Computers in Human Behavior 181 (2026) 108975

Contents lists available at ScienceDirect

Computers in Human Behavior

journal homepage: www.elsevier.com/locate/comphumbeh

Not seeing eye to eye: The effects of perceptual conflicts during social interactions in mixed reality

Eugy Han [a,*], Portia Wang [b], Monique Santoso [b], Keshav Rastogi [c], Jeremy N. Bailenson [b]

[a] *Department of Media Production, Management, and Technology, University of Florida, FL, United States*
[b] *Department of Communication, Stanford, CA, United States*
[c] *University of California, Berkeley, CA, United States*

ARTICLE INFO

Handling Editor: Prof. Nicolae Nistor

Keywords:
Mixed reality
Social interactions
Perceptual conflict
Expectancy violation

ABSTRACT

Not all cues in immersive virtual environments are consistent with physical-world expectations, but can be transformed, resulting in different perceived versions of the world for simultaneous users. Such perceptual conflicts may disrupt communication. In this study, dyads ($n = 104$ participants, 52 dyads) sat face-to-face while wearing mixed reality headsets and memorized a set of nine virtual objects varying in shape and color. In the first trial, both members saw identical arrangements. In the second trial, they saw a new set of virtual objects. This time, either one or two pairs of virtual objects swapped positions, so each member saw a different arrangement. After each trial, members discussed their observations and agreed on a final recalled set. Results showed that when the discrepancy was greater, participants demonstrated less nonverbal synchrony with their dyad partner, made more changes from individual to collaborative recall, and reported feeling greater cognitive load. While there was no effect of the manipulation on participants' confidence in self or partner's recall, when the discrepancy was greater, participants felt less confident in their collaborative recall and less willing to submit their collaborative response. The manipulation did not affect trust in partners or the technology. This study offers an initial examination of perceptual conflict, with implications for how technologies designed to foster collaboration and intergroup relations may inadvertently undermine these goals when users' perceptual experiences diverge and reaching a shared understanding is disrupted.

1. Introduction

Imagine you and your coworkers are wearing mixed reality (MR) headsets to visualize information during a brainstorming session. Throughout the session, your coworker references a green bar and gestures towards it, but in that spot you only see a blue bar. At first, you assume they just pointed in the wrong place. However, as they keep talking, you lose your focus and end up leaving the session confused. Later, you remember your settings were customized to replace green with blue for your colorblindness. You realize both of you were referring to different parts of the graph the entire time.

This scenario, while strange, illustrates how communication partners can experience discrepant mediated realities. Such discrepancies likely violate people's expectations about the ways others will communicate – expectations that make joint construction of reality and action possible. This is the central argument of the Expectancy Violations Theory (EVT), which posits that when communication patterns deviate from expectations, they lead to violations that distract the recipient, causing an attentional shift to the communicator, relationship, and the characteristics and meaning of the violation. These shifts have downstream effects, influencing the communication process at multiple levels, from nonverbal involvement to perceptions of comprehension and credibility (Burgoon, 1993).

These violations often hinge on perception, particularly visual cues such as gaze, conversational distance, immediacy, and nonverbal involvement. Mixed reality is especially suited to studying such violations, as it transforms physical cues with virtual ones, creating a seamless blend of the two, ultimately creating scenarios where it is possible for discrepancies in external referents to lead to violations in expectations of nonverbal and verbal communication. Unlike other similar media like virtual reality (VR), where everything is completely virtual and the physical world is invisible, MR allows people to interact with

This article is part of a special issue entitled: ITIR published in Computers in Human Behavior.

* Corresponding author. University of Florida, Box 118400 1885, Stadium Road, Gainesville, FL, 32611-7001, United States.
E-mail address: eugyoung.han@ufl.edu (E. Han).

https://doi.org/10.1016/j.chb.2026.108975
Received 9 October 2025; Received in revised form 28 February 2026; Accepted 5 March 2026
Available online 6 March 2026
0747-5632/© 2026 Published by Elsevier Ltd.

virtual content while remaining aware of their physical surroundings. In MR interactions, there are cues expected to be computer-generated (i.e., virtual cues) and cues that remain constant and shared for all interactants (i.e., physical cues). For each user, their MR-mediated reality can be either unintentionally transformed through technological artefacts (e.g., objects lagging) or intentionally transformed by the programmer (e.g., different content displayed for each user). Because such artefacts render different realities for each user, these discrepancies may violate expectations when multiple users come together. Despite this potential, the spatial and environmental dimensions of immersive contact have rarely been theoretically examined – a notable gap in relevant research, including intergroup interventions, that MR is uniquely positioned to address (Tassinari et al., 2026).

From an application-standpoint, such violations are plausible, as virtual experiences are becoming increasingly personalized for each individual user. Research on social media has shown that no two users have the same experience on a platform (Kalyanaraman & Sundar, 2006). Similarly, virtual environments can be adapted for each user in real-time, based on information such as object metadata and user feedback (e.g., Valmorisco et al., 2024). However, such discrepant realities may lead individuals to inhabit curated versions of reality, and this fragmentation poses a threat to something fundamental to communication itself: common ground, or the mutual knowledge and assumptions that interactants coordinate to reach shared understanding (Clark & Brennan, 1991). In realities accessed through technologies like MR, where elements of the physical and virtual overlap, discrepant realities can fragment and disrupt interactants from reaching a shared understanding. What occurs when two or more users with these altered perceptions come together, and how such disruptions shape the communication process and its outcomes, remains less clear.

