



HAL
open science

Understanding Social Interactions in Reality Versus Virtuality

Aurélien Léchappé, Ross Johnstone, Aurélien Milliat, John H Williamson, Mathieu Chollet, Julie R Williamson

► To cite this version:

Aurélien Léchappé, Ross Johnstone, Aurélien Milliat, John H Williamson, Mathieu Chollet, et al.. Understanding Social Interactions in Reality Versus Virtuality. Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems, Apr 2025, Yokohama, Japan. pp.1 - 18, <10.1145/3706598.3713628>. <hal-05317038>

HAL Id: hal-05317038

<https://imt-atlantique.hal.science/hal-05317038v1>

Submitted on 16 Oct 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY-SA 4.0 - Attribution - ShareAlike - International License



Understanding Social Interactions in Reality Versus Virtuality

Aurélien Léchappé
LS2N, UMR CNRS 6004
IMT Atlantique
Nantes, France
Lab-STICC, UMR CNRS 6285
IMT Atlantique
Brest, France
aurelien.lechappe@imt-atlantique.fr

Ross Johnstone
Advanced Research Centre
University of Glasgow
Glasgow, United Kingdom
ross.johnstone@glasgow.ac.uk

Aurélien Milliat
LS2N, UMR CNRS 6004
IMT Atlantique
Nantes, France
aurelien.milliat@imt-atlantique.fr

John H Williamson
School of Computing Science
University of Glasgow
Glasgow, United Kingdom
johnh.williamson@glasgow.ac.uk

Mathieu Chollet
School of Computing Science
University of Glasgow
Glasgow, United Kingdom
mathieu.chollet@glasgow.ac.uk

Julie R. Williamson
Advanced Research Centre
University of Glasgow
Glasgow, United Kingdom
Julie.Williamson@glasgow.ac.uk

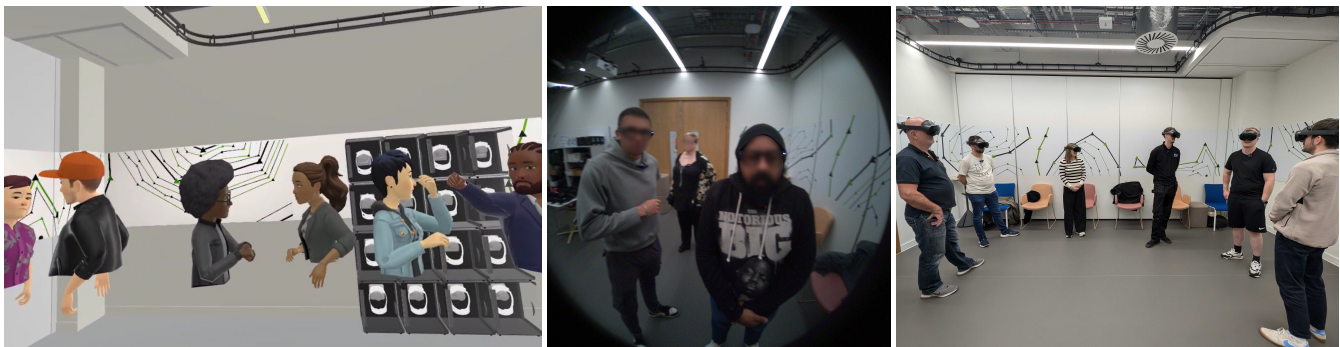


Figure 1: Participants completed social tasks in instrumented physical and virtual environments. **Left:** Participants complete conversations in pairs in a virtual environment, seen from within Meta Quest Pro. **Centre:** Participants play a game of charades as a team of three, seen from within Project Aria device. **Right:** Participants were photographed virtually playing two truths one lie in the physical environment, each wearing a Meta Quest Pro.

Abstract

Immersive realities enable social interactions that are radically different from traditional communication technologies, but how we experience immersion together is not yet indistinguishable from face-to-face interactions. Some social signals are not stable across realities, may change in semantics, or are missing all together. Understanding how social signals impact behaviours and experiences of social connection in immersive environments is key to creating experiences that are meaningful, satisfying, and productive. We completed a lab study where 6 groups of 6 participants (N=36) completed co-located social tasks in an instrumented face-to-face environment and its digital twin, creating a rich open dataset of 1.8

million rows across 45 columns. Our quantitative results demonstrate the stability of position as a social signal, measure lower social synchronisation in XR compared to face-to-face, and propose a method for bench marking XR against face-to-face interactions. This enables direct quantitative comparisons between experiences of co-located physical and virtual interactions for the first time.

CCS Concepts

• **Human-centered computing** → **Human computer interaction (HCI)**.

Keywords

Social XR, Social Signal Processing



This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

CHI '25, Yokohama, Japan

© 2025 Copyright held by the owner/author(s).

ACM ISBN 979-8-4007-1394-1/25/04

<https://doi.org/10.1145/3706598.3713628>

ACM Reference Format:

Aurélien Léchappé, Ross Johnstone, Aurélien Milliat, John H Williamson, Mathieu Chollet, and Julie R. Williamson. 2025. Understanding Social Interactions in Reality Versus Virtuality. In *CHI Conference on Human Factors in Computing Systems (CHI '25)*, April 26–May 01, 2025, Yokohama, Japan. ACM, New York, NY, USA, 18 pages. <https://doi.org/10.1145/3706598.3713628>

1 Introduction

Immersive realities (collectively called *XR*) allow us to connect with each other beyond the possibilities of traditional communication technologies [4], but what happens when we try to transpose our social interactions into immersive spaces? While XR opens up new forms of social connection, we do not yet fully understand how our face-to-face communications are changed or lost when we interact together in XR. This paper explores the differences between interactions face-to-face versus in XR by creating a novel dataset of social interactions in directly comparable instrumented physical and virtual environments during co-located experiences. This new approach to instrumenting behaviour lays a foundation for understanding how we transpose ourselves into social XR.

Harnessing the power of XR as a social technology is a growing area of research and innovation. Researchers have explored how we collaborate together in virtual environments [6, 11, 21], how we move through virtual spaces [17, 25, 42], and how we establish norms and social groups [12, 38, 43]. Commercial investments in immersive social worlds, like Meta Horizon ¹, Spatial ², and Bigscreen ³ are creating products that bring people together in immersive environments on a growing scale. The opportunity for connection and experience is immense, but there are still many unknowns for how we can best experience immersive realities together.

Social signals from face-to-face interactions provide a fundamental baseline that can inform how we can experience, design, and augment social signals for XR. A lifetime exposure to social encounters makes us experts at picking up subtle social cues and adopting the right verbal and non-verbal behaviour to maintain successful interactions [18]. Groups arrange themselves spatially in a seamless manner depending on occasion, relationships, and spatial environment [29]; individuals negotiate conversations by sharing the speaking floor and taking turns according to complex yet highly integrated social rules and patterns of behaviour [52], and we give and perform joint attention through mutual gaze [55]. Behavioural and psychological sciences have explored these patterns of behaviour, and provide us with a wealth of knowledge and models to describe behaviour and infer information about individuals, dyads, or groups during interaction. Our work explores these observable social signals and their stability across realities, aiming to identify how social signals translate across realities, which ones change, are missing, or break down in other ways, and how interaction difficulties can be observed in social signals. For example, detailed facial expressions are often missing or of a lower fidelity (compared to face-to-face interaction) in XR. What impact does this have on the social signals we do have available in XR, like conversational fluency?

To explore how our face-to-face interactions are transposed into social XR, we completed a lab study to directly compare face-to-face and XR-mediated interactions. We recruited six groups of six participants (N=36) to complete social tasks in fully instrumented physical and virtual environments. In the face-to-face condition, participants completed social tasks while using the Project Aria device [14]. The Project Aria device is a wearable sensing device in

a glasses form factor which replicates the sensors typically available on XR headsets, including spatial, hand, and gaze tracking. In the XR condition, participants are physically co-located and completed the same social tasks in a digital twin of the face-to-face environment using a Meta Quest Pro headset. The Quest Pro is a head mounted display which includes spatial, hand, gaze, and face tracking.

The resulting dataset is composed of over 1.8 million rows at 30 FPS of rich spatial, gaze, and hand tracking data for groups of six people and a facilitator interacting face-to-face and XR. This dataset is the first of its kind to enable direct comparisons between reality and virtuality for co-located individuals, and lays a foundation for additional comparisons in different configurations (for example, fully remote interactions [58]), different hardware affordances (for example, rich facial tracking [33]), and design interventions (for example, visualising conversation balance [35]). Our results expose new insights into designing more successful, satisfying, and productive social interactions in XR. We found that position [31, 56–58] is a notably stable signal between realities, but there is an experience gap because individuals do not *feel* like they are using space the same way between realities. We also measured social synchronisation [13, 29, 41, 59], and demonstrate lower social synchronisation in XR versus face-to-face, highlighting the challenges people face when building rapport in XR. We link the challenges in social synchronisation to conversational metrics like cross-talk, conversational equity, and silence to demonstrate that lower synchronisation leads to less fluent conversations. We discuss the implications for these results in how we can design more successful interactions in XR and additional future directions for using this approach to analyse more social signals, more types of social interactions, and more configurations of social experiences.

This paper makes the following contributions:

- A novel evaluation where participants (N=36) completed conversational social tasks in directly comparable and highly instrumented face-to-face and XR-mediated conditions.
- Analysis of social signals in XR, identifying and measuring signal stability and social synchronisation across realities for co-located interactions.
- An open source dataset and suite of analysis tools for analysing new social metrics from a standard data schema, which is the first open dataset of its kind and readily extensible.

2 Background

2.1 Being Social in XR

One of the core promises of Social XR is to enable “familiar, meaningful, and strategically heightened social interaction [7].” Bailenson highlights that social XR is a radical change from traditional digital communication technologies because there is a significant difference between seeing and doing [4]. Social XR experiences promise to enable users to remotely engage with one another, communicate and socialise as if it were (or augmenting) a normal face-to-face situation. Core to this goal is enabling users to rely on familiar social cues, including verbal and non-verbal signals reproduced on avatars embodying users in the XR environment. 6-DoF headsets enable users to move and orient themselves, facial tracking enable to reproduce audio-visual speech (lip-sync), facial expressions of emotions and gaze; hand tracking enables to reproduce complex

¹Online Virtual World Meta Horizon: <https://horizon.meta.com/>

²Online Gaming Platform: <https://www.spatial.io/>

³Social Entertainment Experiences: <https://www.bigscreenvr.com/experiences>

gestures on avatars. Such features of Social XR immersive experiences are meant to reproduce some of the most nuanced and highly integrated aspects of our daily social face-to-face experiences in the XR medium. Does a social interaction in XR *feel* like a face-to-face social interaction? Does a virtual smile observed in XR, generated by picking up on a user's facial expressions through facial tracking and reproducing it on an XR avatar in real-time, convey the same meaning as the original smile?

Social interaction is a complex, nuanced and coordinated activity. Much research, such as the pioneering works of Goffman [19], Kendon [30] and Argyle [2], have examined the many functions and patterns of behaviour in face-to-face social interactions. With the possibility of Social XR thus comes the central research question of whether patterns of face-to-face social behaviour reproduce in XR settings. Some of the earliest studies on the topic examined this very question. For instance, Bailenson reproduced Argyle's "Equilibrium Theory" [5] in VR: this theory postulates that comfortable levels of intimacy are negotiated by individuals through a balance of mutual gaze and interpersonal proximity [5]. An interesting tension arises here as XR becomes both an experimental tool to study social interaction, and an object of study in itself. The first position requires the strong assumption that social behaviour is stable in XR compared to face-to-face, but opens up great possibilities for psychological and behavioural research; the second position questions this assumption. Han et al. conducted an eight-week study using the ENGAGE VR platform, incorporating 192 unique immersive environments [24]. This approach provided a rich understanding of behavioural adaptation over time. The study highlighted how avatar design and environment influenced experiences: uniform avatars felt "costume-like," enhancing fun but reducing realism, while personalised avatars increased self-presence and self-consciousness. Head-motion synchrony decreased over time, with panoramic environments showing the greatest improvement. Miller et al. also explored interaction and behaviour over time [40], showing differences in personal distance and gaze over time, demonstrating increased personal space and mutual gaze as participants gained more experience interacting together in XR.

