own contribution to the topic and their partner's contribution using a 5-point Likert scale ranging from 1 (*not at all*) to 5 (*extremely*). To further assess adherence to the instructions, we assessed the intimacy level of personal topic conversations compared to small-talk conversations and participants rated who was more convincing in discussions involving differing opinions on the same scale. An open question was included to gather participants' perceptions of the transitions between topics.

Avatar Appearance

In order to assess whether the manipulation of the avatar appearance was successful, participants responded to eight items regarding their interaction with and perceptions of their virtual avatars. All items were rated on a 5-point Likert scale from 1 (*not at all*) to 5 (*completely*). The questions assessed the avatar's visual resemblance with the participant's own physical appearance (*similarity*); satisfaction with the avatar's appearance (*satisfaction*); perceptions of the avatar being more or less attractive than oneself (*attractiveness, unattractiveness*); whether friends would be able to recognize the avatar as a representation of oneself (*recognizability*) and assumed ability to re-identify the person they interacted with if encountered again (*re-identification*). Additionally, the clarity of auditory communication with their interaction partner was assessed to inquire if sound transmission functioned properly (*audibility*).

VR Experience

Self presence was assessed with the General Presence and Spatial Presence subscales subscales of the IGroup Presence Questionnaire (IPQ; Schubert et al., 2001). The questionnaire is comprised of six items on a 7-point Likert scale ranging from 1 (*fully disagree*) to 7 (*fully agree*), for example "In the computer generated world I had a sense of being there" and "I felt present in the virtual space". Internal consistency of this questionnaire was acceptable (Cronbach's $\alpha = .64$).

Social presence was assessed with the Social Presence Scale (SPS; Herrera et al., 2018), which included five items using a 7-point Likert scale ranging from 1 (*strongly disagree*) to 7 (*strongly agree*), such as "I felt like I was face-to-face with my partner" and "I felt like my partner was watching me". Internal consistency of this scale was good (Cronbach's $\alpha = .79$).

Embodiment was evaluated using the Body Ownership subscale of the Embodiment Questionnaire (Gonzalez-Franco & Peck, 2018). This questionnaire assesses the participant's sense of identification with their avatar's body with five items on a 7-point Likert scale ranging from 1 (*strongly disagree*) to 7 (*strongly agree*). Examples are "I felt as if the body I saw in the virtual world might be my body" and "I felt as if the virtual body I saw when looking in the mirror was my own body". Internal consistency of this scale was good (Cronbach's $\alpha = .73$).

We additionally included a questionnaire on motion sickness in VR. Participants who answered yes to a screening question ("Did you feel any discomfort in VR?") were given the VR Sickness Questionnaire (VRSQ; Kim et al., 2018). Participants were asked to report on nine common cybersickness symptoms on a 4-point Likert scale ranging from 0 (*not at all*) to 3 (*very much*), which reflect two components of motion sickness (oculomotor and disorientation). The total score is calculated as an average of the two components, ranging from 0 to 100. Internal consistency of the total scale was good (Cronbach's $\alpha = .74$). 23 out of 128 (18%) participants replied that they felt some discomfort in VR. Mean cybersickness score of these participants were low ($M = 26.89$, $SD = 13.67$, ranging from 6.67 to 66.89), indicating that the VR experience was well-tolerated by the majority of participants.

Data Analysis

The influence of conversational topics and experimental condition on behavioral measures were examined using linear mixed-effects models. The model looked at the fixed effect of the independent variable of interest on the behavioral variables while accounting for the variability of the conversation topic within individuals, as reflected by random intercepts for each participant ID and random slopes for conversational topics nested within ID. Significance of the effects were determined with an *F*-Test with an α of 5%. We additionally report adjusted *p*-value computed using the Benjamini and Hochberg (1995) procedure, which controls for the False Discovery Rate (FDR). Effects of conversation topics were followed up using post-hoc pairwise comparisons using estimated marginal means (EMM). Reported *p*-values were adjusted for multiple comparisons using the Tukey's HSD (Honestly Significant Difference). The effect of condition on self-report measures were inspected using linear regression models.

Data were analysed using R (version 4.3.2, R Core Team, 2024). We used the nlme package (Pinheiro et al., 2023) to fit mixed-effects models with nested random effects and the emmeans package (Lenth, 2024) for pairwise comparisons. All dependent variables were standardized and centered prior to regression in order to facilitate comparison of effects. For effect sizes of differences between conversation topics, Cohen's *d* was calculated with the mean difference between paired observations, divided by the adjusted standard deviation (taking the variances of the measures and their correlation into account). For effect sizes of differences between conditions, Cohen's *d* was calculated for independent groups, where the mean difference was divided by the pooled standard deviation.

Results

Coordination of Speech and Gaze across Participants

Descriptive results of the behavioral variables aggregated over all three conversation topics are shown in Table 1. Participants gazed at their partner significantly less during speaking than listening ($\beta = -26.41$, 95% CI [-27.71, -25.10], $t(639) = -39.78$, $p < .001$, $d = 1.80$, 95% CI [-1.97, -1.63]), indicating that participants spent 26.41% less time gazing at their partner during

speaking compared to listening. Additionally, participants were more likely to gaze at their partner at the end of a speaking turn compared to the beginning of a turn ($\beta = 17.24$, 95% CI [15.99, 18.49], $t(639) = 27.07$, $p < .001$, $d = 1.28$, 95% CI [-1.44, -1.12]), showing that participants gazed at their partner 17.24% more at the end of a speaking turn.

Influence of Conversation Topics on Interaction Behavior

Analyses on the quality of interaction indicated that on average, participants reported to have engaged effectively with their partners on each topic (for details see Supplementary Results sec. Quality of Interaction). Linear mixed models examining the effect of conversation topic on each behavioral measure were adjusted for multiple testing and were significant for all measures. Based on the significance of models, post hoc pairwise comparisons were conducted between the conversation topics for each measure. Figure 4 presents the means and standard errors of the measures for each conversation topic (for details see Supplementary Table S5).

We found a significant effect of conversation topic on gaze while speaking ($F(2, 254) = 6.70$, $p = .002$, $p_{adj} = .002$). Participants gazed significantly more at their partner's eye region while speaking in small talk compared to opinion talk ($p = .007$, $d = 0.32$, 95% CI [0.14, 0.50]). The difference in gaze between small talk and personal talk was marginally significant ($p = .056$, $d = 0.20$, 95% CI [0.03, 0.38]). There was no significant difference in gaze between personal talk and opinion talk ($p = .174$, $d = 0.16$, 95% CI [-0.33, 0.02]).

We observed a significant effect of conversation topic on gaze while listening ($F(2, 254) = 8.96$, $p < .001$, $p_{adj} < .001$). Participants gazed significantly more at their partner while listening in personal talk compared to small talk ($p = .007$, $d = 0.27$, 95% CI [-0.45, -0.09]) or opinion talk ($p = .001$, $d = 0.31$, 95% CI [-0.49, -0.13]). There was no significant difference in gaze between small talk and opinion talk ($p = .835$, $d = 0.05$, 95% CI [-0.12, 0.22]).

The effect of conversation topic on gaze at the beginning of a turn was significant ($F(2, 254) = 38.80$, $p < .001$, $p_{adj} < .001$). Post hoc pairwise comparisons were all significant, such that participants gazed significantly more at their partner in the first second of their turn in small talk compared to personal talk ($p < .001$, $d = 0.33$, 95% CI [0.15, 0.51]) or opinion talk ($p < .001$, $d =$

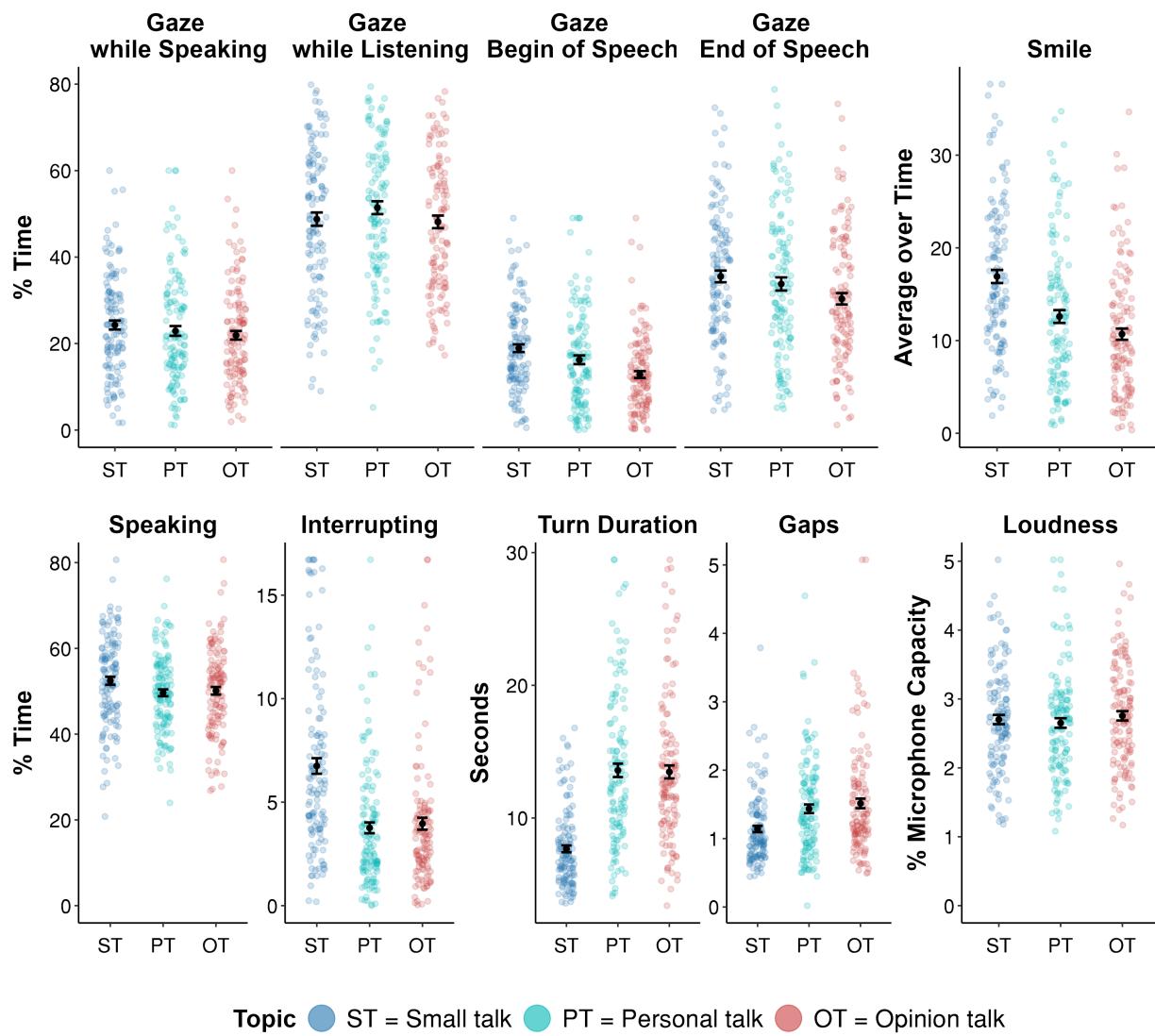
Table 1*Means and Standard Deviations For Behavioral Variables*