This gap is consequential both for research and for the rapid deployment of immersive technologies in social contexts. From a research standpoint, interactions between members of social groups can be complicated by biases and differences in perceptions and attitudes, which may lead to exclusion and social fragmentation (Tassinari et al., 2026). Immersive technologies have increasingly been leveraged to instantiate the conditions thought to facilitate positive intergroup contact, such as shared goals and intergroup cooperation (Miler & Steinfeld, 2025). While various empirical strategies have been used to structure such interventions, they may only go so far when the common ground shared between interactants is made malleable by the medium itself. Scholars have called for closer attention to the moderators and boundary conditions that shape when and for whom such interventions work (Tassinari et al., 2022). This is particularly salient given the growing importance of hybrid spatial ecologies and constantly shifting boundaries between the physical and virtual. However, immersive technologies have rarely been used to explicitly model such spatial organizations in this context (Tassinari et al., 2026). To this end, perceptual discrepancy represents one such spatial boundary condition, such that it is not a social or attitudinal one, but a feature of the visible environment itself (i.e., an external referent), that has yet to be examined.

From an application standpoint, this gap is amplified by the rapid industry deployment of similar immersive technologies like Augmented Reality Artificial Intelligence (AR-AI) wearables (e.g., Meta Displays [Meta, 2025]; Snapchat Spectacles [Snapchat, 2024]), which fall under the broader category of computing systems that integrate directly into users environments rather than requiring users to adjust their behaviors to fit the systems (Zhao et al., 2023). Like MR, these devices merge physical surroundings with virtual elements, and have been envisioned to become tools that will shift how people communicate, work, and play (Meta Orion, 2025). Yet, while industries promote seamless integration of these technologies into work and social settings, how they may introduce further challenges remains unexplored.

This concern is particularly striking given that recent research has documented broader social consequences of MR devices, such as users wearing MR headsets experiencing "social absence" where other people physically present in the room feel less socially present (Santoso & Bailenson, 2024). Mixed reality therefore provides an ideal tool for examining how these perceptual conflicts arise in such social contexts, as it allows for experimental control over visual discrepancies while maintaining the in-person, group context. Understanding how expectancy violations manifest in MR settings can inform the design and deployment of similar technologies, helping anticipate communication breakdowns before these devices become ubiquitous in collaborative and intergroup settings.

2. Background and previous work

2.1. Expectancy violations in the virtual space

Central to communication are the expectations individuals hold for how others will behave (Burgoon, 1993). These expectations are grounded in shared understandings and rules that make joint construction of reality and coordinated action possible (Clark & Brennan, 1991). When behaviors – verbal or nonverbal – conform to expectations, the interaction can be smooth and predictable. These expectancies can also serve as perceptual filters, shaping how an individual processes social information and interprets another's actions. Consequently, when behaviors deviate, these violations can be disruptive. Depending on the direction and magnitude of the discrepancy, these violations can be perceived as positive, negative, or ambiguous, leading to either favorable or detrimental outcomes on message comprehension, persuasion, attraction, credibility, and reciprocal nonverbal involvement (Burgoon et al., 1995).

Within the context of technology, such violations, often resulting from asymmetries in content and media or lack of shared space, are considered a challenge that should be addressed (Gergle et al., 2012; Voida et al., 2008). One example is the "What You See is What I See" paradigm, which expects that interactants in a system share the same visual perception and access the same information (Stefik et al., 1987). In practice, however, there are several factors that can disrupt this expectation in technologically-mediated communication, such as differences in modality (i.e., device for input and output), mobility (i.e., ability or need to move while having access to the interaction device), and content access (i.e., users having different visibility and ability to act on content), creating potential violations of any shared expectations (Bréhault et al., 2025).

These factors can also arise unintentionally, such as through technological artefacts. To illustrate an example, Seuren and colleagues (2021) investigated how latency (i.e., transmission delays) during videoconferencing can lead to expectancy violations. Analyzing 25 video consultations between patients and clinicians, they observed noticeable silences and overlapping speech. These delays had procedural consequences, such that interactants treated these silences and overlaps as turn-taking problems, not latency problems. Latency often went unnoticed until it became problematic. Until then, participants proceeded to interact with a non-mutual reality, but acted as though they had a shared reality. Even when aware of such violations, participants struggled to recognize and adjust their meaning-making in real time.

Similarly, research in social psychology sheds light on the consequences of how individuals navigate violations of expectations. Mori and Arai (2010) replicated classic studies by Asch (1951), where participants matched the length of a line to three options. Using polarizing glasses, one participant (the minority) perceived a different stimulus than the other three (the majority). When asked to announce judgements aloud, women in the minority conformed more often, though men did not. Comparable effects have appeared in online spaces as well, in which the actions of others can transform users' original perceptions, turning otherwise individual experiences into social ones (Krämer et al., 2021). For instance, contradictory comments on social media have shown to undermine the persuasive impact of original messages, highlighting how dissenting perceptions strongly shape individual responses

(Winter et al., 2015). Beyond conformity, people are also highly sensitive to the social cues that accompany such disagreements, often relying on behavioral, verbal, and nonverbal signals to continuously monitor their social environment (Leary & Baumeister, 2000). When perceptual conflicts arise in immersive settings, cues that typically signal shared attention may break down, and behaviors that appear ignoring norms or attending to something else, may lead to exclusionary consequences and feelings of ostracism (Kerr & Levine, 2008).

Together, past research on social influence suggests that some individuals may change their responses when their judgements do not align with their social others, and that the social cues accompanying such discrepancies may shape the interaction in consequential, often negative, ways, though individual differences may play a role.