Collaboration in XR comes with its own set of challenges regarding social interaction compared to face-to-face situations. Quality of social interaction in XR can be affected by the sense of presence (e.g., the feeling of being in the virtual environment) and co-presence (e.g., the sense of being with others in the virtual environment) [23]. The type and quality of users' avatars can impact social presence: avatars with realistic behavioural features can enhance social presence [44]. Emotionally expressive avatars further contribute to a higher sense of presence, leading to greater engagement and improved task performance [15]. In contrast, avatars lacking non-verbal cues, such as postural, gestural, or facial expressions, may result in a weaker sense of social presence in XR, affecting individual behaviour and social interaction compared to real-life collaborative situations. A key challenge for social interaction in XR is to provide the behavioural cues that foster the social presence and allow the user to interact in the same state as in a face-to-face situation. An important question is then to define which indicators and in which format are relevant and necessary for social interaction in XR.

2.2 Communication and Collaboration

Collaboration in virtual spaces can be classified using the Time Matrix framework [27], which categorises collaborative activities based on whether they are performed synchronously or asynchronously and whether individuals are co-located or distributed. Asynchronous or distributed collaborative context limits verbal and non-verbal signals, thereby affecting team efficiency [16]. Within this continuum, XR emerges as a valuable tool, enabling people to gather in the same space, both spatially and synchronously allowing them to collaborate and have direct exchanges. This makes XR a powerful medium for fostering interaction with others [9], and enhancing social interactions by facilitating spontaneous group discussions and interactions with the virtual environment [23].

In social settings, interactions can be broadly categorised into two types: focused and unfocused. Unfocused interactions occur when people are co-present but not engaged in a shared activity, such as wandering about during a networking event. These interactions are primarily non-verbal, with individuals managing co-presence by displaying cues through posture and facial expressions to indicate availability for interaction [20, 30]. Focused interactions involve groups engaged in coordinated activities, such as conversing as a group or listening to a presentation [20]. This category covers a spectrum of social activities, some with highly formal rules around access, turn-taking, attention, and participation, and others that are less structured [20].

Compared to traditional interfaces, XR has a direct impact on fostering direct social communication and enhancing user presence and co-presence; however, although XR fosters communication and collaboration compared with other virtual media, how do XR experiences compare with real-life collaborative situations? Some aspects of immersive experiences in XR such as the quality of user's avatars, can play a critical role in collaboration and communication.

In this context, communication is a key component of the collaboration in XR, acting in conjunction with other dimensions on the success or otherwise of the collaborative activity, especially in tasks that rely on conversation. Metrics commonly used to assess communication include both speech content (e.g., dialogues, verbal discourse) and communication dynamics (e.g., equality, speaking time participation, proportion of silence and overlap of speech, turn-taking events, etc.).

2.3 Digital Proxemics and Quantitative Behaviour Analysis

To understand how individuals "synchronise" during social interactions, various methods analysing both verbal and non-verbal cues have been employed. Two foundational studies explored the concept of *interactional synchrony*, referring to how people naturally match their movements, gestures and expressions with others during social interactions [13, 29]. Using high-speed film and polygraph recordings of interactions between speakers and listeners, Condon and Ogston saw a correlation between body motion and speech in speakers, and that listeners' reactions exhibited synchronised patterns. This revealed that communication involves a complex exchange of verbal cues and physical responses, able to be characterised into predictable patterns and systematically analysed [13]. Kendon's work focused on how body movements correlate

with speech patterns through time-motion analysis. His study highlighted that listeners' body movements often mirror the speakers', and more synchrony suggested better engagement and attention [29, 30].

High fidelity body-tracking technologies have made analysis of behaviour easier to conduct. Using Microsoft Kinect technology, Won et al. explored nonverbal synchrony in collaborative settings [59]. They were able to accurately track body movements among dyads with real-time data on the positions of users' arms, legs, torsos and heads. A synchrony score was calculated by analysing and correlating the timing, coordination and mirroring of body movements between participants. This was further refined with machine learning algorithms to predict interaction patterns and assess synchrony. A key finding was that higher synchrony in body movements between pairs was associated with more harmonious interactions, with increased understanding and mutual engagement [59].

Krell and Wettmann studied nonverbal synchrony in the context of VRChat [34] using qualitative methods to assess how body synchronisation affects user experience. Interviews and field notes revealed that through alignment with their avatars' movements, VR interactions evoked heightened emotional engagement similar to real-life intimacy. This "corporeal synchronisation" enables practices such as virtual cuddling and holding hands, effectively replicating real-world intimacy despite physical separation [34]. Their findings demonstrate how nonverbal synchrony in virtual environments impacts social dynamics and emotional engagement, aligning with Goffman's theories on social interaction and co-presence [19]. This research illustrates how platforms like VRChat can foster meaningful social interactions and intimacy, turning digital spaces into venues for human connection.

Spatial social behaviour studies in the VR game *Second Life* demonstrate that users adjust movements based on avatar proximity, emphasising the significance of spatial context [17]. Research on co-presence in walkable shared VEs shows that physical co-location enhances the sense of being together and social presence, with task nature and gender influencing social dynamics [45]. Building on this research Williamson et al. introduced "digital proxemics," which examines how social proximity affects interactions within virtual environments (VEs). Unlike physical settings, VEs allow for dynamic manipulation of space, voice accentuation, and user representation, offering unique opportunities to study and influence social cues [58]. The study of proxemics in VEs also faces challenges, such as the absence of traditional social and environmental constraints, and the difficulty of incorporating olfaction and haptics, which are critical in physical proxemics. Furthermore, auditory experiences through HMDs differ from real-world perceptions [54]. Despite these challenges, VEs offer the intriguing potential to manipulate proxemic cues to shape virtual social dynamics effectively. Recent work on the injectX prototype [51] shows how behaviour augmentation can improve XR interactions. InjectX tracks and adjusts users' body movements, facial expressions, and gaze to make virtual interactions feel more natural and real. This new approach builds on earlier avatar systems and adds more sophisticated, real-time behaviour modifications, offering new ways to create meaningful social connections in immersive realities.

2.4 Open Challenges for Social Signals Across Realities

We aim to study how social interactions in XR differ from those in the real world, focusing specifically on social signals that characterise these interactions. Previous research has identified key aspects of face-to-face interactions, such as proxemics [22, 37, 58] (spatial arrangements and proximity), social synchronization [41, 59], conversation dynamics [1], and joint attention [40, 55], all of which influence social interactions depending on individual and contextual factors. Building on this, our study uses these established social signals as a baseline and explores their translation into XR environments. Research shows that some of these signals are relevant for conducting interaction, but there are many unknowns for how these translate into immersive environments.

XR environments often face challenges such as the absence of olfaction and haptic cues, as well as degraded non-verbal signals, which we expect will impact key social metrics like proxemics, social synchronisation, and conversational dynamics. Our study leverages a physical co-located XR design with face, hand and eye-tracking features that lays a foundation for direct comparisons between face-to-face and XR to begin understanding what we can learn from these social signals in XR. This research aims to identify which of these social metrics remain stable across both real and virtual contexts, and which may vary, and how this will impact social interaction quality. Through a dataset designed for direct comparison of social interactions in both settings, we will assess changes in these metrics in relation to interaction difficulties (e.g., social friction, communication breakdowns). By comparing these contexts, we will identify stable social signals that can be reliably used across contexts, as well as those that may require adaptation in virtual environments to improve social interaction and reduce related difficulties.

3 Evaluation

To create a dataset where we could make direct comparisons of social interactions between reality and virtuality, we needed to establish an experimental setup where we could instrument face-to-face and XR mediated interactions with comparable levels of fidelity and data streams. Previous works have explored purely virtual social interactions quantitatively [17, 31, 45, 57], and manually annotated face-to-face interactions [13, 30], but providing equivalent quantitative instrumentation across these contexts is unexplored.

While immersive experiences provide rich social signals like body position, body orientation, gesture, gaze, and speech, through the headset-embedded tracking systems, the same range of social signals requires significant instrumentation to capture during face-to-face interactions. Unlocking a direct comparison opens up new avenues of research, for example validating previous work that was guided by theory, benchmarking virtual interactions against their face-to-face counterparts, assessing social signal stability across realities, and evaluating which signal instabilities contribute to experiential differences between realities.

To better understand social signals between realities, we completed a lab evaluation (N=36) where participants completed social activities in groups of six in a physical lab (face-to-face) and its

digital twin (XR mediated). In our face-to-face condition, we instrumented participants using the Project Aria device [14], a wearable sensing device in a glasses form-factor. The Project Aria device provides the same data streams as typical immersive tracking systems, allowing a direct comparison to data collected using immersive headsets. In our XR mediated condition, participants completed the same social tasks using a Meta Quest Pro headset in a digital twin of the physical lab while being physically co-located in the same physical space enabling direct communication between individuals. The resulting dataset provides a novel direct comparison between social signals in physical and virtual environments.

This work is driven by two core research questions:

- **Which social signals are stable across realities?** This broad area encapsulates more social signals and contexts than can be explored in a single evaluation. For the scope of our own work, we focus on co-located interaction in reality and mediated by XR, and social signals for proximity and speech.
- **Which social signals and metrics indicate interaction difficulties across realities?** We still have limited metrics for measuring how interaction is unfolding using quantitative techniques across reality and virtuality. Our instrumented approach enables a new baseline for measuring interaction with reality as a baseline, where our first steps focus on established concepts of conversational metrics and interaction synchrony.

We explore the stability of social signals across realities because social signals form the foundation of how we conduct successful interactions. This is inspired by Garfinkle’s “breaching experiments” in ethnomethodology [18], which aim to expose behaviours that establish normality by intentionally disrupting interaction. For example, if you stand too close to someone they will step back to re-establish and define their comfortable standing distance. The behaviours of re-establishing normality are observable in reality, but if these social signals these are missing, incongruent, or unstable in XR, we hypothesis that these missing signals will disrupt interaction and indicate interaction difficulties.

3.1 Experimental Procedure

We designed four social tasks that could be completed face-to-face or mediated by XR for groups of 6 in a lab setting with the help of a facilitator. These tasks covered a range of different social constraints, including tasks involving all six participants, tasks which require unstructured simultaneous conversations, and breakout groups. Each group completed all four tasks face-to-face and mediated by XR. The order of the tasks was counterbalanced using a Latin Square, and half of the groups (3 groups of 6, N=18) completed the face-to-face tasks first, and half of the groups (3 groups of 6, N=18) completed the tasks mediated by XR first. A facilitator was present during all tasks to lead the experiment, manage timing, and ensure progress through the tasks.

Each session began with an overview of the evaluation, reading the information sheet, and completing the consent form. Participants were assigned a colour (Red, Yellow, Green, Purple, Blue, Orange), which identified them in both conditions. Each user completed a short tutorial of how to use the Meta Quest Pro headset and

completed an eye tracking calibration. Eye tracking calibration was not required for the Project Aria devices. After completing the four tasks described below, the participants completed a focus group interview. Each group lasted one hour in total. The experimental protocol was approved by the local ethical authority (Application number 300230229).

3.1.1 Task: Two Truths and One Lie. Participants were asked to make three statements about themselves, introducing two truths and one lie. This common ice breaking activity is a focused interaction [20] around the main speaker while the others try to assess which statement is the lie by asking questions and discussing the statements as a group. All participants had a turn to share their two truths and a lie, with three sharing their statements face-to-face, and the remaining three sharing their statements mediated by XR.