Variable	Description	M	SD	95% CI
Gaze While Speaking	Proportion of one's own speaking time which is spent gazing towards the partner's eye region	23.05	11.96	[21.84, 24.25]
Gaze While Listening	Proportion of one's own listening time which is spent gazing towards the partner's eye region	49.45	16.96	[47.75, 51.15]
Gaze Beginning of Turn	Proportion of the first second of one's own speaking turn which is spent gazing towards the partner's eye region	16.01	10.49	[14.95, 17.06]
Gaze End of Turn	Proportion of the last second of one's own speaking turn which is spent gazing towards the partner's eye region	33.23	15.93	[31.63, 34.83]
Smile	Average magnitude of smiling facial expression over time	13.40	7.99	[12.60, 14.20]
Speaking	Proportion of time spent speaking	50.73	9.90	[49.74, 51.72]
Interrupting	Proportion of time spent interrupting the partner's speech	4.82	3.79	[4.44, 5.20]
Turn Duration	Average length of one's speech turn duration in seconds	11.58	5.61	[11.02, 12.14]
Gaps	Average length of pause before responding in seconds	1.36	0.70	[1.29, 1.43]
Loudness	Average amplitude of speech relative to microphone maximum recording capacity	2.70	0.77	[2.63, 2.78]

Note. *M* = mean. *SD* = standard deviation.

Figure 4

Means and standard errors of gaze and speech behavior in individual conversation topics.



Note: Error bars represent standard errors.

0.78, 95% CI [0.58, 0.98]), and in personal talk compared to opinion talk ($p < .001$, $d = 0.38$, 95% CI [-0.56, -0.20]).

The effect of conversation topic on gaze at the end of a turn was significant ($F(2, 254) = 9.04$, $p < .001$, $p_{adj} < .001$). In the last second of their turn, participants gazed significantly less at their partner in opinion talk compared to small talk ($p < .001$, $d = 0.37$, 95% CI [0.19, 0.55]) or personal talk ($p = .007$, $d = 0.27$, 95% CI [-0.45, -0.09]). There was no difference between small talk and personal talk ($p = .220$, $d = 0.15$, 95% CI [-0.03, 0.32]).

The effect of conversation topic on the amount of smiling was significant ($F(2, 254) = 94.12$, $p < .001$, $p_{adj} < .001$). Post hoc pairwise comparisons were all significant, such that participants smiled more in small talk compared to personal talk ($p < .001$, $d = 0.94$, 95% CI [0.73, 1.15]) or opinion talk ($p < .001$, $d = 1.19$, 95% CI [0.96, 1.42]), and in personal talk compared to opinion talk ($p < .001$, $d = 0.46$, 95% CI [-0.64, -0.27]).

We found a significant effect of conversation topic in predicting interrupting ($F(2, 254) = 64.93$, $p < .001$, $p_{adj} < .001$). Participants interrupted significantly more in small talk compared to personal talk ($p < .001$, $d = 0.99$, 95% CI [0.78, 1.21]) or opinion talk ($p = .018$, $d = 0.83$, 95% CI [0.63, 1.03]). There was no difference between personal talk and opinion talk ($p = .608$, $d = 0.08$, 95% CI [-0.09, 0.26]).

We found a significant effect of conversation topic in predicting turn duration ($F(2, 254) = 129.49$, $p < .001$, $p_{adj} < .001$). Compared to small talk, participants had longer turns in personal talk ($p < .001$, $d = 1.23$, 95% CI [-1.46, -1.00]) and opinion talk ($p < .001$, $d = 1.10$, 95% CI [-1.32, -0.88]). Analyses revealed no significant differences between personal talk and opinion talk ($p = .971$, $d = 0.02$, 95% CI [-0.19, 0.15]).

We found a significant effect of conversation topic on gaps ($F(2, 250) = 22.86$, $p < .001$, $p_{adj} < .001$). Compared to small talk, participants took longer to reply in personal talk ($p < .001$, $d = 0.49$, 95% CI [-0.67, -0.30]) and in opinion talk ($p < .001$, $d = 0.50$, 95% CI [-0.68, -0.31]). The difference between personal talk and opinion talk was non-significant ($p = .538$, $d = 0.09$, 95% CI [-0.08, 0.27]).

The overall effect of conversation topic on loudness was significant ($F(2, 254) = 7.47, p < .001, p_{adj} = .001$). Participants spoke significantly louder in opinion talk compared to personal talk ($p < .001, d = 0.34, 95\% \text{ CI } [0.16, 0.52]$). Differences between other topics were non-significant (small talk vs personal talk, $p = .257, d = 0.14, 95\% \text{ CI } [-0.04, 0.31]$; small talk vs opinion talk, $p = .160, d = 0.16, 95\% \text{ CI } [-0.34, 0.01]$).

Comparison Between Personalized and Generic Avatar

Comparison of the quality of interaction between participants embodying generic or personalized avatars showed no differences between the two conditions (see Supplementary Results sec. Quality of Interaction). We then examined the efficacy of the manipulation of the avatars for the personalized and generic conditions. When embodying personalized as compared to generic avatars, participants reported higher values on avatar similarity ($F(1, 125) = 30.74, p < .001, p_{adj} < .001, d = 0.98, 95\% \text{ CI } [0.61, 1.36]$), avatar recognizability ($F(1, 125) = 62.39, p < .001, p_{adj} < .001, d = 1.40, 95\% \text{ CI } [1.01, 1.79]$) and avatar re-identification ($F(1, 125) = 26.32, p < .001, p_{adj} < .001, d = 0.91, 95\% \text{ CI } [0.54, 1.28]$). There were no significant differences for avatar satisfaction ($p = .271, p_{adj} = .590, d = 0.20, 95\% \text{ CI } [-0.16, 0.55]$), attractiveness ($p = .295, p_{adj} = .590, d = 0.19, 95\% \text{ CI } [-0.54, 0.17]$), unattractiveness ($p = .070, p_{adj} = .280, d = 0.32, 95\% \text{ CI } [-0.03, 0.68]$) and audibility ($p = .745, p_{adj} = .745, d = 0.06, 95\% \text{ CI } [-0.29, 0.41]$). Compared to participants embodying generic avatars, participants embodying personalized avatars reported that the avatar appearance was more similar to their own physical appearance; reported that a friend would be more likely to recognize them; and stated that they would be able to recognize the interaction partner if they met her again outside of social VR, indicating that the manipulation of the conditions were overall successful.

Analyses on the experience of VR revealed no significant effect of condition, such that participants embodying a personal or generic avatar did not differ in experiences of presence ($p = .693, p_{adj} = .693, d = 0.07, 95\% \text{ CI } [-0.28, 0.42]$), social presence ($p = .180, p_{adj} = .539, d = 0.24, 95\% \text{ CI } [-0.11, 0.59]$) nor body ownership ($p = .283, p_{adj} = .565, d = 0.19, 95\% \text{ CI } [-0.16, 0.54]$) (for details see Supplementary Figure S4).

The influence of condition on behavioral measures were examined using linear mixed-effects models. Figure 5 display the means and standard errors of the behavioral variables between the conditions. Initial analyses found a significant effect of condition on turn duration with longer turns in participants embodying personalized avatars, but this effect did not withstand Benjamini-Hochberg's corrections for multiple testing ($p = .029$, $p_{adj} = .319$), indicating that the difference is negligible. No other difference between the two conditions were statistically significant, even when not correcting for multiple testing. Supplementary Figure S5 shows the means and standard errors of the measures for each conversation topic in both conditions.