2.2. *Perceptual conflicts in virtual reality*

Immersive media provide an especially fertile environment for expectancy violations, as there are many opportunities for asymmetric experiences for each user in a shared setting. Zhang and Furnas (2002) note how asymmetries in collaborative VR may form as a result of being located in different places and seeing different things, embodying avatars of different sizes yielding different mobilities and different perspectives and scales of the same object – factors that ultimately interfere with participants' ability to communicate.

Past studies have investigated asymmetries of this nature, such as scenarios in which two people occupy the same virtual seat simultaneously while each believes the other person is sitting in a different location (Bailenson et al., 2008; Hasenbein et al., 2022); physically co-located participants collaborate while both are in the same or different mediums, which provide different abilities and information (VR and mobile AR, Drey et al., 2022; Grandi et al., 2019; Jeong et al., 2019; Numan & Steed, 2022); or different angles and perspectives (Hoppe et al., 2021; Zhou & Won, 2024).

Beyond VR, perceptual conflicts have also been examined in other immersive technologies, such as CAVE systems (i.e., projection-based displays). Chen and colleagues (2014) created a scenario where participants received instructions from both a physical experimenter and their avatar. By varying the angle between them and the instruction type, the researchers found participants took longer to execute the task when the angle discrepancy between the physical experimenter and their avatar was larger. Additionally, the physical presence of the experimenter had an influence on participants, such that a significant portion chose to follow the instructions based on the physical experimenter rather than their avatar.

This study by Chen and colleagues (2014) illustrates how, when both virtual and physical cues are presented in what would normally be considered a virtual-centered experience (e.g., CAVE, immersive VR, AR, MR), it can not only make the physical cues especially salient, but it can also cause confusion. Furthermore, their study suggests that participants may rely on or prioritize physical cues more heavily than virtual ones. While this study is one such example of how perceptual conflicts can emerge, there are other types of conflicts that may arise across immersive technologies. The same referent, for instance, may differ for each interactant in its characteristic (e.g., shape, color), semantic (e.g., category), numeric (e.g., quantity), dimensional (e.g., size), or locational (e.g., spatial orientation) properties. This list is not exhaustive, but illustrates the range of possible scenarios that may occur. We illustrate a few possibilities in Fig. 1.

This study focuses on one of these properties, namely the degree of perceptual discrepancy. Returning to the theoretical framework, Expectancy Violations Theory posits that the valence of a violation is a function of several factors, namely the direction (i.e., favorable or unfavorable) and magnitude of the discrepancy, the latter of which maps onto the degree of perceptual conflict examined here.

3. Current study

Mixed reality raises a novel setting in which interactants in the same physical space are simultaneously aware of both the physical and virtual world. In other words, unlike VR, where interactants are more aware of the virtual world than their physical world, in MR, the physical world is also salient. Sharing the same physical space allows interactants to readily see and hear what each other is doing and looking at (Clark & Brennan, 1991). However, the physical space may not be "fixed and readily accessible to co-present interlocutors," but is "uncertain and unfolding, [and] the relationship between presentation and acceptance may be muddled" (Koschmann & LeBaron, 2003, p. 10).

To investigate how these violations of expectations, through perceptual conflict, play out in a collaborative setting in MR, we investigate the following question: when perceptual conflicts arise in participants' virtual experience, violating expectations, how will they shape the communication process? Given prior literature suggests that expectancy violations can influence multiple at multiple levels, we investigate several types of outcomes using multimodal measurement approaches beyond self-report measures alone (Tassinari et al., 2026).

At the social behavioral level, perceptual conflicts may disrupt interpersonal processes that sustain coordination, such as nonverbal synchrony, or being in similar states or having similar behaviors at similar times (Condon & Ogston, 1966). The tendency to synchronize and adapt to conversational partners often emerges as an automatic, instinctive response, but it can also be adjusted to regulate interactions, signal interest, involvement, rapport, or to exert social influence (Burgoon et al., 2010). Given this, the present study operationalized social behavioral outcomes through dyadic nonverbal synchrony.

At the individual behavioral level, past research suggests that discrepant perceptions can lead participants to adjust judgements or decisions under social influence (Mori & Arai, 2010). Accordingly, we operationalized this outcome as response change, reflecting how much participants altered their initial responses.

At the individual attitudinal level, past research suggests that perceptual conflicts may lead to uncertainty, discomfort, and increased effort in coordinating with others (Burgoon et al., 1995; Seuren et al., 2021). Research in immersive social applications, such as social VR, shows that the quality of social presence shapes psychological outcomes (Barreda-Ángeles & Hartmann, 2022). In MR, disruptions to this social process may affect cognitive demands, confidence in perception and memory, and willingness to defend or defer to another. Accordingly, we measured self-reported cognitive load, willingness to submit, confidence, and trust to capture how perceptual conflicts shape individuals' experiences.

While prior technology-mediated communication research has examined similar phenomena, such as asymmetry in media (Voida et al., 2008), this study extends that foundation by empirically testing the degree of perceptual discrepancy, examining not whether conflicts matter, but how their magnitude shapes outcomes. This approach establishes a framework for future investigations that can systematically vary relevant factors, including but not limited to, task type, group size, intergroup dynamics, and discrepancy scenarios, to build a comprehensive understanding of the effects of perceptual conflicts brought about by immersive technologies such as MR.