3.1.2 Task: Milling Line Up. Participants were asked to line up according to information about themselves that would require conversation, either by order of birth date within the year or by the farthest they had travelled from the experimental location. This unfocused interaction [20] required participants to move through the space in order to line up, and ask questions of the other participants in an unstructured manner to determine the correct order.

3.1.3 Task: Introduction in Pairs. Participants were asked to find a partner and spend two minutes introducing themselves to each other. These paired conversations happened simultaneously, creating a cocktail party effect during this task [49, 60]. Conversation in pairs is also a common benchmark for behavioural and conversation analysis [59]. When repeated face-to-face or mediated by XR, participants were asked to find a different partner.

3.1.4 Task: Charades. Participants were asked to form teams of three and nominate one person to act out charades, creating two breakout groups which could observe each other playing. The nominated “charader” was given a topic (Animals or Sports) and given one minute to act out as many ideas for their fellow team members, becoming the focus point of the interaction [20]. While one team was playing, the other team observed, making the charader an explicit “performer” during this task [19].

3.2 Physical Environment and Hardware

A key innovation of this evaluation is the instrumentation of the face-to-face interactions to make direct comparison to the XR mediated interactions. We also matched the physical and virtual environments using the digital twin of our lab environment for the XR mediated condition. The physical lab where the evaluation was conducted is a dedicated XR room with a dimension of 6.4 m x 6.4m, as shown in Figure 2. The physical space has limited furniture and obstructions to allow for easy movement in the space during XR experiences. To time synchronise the Project Aria devices, we used the TICSync library available in the Aria Research Kit ⁴, which synchronised the devices to a common time reference over a WiFi 6 router.

⁴TICSync in Aria Research Kit: https://facebookresearch.github.io/projectaria_tools/docs/ARK/sdk/ticsync

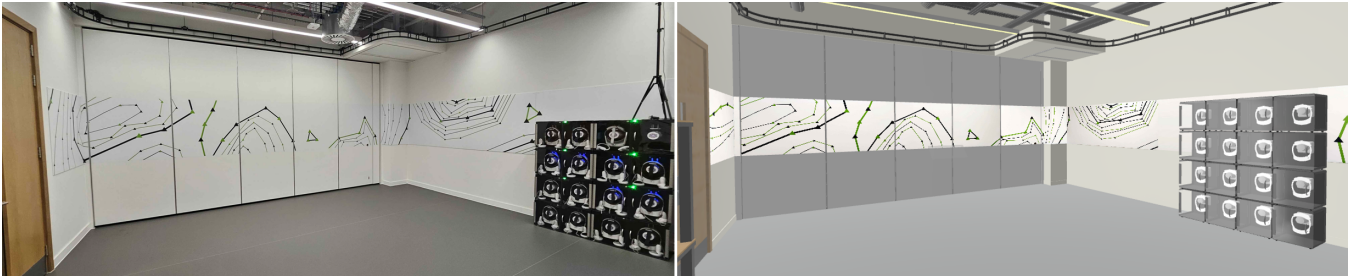


Figure 2: The physical lab setting (left) and its digital twin (right) provide the same environmental setting in the physical and virtual conditions. The space is a square room and a dedicated XR lab with limited furniture and obstructions to allow free movement throughout the space.



Figure 3: The Project Aria device (left) is a wearable sensing device in a glasses form factor used in our face-to-face condition. The device is for sensing only and has no display capabilities. The Quest Pro (right) is a head-mounted display which fully occluded reality in our XR mediated condition. Participants used hand tracking and were not provided with controllers.

3.2.1 Hardware. To instrument the physical environment, participants wore Project Aria devices [14], a wearable sensing device in a glasses form factor, as shown in Figure 3. The device includes five cameras, two IMUs, magnetometer, barometer, seven channel microphone, GPS, WiFi, and Bluetooth. This makes it possible to track the participant’s position, orientation, gaze, and hands in world coordinates, providing the same social signal data streams as our XR condition. The audio was recorded using Rode Wireless Go II kits, comprising one receiver and two transmitters. Four kits were employed to record audio from the six participants and the facilitator.

3.3 Virtual Environment and Hardware

A Unity environment was developed to serve as a 1:1 scale digital twin of the physical lab where the evaluation was conducted. Key elements from the physical setting, such as computers, headset boxes, and storage units, were added in the virtual environment as shown in Figure 2. Participants were represented by virtual avatars. Each participant could select an avatar from a diverse set of 32 options in the Meta Avatar sample pack, ensuring that no two individuals share the same avatar. To achieve synchronisation and co-location of participants in both physical and virtual spaces, we used spatial anchors provided by the Meta All-in-One SDK. For the networking component, Unity Netcode was employed to manage individuals and real-time interactions in the virtual environment.

3.3.1 Hardware. The setup included seven Meta Quest Pro headsets, as shown in Figure 3, all with face, hand, and eye tracking enabled. The audio was recorded using the same microphones as the face-to-face condition. Two computers were used for the XR mediated condition: one used as the Unity server and the other for data collection. The Unity server managed various tasks, including initiating and concluding data collection, starting audio recording, and managing the connections between all HMD clients. The data collection server was responsible for gathering data from the Unity server, each HMD, and the external microphone worn by each participant. To support the connectivity between all the devices, a Wi-Fi 6 router was used.

3.4 Participants

We recruited 36 participants from the local population using a recruitment agency. Participants were between the ages of 18 and 50 (average age 31.5), of which 18 were female. Because of the tracking devices used, all participants were required to be able to read without wearing glasses. Most participants had no prior XR experience, as assessed using a 1-7 Likert scale (average score 2.5).

3.5 Data Processing and Analysis Techniques

Table 1: A summary of the social features used for this analysis, including a breakdown of which features were present during interaction and how these were logged.

Feature	face-to-face		XR	
	Present	Logged	Present	Logged
Head Position	Yes	Yes	Yes	Yes
Head Orientation	Yes	Yes	Yes	Yes
Eye Gaze	Yes	No	Yes	No
Facial Expression	Yes	No	Yes	No
Hand Pose	Yes	No	Yes	No
Hand Position	Yes	Yes	Yes	Yes
Hand Orientation	Yes	Yes	Yes	Yes
Body Pose	Yes	No	No	No
Vocalisation	Yes	Yes	Yes	Yes

All of the raw data was organised into a dataframe using Python notebooks, as described in Appendix D. All the notebooks for organising and processing raw data files, generating the visualisations

in the paper, and the raw data and processed dataframe, will be provided as artefacts underpinning this paper, as described in Appendix B. Raw camera data, audio, and audio transcripts are not provided for privacy purposes. To process this data into a time series indexed dataframe, the face-to-face data and XR mediated data were read in to individual dataframes, all orientation vectors and world coordinates harmonised into a consistent convention, and the data was resampled to 1, 15, and 30 FPS for analysis. An overview of all the social signals present and logged during this study is shown in Table 1.

The raw data represents a unique and rich dataset on interactions that can be compared directly between reality and virtuality. The raw data is composed of 953,641,290 rows of positional tracking data, 32,464,680 rows of gaze data, and 41,124 rows of hand tracking data for the face-to-face condition. There are 38,691,640 rows of positional tracking data, 32,085,744 rows of gaze data, and 306,488 rows of hand tracking data for the XR condition, for a total of 1,057,230,966 rows of raw data for position, gaze, and hands in the total raw dataset. Note that in both conditions the hand tracking data is sparse as hands were not tracked unless they were within the field of view for both conditions. When resampled at 30 FPS, this is 1,858,509 rows of data in 45 columns, totalling 83,632,905 data points. This dataset is the first of its kind, with significant potential for secondary analysis, future bench marking, and dataframe standardisation. A summary of the social metrics derived from the social signals is shown in Table 2. Further detail on how each metric was calculated is in the relevant results sections.

3.5.1 Face-to-Face Data Processing. Data for the face-to-face condition was recorded using the Project Aria devices using Recording Profile 28 and processed using the Aria Research Kit⁵. The raw data recordings are formatted as VRS files, a file format designed to encode multimodal sensor streams into time sorted records⁶. These raw data files were processed using the Machine Perception Services available in the Aria Research Kit, which produce CSV files of SLAM tracking for position and orientation of the device, gaze tracking, and hand tracking. SLAM tracking is generated at 1000Hz, gaze tracking at 100Hz, and hand tracking at 30Hz.

3.5.2 XR Mediated Data Processing. Data for the XR mediated condition was collected using the open-source \psi framework [10]. This framework is designed to gather, synchronise, and record multimodal data signals, making it particularly useful for studying collaboration processes. In our study, we collected both continuous and discrete data. Continuous data included audio, eyes movements and position and rotation of the head and wrist of the participants. Discrete data captured gaze events, specifically identifying when a participant started or stopped looking at another person. Within the \psi framework, these gaze events were processed in real-time to filter out instances where one individual looked at another, termed "gaze on peers" events [26]. Continuous data were recorded at a frequency of 30 Hz. CSV files were generated during the XR activities to store head and hand position and orientation, as well as gaze data. The format of these CSV files was designed to match

that of the Aria Research Kit CSV files, facilitating easy comparison between the two conditions.

4 Results

These results are based on the significant dataset generated from instrumented face-to-face and XR mediated interactions and audio transcripts during the tasks and focus group interviews. This dataset encompasses data from six collaborative sessions conducted under both physical and virtual conditions, utilising Aria Glasses and Quest Pro devices, totalling 1,858,509 rows of data when resampled to 30 FPS. The combined duration of these sessions amounts to 3 hours and 18 minutes of task-related activity. Additionally, the synchronised audio recordings from these sessions, including 56 minutes of interview audio, result in a total of 4 hours and 14 minutes of audio data. The individual audio streams from participants collectively span 30 hours 30 minutes in duration. All of the data and the scripts for processing and analysing the data will be made available as an artefact with this paper.

4.1 Qualitative Results

As outlined in the interview protocol (see Appendix C), participants were asked a series of questions to compare their social experiences in both environments. The qualitative results were analysed from the focus group transcripts using a three stage coding process [53]. The transcript data includes 56 minutes of discussion with the 36 participants in groups of 6. Participant identifiers are shown as the group number and participant ID (Red, Yellow, Green, Purple, Blue, Orange).

4.1.1 Movement. Overall, the choice to use a digital twin of the physical lab helped participants feel oriented and comfortable within the space. *"I really enjoyed that the scenario was in the VR space. It's the same room with the wall and the VR headsets on the shelf. It really helped me orientate myself in the space"* (5O). Trust in the virtual room's layout helped some participants to navigate better: *"I kind of trusted the room because it was obviously the same as this."* (6O).

However, participants experienced several challenges and adaptations while interacting with the virtual environment and others in it. Participants generally felt they were cautious and concerned about collisions as they moved around: *"You feel a bit restricted"* (2P), *"I found it a bit disorientating"* (6Y). Perceived mismatches between reality and virtuality seemed to cause difficulties in spatial awareness and contributed to a scepticism of the space. 6P stated *"I don't know why but I didn't fully trust the space"*. 5G noted this mismatch: *"It's like when you've got a glass of water and you're seeing this false image, I could see people's feet in my periphery and know they were there. But the avatar would be ever so slightly different. It made me sort of second guess when I was walking around."* Without the ability to see their lower halves, gauging proximity and spatial positioning became difficult. 1B put this simply: *"See because you didn't have legs, it was hard to know where they were."* This combined with the responsiveness of the avatars lead to a heightened sense of caution and nervousness about potential collisions: *"I was worried about bumping into people, you know?"* (6G). Others had more visceral reactions: *"I can't fully move about without worrying you're going to hit someone or injure someone"* (4G). In short, latency, limited body

⁵Aria Research Kit: https://facebookresearch.github.io/projectaria_tools/docs/ARK/about_ARK

⁶VRS file format: <https://github.com/facebookresearch/vrs>

Table 2: From the raw social signals organised into a dataframe, we calculated additional social metrics. These metrics form the basis of our analysis and results.