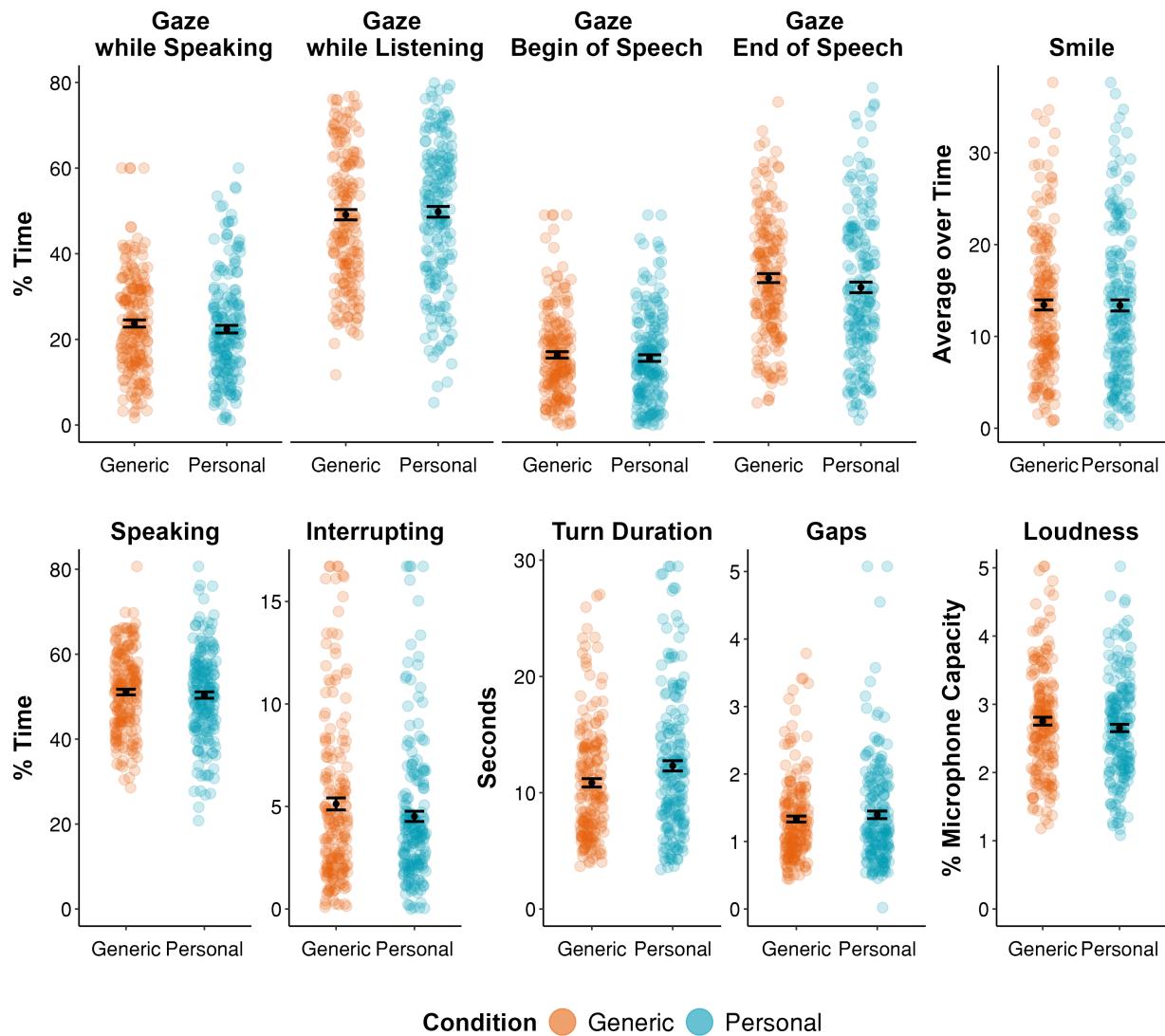
Discussion

The present study examined verbal and nonverbal behavior in real-time dyadic interactions in social VR while participants embodied either personalized or generic avatars. Interaction behavior shared several similarities with face-to-face dyadic interactions (Degutute & Astell, 2021; Hessels et al., 2019; Ho et al., 2015; Whittaker, 2003): participants gazed at their interaction partners' eye region more while listening compared to while speaking, and when speaking, participants gazed at their interaction partners' eye region more at the end of a speaking turn compared to the beginning of a turn. This coordination of speech and gaze behavior, which is known to help people regulate turn-taking in conversations (Hessels et al., 2019; Ho et al., 2015; Kendon, 1967), was observed as statistically large effect in each of the three conversation topics and in both participant groups. These observed patterns of gaze behavior in social VR are consistent with those found in face-to-face interactions. Multiple previous studies observed that individuals typically spend about half as much time gazing at their partner during speaking compared to listening in face-to-face interactions, typically ranging from 20-65% while speaking to 30-80% while listening (Argyle & Cook, 1976; Kendon, 1967; Levine & Sutton-Smith, 1973). Our results align with these findings, demonstrating a similar reduction in gaze during speaking turns and an increase at the end of speaking turns, suggesting that social VR can effectively replicate the natural coordination of speech and gaze observed in face-to-face communication.

Moreover, our study supports the findings of Abdullah et al. (2021), where participants

Figure 5

Means and standard errors of gaze and speech behavior in the two conditions.



Note: Error bars represent standard errors.

similarly showed decreased gaze towards interaction partners while speaking in triad interactions in social VR (28% compared to 23% in the current study). These results suggest that the function of eye gaze remains constant for conversations regardless of the interaction occurring entirely in a virtual environment and highlight the importance of integrating nonverbal behavior in virtual social interactions. Similarly, as in real-life interactions, users of advanced social VR setups can view each others' head and eye movements, hand and finger movements as well as facial expressions and lip movements. The situation may be contrasted with communication via videoconferencing (i.e. using Zoom), where non-verbal communication is detached from its natural integration in social space, sometimes disrupting the communication flow and causing irritation and fatigue (Bailenson, 2021). Note that while people using social VR may easily monitor who is looking at and attending to them similarly as in real life, videoconferencing setups give the visual impression that others are constantly looking in one's direction (as they are facing their own computer monitor). When communicating via videoconferencing as compared to social VR, Abdullah and colleagues (2021) found that participants spent more time looking at others' faces, used less deictic gestures, had longer conversational turns and showed more backchannelling behavior, possibly suggesting the need for active efforts to maintain the social connection and a less natural and lively communication in videoconferencing. In a study by Rubo and Gamer (2021), participants more specifically reacted to an agent's smiling in VR compared to when viewing the same scene on a computer monitor, similarly suggesting a more natural sense of social involvement in a shared social space.

Differences between conversation topics

Participants showed a robust and similar coordination of speech and gaze across conversations, but also exhibited behavioral adaptations to specific conversation topics. In small-talk conversations, where participants were instructed to discuss everyday topics, participants spent significantly more time speaking, interrupted each other more, had shorter turn durations, responded faster, smiled more, and gazed at their conversation partners' eye region more at the beginning of turns compared to all other conversation topics. These findings show that

the relatively fast-paced and dynamic exchange in face-to-face small-talk conversations is similarly actualized when people interact with each other in social VR (Eggins & Slade, 2004).

In contrast, conversations on more personal topics appeared to foster more reflective communication behavior, as indicated through longer turn durations and increased gaps before replying. In these conversations, participants spent the longest time gazing at their partners' eye region while listening, which may indicate heightened attentiveness and engagement, likely due to the personal and meaningful nature of the topic. Participants furthermore smiled less compared to small-talk conversations. During opinion talk, where participants were asked to convince their interaction partner on controversial issues, participants spent the shortest amount of time gazing at their partner's eyes during the beginning and end of turns and showed the least smiling of the three conversations. Similar to personal talk, participants spoke less frequently, interrupted less often, and had longer turns and gaps compared to small talk, indicating a more contemplative and thoughtful communication style. These findings align with the idea that discussions about differing opinions may lack the fluency of other conversations, but also disagree with an observation by Rutter et al. (1978) where participants gazed more at the end of a turn during a competitive conversation compared to a cooperative one. However, direct comparison is difficult as we did not include a cooperative condition where participants discussed common ground. The observed relative gaze avoidance at the beginning and end of turns when discussing differing opinions may indicate increased shyness or discomfort among participants or may suggest a need for increased effort to manage the conversational flow in such a confrontational context.

Effects of Avatar Personalization

Participants showed overall similar verbal and non-verbal behavior when embodying generic avatars as when embodying personalized avatars. In particular, participants embodying generic avatars did not show prolonged gaze to the partner's eye region, louder speaking or more interrupting behavior, which were hypothesized signs of disinhibition. Furthermore, there were no differences in reported social presence, self-presence or body ownership between the two conditions. These observations contrast both previous observations from single-user VR studies

where participants reported on higher levels of presence and body ownership when embodying personalized avatars (e.g., Mottelson et al., 2023; X. Pan & Hamilton, 2018; Wei et al., 2022; Weidner et al., 2023) and our hypothesis that generic avatars would be associated with signs of social disinhibition due to their relative concealment of participants' identities in our social VR setup (Suler, 2004).

Several factors may explain these discrepancies. With regards to self-reported levels of presence, previous studies typically examined effects of avatar appearance in within-subject designs where the participants repeatedly acted in isolation while embodying varying avatars (e.g., Latoschik et al., 2017; Salagean et al., 2023; Waltemate et al., 2018). While such setups can exhibit high statistical power, their drawing of participants' attention to a single varying factor – their own appearance in a virtual mirror – may also introduce a risk of expectancy effects, where participants may relatively prefer the avatar which was associated with the highest production effort. The between-subject design used in the present study alleviates the risk for expectancy effects at the cost of statistical power. In addition, participants in dyadic interactions may be less strongly influenced by their own appearance compared to participants in single-user VR environments: although participants in our setup spent a short period of time looking at their own avatar in a virtual mirror prior to the interaction, their sense of presence and social presence may then have been determined by the vividness of the interaction with their partner which may not strongly depend on the personalization of avatars. Consistent with our findings, other studies likewise did not observe differences in presence or social presence between personalized and generic avatars when participants were engaged in a social interaction. In a study by Suk and Laine (2023), participants interacting with a virtual agent reported higher embodiment but not presence when embodying personalized avatars. In a study by Han et al. (2022), the use of personalized avatars was associated with only a slight increase in self-presence and no differences in social or spatial presence. Similarly, visual appearance of the avatar may be less crucial for the sense of body ownership. Instead, high movement fidelity may suffice in fostering a sense of body ownership as suggested by recent meta-analytical findings (Mottelson et al., 2023).

With regards to behavioral disinhibition due to a relative anonymity when embodying generic avatars, effects may have been reduced in our setup since most participants were drawn from the same student population (albeit consisting of several cohorts with a total of more than one thousand potential interaction partners). Recognition was in principle possible since voices were not distorted during the interaction, but as participants in the condition with generic avatars rated their similarity and recognizability, but also their partner's identifiability substantially lower compared with participants in the condition with personalized avatars, the use of generic avatars did entail a relative anonymization of the situation. Nonetheless, a sense of belonging to a common overall group may have contributed to a diminishing effect of anonymity seen in communication with strangers on the internet. Alternatively, it is possible that effects of disinhibition known from anonymous written communication are substantially alleviated in social VR due to the situation's inherent naturalness and intimacy. Note that in social VR, even anonymous interaction partners engage in eye-contact with each other, potentially eliciting a sense of familiarity and affiliation which is associated with eye-contact in face-to-face interactions (e.g., Argyle & Cook, 1976; Broz et al., 2012; Kellerman et al., 1989). In a study by Lapidot-Lefler and Barak (2012), eye contact reduced subsequent negative behavior in written communication, further highlighting the impact of eye contact on social behavior.

An absence of disinhibition effects of anonymous communication would allow to use social VR for anonymous interactions without risking negative side effects. Note that while some future use cases of social VR may require the implementation of personalized avatars, other situations may profit from being able to communicate via generic avatars. For instance, the use of generic avatars may more easily comply with stricter privacy needs, allowing users to interact with a full range of nonverbal communication while maintaining privacy, which is particularly crucial for marginalized users (Lin & Latoschik, 2022; Maloney et al., 2020). In recruiting interviews, using social VR with generic avatars may allow for fairer, less biased assessment processes. For example, a study using Swedish data demonstrated that anonymous application procedures increased the chances of women and individuals of non-Western origin advancing to the interview

stage and, for women, improved job offer rates (Åslund & Skans, 2012). In therapeutic settings, vulnerable participants may benefit from communicating anonymously with a therapist or with each other, or may be more willing to engage in therapy in the first place. Previous studies examining social media communities have found that anonymity associated with online written communication can facilitate open and honest discussions about mental health, thereby improving access to social support and reducing stigma (De Choudhury & De, 2014; Eysenbach et al., 2004). Using generic avatars in social VR for therapeutic purposes could similarly promote a safe environment for self-disclosure and support, encouraging more participants to engage in therapy.

The naturalistic interaction patterns observed in social VR holds multiple significant implications. These findings show that social VR can provide a naturalistic environment for researching social behavior, enabling more accurate observations of how individuals interact in realistic scenarios. The capacity of VR to automatically collect rich behavioral data will allow researchers to capture complex interactions that have remained often difficult and time-consuming to analyze with conventional methods (Yaremych & Persky, 2019). As such, the study of social VR holds significant promise to deepen our understanding of social dynamics in ecologically valid situations, making it a powerful new measurement tool. However, it is crucial for social VR users to be aware of the fine-grained behavioral data which is generated in VR. Such data should be managed with strict adherence to privacy protocols, requiring informed consent from users before any collection or analysis occurs.