The present study[1] examines the effects of asymmetric virtual experiences, or perceptual conflicts. Dyads met in a shared physical environment and completed two trials of a memory task displayed

[1] A version of this study is pre-registered at OMITTED. The present manuscript deviates from the original plan in two ways: first, we dropped an independent variable based on media literacy training due to lack of significant findings (see section 4.4 for more details), and second, we shifted from a repeated-measures ANOVA to a linear mixed-effects model to better fit the data structure (see section 5 for more details).

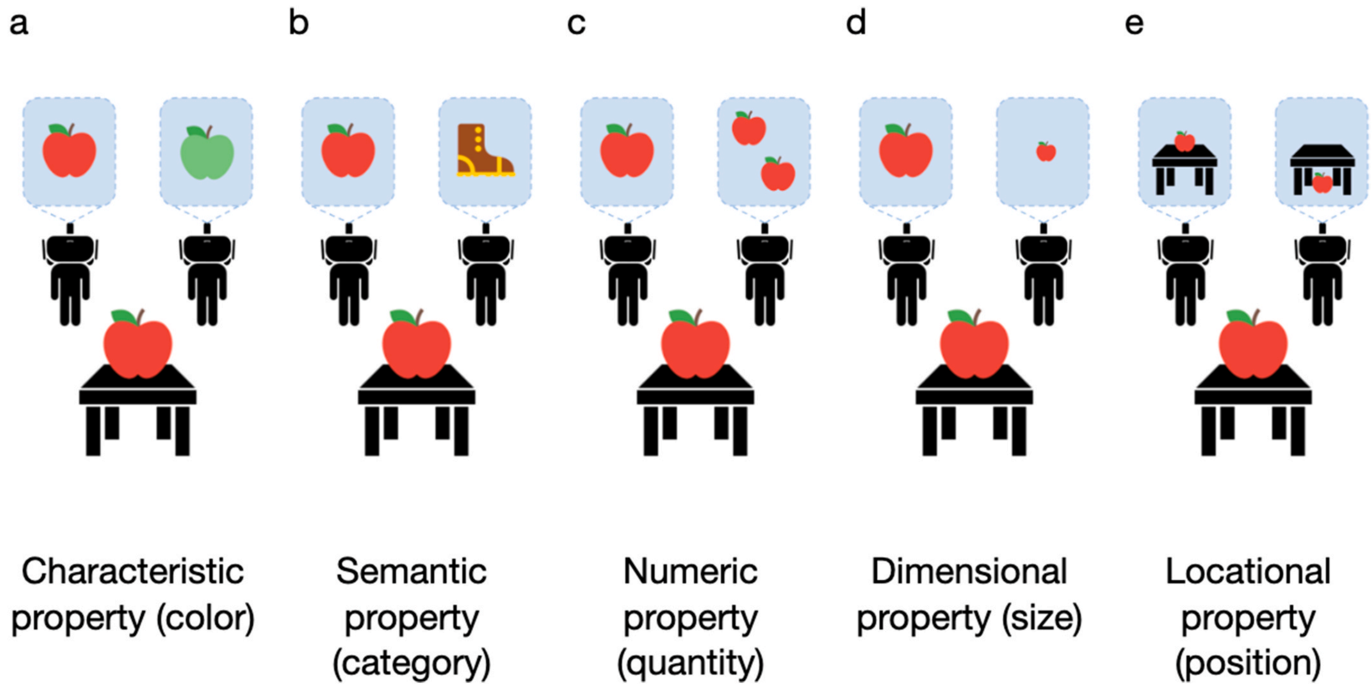

Fig. 1. A non-exhaustive list of scenarios of potential visual perceptual conflicts. These perceptual conflicts pertain to the virtual object's characteristic, semantic, numeric, dimensional, and locational properties.

through MR. They memorized nine virtual objects varying in shape and color, presented on a physical table. In the first trial, both members saw the same stimuli. In the second trial, the stimuli differed, such that either one pair (small number of discrepancies) or two pairs (large number of discrepancies) of objects swapped locations. Participants were first asked to list what they remembered individually, and then discuss with their partner on their answers and agree on a final list of seen virtual objects. Various aspects of the participants' attitudes and behaviors were measured.

4. Methods

4.1. Participants

Participants were recruited from a medium-sized private university via the department pool, email, flyers, word-of-mouth, and in-person recruitment and were paid $30. Based on a pre-registered power analysis, we aimed for 96 participants (48 dyads) and recruited 104 participants (52 dyads, *female-female* = 11, *female-male* = 30, *male-male* = 10, *female-non-binary/third-gender* = 1). Participants (*female* = 53, *male* = 50, *non-binary/third-gender* = 1) were between 18 and 64 years old (M = 25.48, SD = 7.58) and identified as Asian or Asian-American (n = 59), White (n = 21), Hispanic or Latinx (n = 8), bi- or multi-racial (n = 8), African, African-American, or Black (n = 3), Middle Eastern (n = 2), a racial group not listed (n = 1), or did not answer (n = 2). Participants had varying levels of experience with extended reality (XR) technology (27.2% no experience, 64% limited experience, and 8.7% a lot of experience) and XR programming, as well as prior familiarity with the other dyad member/partner. No participant had any form of color blindness.

4.2. Equipment and setup

Each participant wore an Oculus Quest 3 headset (standalone HMD; 2064 × 2208 pixels per eye, 515 g, 110° horizontal FOV, 96° vertical FOV, six-degree-of-freedom inside-out head and hand tracking). Headsets were connected to a PC via a Quest link cable. Three DSLR cameras were placed around the setup: one central camera capturing participants from the side and one facing each participant's face. A microphone at the center edge of the table recorded verbal conversation, and a tablet displayed a 5-min countdown during the discussion portion of the task (Fig. 2).