Metric derived	Type	Features
Distance between individuals [22]	Proxemics	Head Position & Orientation
Synchrony score [41, 59]	Social Synchronisation	Head Position & Orientation
Equality index [3, 32]	Conversational	Vocalisation
Speaker participation [35] as speaker rank	Conversational	Vocalisation
Total Speaking Time (inversely silence) [1]	Conversational	Vocalisation
Overlap in speech (Cross talk) [1]	Conversational	Vocalisation

representation, and being co-located in the same physical space made users feel disoriented and more cautious.

Finally, participants faced issues with hand-tracking during the experiment, often finding it disorienting when tracking was lost. One described how “suddenly (peoples) arms would be right in front of you” (6P), while another noted that small hand movements sometimes felt exaggerated, as if “she’s going to punch me in the face” (5P). This disrupted immersion for some: “I felt as if I had another pair of arms that I was trying to keep under control” (2P). These distortions led participants to adapt their gestures. 6P put it this way: “You had to think creatively to translate real-world actions into the virtual world”, implying a concurrent secondary high-level task of managing your own bodily behaviour, which is of course absent in face-to-face. However, some participants praised hand-tracking for enhancing their experience, noting it was valuable for “picking up social cues” (5O) and appreciating the tracking quality: “I thought it was surprising how well you were able to use your hands” (4O). Participants found hand-tracking in the experiment both challenging and beneficial, highlighting a dual experience of both difficulties and advantages in virtual interactions.

4.1.2 Communication. The ability to see real facial features rather than animated ones was a key advantage in communication. As one participant noted, “It’s because you’re seeing their features rather than an animated face” (4O). In-person interactions also provided clearer emotional and social cues. For example, it was easier to understand humour face-to-face compared to XR: “You could tell when something was meant to be funny. It was harder to pick up on that in the virtual world” (5G). These issues were particularly apparent at the beginning of XR sessions, where the lack of clear cues led to confusion. Participants struggled with avatar selection and determining who should go first, a problem they felt would be easily resolved in person: “Nobody knew who would go first. We could have sorted that out easily in real life” (4P). Our replication of social signals (hand-tracking, face-tracking, eye-tracking) just failed to match up to the real thing when it comes to effectiveness in communication.

A key theme in the focus groups was the lack of detail in facial expressions. While the Quest Pro headset features real-time face-tracking, participants found it lacking, believing it doesn’t fully replicate human expressions. For example, avatars’ smiles during laughter “it didn’t feel real, but it felt more real than without it” (6O). This issue was especially evident in games like Two Truths One Lie, where participants struggled to detect lies due to missing social cues: “You can’t pick up on all their facial movements and have to

really concentrate” (1B). Groups found it challenging to discern lies and score points in XR due to these imitated social cues: “Charades is definitely harder in the headsets because of the arms” (2B). This extended to micro-mannerisms: “Tiny social cues you pick up on in person but miss in VR” (5P). Issues with eye contact and facial details were also apparent: “It’s the whole eye contact being removed” (2B) and “In VR, you don’t see the eyes or lips moving” (2Y). Additionally, the absence of full-body visibility reduced body language cues: “It’s easier in the real world to read body language compared to an avatar” (6Y). These limitations highlight the ongoing challenge of replicating nuanced human interactions in virtual settings.

Communication in XR presented challenges due to the absence of social cues typically in person. The lack of detailed visual feedback from avatars made it harder to gauge reactions and emotions, pushing participants to rely more on auditory cues. One participant noted, “I actually had to close my eyes, to listen, to try to make up for what I was missing” (5O). This reliance on listening stemmed from a lack of trust in the visual information provided by XR. This shift in communication dynamics even led some to alter their behaviour, as another participant found, “I found myself shouting more” (5O), highlighting their efforts to compensate for the missing cues and adjust to the communication dynamics of virtual environments.

Not only was there significant effort involved in expressing yourself, managing your avatar, but also in understanding others. There are two possible approaches, a photo-realistic avatar or an avatar with augmented behaviour. [51] However, some participants felt hyper-realistic avatars would be unsettling: “A super realistic avatar would be even more uncomfortable than a cartoonish one” (5G). This underscores the challenge of replicating human nature in VR: “You’re trying to replicate human nature, and human nature is already there” (6P). Achieving true expressiveness and understanding in virtual environments remains a significant endeavour.

4.1.3 Community in Anonymity. Participants felt a sense of detachment from identities in XR since their avatars were not customisable. As one noted, “I took my headset off and that’s what you actually look like. Totally weirded me out” (5P). Another added, “I didn’t realise how young you were until we took off the headsets” (4O). Common issues included difficulty in matching voices to avatars and identifying individuals: “You don’t even know who you’re talking to” (4P). In some ways this boosted confidence, making interactions more relaxed: “I’m under the presumption people are more self-conscious when they didn’t have the headsets on. It made it less “everyone’s watching me”” (5R). While some found this detachment increased

their confidence by reducing self-consciousness, others felt less confident and disoriented due to the lack of full-body representation: *“I found myself sort of quieter... My brain was still trying to figure out what was going with the lack of low body and everything”* (5G).

The digital twin of the physical lab, combined with the experimental protocol, quickly fostered a sense of community: *“We were just all in it together”* (2Y). Participants also enjoyed the virtual world’s playful elements: *“Something in real life wouldn’t be as much fun”* (6P) and *“It was more fun in the virtual world because there was another element”* (6O). This enjoyment might have influenced the data. One participant said, *“I had to focus myself because I couldn’t stop laughing”* (6P). Distractions from unfamiliar avatars and hand-tracking were noted: *“Felt like I was staring a lot”* (2G) and *“It was a novelty, just looking at your hands”* (6O).

4.2 How We Use Physical and Virtual Space

Digital proxemics describes how we use virtual space using metaphors from physical proxemics [22, 58], but our data provides the first quantitative validation of previous results and uncovers additional nuance between how we use physical and virtual space when colocated. To analyse the proxemic distances, we calculate a matrix of pairwise Euclidean distances for each frame in the time series use the upper triangle (excluding the diagonal) to generate a histogram. Figure 4 gives an overview of the proxemic plots during face-to-face and XR mediated activities, which were calculated using the data resampled at 15 FPS.

Proxemic distances were notably stable between reality and virtuality, with participants maintaining comparable probability distribution of social distances across all of the tasks in our evaluation. Even during the “Milling Line Up Task,” which required movement through the space, there are no significant differences in the probability distributions shown in Figure 4. This is an important validation of related works, where proxemic analysis has drawn from theory, but not been quantitatively compared to face-to-face interactions. The probability distributions show some variations between tasks, in particular the “Introduction in Pairs” and the “Two Truths and a Lie” tasks, where participants were more likely to make use of the personal distance.

In both face-to-face and XR interactions, there is a clear drop in likelihood to stand within the ambiguous zone at the edge of personal and social distances. The data demonstrates a strong preference to stand clearly within or outside of the personal zone, with limited time spent standing at the edge of the personal zone. A notable difference between reality and virtuality is the presence of outliers within the social distances that can be seen in the tail end all the plots for the XR condition - in XR someone is always “standing on the fringes.” Given that this data was collected within subjects, the same people did not exhibit these behaviours when interacting face-to-face, as in XR. The “anonymity” of XR as highlighted in the qualitative data can create a situation where individuals disengage, stand back from the group, or otherwise remain on the fringes of interaction in ways they would not face-to-face.

The stability of proxemic distances across realities highlights a critical experiential gap: participants felt they were using space differently in XR but they were not. The qualitative data highlights many of the technical limitations of tracking like loss of hand

tracking and potential differences in the angular field of view an avatar occupies compared to one’s actual body [22]. But these issues did not result in using space differently, just in the perception of using space differently. Closing this experience gap is a crucial open challenge to improving the experience of being in XR together.

4.3 Synchronisation and Social Cues

Social synchronisation describes how we synchronise our movements when building rapport, and quantifying this continuous social behaviour can provide new insights on behavioural data. We followed the approach described by Won et al. [59] to calculate a synchrony score based on the change in angular movements between interacting individuals. We focused on the head movements synchronisation, one of several body locations described by Won et al., because this signal had the best stability between reality and virtuality and did not include tracking gaps seen in the hand tracking data. We calculated the synchrony scores by measuring the change in head vector frame by frame over 1 second intervals, computing the Pearson R correlation of the x , y , and z head orientation vectors individually for the duration of the interaction, and adding these correlation coefficients together for an aggregate synchrony score. Scores near 0 indicate no synchronisation, negative scores indicate de-synchronisation, and positive scores indicate synchronisation. Figure 5 shows the synchronisation scores for selected activities as a boxplot comparing synchronisation face-to-face and mediated by XR.

The results demonstrate that the participants were more successful at synchronising movements face-to-face compared to XR, where XR synchronisation was often actively desynchronised with negative correlations. Using a t-test to compare the means for the synchronisation scores between face-to-face and XR mediated for each task, we found significant differences for all tasks (Partner Conversation ($t=3.1$, $p<0.01$), Milling Line Up ($t=2.0$, $p<0.01$), Two Truths and a Lie ($t=2.8$, $p<0.01$)). Figure 5 also demonstrates that in all conditions some individuals were highly or poorly synchronised. This is not surprising, as a distribution of “open” and “closed” individuals will be present in groups of strangers. But the overall ability of the participants to synchronise in XR was statistically significantly lower when compared to face-to-face interactions. We see the highest variation during the “milling line up” task, which did not have a single focus point and would have resulted in more divergent behaviour compared to conversations in pairs or focused interaction in a group during “two truths and a lie.”

4.4 Conversation

Vocalisation metrics were derived processing the audio data using the voice activity detector (VAD) provided in the \psi framework. VAD is a speech processing technique to detect the presence or absence of human speech. Noise and false detections were manually filtered. For our conversational analysis, we focused on two tasks where speech and coordination were important: Milling Line Up and Two Truths and A Lie. The Milling Line Up task is distinguished by the other due to its unstructured interactions and the absence of an explicit leader or focus point. In this context, analysing conversational parameters such as speech equality, and the proportion of “cross talk” and silence during the activity provides valuable