Limitations

While the present study provides detailed new insights into verbal and non-verbal behavior of individuals interacting in social VR, several factors limit the generalizability of our results. Firstly, participants were primarily drawn from a student population in Switzerland. As many aspects of social behavior are culturally learned, applicability of our results may decrease with increasing cultural distance of a population to the one tested in the present study (Henrich et al., 2010). For instance, maintaining eye contact may be more important and socially encouraged in Western cultures compared to East Asian cultures (Argyle & Cook, 1976;

Argyle et al., 1986), which have been shown in gaze avoidance in Japanese individuals (e.g., Japanese vs Finnish; Akechi et al., 2013; Senju et al., 2013). These results highlight the complexity of cultural influences on non-verbal communication and indicate the importance of taking cultural background into consideration.

Additionally, more women than men participated in the present study. Since gender composition in dyads were observed to modulate behavioral markers such as behavioral synchrony (Fujiwara et al., 2019) and gaze behavior can be influenced by the level of attraction to an interaction partner (C. Hall et al., 2010), future research should systematically vary the gender composition of dyads, as well as to account for sexual orientation of the participants. Recruiting participants from a wider population and implementing assessment techniques in social VR environments as people use them at home or at work – outside of a scientific laboratory – may additionally allow to elicit stronger senses of anonymity in participants embodying generic avatars.

Participants in the present study only interacted with one other participant. While behavioral markers such as the overall coordination of speech and gaze may likely generalize in principle across interactions, it is noteworthy that dyadic conversations, in contrast to reactions to pre-defined stimuli such as images, are by nature interactive and include reactions to a partner's behavior (Patterson, 2019). Previous studies utilizing the social relations model have indicated that individuals show moderate consistency in their behavior across various interaction partners, however, distinct responses may be elicited depending on their partner (Kenny et al., 2001). To further investigate the stability of a person's behavior across interactions (e.g., the amount of gaze towards the partner's eye region) as well as the extent to which behavior is influenced by the partner, future research should assess interaction behavior in participants across multiple interactions with different partners.

The comparison of behavior between participants embodying generic or personalized avatars may have been compromised by limited quality of the personalization process. Although our manipulation check indicated a substantial difference between the two conditions - with

personalized avatars rated higher in similarity, recognizability, and re-identifiability - participants embodying personalized avatars may have shown different behavior if their avatars had more closely resembled their physical appearance. Additionally, ratings on avatar satisfaction and attractiveness were moderate. While the present study used a state-of the art tool for avatar personalization, future studies should make use of technical improvements to allow participants to embody even more closely personalized avatars. Emerging research is focused on developing highly realistic avatars, capturing fine features such as hair strands and skin pores (e.g., Saito et al., 2024). Such advancements will significantly improve the realism and effectiveness of social VR interactions and may soon be accessible for academic use.

Similarly, the visual quality and fidelity of replicating participants' facial expressions on their avatars may be improved. In particular, we employed face-tracking and avatar technology which allowed to distinguish a range of different facial movements such as a pulling of each corner of the mouth in creating a smile, but facial expression replications may not have accounted properly for inter-individual variation. A real person's characteristic smile may not only be defined by the composition of muscle contraction intensities, but also by idiosyncrasies in how her skin is distorted or the tip of the nose is pulled upwards (Schmidt & Cohn, 2001). Improvements in facial expression replication should be used in future studies and may likely influence interaction behavior. While we focused on smiling behavior as a common and particularly salient facial expression used in communication, a more detailed assessment may allow to investigate how different types of smiles (e.g., polite or embarrassed smiles; Ambadar et al., 2008) can convey different emotional nuances in social VR communication.

Conclusion

This present study examined verbal and non-verbal behavior in a dyadic conversation in embodied social VR across three different conversation topics. Participants robustly showed a fine-grained coordination of speech and gaze behavior known from real face-to-face interactions – with more gaze to the partner's eyes while listening compared to while speaking and at the end of a speaking turn compared to the beginning of a speaking turn – indicating an overall natural

interaction behavior in social VR. Characteristic modulations in interaction behavior depending on conversation topic further corroborate this view. Contrary to our expectation, the use of generic as compared to personalized avatars did not result in disinhibited social behavior known from anonymous online communication. The possibility for direct eye-contact, which is known to regulate interaction behavior, may have alleviated effects of social disinhibition due to anonymity.

Overall, our findings suggest that social VR elicits natural interaction behavior and may be used for anonymous interactions without negative side-effects known from other technology-mediated forms of communication. As communication in social VR become more widespread, it is increasingly important to understand the patterns of interaction behavior in social VR and to examine how different factors such as avatar appearance and different topics of conversation may influence behavior. These results also highlight the prospect of utilizing social VR as an ecologically valid tool in the investigation of complex multi-modal interaction situations.

References

- Abdullah, A., Kolkmeier, J., Lo, V., & Neff, M. (2021). Videoconference and embodied vr: Communication patterns across task and medium. *Proceedings of the ACM on Human-Computer Interaction*, 5, 29. <https://doi.org/10.1145/3479597>
- Akechi, H., Senju, A., Uibo, H., Kikuchi, Y., Hasegawa, T., & Hietanen, J. K. (2013). Attention to eye contact in the west and east: Autonomic responses and evaluative ratings (A. Mesoudi, Ed.). *PLoS ONE*, 8(3), e59312. <https://doi.org/10.1371/journal.pone.0059312>
- Ambadar, Z., Cohn, J. F., & Reed, L. I. (2008). All smiles are not created equal: Morphology and timing of smiles perceived as amused, polite, and embarrassed/nervous. *Journal of Nonverbal Behavior*, 33(1), 17–34. <https://doi.org/10.1007/s10919-008-0059-5>
- Andrist, S., Mutlu, B., & Gleicher, M. (2013). Conversational gaze aversion for virtual agents. In *Lecture notes in computer science* (pp. 249–262). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-40415-3_22
- Argyle, M., & Cook, M. (1976). Gaze and mutual gaze.
- Argyle, M., Henderson, M., Bond, M., Iizuka, Y., & Contarello, A. (1986). Cross-cultural variations in relationship rules. *International Journal of Psychology*, 21(1–4), 287–315. <https://doi.org/10.1080/00207598608247591>
- Aron, A., Melinat, E., Aron, E. N., Vallone, R. D., & Bator, R. J. (1997). The experimental generation of interpersonal closeness: A procedure and some preliminary findings. *Personality and Social Psychology Bulletin*, 23, 363–377. <https://doi.org/10.1177/0146167297234003>
- Åslund, O., & Skans, O. N. (2012). Do anonymous job application procedures level the playing field? *ILR Review*, 65(1), 82–107. <https://doi.org/10.1177/001979391206500105>
- Bailenson, J. N. (2021). Nonverbal overload: A theoretical argument for the causes of zoom fatigue. *Technology, Mind, and Behavior*, 2. <https://doi.org/10.1037/tmb0000030>
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B*:

- Statistical Methodology*, 57, 289–300.
<https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>
- Broz, F., Lehmann, H., Nehaniv, C. L., & Dautenhahn, K. (2012). Mutual gaze, personality, and familiarity: Dual eye-tracking during conversation. *2012 IEEE RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication*.
<https://doi.org/10.1109/roman.2012.6343859>
- Chen, N. T., & Clarke, P. J. (2017). Gaze-based assessments of vigilance and avoidance in social anxiety: A review. *Current Psychiatry Reports*, 19, 1–9.
<https://doi.org/10.1007/S11920-017-0808-4/METRICS>
- De Choudhury, M., & De, S. (2014). Mental health discourse on reddit: Self-disclosure, social support, and anonymity. *Proceedings of the International AAAI Conference on Web and Social Media*, 8(1), 71–80. <https://doi.org/10.1609/icwsm.v8i1.14526>
- Degutyte, Z., & Astell, A. (2021). The role of eye gaze in regulating turn taking in conversations: A systematized review of methods and findings. *Frontiers in Psychology*, 12, 616471.
<https://doi.org/10.3389/fpsyg.2021.616471>
- DeWall, C. N., Maner, J. K., & Rouby, D. A. (2009). Social exclusion and early-stage interpersonal perception: Selective attention to signs of acceptance. *Journal of Personality and Social Psychology*, 96, 729–741. <https://doi.org/10.1037/a0014634>
- Eggins, S., & Slade, D. (2004). *Analysing casual conversation*. Equinox Publishing Ltd.
- Exline, R., Gray, D., & Schuette, D. (1965). Visual behavior in a dyad as affected by interview content and sex of respondent. *Journal of Personality and Social Psychology*, 1(3), 201–209. <https://doi.org/10.1037/h0021865>
- Eysenbach, G., Powell, J., Englesakis, M., Rizo, C., & Stern, A. (2004). Health related virtual communities and electronic support groups: Systematic review of the effects of online peer to peer interactions. *BMJ*, 328(7449), 1166. <https://doi.org/10.1136/bmj.328.7449.1166>

- Fujiwara, K., Kimura, M., & Daibo, I. (2019). Gender differences in synchrony: Females in sync during unstructured dyadic conversation. *European Journal of Social Psychology, 49*(5), 1042–1054. <https://doi.org/10.1002/ejsp.2587>
- Glenberg, A. M., Schroeder, J. L., & Robertson, D. A. (1998). Averting the gaze disengages the environment and facilitates remembering. *Memory & Cognition, 26*(4), 651–658. <https://doi.org/10.3758/bf03211385>
- Gonzalez-Franco, M., & Peck, T. C. (2018). Avatar embodiment. towards a standardized questionnaire. *Frontiers in Robotics and AI, 5*, 343504. <https://doi.org/10.3389/frobt.2018.00074>
- Hall, C., Hogue, T., & Guo, K. (2010). Differential gaze behavior towards sexually preferred and non-preferred human figures. *The Journal of Sex Research, 48*(5), 461–469. <https://doi.org/10.1080/00224499.2010.521899>
- Hall, J. A., Coats, E. J., & LeBeau, L. S. (2005). Nonverbal behavior and the vertical dimension of social relations: A meta-analysis. *Psychological Bulletin, 131*, 898–924. <https://doi.org/10.1037/0033-2909.131.6.898>
- Hall, J. A., Horgan, T. G., & Murphy, N. A. (2019). Nonverbal communication. *Annual Review of Psychology, 70*(1), 271–294. <https://doi.org/10.1146/annurev-psych-010418-103145>
- Hall, J. A., & Knapp, M. L. (Eds.). (2013). *Nonverbal communication*. De Gruyter Mouton. <https://doi.org/doi:10.1515/9783110238150>
- Han, E., Miller, M. R., Ram, N., Nowak, K. L., & Bailenson, J. N. (2022). Understanding group behavior in virtual reality: A large-scale, longitudinal study in the metaverse. *72nd Annual International Communication Association Conference, Paris, France*.
- Hennig-Thurau, T., Aliman, D. N., Herting, A. M., Cziehso, G. P., Linder, M., & Kübler, R. V. (2023). Social interactions in the metaverse: Framework, initial evidence, and research roadmap. *Journal of the Academy of Marketing Science, 51*, 889–913. <https://doi.org/10.1007/s11747-022-00908-0>

- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, 33(2–3), 61–83. <https://doi.org/10.1017/s0140525x0999152x>
- Herrera, F., Oh, S. Y., & Bailenson, J. N. (2018). Effect of behavioral realism on social interactions inside collaborative virtual environments. *Presence: Teleoperators and Virtual Environments*, 27, 163–182. https://doi.org/10.1162/pres_a_00324
- Hessels, R. S., Holleman, G. A., Kingstone, A., Hooge, I. T., & Kemner, C. (2019). Gaze allocation in face-to-face communication is affected primarily by task structure and social context, not stimulus-driven factors. *Cognition*, 184, 28–43.
<https://doi.org/10.1016/j.cognition.2018.12.005>
- Ho, S., Foulsham, T., & Kingstone, A. (2015). Speaking and listening with the eyes: Gaze signaling during dyadic interactions (K. Guo, Ed.). *PLOS ONE*, 10, e0136905.
<https://doi.org/10.1371/journal.pone.0136905>
- Ichino, J., Ide, M., Yoshiki, T., Yokoyama, H., Asano, H., Miyachi, H., & Okabe, D. (2023). How gaze visualization facilitates initiation of informal communication in 3d virtual spaces. *ACM Trans. Comput.-Hum. Interact.*, 31(1). <https://doi.org/10.1145/3617368>
- Karras, T., Laine, S., & Aila, T. (2018). A style-based generator architecture for generative adversarial networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43, 4217–4228. <https://doi.org/10.1109/TPAMI.2020.2970919>
- Kellerman, J., Lewis, J., & Laird, J. D. (1989). Looking and loving: The effects of mutual gaze on feelings of romantic love. *Journal of Research in Personality*, 23(2), 145–161.
[https://doi.org/10.1016/0092-6566\(89\)90020-2](https://doi.org/10.1016/0092-6566(89)90020-2)
- Keltner, D., Gruenfeld, D. H., & Anderson, C. (2003). Power, approach, and inhibition. *Psychological Review*, 110(2), 265–284. <https://doi.org/10.1037/0033-295x.110.2.265>
- Kendon, A. (1967). Some functions of gaze-direction in social interaction. *Acta Psychologica*, 26, 22–63. [https://doi.org/10.1016/0001-6918\(67\)90005-4](https://doi.org/10.1016/0001-6918(67)90005-4)

- Kenny, D. A., Mohr, C. D., & Levesque, M. J. (2001). A social relations variance partitioning of dyadic behavior. *Psychological Bulletin*, 127(1), 128–141.
<https://doi.org/10.1037/0033-2909.127.1.128>
- Kim, H. K., Park, J., Choi, Y., & Choe, M. (2018). Virtual reality sickness questionnaire (vrsq): Motion sickness measurement index in a virtual reality environment. *Applied Ergonomics*, 69, 66–73. <https://doi.org/10.1016/j.apergo.2017.12.016>
- Kimmel, S., Jung, F., Matviienko, A., Heuten, W., & Boll, S. (2023). Let's face it: Influence of facial expressions on social presence in collaborative virtual reality. *Conference on Human Factors in Computing Systems - Proceedings*. <https://doi.org/10.1145/3544548.3580707>
- Krumhuber, E., Manstead, A. S. R., Cosker, D., Marshall, D., Rosin, P. L., & Kappas, A. (2007). Facial dynamics as indicators of trustworthiness and cooperative behavior. *Emotion*, 7, 730–735. <https://doi.org/10.1037/1528-3542.7.4.730>
- Lapidot-Lefler, N., & Barak, A. (2012). Effects of anonymity, invisibility, and lack of eye-contact on toxic online disinhibition. *Computers in Human Behavior*, 28, 434–443.
<https://doi.org/10.1016/J.CHB.2011.10.014>
- Latoschik, M. E., Roth, D., Gall, D., Achenbach, J., Waltemate, T., & Botsch, M. (2017). The effect of avatar realism in immersive social virtual realities. *Proceedings of the ACM Symposium on Virtual Reality Software and Technology, VRST, Part F131944*.
<https://doi.org/10.1145/3139131.3139156>
- Lenth, R. V. (2024). *Emmeans: Estimated marginal means, aka least-squares means* [R package version 1.10.1]. <https://CRAN.R-project.org/package=emmeans>
- Levine, M. H., & Sutton-Smith, B. (1973). Effects of age, sex, and task on visual behavior during dyadic interaction. *Developmental Psychology*, 9(3), 400–405.
<https://doi.org/10.1037/h0034929>
- Lin, J., & Latoschik, M. E. (2022). Digital body, identity and privacy in social virtual reality: A systematic review. *Frontiers in Virtual Reality*, 3, 974652.
<https://doi.org/10.3389/FRVIR.2022.974652/BIBTEX>

- Ma, F., & Pan, X. (2022). Visual fidelity effects on expressive self-avatar in virtual reality: First impressions matter. *2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. <https://doi.org/10.1109/vr51125.2022.00023>
- Maloney, D., Freeman, G., & Wohn, D. Y. (2020). "talking without a voice". *Proceedings of the ACM on Human-Computer Interaction*, 4. <https://doi.org/10.1145/3415246>
- Matz, D. C., & Wood, W. (2005). Cognitive dissonance in groups: The consequences of disagreement. *Journal of Personality and Social Psychology*, 88, 22–37. <https://doi.org/10.1037/0022-3514.88.1.22>
- McVeigh-Schultz, J., Kolesnichenko, A., & Isbister, K. (2019). Shaping pro-social interaction in vr an emerging design framework. *Conference on Human Factors in Computing Systems - Proceedings*. <https://doi.org/10.1145/3290605.3300794>
- Miller, M. R., DeVeaux, C., Han, E., Ram, N., & Bailenson, J. N. (2023). A large-scale study of proxemics and gaze in groups. *2023 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*, 409–417. <https://doi.org/10.1109/VR55154.2023.00056>
- Mottelson, A., Muresan, A., Hornbæk, K., & Makransky, G. (2023). A systematic review and meta-analysis of the effectiveness of body ownership illusions in virtual reality. *ACM Transactions on Computer-Human Interaction*, 30, 76. <https://doi.org/10.1145/3590767>
- Oh, S. Y., Bailenson, J., Krämer, N., & Li, B. (2016). Let the avatar brighten your smile: Effects of enhancing facial expressions in virtual environments. *PLOS ONE*, 11, e0161794. <https://doi.org/10.1371/JOURNAL.PONE.0161794>
- Pan, X., & Hamilton, A. F. C. (2018). Why and how to use virtual reality to study human social interaction: The challenges of exploring a new research landscape. *British Journal of Psychology*, 109, 395–417. <https://doi.org/10.1111/bjop.12290>
- Pan, Y., & Steed, A. (2017). The impact of self-avatars on trust and collaboration in shared virtual environments (M. Lappe, Ed.). *PLOS ONE*, 12, e0189078. <https://doi.org/10.1371/journal.pone.0189078>

- Parsons, T. D. (2015). Virtual reality for enhanced ecological validity and experimental control in the clinical, affective and social neurosciences. *Frontiers in Human Neuroscience*, 9. <https://doi.org/10.3389/fnhum.2015.00660>
- Patterson, M. L. (2019). A systems model of dyadic nonverbal interaction. *Journal of Nonverbal Behavior*, 43(2), 111–132. <https://doi.org/10.1007/s10919-018-00292-w>
- Pinheiro, J., Bates, D., & R Core Team. (2023). *Nlme: Linear and nonlinear mixed effects models* [R package version 3.1-164]. <https://CRAN.R-project.org/package=nlme>
- R Core Team. (2024). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. <https://www.R-project.org/>
- Rogers, S. L., Broadbent, R., Brown, J., Fraser, A., & Speelman, C. P. (2022). Realistic motion avatars are the future for social interaction in virtual reality. *Frontiers in Virtual Reality*, 2, 750729. <https://doi.org/10.3389/FRVIR.2021.750729/BIBTEX>
- Rubo, M., & Gamer, M. (2021). Stronger reactivity to social gaze in virtual reality compared to a classical laboratory environment. *British Journal of Psychology*, 112(1), 301–314.
- Rutter, D. R., Stephenson, G. M., Ayling, K., & White, P. A. (1978). The timing of looks in dyadic conversation. *British Journal of Social and Clinical Psychology*, 17(1), 17–21. <https://doi.org/10.1111/j.2044-8260.1978.tb00890.x>
- Saito, S., Schwartz, G., Simon, T., Li, J., & Nam, G. (2024). Relightable gaussian codec avatars. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 130–141.
- Salagean, A., Crellin, E., Parsons, M., Cosker, D., & Fraser, D. S. (2023). Meeting your virtual twin: Effects of photorealism and personalization on embodiment, self-identification and perception of self-avatars in virtual reality. *Conference on Human Factors in Computing Systems - Proceedings*, 16. <https://doi.org/10.1145/3544548.3581182>
- Schmidt, K., & Cohn, J. (2001). Dynamics of facial expression: Normative characteristics and individual differences. *IEEE International Conference on Multimedia and Expo, 2001. ICME 2001*. <https://doi.org/10.1109/icme.2001.1237778>

- Schubert, T., Friedmann, F., & Regenbrecht, H. (2001). The experience of presence: Factor analytic insights. *Presence: Teleoperators and Virtual Environments*, 10, 266–281.
<https://doi.org/10.1162/105474601300343603>
- Senju, A., Vernetti, A., Kikuchi, Y., Akechi, H., Hasegawa, T., & Johnson, M. H. (2013). Cultural background modulates how we look at other persons' gaze. *International Journal of Behavioral Development*, 37(2), 131–136. <https://doi.org/10.1177/0165025412465360>
- Smith, H. J., & Neff, M. (2018). Communication behavior in embodied virtual reality. *Conference on Human Factors in Computing Systems - Proceedings, 2018-April*.
<https://doi.org/10.1145/3173574.3173863>
- Suk, H., & Laine, T. H. (2023). Influence of avatar facial appearance on users' perceived embodiment and presence in immersive virtual reality. *Electronics*, 12(3), 583.
<https://doi.org/10.3390/electronics12030583>
- Suler, J. (2004). The online disinhibition effect. *CyberPsychology I& Behavior*, 7(3), 321–326.
<https://doi.org/10.1089/1094931041291295>
- Tarnec, H. L., Bevacqua, E., Augereau, O., & Loor, P. D. (2023). Effect of avatar facial expressiveness on team collaboration in virtual reality. *Proceedings of the 23rd ACM International Conference on Intelligent Virtual Agents*, 1–8.
<https://doi.org/10.1145/3570945.3607330>
- Waltemate, T., Gall, D., Roth, D., Botsch, M., & Latoschik, M. E. (2018). The impact of avatar personalization and immersion on virtual body ownership, presence, and emotional response. *IEEE Transactions on Visualization and Computer Graphics*, 24, 1643–1652.
<https://doi.org/10.1109/TVCG.2018.2794629>
- Wei, X., Jin, X., & Fan, M. (2022). Communication in immersive social virtual reality: A systematic review of 10 years' studies. *Proceedings of the Tenth International Symposium of Chinese CHI*, 27–37. <https://doi.org/10.1145/3565698.3565767>
- Weidner, F., Boettcher, G., Arboleda, S. A., Diao, C., Sinani, L., Kunert, C., Gerhardt, C., Broll, W., & Raake, A. (2023). A systematic review on the visualization of avatars and

agents in ar & vr displayed using head-mounted displays. *IEEE Transactions on Visualization and Computer Graphics*, 29, 2596–2606.