4.3. Virtual stimulus

The virtual stimuli were built using Unity (Fig. 3). Several versions varying in the number of colors and objects (e.g., singular color or shape, three colors and shapes) were pilot-tested using the same task outlined in the Procedure section on Qualtrics (n = 16). The final selection of colors and objects was selected for the task to be challenging enough to allow meaningful disagreements between participants, but not so difficult that the participant would defer to the other. All the virtual objects used throughout the study were primitive objects (sphere or cube) to prevent any individual familiarity with physical-world objects from interfering with the memorization processes. All the virtual objects were blue or red to make the colors distinguishable, in particular for those who may have any type of colorblindness. Nine virtual objects were laid out in a 3x3 grid. Their spatial orientation was randomized for all dyads and structured such that there were no more than five virtual objects of the same shape or color, and no more than three of the same color and shape combinations (e.g., red sphere), to prevent any clear patterns from emerging (e.g., three red spheres in a row). A countdown timer was displayed at the bottom to indicate how time participants had left during the memorization stage.

4.4. Independent variable

All dyads first underwent the no manipulation trial, in which both participants saw and reported on the same set of virtual objects. Dyads underwent the same base trial to familiarize themselves with the task procedures and XR environment, and to establish baseline measurements in attitudes and behaviors prior to introducing discrepancies. Dyads were then randomly assigned to either a *small number of discrepancies* or *large number of discrepancies* trial, where each member saw an

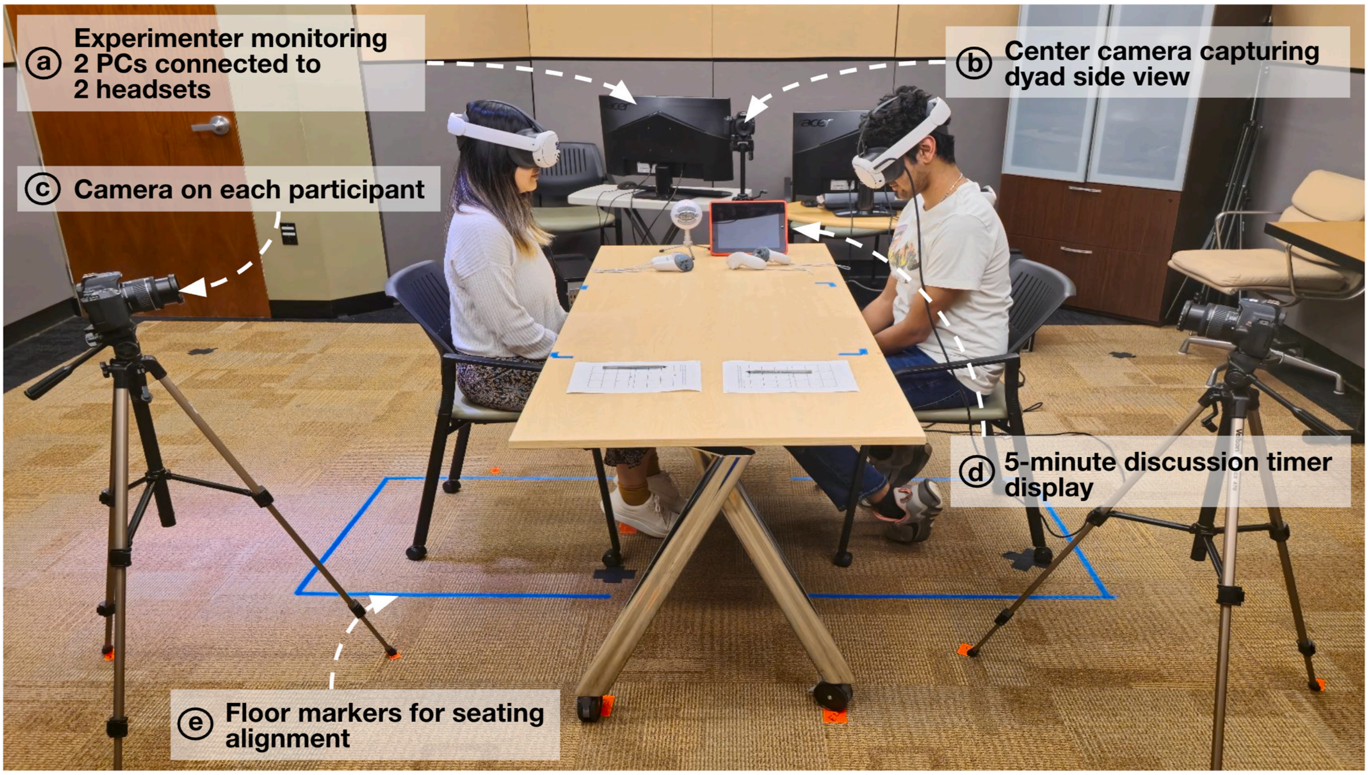

Fig. 2. The physical setup of the main experiment room, shown with research personnel standing in as users.