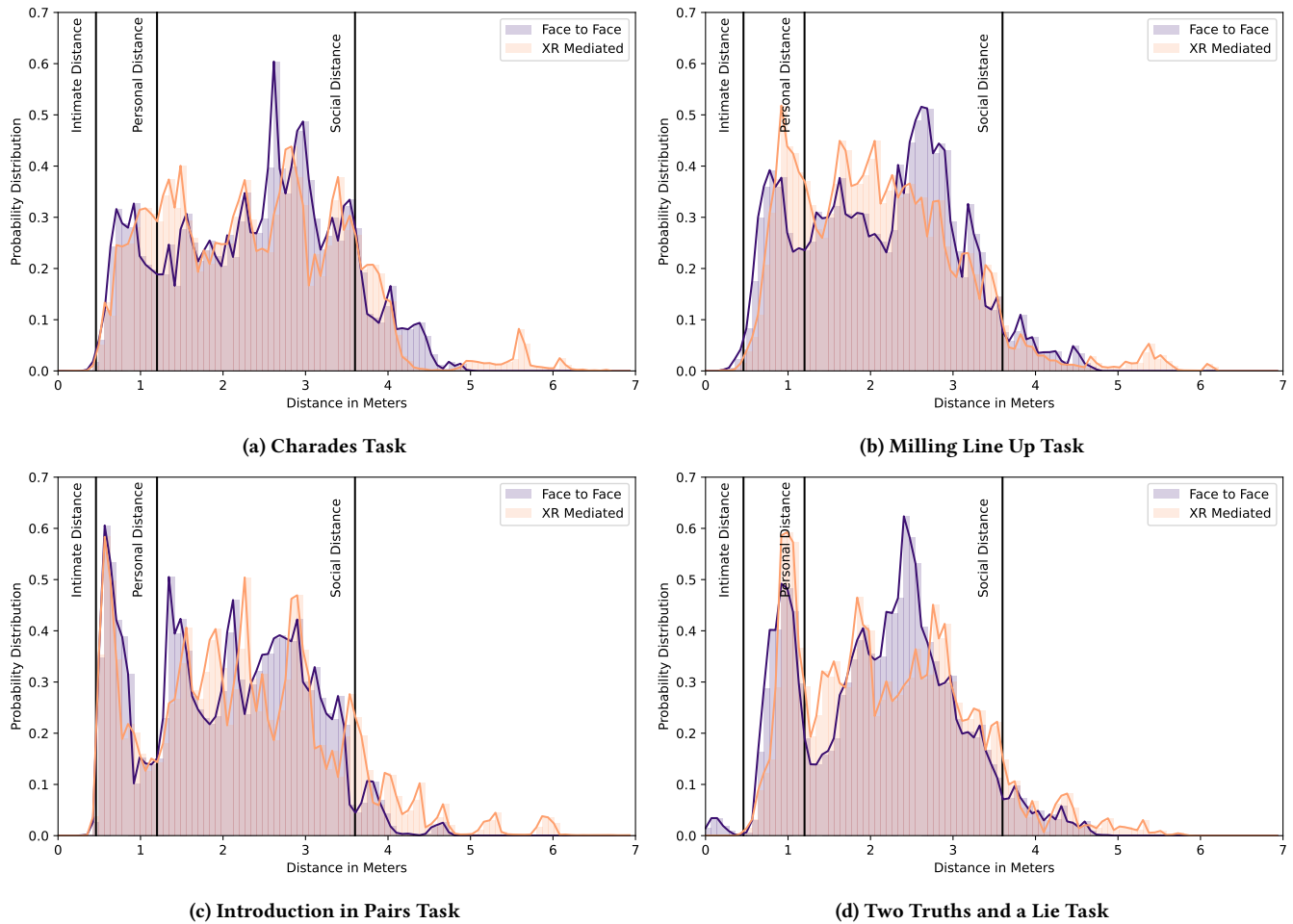


Figure 4: Proxemic analysis of position during face-to-face and XR mediated activities shows a probability distribution of pairwise comparisons of standing distances, with the intimate, personal, and social zones marked [22]. Overall proxemic distances are notably stable between reality and virtuality.

insights into how participants collaborate across face-to-face and XR contexts.

Speech equality is an index computed on the total speaking time of each collaborator during the task. A high participation equality (represented by an index near 0) means less dominance and is related to better collaboration quality [3, 32]. The proportion of cross-talk and silence are relevant to evaluate the quality of the communication dynamics and conversational flow between individuals: a low proportion of cross-talk (i.e. interruptions or overlaps in speech) is associated with better turn-taking quality [1], while a high presence of silence can indicate poor collaboration quality [48].

The non-parametric Wilcoxon Signed-Rank tests were performed in order to compare the non verbal metrics between the face-to-face and XR conditions. Comparing speech equality, proportion of cross-talk, and proportion of silence between face-to-face and

XR conditions, we only found significant differences in the proportion of silence ($W=3.0$, $p<0.01$), indicating a particular challenge for turn-taking in XR. Abdullah et al. observed an increase in cross-talk in VR [1], indicating that missing social signals in VR made it challenging to know when to interject. Our observed increase in the proportion of silence indicates a related problem: lack of an attempt to interject. Our conversational task was different than Abdullah et al. [1], which may explain the different conversational challenges we observed. Given our qualitative results around communication challenges, it's surprising we do not see more of these challenges represented in the quantitative data. Again, we observe an experience gap, where participants perceive differences in behaviour between XR and face-to-face, when we do not observe these differences quantitatively.

We also explored how the conversation metrics correlated with the synchronisation scores, as shown in Figure 6. Although our small sample sizes for these metrics calculated as overall values for a

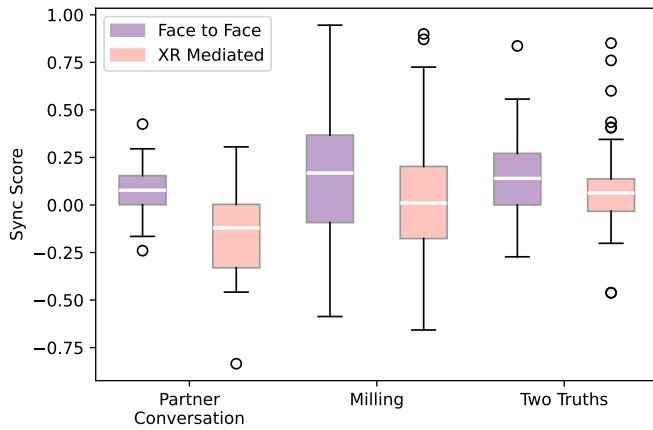


Figure 5: The synchronisation score describes how correlated head movements are during different social activities. Participants are significantly less synchronised in XR compared to face-to-face.

given task are small, these initial results suggest further exploration of the relationship between synchronised behaviour and conversational outcomes is needed. Figure 6 shows positive correlation between synchronisation score and speech equality, synchronisation score and silence, and a negative correlation between synchronisation score and speech overlap. If analysed with larger samples of repeated trials, this approach could provide valuable benchmarks for measuring improvements to behavioural and conversational metrics in the future.

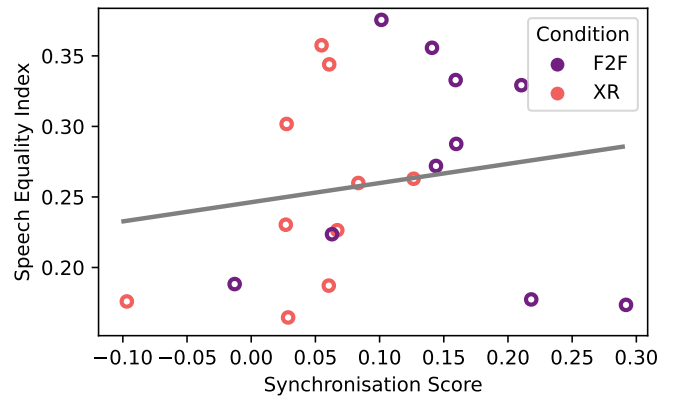
5 Discussion

Our findings contribute to the growing domain of social XR behavioural analysis and our novel approach to understanding social signals across realities lays the foundation for rich analysis between reality and virtuality, establishing new benchmarks for social XR. The results of our evaluation explore social signal stability between realities during co-located experiences, and where we see differences in social signals we have triangulated this with qualitative data and theory to analyse the signal in terms of interaction challenges. We also contribute our methodology and a significant open dataset created during this experiment, with the goal that this approach inspires new techniques for benchmarking social XR against face-to-face interactions and our data schema provides structure for the community to build a corpus of open social XR data.

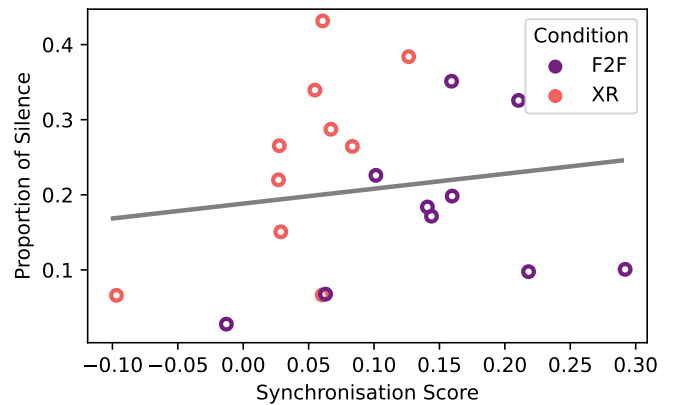
5.1 Experience Gaps in Current Social XR

We observed that qualitative interaction challenges and the resulting social synchronisation scores were clearly observable in the quantitative data, while other issues like proxemics and personal space were stable across realities but different in qualitative experience. We identified these experience gaps, but the root cause and how to close these gaps is still an open challenge.

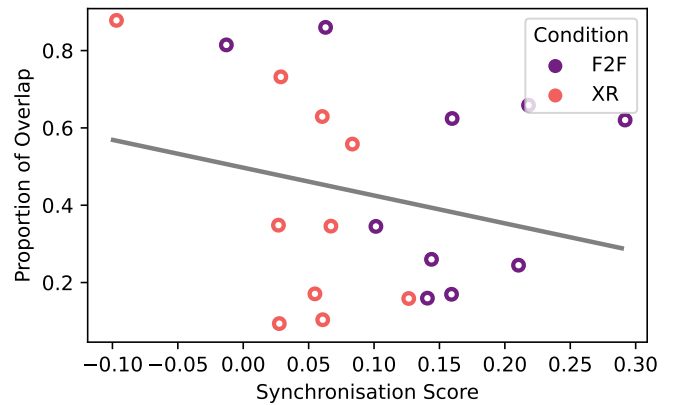
We focused on digital proxemics because this signal is part of growing body of work [17, 31, 58] and is well grounded theoretically [22] and technically [50]. In XR, participants described feelings of



(a) Synchronisation Score and Speech Equality



(b) Synchronisation Score and proportion of silence



(c) Synchronisation Score and proportion of overlap

Figure 6: Plots showing the relationship between vocalisation metrics and synchronisation scores with a fitted trend-line.

uncertainty and nervousness over collisions and took care when navigating the virtual space. However, quantitative proxemic analysis didn't show any behavioural difference in how people used space between reality and virtuality. This discrepancy suggests an experience gap, where the subjective feeling of caution and careful movement did not translate into any measurable changes in behaviour. Such experience gaps present a challenge for HCI researchers - how to identify, improve, and measure design interventions in XR where quantitative data doesn't measure changes in experience? While mixed-methods approaches are popular because many questions can't be answered by qualitative or quantitative data alone, an experience gap like the one we've identified exposes a particular challenge.

Given the current capabilities of social XR technology and our experimental design, there are several potential causes of and solutions for the proxemic experience gap we identified. The physical and virtual co-location could've created a sense of unease about spacial orientation and interaction. Being unable to see avatars' legs led to participants glancing outside of the headset to see the real world. Noticing mismatch between the visual appearance of avatars and their actual bodies, participants struggled to gauge proximity and movement of others. Podkosova et al. compared closest passing distances for a mixed of co-located and distributed users [45], and found small differences while passing others who are co-located compared to distributed. However, this focus on passing differences rather than continuous position may skew the data because proxemics in physical spaces is a continuous action which is constantly adjusted.

A commonly discussed issue in social XR is the realism of avatars and avatar embodiment. Although avatar embodiment is not the focus of this work and thus we did not allow for customisation beyond selection from a closed set, avatar issues came up in the qualitative data and could explain some of the proxemic experience gap. While participants brought up the mismatch between the realism of face-to-face and avatars, photo realistic avatars risks falling into the "uncanny valley." Research, including work by Bailenson [4], suggests that photo-realistic avatars are not ideal for social interaction. Instead, avatars with augmented behaviour such as improved facial tracking and expressive gestures can enhance virtual interactions. Virtual interaction can become more familiar and engaging without falling into the uncanny valley [51].

Another issue with avatar representation is that the shape of the avatar may not be congruent with the physical person, creating a mismatch between the visual angle taken up by the avatar and the person. Since visual angle is a key cue for physical proxemics [22], this parameter may need to be adjusted or actively manipulated in order for proxemic distances to "feel" the same in reality and virtuality. Such manipulations could be initially informed by theory, but would require controlled experimentation to understand the perceptual and experiential thresholds.