<https://doi.org/10.1109/TVCG.2023.3247072>

Whittaker, S. (2003, April). Theories and methods in mediated communication. In A. C. Graesser, M. A. Gernsbacher, & S. R. Goldman (Eds.). Routledge.

<https://doi.org/10.4324/9781410607348>

Xenakis, I., Gavalas, D., Kasapakis, V., Dzardanova, E., & Vosinakis, S. (2022). Nonverbal communication in immersive virtual reality through the lens of presence: A critical review. *PRESENCE: Virtual and Augmented Reality*, 31, 147–187.

https://doi.org/10.1162/pres_a_00387

Yaremych, H. E., & Persky, S. (2019). Tracing physical behavior in virtual reality: A narrative review of applications to social psychology. *Journal of Experimental Social Psychology*, 85, 103845. <https://doi.org/10.1016/j.jesp.2019.103845>

Acknowledgements

We thank Luca Panico, Susanne Felder, Michelle Bürki, Ana Rita Calçada da Silva and Michael Mozhacherry for their help in data collection.

Author contributions

Gayoung Son: Conceptualization, Project Administration, Formal Analysis, Visualization, Writing - Original Draft, Writing - Review & Editing.

Marius Rubo: Conceptualization, Funding Acquisition, Project Administration, Software, Supervision, Writing - Review & Editing.

Research Transparency Statement

Conflicts of interest: The authors declare no conflict of interest.

Funding: This research was supported by Swiss National Science Foundation to Marius Rubo (SNSF, Grant Number PZ00P1_208909).

This study's sample size, obtained measures and analysis approach were preregistered (<https://osf.io/p53fu>). The study conformed to the principles expressed in the Declaration of Helsinki and was approved by the local ethics committee at the University of Bern (Ref-No. 2023-09-01). Data that support the findings of this study are available at <https://osf.io/fyr2q/>

Supplementary Materials

Conversation Topics

Table S1

List of Conversation Topics and Questions in English and German

Topic	Question
Small Talk	1. How has today been for you so far? <i>Wie ist der heutige Tag für Sie bisher verlaufen?</i>
	2. What do you think of the weather today? <i>Wie finden Sie das Wetter heute?</i>
	3. How did you arrive for this examination? <i>Wie sind Sie zu dieser Untersuchung angereist?</i>
Personal Talk	1. What do you feel most grateful for in life? <i>Wofür fühlen Sie sich im Leben am meisten dankbar?</i>
	2. What does friendship mean to you? <i>Was bedeutet Freundschaft für Sie?</i>
	3. If you could change anything about your childhood, what would it be? <i>Wenn Sie etwas an Ihrer Erziehung als Kind ändern könnten, was wäre es?</i>
Opinion Talk	1. Should cannabis be legal? <i>Sollte Cannabis legal gekauft werden können?</i>
	2. Should organs be allowed to be removed from the deceased if the person has not objected? <i>Sollten Organe von Verstorbenen entnommen werden dürfen, wenn die Person dem nicht widersprochen hat?</i>
	3. Should you eat animals? And if so, where is the line between animals that can or cannot be eaten? <i>Darf man Tiere essen? Und wenn ja, wo verläuft die Grenze zwischen Tieren, die man essen oder nicht essen darf?</i>

Questionnaires

Table S2

Measures on Avatar Appearance and Interaction in English and German

Variable	Item
Avatar Appearance*	1. I think my virtual avatar looked similar to me. <i>Ich finde, mein virtueller Avatar sah ähnlich aus als ich.</i>
	2. I think my friends would recognize that my avatar is supposed to represent me. <i>Ich denke, meine Freunde würden erkennen, dass mein Avatar mich darstellen soll.</i>
	3. I am satisfied with the appearance of my avatar. <i>Ich bin zufrieden mit dem Aussehen meines Avatars.</i>
	4. I think my avatar looked more attractive than me. <i>Ich finde, mein Avatar sah attraktiver aus als ich.</i>
	5. I think my avatar looked less attractive than me. <i>Ich finde, mein Avatar sah weniger attraktiver aus als ich.</i>
	6. If I saw the person I interacted with again (e.g., in the hallway of this building or in the city), I would recognize them. <i>Wenn ich die Person, mit der ich interagiert habe, wieder sehen würde (z.B. im Gang dieses Gebäudes oder in der Stadt), dann würde ich sie wiedererkennen.</i>
	7. I could hear my partner well. <i>Ich konnte mein Gegenüber akustisch gut verstehen.</i>
Relationship	1. I know the person I interacted with. (Yes, No) <i>Ich kenne die Person, mit der ich interagiert habe. (Ja, Nein)</i>
	2. How well do you know the person? (Acquaintance, Colleague, Friend, Other) <i>Wie gut kennst du die Person? (Bekannte, Kolleg*in, Freund*in, Andere)</i>

Note: * = Answer options ranged from 1 = Not at all (*Überhaupt nicht*) to 5 = Completely (*Vollkommen*).

Table S3*Measures on Quality of Interaction in English and German*

Topic	Item
Small Talk	<p>In one phase of the conversation, you were asked to talk about personal topics. (1 = <i>Not at all</i>, 5 = <i>Completely</i>)</p> <p><i>In einer Phase des Gesprächs wurden Sie gebeten, sich über persönliche Themen auszutauschen. (1 = Trifft überhaupt nicht zu, 5 = Trifft von und ganz zu)</i></p> <ol style="list-style-type: none"> 1. Were you successful in discussing this topic together? <p><i>Ist es Ihnen als Zweiergruppe gelungen, sich zu diesem Thema auszutauschen?</i></p> <ol style="list-style-type: none"> 2. Were you able to contribute to this topic? <p><i>Konnten Sie zu diesem Thema etwas beitragen?</i></p> <ol style="list-style-type: none"> 3. Was your partner able to contribute to this topic? <p><i>Konnte Ihr Partner zu diesem Thema etwas beitragen?</i></p>
Personal Talk	<p>In one phase of the conversation, you were asked to talk about personal topics. (1 = <i>Not at all</i>, 5 = <i>Completely</i>)</p> <p><i>In einer Phase des Gesprächs wurden Sie gebeten, sich über persönliche Themen auszutauschen. (1 = Trifft überhaupt nicht zu, 5 = Trifft von und ganz zu)</i></p> <ol style="list-style-type: none"> 1. Were you successful in discussing this topic together? <p><i>Ist es Ihnen als Zweiergruppe gelungen, sich zu diesem Thema auszutauschen?</i></p> <ol style="list-style-type: none"> 2. Were you able to contribute to this topic? <p><i>Konnten Sie zu diesem Thema etwas beitragen?</i></p> <ol style="list-style-type: none"> 3. Was your partner able to contribute to this topic? <p><i>Konnte Ihr Partner zu diesem Thema etwas beitragen?</i></p> <ol style="list-style-type: none"> 4. Did you feel that the personal conversation was more intimate than the small talk conversation? <p><i>Hatten Sie das Gefühl, dass die persönliche Gesprächssituation intimer war als die alltägliche Gesprächssituation?</i></p>

Table S3*Measures on Quality of Interaction in English and German (continued)*

Topic	Item
	In one phase of the conversation, you were asked to talk about personal topics. (1 = <i>Not at all</i> , 5 = <i>Completely</i>)
	<i>In einer Phase des Gesprächs wurden Sie gebeten, sich über persönliche Themen auszutauschen. (1 = Trifft überhaupt nicht zu, 5 = Trifft von und ganz zu)</i>
	1. Were you successful in discussing this topic together? <i>Ist es Ihnen als Zweiergruppe gelungen, sich zu diesem Thema auszutauschen?</i>
Opinion Talk	2. Were you able to contribute to this topic? <i>Konnten Sie zu diesem Thema etwas beitragen?</i>
	3. Was your partner able to contribute to this topic? <i>Konnte Ihr Partner zu diesem Thema etwas beitragen?</i>
	4. Who was better able to convince the other of their opinion? (Myself–Partner) <i>Wer konnte den anderen besser von ihrer oder seiner Meinung überzeugen? (Ich–Gegenüber)</i>

Note.: Items in small talk and personal talk were adapted from Aron et al. (1997). Items in opinion talk was selected from a pilot survey of university students, which showed the highest variance in response distributions, in order to increase the probability that the dyad would hold divergent opinions to at least one of the proposed topics.