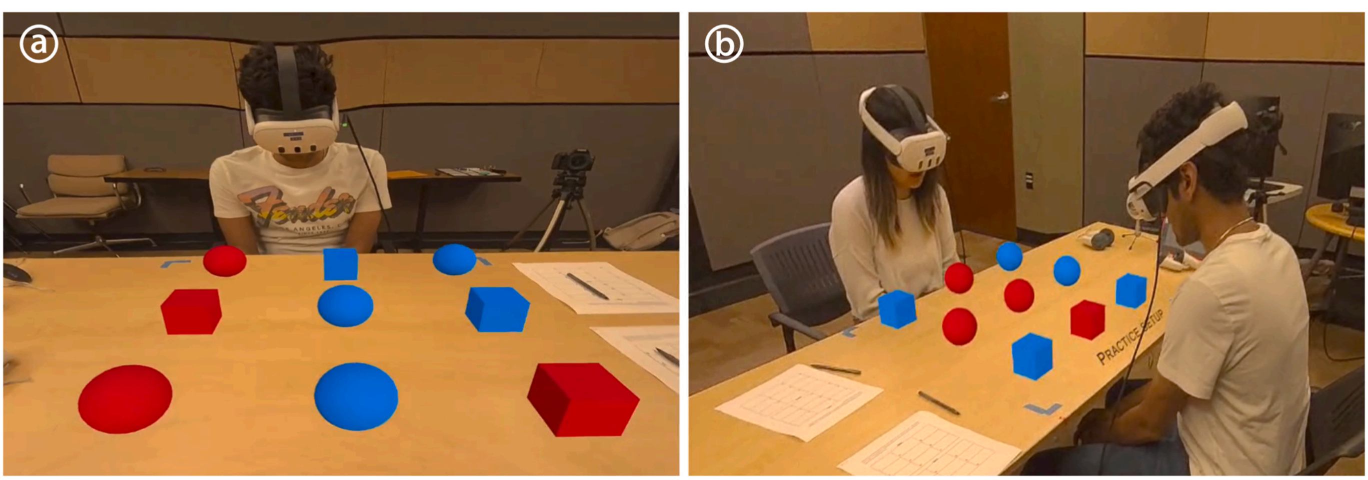

Fig. 3. Virtual objects on the table as seen through the headset (panel a). Objects were oriented from each user's perspective (e.g., the row closest to user A appears farthest from user B; panel b). These objects are for demonstration only and do not reflect actual trial conditions or discrepancies. Research personnel are standing in as users.

arrangement of virtual objects that differed from the first trial and from each other's view.[2] When there was a *small number of discrepancies*, one

[2] There was originally an additional independent variable in which each member of the dyad received different information about what could be expected from the virtual content through an instructional video about the task. One video detailed how virtual content could look ambiguous, shifted, or floating in incorrect positions as the technology was still new and developing, while the other video detailed how virtual content could look blurry, pixelated, or otherwise unclear as part of the technology. However, this independent variable had no effect on any of the outcome variables and was trimmed from the models and study. Details of the instructional video are consequently omitted from the manuscript for clarity and to prevent any confusion.

pair of objects was in different spatial locations for each member (Fig. 4a and b). When there was a *large number of discrepancies*, two pairs of objects were in different spatial locations for each member (Fig. 4c and d). The manipulated objects were objects that were directly left, right, up, down, or diagonal to one another.

4.5. Procedure

After signing consent forms, participants were led to separate rooms to complete a pre-test questionnaire and receive study instructions. They were then brought to the main experiment room. In the center of the room was a table with a chair and a Meta Quest 3 headset on each opposite end (Fig. 3). A corner table allowed research personnel to monitor and control the stimuli from two computers, which were not visible to participants.

After sitting at opposite ends of the table, participants were asked to put on the headsets and confirm they could see a virtual cube. If not, research personnel helped troubleshoot. They were told that the cube was not interactable. Once confirmed, instructions were delivered, and participants began the first no-manipulation trial.

The task was a memory task in which participants had 1 min to observe and memorize nine virtual objects on a table. Specifically, a memory task was chosen, as it relies on basic cognitive processes with minimal prior knowledge, providing a baseline sensitive to experimental manipulations without being confounded by skill differences (Gevins et al., 2011). In the first no-manipulation trial, all objects were identical for both participants from their seating perspective. During the minute, participants could not verbally communicate. When time was up, the objects disappeared, and participants removed their headsets.

A piece of paper with a 3x3 grid, similar to how the virtual objects had been oriented, was placed next to each participant. Each cell of the 3x3 grid was labelled with an alphabet letter (i.e., “A-I”) to help the participants orient the virtual objects from their perspective (e.g., for each dyad, one participant's “A” on the 3x3 grid matched the location of the “A” in the other participant's 3x3 grid). Participants silently recorded each object's shape and color. Afterwards, dyads had 5 min to verbally discuss and finalize their responses on a whiteboard. This memory-based task structure required collaboration, rather than cooperation, such that participants could not divide and assign subtasks but instead had to reconcile potentially discrepant observations and negotiate a joint decision through discussion. To increase participants' incentive to engage in the collaboration (Abdelazeem et al., 2022), they had the opportunity to earn additional compensation based on the accuracy of their final response on the whiteboard. If they got all nine of the virtual objects correct, they would receive $2.50 USD.

After the control trial, participants returned to separate rooms to complete a questionnaire. They were then brought back to the main room and put on the headsets for the second, manipulated trial. The procedure was the same as the first trial, but participants saw different object layouts with either one or two pairs of objects swapping locations (small or large number of discrepancies). Following the manipulated condition trial, participants returned to separate rooms to complete a final questionnaire. Afterwards, research personnel debriefed the participants.