We chose to use a synchronous co-located XR design, characterised by users immersed in a virtual space while being physically present in the same location at the same time. Users' positions in the virtual and physical worlds are the same and all movements within the physical and virtual environments are matched. The impact of this type of co-location on social interaction and collaboration has been less explored. Validated design criteria for promoting effective

collaboration and social interaction in co-located social XR experiences are still lacking due to insufficient research. Recent studies suggest that being physically co-located in XR enables direct communication and helps users communicate more naturally [8, 36], thereby enhancing the social experience of users and promoting better collaboration. Feng et al. [16] identified 4 key design considerations for XR spaces that support effective collaboration: *role script design, task design, collaboration mechanism design, and feedback and communication design*. In relation to these considerations, they explored factors of collaborative experiences that affect collective efficacy and social experiences in a synchronous co-located five-player VR tower defence game. They found that spatial presence and natural feedback when interacting in the virtual environment, are two important factors that affect such XR design experiences, and that synchronous and co-located virtual collaboration puts high demands on individual skills, such as motor skills. Our study also contributes to identify factors mediating user experiences in co-located XR settings, especially with respect to the '*Feedback and communication design*' category: Individuals can have difficulty to gauge proximity, due to a mismatch between collaborators' avatars and their physical appearance. This observation highlights a challenge that needs to be explored to investigate its impact on social interaction in XR.

Our findings highlight interaction difficulties for specific immersive experiences, where users are exposed to both virtual and real-world social cues. These results may require new interpretation when applied to fully virtual settings, where physical cues are absent. However, previous work would indicate more similarities than dependencies when immersive interactions are fully remote. For example, our previous work explored fully remote social interactions, demonstrating that HMD users were still conscious of personal space, avoided collisions, oriented towards others, and kept the active speaker in view [58]. A key difference making direct comparison more complicated is how movement is constrained in our co-located interactions compared to the unconstrained movements in remote interactions. This results in a mismatch between physical and virtual social signals that remains an open challenge. Furthermore, concerns about avatar realism and embodiment may carry different weight when users do not see their real-world counterparts. While our study focuses on co-located interactions, future work should explore whether similar experience gaps emerge in fully remote VR or other XR modalities along the reality-virtuality continuum [39].

5.2 Social Synchronisation and Interaction Difficulties

A key aspect of social interaction is the presence of non-verbal behaviour. Having access to these cues enables us to interpret and make sense of the actions and emotions of others. In the literature on social interaction in immersive realities, it is known that transcribing these cues can be complex. The absence of non-verbal cues can degrade interactions and affect the quality of interaction. To compensate for this, we have seen the emergence of virtual avatars capable of replicating increasingly more behaviours of individuals, for example by tracking hands, body, legs or facial expressions. In this study, we used Meta avatars that have hand, eye and face

tracking, with the aim of offering social interactions as close as possible to what a face-to-face collaborative situation might be.

Despite this, we observed low social synchronisation in XR compared to face-to-face and qualitative analysis revealed communication difficulties in XR. Participants cited the lack of fidelity of some social signals in the XR condition in terms of hand tracking, eyes, lips sync, or micro-mannerisms. This raises open questions of how well these non-verbal signals can be transferred into XR using current technologies, and what impact they have on social interactions. It would also be interesting to consider the extent to which the challenges reported in this study are due to technical limitation of the current technology, or of avatar fidelity toward reality. Participants also indicated that to overcome this lack of social cues, they relied more on other cues, such as voice, focusing on other participants' speech. Future research should consider assessing the stability of different levels of fidelity of non-verbal cues in the XR environment compared to the face-to-face condition, and testing their impact on social signals and metrics.

If we cannot completely capture and display social signals in XR, perhaps manipulating the social signals we currently have access to could improve social synchronisation. Bailenson describes a potential manipulation where one's body movements could be transposed onto another to simulate social synchrony when in reality their movements might be de-synchronised [4]. Would this "manipulated" social synchronisation result in observable differences in behaviour and experience?

Users may also learn to adapt their behaviour and establish new norms based on the interaction medium, similar to how we develop norms for video teleconferencing to try to stabilise these interactions in the absence of social cues and technical issues like latency. Our qualitative analysis suggests that users focused more on discussion to counterbalance the lack of expected social signals, for example listening closely to vocal cues during the "two truths and a lie" task in the absence of expected facial cues. Further social metrics, for example detailed turn taking analysis or mutual gaze analysis, could expose the adaptations between reality and virtuality and open up new areas for designing for social XR.

5.3 Data Schema and Future Extensions

We have only scratched the surface of possible research questions and directions enabled by our methodology and resulting dataset. We focused on co-located interactions face-to-face and mediated by XR, exploring specific signals and metrics for the scope of this paper, but additional analysis on gaze, hands, interpersonal social signals, and new social metrics we have not yet thought of are all possible using our approach and dataset. The scale and richness of our dataset establishes a solid baseline for new opportunities for further analysis and extension with additional conditions. For example, our work does not yet include a fully virtual condition, but extending this dataset by adding additional factors and conditions within a standard schema is vital to addressing the broader challenges of understanding how we experience immersive realities together. By releasing our dataset, analysis, and visualisation code, we hope to encourage secondary analysis and extension of our dataset by others.

In developing this dataset, we also defined a data schema which could standardise open datasets created from any XR experiment, in line with related works that define datasets for single users in immersive environments [47]. We define a series of columns based on readily available signals used in XR tracking and a growing set of social metrics that can be calculated from these columns. We describe this data schema in Appendix D and give clear conventions for positional and vector data needed to make comparisons between open datasets. Standardising the format (and in particular the coordinate conventions) and building a corpus of XR experimental data for multi-user interactions is needed to accelerate HCI research in XR. While searching for multi-user datasets in XR to further contextualise this work, we found a notable gap for publicly available datasets that would enable any direct comparisons. There are publicly available datasets for single users, for example Liebers et al. datasets on identifying users based on motion dynamics [28], and game-play data also aiming to identify users based on motion dynamics [46], but our interpersonal social signals like digital proxemics and social synchronisation are not possible with single-user datasets. Where datasets do include multiple users, for example the Williamson et al. digital proxemics dataset [58], many factors vary between that experimental setup and our own work making direct comparisons meaningless. If we want to design more productive, pleasurable, and successful social XR experiences, a principled and data-driven approach driven by the research community could be transformative. This requires the establishment and growth of a significant corpus of compatible data capturing every aspect of how we interact together in immersive realities.

6 Conclusion

Being social in XR has the potential to change the way we connect with each other, but social XR is still far from our experiences of face-to-face interactions. In this paper, we lay a foundation for analysing directly comparable social signals in reality and virtuality as a method for understanding how we transpose our face-to-face interactions into XR. Our results demonstrate how social signals can provide insights into interaction challenges when social synchronisation can be measured, where current XR technologies result in low synchronisation compared to face-to-face interactions. The benchmarks we measured of synchronisation behaviours face-to-face are a valuable tool to assess future design interventions in XR to improve social synchronisation and conversational metrics. We also identified position/proxemics as a stable social signal across realities, but there is an experience gap because participants perceived that they used space differently in XR when the quantitative data says otherwise. Finally, we produced a significant dataset with over 1.8 million rows at 30 FPS from our instrumented face-to-face and XR mediated interactions. We have only scratched the surface of possible analysis directions in this paper and the dataset presents many opportunities for secondary analysis, extension, and future directions. The data schema we designed for this dataset also provides a standard schema that will enable comparison between any social XR dataset using this format and its derived metrics, which can be readily created from any XR experiment that logs tracking data. The future of XR is social, but without principled design and

rich data about interaction across realities, we may miss the social connections that XR makes possible.

Acknowledgments

All of the code, data, and tools for analysis used in this paper are openly available on GitHub: <https://github.com/julierthanjulie/fusion-project-chi-2025>.

This research would not be possible without the participants who supported this research by joining our evaluation and contributing to the open dataset shared with this paper. This research was supported by the Project Aria Research Kit [14], which made our face-to-face data collection possible.

This research was supported by ERC Consolidator Grant FUSION, proposal 101126024, (funded by the UK Horizon guarantee scheme EPSRC project EP/Z000432/1). John H. Williamson was supported by ERC Advanced Grant DIFAI, proposal 101097708, (funded by the UK Horizon guarantee scheme EPSRC project EP/Y029178/1). Aurélien Léchappé was supported by a grant from the SPIN doctoral school from IMT Atlantique.