Supplementary Results

Outlier Correction

Table S4

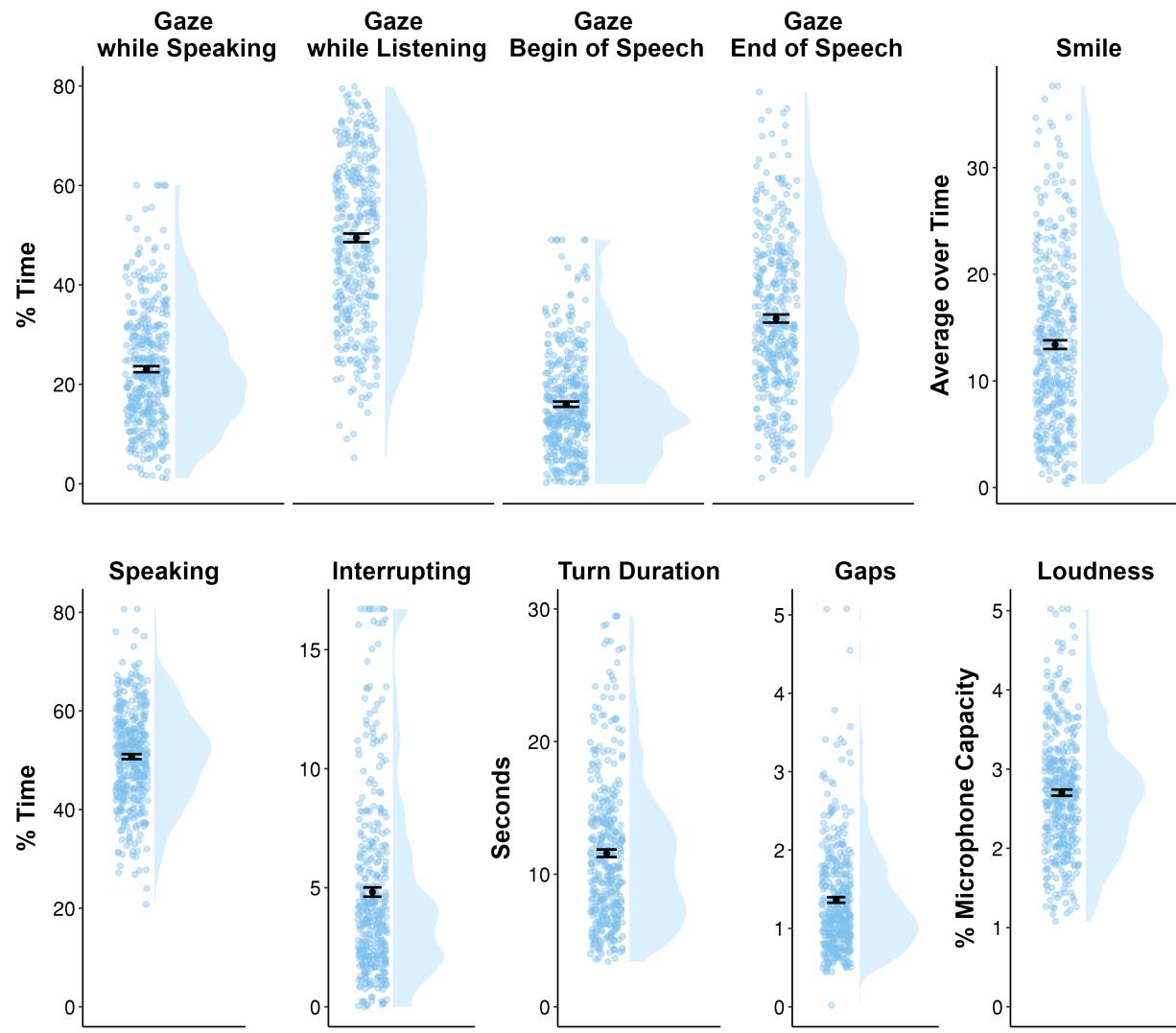
Percentage of Corrected Outliers of Behavioral Variables

Variable	Percentage Outliers (%)
Gaze While Speaking	1.04
Gaze While Listening	0
Gaze Beginning of Turn	1.56
Gaze End of Turn	0.26
Smile	0.52
Percent Speaking	0.78
Percent Interrupting	2.08
Turn duration	0.78
Gaps	0.53
Loudness	0.78

Distribution

Figure S1

Means and Standard Errors of Gaze and Speech Behavior.



Note: Error bars represent standard errors.

Table S5

Means, Standard Deviations and 95% Confidence Intervals of Behavior in Individual Topics

Measure	Topic	M	SD	95% CI
Gaze while Speaking	Small Talk	24.29	11.76	[22.23, 26.35]
	Personal Talk	22.93	12.71	[20.71, 25.15]
	Opinion Talk	21.92	11.35	[19.93, 23.90]
Gaze while Listening	Small Talk	48.78	17.45	[45.73, 51.83]
	Personal Talk	51.43	16.88	[48.47, 54.38]
	Opinion Talk	48.15	16.48	[45.26, 51.03]
Gaze Beginning of Speech	Small Talk	18.92	10.08	[17.16, 20.69]
	Personal Talk	16.25	11.46	[14.25, 18.26]
	Opinion Talk	12.84	8.94	[11.28, 14.41]
Gaze End of Speech	Small Talk	35.53	15.29	[32.86, 38.21]
	Personal Talk	33.79	17.21	[30.78, 36.80]
	Opinion Talk	30.36	14.89	[27.75, 32.96]
Smile	Small Talk	16.91	7.99	[15.51, 18.31]
	Personal Talk	12.60	7.83	[11.23, 13.97]
	Opinion Talk	10.69	6.85	[9.49, 11.89]
Speaking	Opinion Talk	50.11	10.04	[48.36, 51.87]
	Personal Talk	49.64	8.97	[48.07, 51.20]
	Small Talk	52.44	10.48	[50.61, 54.28]
Interrupting	Small Talk	6.75	4.27	[6.00, 7.49]
	Personal Talk	3.76	2.95	[3.24, 4.28]
	Opinion Talk	3.96	3.27	[3.39, 4.53]
Turn Duration	Small Talk	7.68	2.92	[7.17, 8.19]
	Personal Talk	13.59	5.70	[12.59, 14.59]
	Opinion Talk	13.47	5.55	[12.50, 14.44]
Gaps	Small Talk	1.14	0.52	[1.05, 1.23]
	Personal Talk	1.44	0.71	[1.31, 1.56]
	Opinion Talk	1.51	0.79	[1.38, 1.65]
Loudness	Small Talk	2.70	0.76	[2.57, 2.83]
	Personal Talk	2.65	0.79	[2.51, 2.79]
	Opinion Talk	2.76	0.77	[2.62, 2.89]

Note: Gaze refers to gaze fixation on the eye region of the partner.

Quality of Interaction

Comparison Between Conversation Topics

Table S6 show the means, standard deviations, 95% confidence intervals of ratings on the quality of interaction for each conversation topic. Mean ratings were high on all three measures, indicating that on average, participants engaged effectively with each other in each conversation topic. We further examined differences in ratings between topics through linear mixed models. The effect of conversation topic on each behavioral measure were significant for all measures. Based on the significance of models, post hoc pairwise comparisons were conducted (adjusted for multiple testing using Tukey's HSD method) between the conversation topics for each measure. Figure S2 present the mean and standard errors of the ratings on the quality of interaction.

Table S7 show the effect sizes of pairwise comparisons as Cohen's d and 95% confidence intervals. There was a significant effect of conversation topic on talk success ($F(2, 254) = 25.45$, $p < .001$, $p_{adj} < .001$), self contribution ($F(2, 254) = 16.02$, $p < .001$, $p_{adj} < .001$) and partner contribution ($F(2, 254) = 18.05$, $p < .001$, $p_{adj} < .001$). Participants rated the success of the discussion, their personal contribution to the topic, and their partner's contribution significantly lower in opinion talk compared to personal talk or small talk ($p < .001$ for mentioned comparisons). The difference between small talk and personal talk was not significant in all three ratings (talk success, $p = .207$; self contribution, $p = .720$; partner contribution, $p = .794$).

Comparison Between Personalised and Generic Avatar

We examined the effect of condition on the quality of interaction. The linear regression did not indicate any significant differences between conditions for talk success ($p = .186$, $p_{adj} = .372$, $d = 0.17$, 95% CI [-0.08, 0.41]), self contribution ($p = .127$, $p_{adj} = .372$, $d = 0.19$, 95% CI [-0.44, 0.06]) or partner contribution ($p = .827$, $p_{adj} = .827$, $d = 0.03$, 95% CI [-0.27, 0.22]).

Table S6

Means, Standard Deviations and 95% Confidence Intervals of Ratings on Quality of Interactions in Individual Topics

Measure	Topic	M	SD	95% CI
Talk Success	Small Talk	4.49	0.74	[4.36, 4.62]
	Personal Talk	4.61	0.64	[4.50, 4.72]
	Opinion Talk	3.86	1.20	[3.65, 4.07]
Self Contribution	Small Talk	4.60	0.63	[4.49, 4.71]
	Personal Talk	4.65	0.60	[4.54, 4.75]
	Opinion Talk	4.23	0.94	[4.06, 4.39]
Partner Contribution	Small Talk	4.70	0.48	[4.62, 4.79]
	Personal Talk	4.73	0.52	[4.64, 4.83]
	Opinion Talk	4.31	0.82	[4.17, 4.46]

Note: M = Mean, SD = Standard deviation.

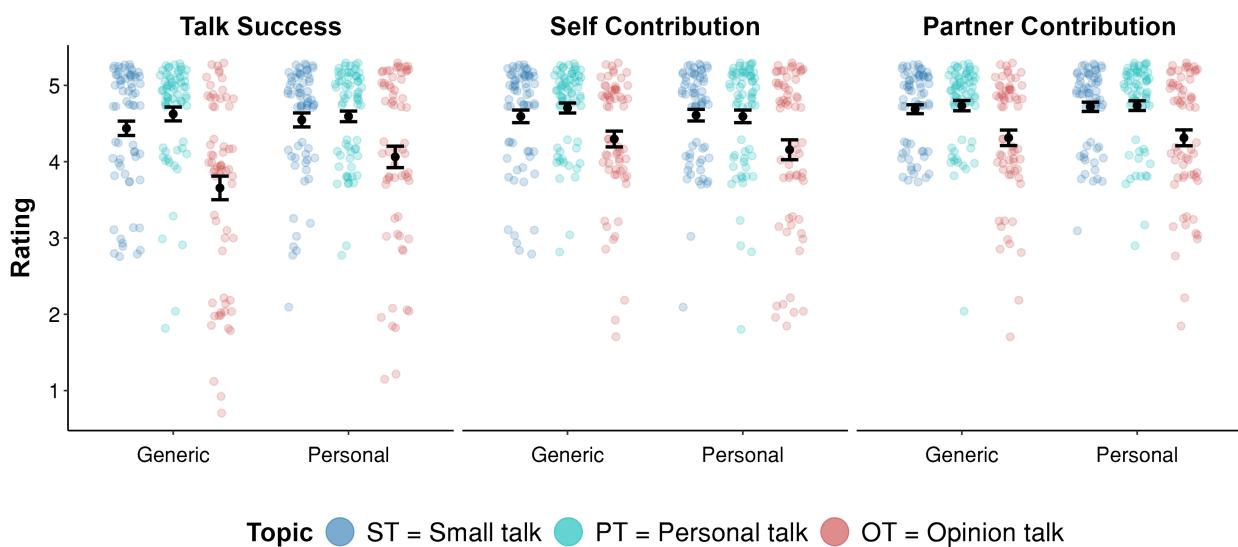
Table S7

Effect Sizes of Pairwise Comparisons of Quality of interaction in Individual Topics