4.6. Measures

Multiple aspects of the individuals' attitudes were measured at the start of the study (pre-test), after the no-manipulation trial (baseline), and after the manipulated condition trial (outcome variables). To reduce fatigue, repetitiveness, and burden, the questionnaire was purposely designed to be brief. Behavioral data was captured through the video recordings, verbal conversation through the audio recordings, and individual and final collaborative response through the 3x3 grid papers and whiteboards. Descriptive statistics of the overall totals in each trial are provided at the end of each measure's description, and condition-

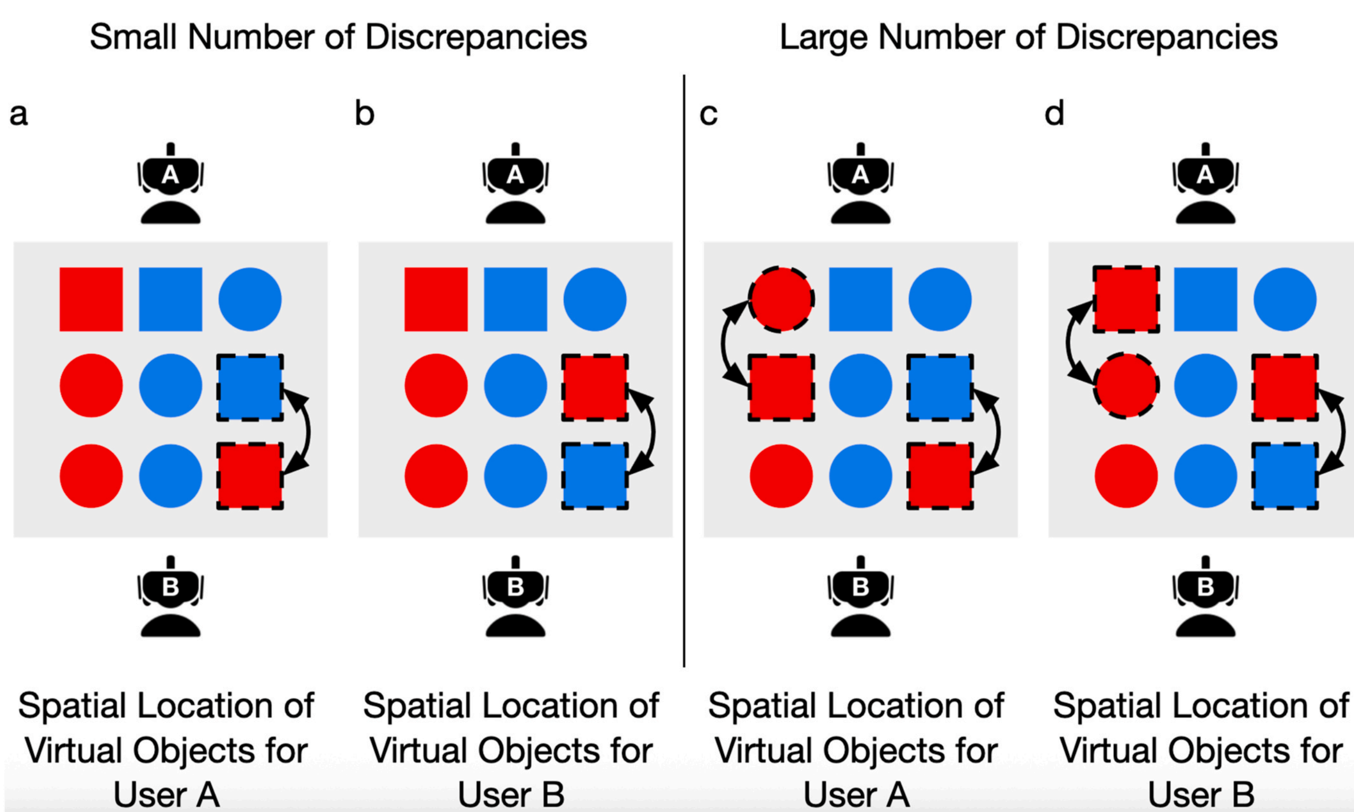

Fig. 4. Exemplar spatial location of the virtual objects per dyad member in the small (panels a, b) and large (panels c, d) number of discrepancies conditions. Arrows indicate the virtual objects that have been manipulated for each user.

specific descriptive statistics are presented in Table 1.

4.6.1. Synchrony

Dyadic synchrony was captured via a camera recording one frame per second from behind the table, showing both participants' faces. Each frame was extracted using OpenCV and facial positions were located using Py-Feat (Cheong et al., 2021) and Mediapipe (Lugaresi et al., 2019). In instances where the model failed to detect a face (e.g., covered by hair), the researchers manually selected an approximate facial location. To account for camera warp, two points on a plane similar to that of the participants, specifically the table between the dyad members, were used as a known reference distance. This known distance was then used to compute a ratio against the pixel distance. For each participant, a time series was generated to track the absolute change in head position across a trial (Han et al., 2023; Miller, 2023) (small number of discrepancies: $M = 273.60$, $SD = 57.40$ s; large number of discrepancies: $M = 300.83$, $SD = 38.29$ s).

Spearman's correlation was used to perform a cross-correlation between the time-series data collected every second with an offset of ± 5 s, including 0, following the procedure by Tsuchiya and colleagues (2020), Tschacher and colleagues (2014), and Ramseyer and Tschacher (2014). By extracting the average values from these 11 correlations, a single value was calculated to represent the synchrony of each trial. While this calculation is synchrony calculated based on a 2D representation, it is a representation of time and similarity of movements, and is similar to motion energy analysis, which has been shown to be correlated with 3D motion (Miller, 2023) ($M = 0.08$, $SD = 0.07$, $min = -0.09$, $max = 0.22$).