References

- [1] Ahsan Abdullah, Jan Kolkmeier, Vivian Lo, and Michael Neff. 2021. Videoconferencing and Embodied VR: Communication Patterns across Task and Medium. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (2021). <https://doi.org/10.1145/3479597>
- [2] Michael Argyle. 1972. Non-verbal communication in human social interaction. *Non-verbal communication* 2, 1 (1972).
- [3] Khaled Bachour, Frederic Kaplan, and Pierre Dillenbourg. 2010. An Interactive Table for Supporting Participation Balance in Face-to-Face Collaborative Learning. *IEEE Transactions on Learning Technologies* 3, 3 (July 2010), 203–213. <https://doi.org/10.1109/TLT.2010.18> Conference Name: IEEE Transactions on Learning Technologies.
- [4] Jeremy Bailenson. 2018. *Experience on demand: What virtual reality is, how it works, and what it can do*. W. W. Norton & Company, New York, NY, US. 290 pages.
- [5] Jeremy N. Bailenson, Jim Blascovich, Andrew C. Beall, and Jack M. Loomis. 2001. Equilibrium theory revisited: Mutual gaze and personal space in virtual environments. *Presence: Teleoperators and Virtual Environments* 10, 6 (2001), 583–598. <https://doi.org/10.1162/105474601753272844>
- [6] Steve Benford, John Bowers, Lennart E. Fahlén, Chris Greenhalgh, and Dave Snowden. 1995. User embodiment in collaborative virtual environments. In *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '95*. 242–249. <https://doi.org/10.1145/223904.223935>
- [7] Mark Billinghurst, Pablo Cesar, Mar Gonzalez-Franco, Katherine Isbister, Julie Williamson, and Alexandra Kitson. 2024. Social XR: The Future of Communication and Collaboration (Dagstuhl Seminar 23482). *Dagstuhl Reports* 13, 11 (2024), 167–196. <https://doi.org/10.4230/DagRep.13.11.167>
- [8] Mark Billinghurst, Maxime Cordeil, Anastasia Bezerianos, and Todd Margolis. 2018. Collaborative immersive analytics. *Immersive Analytics* (2018), 221–257.
- [9] Frank Biocca and Jaron Lanier. 1992. An insider's view of the future of virtual reality. *Journal of communication* 42, 2 (1992), 156–157.
- [10] Dan Bohus, Sean Andrist, Ashley Feniello, Nick Saw, Mihai Jalobeanu, Patrick Sweeney, Anne Loomis Thompson, and Eric Horvitz. 2021. <https://doi.org/10.48550/arXiv.2103.15975> Platform for Situated Intelligence..
- [11] John Bowers, James Pycock, and Jon O'Brien. 1996. Talk and embodiment in collaborative virtual environments. In *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '96*. 58–65. <https://doi.org/10.1145/238386.238404>
- [12] Qijia Chen, Andrea Bellucci, and Giulio Jacucci. 2024. "I'd rather drink in VRChat": Understanding Drinking in Social Virtual Reality. *Conference on Human Factors in Computing Systems - Proceedings* (2024). <https://doi.org/10.1145/3613904.3642405>
- [13] W S Condon and W D Ogston. 1967. *A SEGMENTATION OF BEHAVIOR**. Technical Report. 221–235 pages.
- [14] Jakob Engel, Kiran Somasundaram, Michael Goesele, Albert Sun, Alexander Gamino, Andrew Turner, Arjang Talattof, Arnie Yuan, Bilal Souti, Brigid Meredith, Cheng Peng, Chris Sweeney, Cole Wilson, Dan Barnes, Daniel DeTone, David Caruso, Derek Valleroy, Dinesh Ginjupalli, Duncan Frost, Edward Miller, Elias Mueggler, Evgeniy Oleinik, Fan Zhang, Guruprasad Somasundaram, Gustavo Solaira, Harry Lanaras, Henry Howard-Jenkins, Huixuan Tang, Hyo Jin Kim, Jaime Rivera, Ji Luo, Jing Dong, Julian Straub, Kevin Bailey, Kevin Eckenhoff, Lingni Ma, Luis Pesqueira, Mark Schwesinger, Maurizio Monge, Nan Yang, Nick Charron, Nikhil Raina, Omkar Parkhi, Peter Borschowa, Pierre Moulon, Prince Gupta, Raul Mur-Artal, Robbie Pennington, Sachin Kulkarni, Sagar Miglani, Santosh Gondi, Saransh Solanki, Sean Diener, Shangyi Cheng, Simon Green, Steve Saarinen, Suvam Patra, Tassos Mourikis, Thomas Whelan, Tripti Singh, Vasileios Balntas, Vijay Baiyya, Wilson Dreeves, Xiaqing Pan, Yang Lou, Yipu Zhao, Yusuf Mansour, Yuyang Zou, Zhaoyang Lv, Zijian Wang, Mingfei Yan, Carl Ren, Renzo De Nardi, and Richard Newcombe. 2023. Project Aria: A New Tool for Egocentric Multi-Modal AI Research. (8 2023). <http://arxiv.org/abs/2308.13561>
- [15] Marc Fabri and David Moore. 2005. The use of emotionally expressive avatars in collaborative virtual environments. *Virtual Social Agents* 88, April (2005), 88–94.
- [16] Qinna Feng, Heng Luo, Zijian Li, Jiarong Liang, Gege Li, and Yan Yi. 2024. Creating an Immersive Virtual Reality Game Space for Multiuser, Synchronous Co-Located Collaboration: Design Considerations and Influencing Factors. 14, 5 (2024), 2167. <https://doi.org/10.3390/app14052167>
- [17] Doron Friedman, Anthony Steed, and Mel Slater. 2007. Spatial social behavior in second life. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 4722 LNCS (2007), 252–263. https://doi.org/10.1007/978-3-540-74997-4_23
- [18] Harold Garfinkel. 2016. Studies of the Routine Grounds of Everyday Activities Author (s): Harold Garfinkel Published by: Oxford University Press on behalf of the Society for the Study of Social Problems Stable URL: <http://www.jstor.org/stable/798722> REFERENCES Linked referen. 11, 3 (2016), 225–250.
- [19] Erving Goffman. 1959. *The Presentation of Self in Everyday Life*. 259 pages. <https://doi.org/10.2307/2089106>
- [20] Erving Goffman. 1963. *Behavior in Public Places: Notes on the Social Organization of Gatherings*. 248 pages. <https://doi.org/papers3://publication/uuid/4EEC120E-5776-413F-B43F-044C09251F5F>
- [21] A. Guye-Vuilleme, T. K. Capin, I. S. Pandzic, N. Magnenat Thalmann, and D. Thalmann. 1999. Nonverbal communication interface for collaborative virtual environments. *Virtual Reality* 4, 1 (1999), 49–59. <https://doi.org/10.1007/BF01434994>
- [22] Edward Hall. 1969. *The Hidden Dimension : man's use of space in public and in private*. 217 pages.
- [23] Eugy Han, Cyan DeVeaux, Jeffrey T. Hancock, Nilam Ram, Gabriella M. Harari, and Jeremy N. Bailenson. 2024. The influence of spatial dimensions of virtual environments on attitudes and nonverbal behaviors during social interactions. *Journal of Environmental Psychology* 95 (5 2024). <https://doi.org/10.1016/j.jenvp.2024.102269>
- [24] Eugy Han, Mark R Miller, Cyan DeVeaux, Hanseul Jun, Kristine L Nowak, Jeffrey T Hancock, Nilam Ram, and Jeremy N Bailenson. 2023. People, places, and time: a large-scale, longitudinal study of transformed avatars and environmental context in group interaction in the metaverse. *Journal of Computer-Mediated Communication* 28, 2 (01 2023), zmac031. <https://doi.org/10.1093/jcmc/zmac031> arXiv:<https://academic.oup.com/jcmc/article-pdf/28/2/zmac031/48520441/zmac031.pdf>
- [25] Heiko Hecht, Robin Welsch, Jana Viehoff, and Matthew R. Longo. 2019. The shape of personal space. *Acta Psychologica* 193, April 2018 (2019), 113–122. <https://doi.org/10.1016/j.actpsy.2018.12.009>
- [26] Dineshabu Jayagopi, Dairazalia Sanchez-Cortes, Kazuhiro Otsuka, Junji Yamato, and Daniel Gatica-Perez. 2012. Linking speaking and looking behavior patterns with group composition, perception, and performance. In *Proceedings of the 14th ACM international conference on Multimodal interaction*. ACM, Santa Monica California USA, 433–440. <https://doi.org/10.1145/2388676.2388772>
- [27] Robert Johansen. 1988. *GroupWare: Computer Support for Business Teams*. The Free Press.
- [28] Uwe Gruenefeld Jonathan Liebers, Sascha Brockel and Stefan Schneegass. 2024. Identifying Users by Their Hand Tracking Data in Augmented and Virtual Reality. *International Journal of Human-Computer Interaction* 40, 2 (2024), 409–424. <https://doi.org/10.1080/10447318.2022.2120845> arXiv:<https://doi.org/10.1080/10447318.2022.2120845>
- [29] Adam Kendon. 1970. Movement coordination in social interaction: Some examples described. *Acta Psychologica* 32 (1970), 101–125. [https://doi.org/10.1016/0001-6918\(70\)90094-6](https://doi.org/10.1016/0001-6918(70)90094-6)
- [30] Adam Kendon. 1990. *Conducting interaction. Patterns of behaviour in focused encounters*. 308 pages. <http://books.google.com/books?id=7-8zAAAIAAJ>
- [31] Ikhwan Kim and Junghan Sung. 2024. New proxemics in new space: proxemics in VR. *Virtual Reality* 28, 2 (2024), 1–15. <https://doi.org/10.1007/s10055-024-00982-5>
- [32] Taemie Kim, Agnes Chang, Lindsey Holland, and Alex Sandy Pentland. 2008. Meeting mediator: enhancing group collaboration using sociometric feedback. In *Proceedings of the 2008 ACM conference on Computer supported cooperative work*. ACM, San Diego CA USA, 457–466. <https://doi.org/10.1145/1460563.1460636>
- [33] Simon Kimmel, Frederike Jung, Andrii Matvienko, Wilko Heuten, and Susanne Boll. 2023. Let's Face It: Influence of Facial Expressions on Social Presence in Collaborative Virtual Reality. In *Conference on Human Factors in Computing Systems - Proceedings*. Association for Computing Machinery. <https://doi.org/10.1145/3544548.3580707>

- [34] Felix Krell and Nico Wettmann. 2023. Corporeal Interactions in VRChat: Situational Intensity and Body Synchronization. *Symbolic Interaction* 46, 2 (5 2023), 159–181. <https://doi.org/10.1002/symb.629>
- [35] Jialang Victor Li, Max Kreminski, Sean M Fernandes, Anya Osborne, Joshua McVeigh-Schultz, and Katherine Isbister. 2022. Conversation Balance: A Shared VR Visualization to Support Turn-taking in Meetings. In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) (CHI EA '22). Association for Computing Machinery, New York, NY, USA, Article 181, 4 pages. <https://doi.org/10.1145/3491101.3519879>
- [36] Hai-Ning Liang, Feiyu Lu, Yuwei Shi, Vijayakumar Nanjappan, and Konstantinos Papangelis. 2019. Evaluating the effects of collaboration and competition in navigation tasks and spatial knowledge acquisition within virtual reality environments. *Future Generation Computer Systems* 95 (2019), 855–866.
- [37] Joan Llobera, Bernhard Spanlang, Giulio Ruffini, and Mel Slater. 2010. Proxemics with multiple dynamic characters in an immersive virtual environment. *ACM Transactions on Applied Perception* 8, 1 (2010). <https://doi.org/10.1145/1857893.1857896>
- [38] Joshua McVeigh-Schultz, Elena Márquez Segura, Nick Merrill, and Katherine Isbister. 2018. What's It Mean to "Be Social" in VR? *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility - DIS '18 June* (2018), 289–294. <https://doi.org/10.1145/3197391.3205451>
- [39] Paul Milgram and Fumio Kishimo. 1994. A taxonomy of mixed reality. *IEICE Transactions on Information and Systems* 77, 12 (1994), 1321–1329.
- [40] Mark Roman Miller, Cyan Deveaux, Eugy Han, Nilam Ram, and Jeremy N. Bailenson. 2023. A Large-Scale Study of Proxemics and Gaze in Groups. In *Proceedings - 2023 IEEE Conference Virtual Reality and 3D User Interfaces, VR 2023*. Institute of Electrical and Electronics Engineers Inc., 409–417. <https://doi.org/10.1109/VR55154.2023.00056>
- [41] Mark Roman Miller, Neeraj Sonalkar, Ade Mabogunje, Larry Leifer, and Jeremy Bailenson. 2021. Synchrony within Triads using Virtual Reality. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (10 2021). <https://doi.org/10.1145/3479544>
- [42] Robert Moore, E. Hankinson Gathman, and Nicolas Ducheneaut. 2009. From 3D space to third place: The social life of small virtual spaces. *Human Organization* 68, 2 (2009), 230–240. <https://doi.org/10.17730/humo.68.2.q673k16185u68v15>
- [43] B Nardi and J Harris. 2010. Strangers and friends: Collaborative play in World of Warcraft. In *International Handbook of Internet Research*. 395–410. <https://doi.org/10.1007/978-1-4020-9789-8>
- [44] Catherine S Oh, Jeremy N Bailenson, and Gregory F Welch. 2018. A systematic review of social presence: Definition, antecedents, and implications. *Frontiers in Robotics and AI* 5 (2018), 409295.
- [45] Iana Podkosova and Hannes Kaufmann. 2018. Co-presence and proxemics in shared walkable virtual environments with mixed colocation. In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology, VRST*. <https://doi.org/10.1145/3281505.3281523>
- [46] Christian Rack, Tamara Fernando, Murat Yalcin, Andreas Hotho, and Marc Erich Latoschik. 2023. Who is Alyx? A new behavioral biometric dataset for user identification in XR. *Frontiers in Virtual Reality* 4 (2023). <https://doi.org/10.3389/frvir.2023.1272234>
- [47] Christian Rack, Vivek Nair, Lukas Schach, Felix Foschum, Marcel Roth, and Marc Erich Latoschik. 2024. Navigating the Kinematic Maze: Analyzing, Standardizing and Unifying XR Motion Datasets. In *Proceedings - 2024 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops, VRW 2024*. Institute of Electrical and Electronics Engineers Inc., 507–514. <https://doi.org/10.1109/VRW62533.2024.00098>
- [48] Felicia Roberts, Alexander L. Francis, and Melanie Morgan. 2006. The interaction of inter-turn silence with prosodic cues in listener perceptions of "trouble" in conversation. *Speech Communication* 48, 9 (2006), 1079–1093. <https://doi.org/10.1016/j.specom.2006.02.001>
- [49] Bill Rogers, Masood Masoodian, and Mark Apperley. 2018. A virtual cocktail party: Supporting informal social interactions in a virtual conference. *Proceedings of the Workshop on Advanced Visual Interfaces AVI* (2018). <https://doi.org/10.1145/3206505.3206569>
- [50] Thomas Röggla, David A. Shamma, Julie R. Williamson, Irene Viola, Silvia Rossi, and Pablo Cesar. 2024. A Platform for Collecting User Behaviour Data during Social VR Experiments Using Mozilla Hubs. In *MMVE 2024 - Proceedings of the 2024 16th International Workshop on Immersive Mixed and Virtual Environment Systems*. Association for Computing Machinery, Inc, 41–44. <https://doi.org/10.1145/3652212.3652225>
- [51] Daniel Roth, Gary Bente, Peter Kullmann, David Mal, Chris Felix Purps, Kai Vogeley, and Marc Erich Latoschik. 2019. Technologies for social augmentations in user-embodied virtual reality. In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology, VRST*. Association for Computing Machinery. <https://doi.org/10.1145/3359996.3364269>
- [52] Harvey Sacks, Emanuel A Schegloff, and Gail Jefferson. 1974. A simplest systematics for the organization of turn-taking for conversation. *language* 50, 4 (1974), 696–735.
- [53] Anselm Strauss and Juliet Corbin. 1990. Basics of Qualitative Research. , 3–14 pages. <https://doi.org/10.4135/9781452230153>
- [54] David Thery and Brian F. G. Katz. 2021. Auditory perception stability evaluation comparing binaural and loudspeaker Ambisonic presentations of dynamic virtual concert auralizations. *The Journal of the Acoustical Society of America* 149, 1 (2021), 246–258. <https://doi.org/10.1121/10.0002942>
- [55] Michael Tomasello and Malinda Carpenter. 2007. Shared intentionality. *Developmental Science* 10, 1 (2007), 121–125.
- [56] Laurie M. Wilcox, Samuel Elfassy, Cynthia Grelik, and Robert S. Allison. 2006. Personal Space in Virtual Reality. *ACM Transactions on Applied Perception* 3, 4 (2006), 412–428. <https://doi.org/10.1145/1190036.1190041>
- [57] Julie Williamson, Jie Li, Vinoba Vinayagamoorthy, David A. Shamma, and Pablo Cesar. 2021. Proxemics and Social Interactions in an Instrumented Virtual Reality Workshop. In *ACM Conference on Human Factors in Computing Systems (CHI)*. 1–20. <https://doi.org/10.1145/3411764.3445729>
- [58] Julie R Williamson, Joseph O Hagan, John Williamson, John Alexis, Guerra Gomez, Pablo Cesar, and David A Shamma. 2022. Digital Proxemics : Designing Social and Collaborative Interaction in Virtual Environments. *CHI '22: ACM Symposium on Computer Human Interaction, May 01â€¦05, 2022, New Orleans, LA, USA* 1, 1 (2022).
- [59] Andrea Stevenson Won, Jeremy N. Bailenson, Suzanne C. Stathatos, and Wenqing Dai. 2014. Automatically Detected Nonverbal Behavior Predicts Creativity in Collaborating Dyads. *Journal of Nonverbal Behavior* 38, 3 (2014), 389–408. <https://doi.org/10.1007/s10919-014-0186-0>
- [60] Noelle L. Wood and Nelson Cowan. 1995. The Cocktail Party Phenomenon Revisited: Attention and Memory in the Classic Selective Listening Procedure of Cherry (1953). *Journal of Experimental Psychology: General* 124, 3 (1995), 243–262. <https://doi.org/10.1037/0096-3445.124.3.243>