Measure	Contrast	Cohen's d	95% CI
Talk Success	Small Talk vs Personal Talk	0.15	[-0.33, 0.02]
	Small Talk vs Opinion Talk	0.56	[0.38, 0.75]
	Opinion Talk vs Personal Talk	0.62	[-0.81, -0.43]
Self Contribution	Small Talk vs Personal Talk	0.07	[-0.24, 0.11]
	Small Talk vs Opinion Talk	0.39	[0.21, 0.57]
	Opinion Talk vs Personal Talk	0.50	[-0.68, -0.31]
Partner Contribution	Small Talk vs Personal Talk	0.06	[-0.23, 0.12]
	Small Talk vs Opinion Talk	0.50	[0.32, 0.69]
	Opinion Talk vs Personal Talk	0.50	[-0.68, -0.31]

Figure S2

Means and Standard Errors of the Quality of Interaction Ratings in Individual Topics

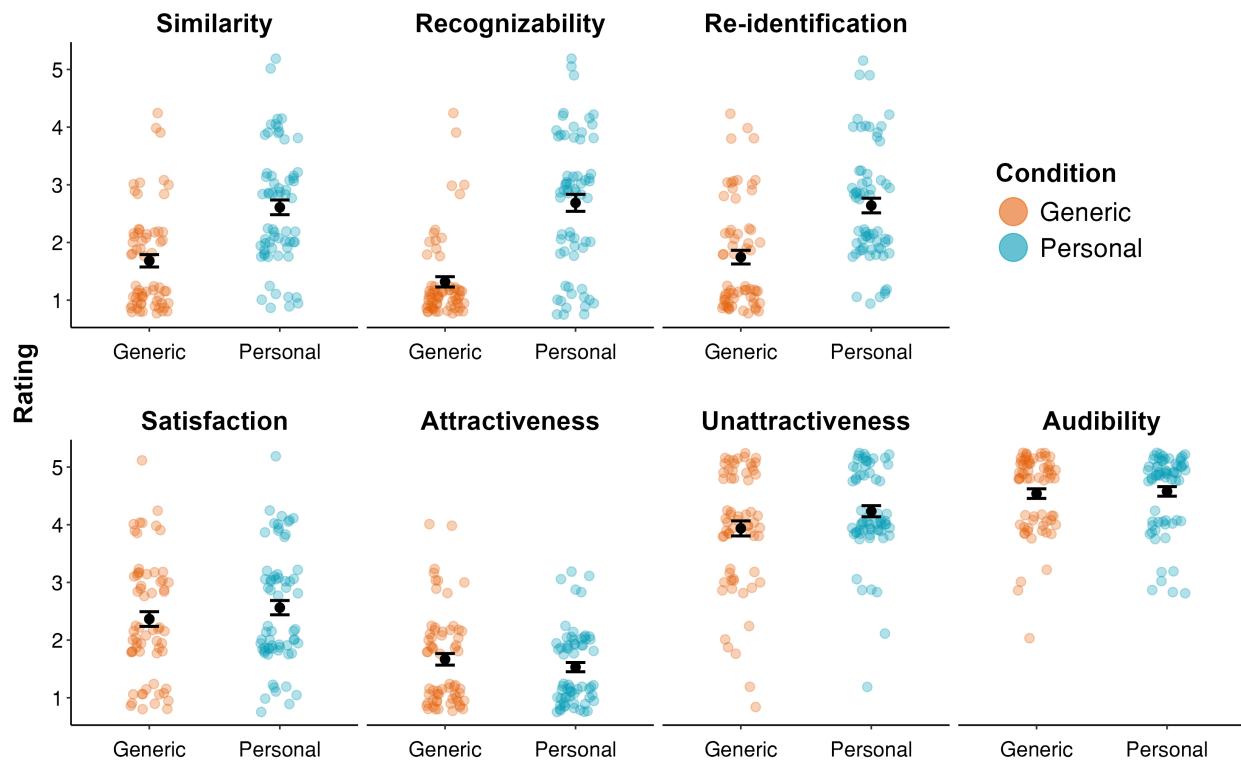


Note: Error bars represent standard errors.

Ratings of Avatar Appearance in the Conditions

Figure S3

Means and Standard Errors of Avatar Appearance Measures in the Two Conditions.



Note: Error bars represent standard errors.

Ratings of VR Experience in the Conditions

Table S8

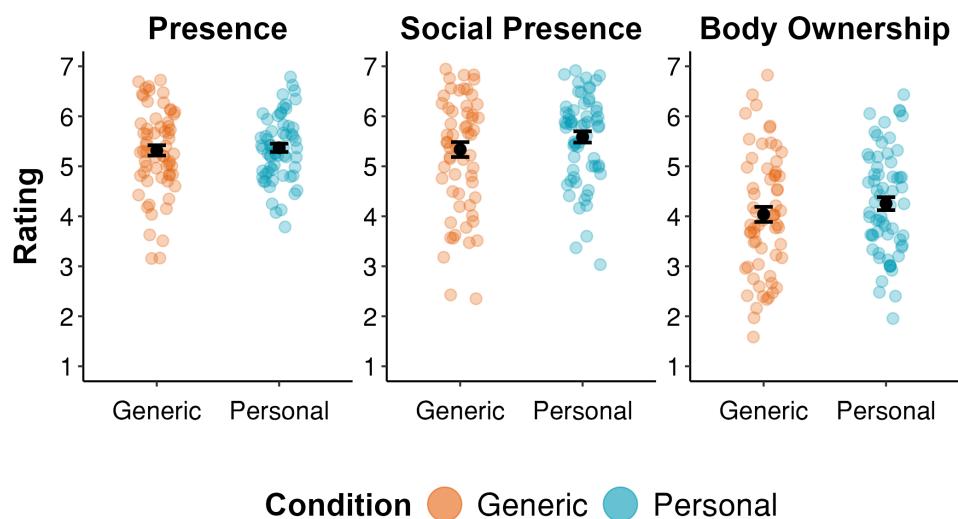
Means, Standard Deviations and 95% Confidence Intervals of Ratings on VR Experience in Individual Topics

Measure	Condition	M	SD	95% CI
Presence	Generic	5.32	0.82	[5.20, 5.43]
	Personalized	5.37	0.66	[5.28, 5.46]
Social Presence	Generic	5.33	1.19	[5.17, 5.50]
	Personalized	5.59	0.90	[5.46, 5.71]
Body Ownership	Generic	4.04	1.20	[3.87, 4.21]
	Personalized	4.25	1.03	[4.11, 4.40]

Note: M = Mean, SD = Standard deviation.

Figure S4

Means and Standard Errors of VR Experience Measures in the Two Conditions.

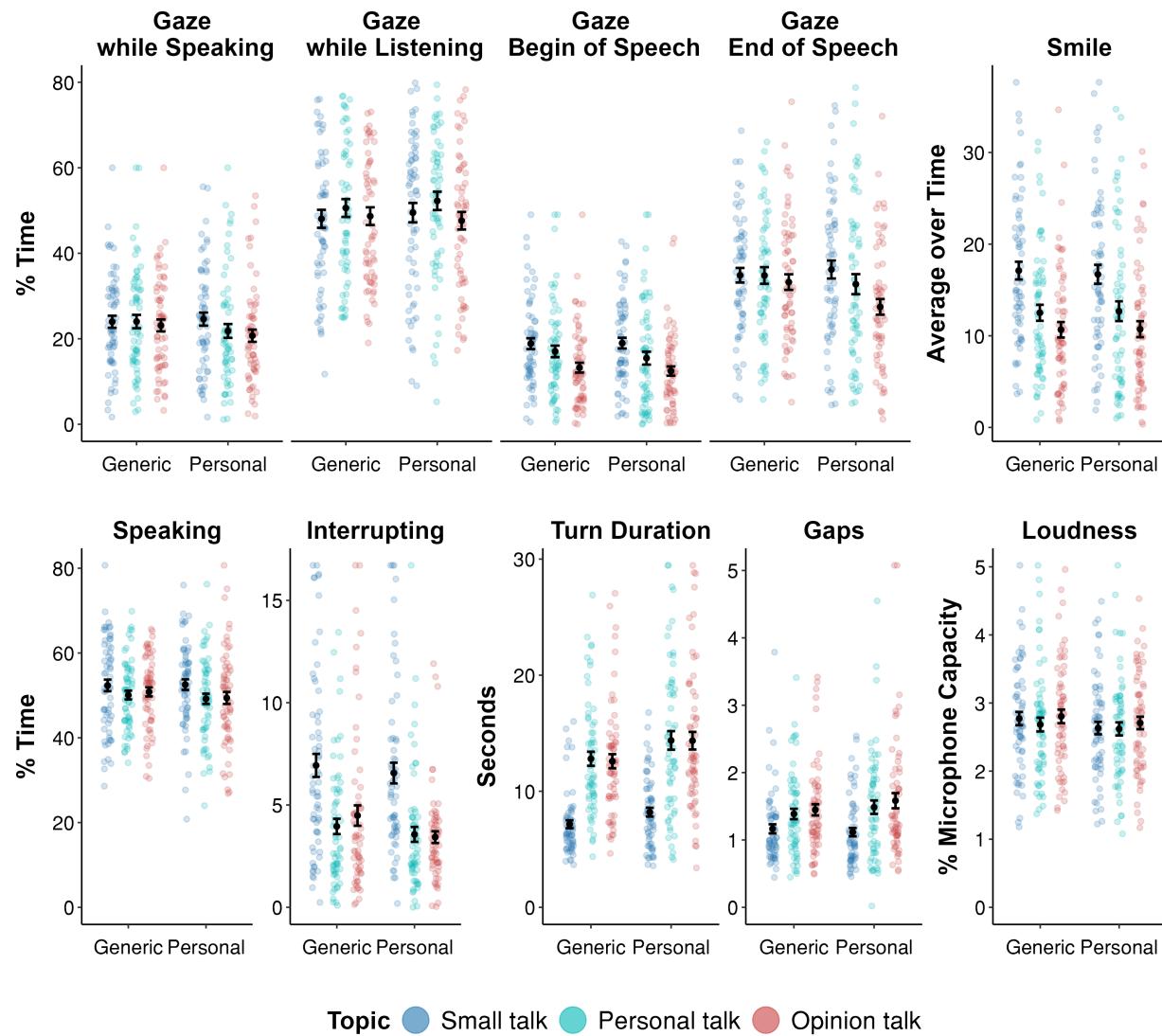


Note: Error bars represent standard errors.

Interaction Behavior in the Conditions and Topics

Figure S5

Means and Standard Errors of Gaze and Speech Behavior in the Two Conditions for Each Individual Conversation Topic



Note: Error bars represent standard errors.