A Data Tables

A.1 Figure 4 Data Tables

The histograms in Figure 4 each include two data series organised into 100 bins. To summarise these histograms in a data table, these are simplified into 10 bins.

Table 3: Data table for Figure 4, Part A, Charades Task

Bin Number	Lower Edge	Upper Edge	Face to Face Probability Distribution	XR Mediated Probability Distribution
1	0.0	0.7	0.04	0.03
2	0.7	1.4	0.25	0.29
3	1.4	2.1	0.24	0.27
4	2.1	2.8	0.34	0.3
5	2.8	3.5	0.34	0.3
6	3.5	4.2	0.17	0.19
7	4.2	4.9	0.05	0.01
8	4.9	5.6	0.0	0.02
9	5.6	6.3	0.0	0.02
10	6.3	7.0	0.0	0.0

A.2 Figure 5

The boxplot in Figure 5 can be summarised using the medians, upper quartiles, and lower quartiles.

B Open Analysis Techniques and Tools

The full GitHub of data processing and analysis techniques will be included here as an artefact underpinning this paper.

All of the code, data, and analysis scripts will be released on a CC-BY-SA License.

Table 4: Data table for Figure 4, Part B, Milling Line Up Task

Bin Number	Lower Edge	Upper Edge	Face to Face Probability Distribution	XR Mediated Probability Distribution
1	0.0	0.7	0.06	0.02
2	0.7	1.4	0.3	0.36
3	1.4	2.1	0.31	0.37
4	2.1	2.8	0.38	0.34
5	2.8	3.5	0.28	0.22
6	3.5	4.2	0.07	0.07
7	4.2	4.9	0.02	0.01
8	4.9	5.6	0.0	0.02
9	5.6	6.3	0.0	0.0
10	6.3	7.0	0.0	0.0

Table 5: Data table for Figure 4, Part C, Introduction in Pairs Task

Bin Number	Lower Edge	Upper Edge	Face to Face Probability Distribution	XR Mediated Probability Distribution
1	0.0	0.7	0.15	0.14
2	0.7	1.4	0.25	0.2
3	1.4	2.1	0.3	0.32
4	2.1	2.8	0.35	0.29
5	2.8	3.5	0.3	0.26
6	3.5	4.2	0.06	0.14
7	4.2	4.9	0.01	0.04
8	4.9	5.6	0.0	0.01
9	5.6	6.3	0.0	0.01
10	6.3	7.0	0.0	0.0

Table 6: Data table for Figure 4, Part D, Two Truths and a Lie Task

Bin Number	Lower Edge	Upper Edge	Face to Face Probability Distribution	XR Mediated Probability Distribution
1	0.0	0.7	0.03	0.01
2	0.7	1.4	0.33	0.34
3	1.4	2.1	0.29	0.34
4	2.1	2.8	0.46	0.31
5	2.8	3.5	0.23	0.28
6	3.5	4.2	0.07	0.09
7	4.2	4.9	0.02	0.04
8	4.9	5.6	0.0	0.01
9	5.6	6.3	0.0	0.0
10	6.3	7.0	0.0	0.0

C Interview Protocol

C.1 Interview Protocol

The following protocol was used during the focus groups. Participants were asked questions about their experiences in a virtual

Table 7: Data Table for Figure 5.

Condition	Median	Upper Quartile	Lower Quartile
Partner Conversation Face to Face	0.01	0.03	-0.04
Partner Conversation XR Mediated	-0.03	0.03	-0.07
Milling Face to Face	0.03	0.14	-0.05
Milling XR Mediated	0.01	0.09	-0.07
Two Truths Face to Face	0.06	0.11	0.0
Two Truths XR Mediated	0.03	0.06	-0.03

environment and how these differed from face-to-face social interactions.

General Experience with Virtual Environment:

- What were your first impressions of trying to do social activities in a virtual environment? Have you done anything like this in a virtual environment before?
- How did it feel to move around the space?
- Were you comfortable or uncomfortable?
- Did you notice the faces were animated? How did this change the experience for you?
- Did you notice the hands were animated? How did this change the experience for you?

Comparison with Face-to-Face Activities:

- How would you compare this to face-to-face activities?
- Which felt easier?
- How were these different?

Specific Activity Comparisons:

Charades Activity:

- Think back to the “Charades” activity we completed.
- How would you compare this in the virtual environment and face to face?
- Which environment felt easier/harder?
- Why?
- What parts of this activity felt socially comfortable/uncomfortable?
- Did this change from environmental conditions?
- Did you have to adapt your behaviour depending on the environment?
- Was there anything surprising about completing this activity today?

Milling Task Activity:

- Think back to the “Milling Task.”
- How would you compare this in the virtual environment and face to face?
- Which environment felt easier/harder?
- Why?

- What parts of this activity felt socially comfortable/uncomfortable?
- Did this change from environmental conditions?
- Did you have to adapt your behaviour depending on the environment?
- Was there anything surprising about completing this activity today?

2 Truths and a Lie Activity:

- Think back to the “2 truths and a lie” where one person spoke and the others had to guess which statement was false.
- How would you compare this in the virtual environment and face to face?
- Which environment felt easier/harder?
- Why?
- What parts of this activity felt socially comfortable/uncomfortable?
- Did this change from environmental conditions?

- Did you have to adapt your behaviour depending on the environment?
- Was there anything surprising about completing this activity today?

Final Thoughts:

- Any other comments about your experience today?

D Open Dataset

D.1 Dataframe Schema

We define a dataframe scheme and standard conventions for XR logging data to support data sharing between experiments and easier secondary analysis of open data sets. These columns represent a minimum set of data points from which many additional social metrics and analyses can be completed.

Column Name	Description	Convention
frame_id	date time	milliseconds
participant_id	A unique identifier	String
position_x	World coordinate, x	X and Y as horizontal plane, Z as vertical plane
position_y	World coordinate, y	
position_z	World coordinate, z	
rotation_x	Rotation Quaternion, x	
rotation_y	Rotation Quaternion, y	
rotation_z	Rotation Quaternion, z	
rotation_w	Rotation Quaternion, w	
device	Descriptor of the device used for this log	String, additional information provided with dataset if needed
trial_id	A unique identifier	String
activity	Description of activity during this log	String, may be qualitative or part of defined tags
forward_x	Vector, x	X and Y as horizontal plane, Z as vertical plane. Directional vectors are all inferred from rotation quaternion and are optional in the dataframe. They are included here for ease during analysis and visualisation.
forward_y	Vector, y	
forward_z	Vector, z	
backward_x	Vector, x	X and Y as horizontal plane, Z as vertical plane
backward_y	Vector, y	
backward_z	Vector, z	
left_x	Vector, x	X and Y as horizontal plane, Z as vertical plane
left_y	Vector, y	
left_z	Vector, z	
right_x	Vector, x	X and Y as horizontal plane, Z as vertical plane
right_y	Vector, y	
right_z	Vector, z	
up_x	Vector, x	X and Y as horizontal plane, Z as vertical plane
up_y	Vector, y	
up_z	Vector, z	
down_x	Vector, x	X and Y as horizontal plane, Z as vertical plane
down_y	Vector, y	
down_z	Vector, z	
gaze_position_x	World coordinate, x	X and Y as horizontal plane, Z as vertical plane. Gaze tracking systems may store this data in different formats (pitch, yaw, depth, 3D vector, etc). Calculating Gaze (x, y, z) in world coordinate mitigates differences in tracking conventions.
gaze_position_y	World coordinate, y	milliseconds
gaze_position_z	World coordinate, z	
left_hand_position_x	World coordinate, x	X and Y as horizontal plane, Z as vertical plane. Hand tracking systems may store this data in many formats (separated joints, hand mesh, etc). Minimum viable features are hand position and vector, although additional columns should be added where available. This dataset used the wrist as the position point.
left_hand_position_y	World coordinate, y	
left_hand_position_z	World coordinate, z	
right_hand_position_x	World coordinate, x	X and Y as horizontal plane, Z as vertical plane
right_hand_position_y	World coordinate, y	
right_hand_position_z	World coordinate, z	
left_hand_direction_x	Vector, x	A vector relative to the hand wrist, where z faces out from the inner wrist as vertical plane.
left_hand_direction_y	Vector, y	
left_hand_direction_z	Vector, z	
right_hand_direction_x	Vector, x	A vector relative to the hand wrist, where z faces out from the inner wrist as vertical plane.
right_hand_direction_y	Vector, y	
right_hand_direction_z	Vector, z	
is_speaking	Utterance heard during this frame	Boolean