4.6.2. Change in response

Change in response was measured by calculating how many items a participant changed from their initial individual recall to the collaborative dyadic response. Items were calculated based on the nine shapes and nine colors, resulting in 18 possible changes ($M = 2.74$, $SD = 2.96$, $min = 0$, $max = 11$).

4.6.3. Cognitive load

Cognitive load was measured with four items adapted from Hart and Staveland's NASA Task Load Index (1988) using a 7-point Likert scale (1 = Strongly disagree, 7 = Strongly agree). The construct includes subdimensions such as mental and temporal demand, and frustration. Sample items include “The task was mentally demanding” and “I felt stressed during the discussion.” Cognitive load scores were calculated as the mean (Cronbach's $\alpha = 0.74$), with higher scores indicating greater cognitive load during the task ($M = 3.95$, $SD = 0.96$, $min = 2.00$, $max = 6.75$).

4.6.4. Willingness to submit collaborative response

Willingness to submit the collaborative response was measured by a single item, “How willing are you to submit the collaborative recall of the objects (i.e., the list you and your partner discussed together)?” using a 7-point Likert scale (1 = Very unwilling, 7 = Very willing). This item captures participants' willingness to submit their finalized, collaborative response, which, depending on its accuracy, would determine their eligibility for a monetary bonus ($M = 5.65$, $SD = 1.37$, $min = 2.00$, $max = 7.00$).

4.6.5. Confidence in memory

Confidence in memory was measured by three questions for the accuracy of the (1) participant's own recall ($M = 3.83$, $SD = 1.14$, $min = 1.00$, $max = 5.00$), (2) their partner's recall ($M = 3.37$, $SD = 1.05$, $min = 1.00$, $max = 5.00$), and (3) their collaborative recall ($M = 3.27$, $SD = 1.20$, $min = 1.00$, $max = 5.00$) using a 5-point Likert scale (1 = Not confident, 5 = Extremely confident).

4.6.6. Trust

Dyadic Trust. Trust within the dyad partner was measured with three items adapted from Larzelere and Huston's (1980) Dyadic Trust Scale using a 7-point Likert scale (1 = Strongly disagree, 7 = Strongly agree). The scale has two pertinent attributions, including benevolence (i.e., partner's interest in only their own or others' welfare) and honesty (i.e., extent to which one can believe a partner about their future intentions). Sample items include “My partner seemed truthful with me” and “I can count on my partner to accomplish other tasks.” Dyadic trust scores were calculated as the mean (Cronbach's $\alpha = 0.57$), with higher scores indicating greater dyadic trust ($M = 4.7$, $SD = 0.48$, $min = 3.33$, $max = 6.00$).

Technology Trust. Trust within the technological medium was measured with three items adapted from Mcknight and colleagues' (2011) Trusting Belief-Specific Technology Scale using a 7-point Likert scale (1 = Strongly disagree, 7 = Strongly agree). The scale reflects beliefs that a specific technology has the attributes necessary to perform as expected in a given situation in which negative consequences are possible, and has three pertinent attributions, including reliability, functionality, and helpfulness. Sample items include “Mixed reality seems to be a reliable piece of technology” and “Mixed reality functioned the way it was supposed to.” Technology trust scores were calculated as the mean (Cronbach's $\alpha = 0.77$), with higher scores indicating greater technological trust ($M = 5.35$, $SD = 0.97$, $min = 3.00$, $max = 7.00$).

5. Data analysis

We used linear mixed-effects models to examine the relationship between our outcome variables and our manipulated variable, controlling for baseline differences by including the outcome variable from the no-manipulation trial as a covariate. This approach was taken over the pre-registered repeated-measures ANOVA because the fixed trial ordering (i.e., no-manipulation baseline trial always preceding the manipulation trial) confounded time and condition effects in a traditional ANOVA framework. Specifically, because all dyads experienced the no-manipulation trial first and their randomly assigned condition second, a repeated-measures ANOVA could not disentangle time and condition effects. The mixed-effects model addresses this issue by treating the baseline trial as a covariate rather than a within-subjects

Table 1
Descriptive statistics of the outcome variables by condition.

Outcome Variable	Small number of discrepancies				Large number of discrepancies			
	Mean	SD	Min	Max	Mean	SD	Min	Max
Synchrony	0.10	0.06	−0.02	0.22	0.05	0.06	−0.09	0.18
Change in Response	2.35	2.97	0.00	11.00	3.13	2.92	0.00	9.00
Cognitive Load	3.79	0.94	2.00	6.25	4.12	0.96	2.50	6.75
Willingness to Submit	5.92	1.17	2.00	7.00	5.36	1.51	2.00	7.00
Confidence in Memory - Self	3.98	1.08	1.00	5.00	3.67	1.20	1.00	5.00
Confidence in Memory - Partner	3.29	0.94	1.00	5.00	3.44	1.16	1.00	5.00
Confidence in Memory - Collaborative	3.58	1.21	1.00	5.00	2.96	1.12	1.00	5.00
Dyad Trust	4.69	0.44	3.33	6.00	4.71	0.52	3.33	6.00
Technology Trust	5.13	1.11	3.00	7.00	5.56	0.77	4.00	7.00

factor, allowing us to isolate the effect of the manipulation while controlling for individual and dyadic differences in initial performance. Model comparisons using likelihood ratio tests are reported to assess whether including baseline covariates improved model fit.

Dyad was included as a random effect to account for variation between dyads. All possible interaction terms between the independent variable and baseline outcomes were tested but excluded due to insignificance across most variables.[3] Various individual- (e.g., familiarity with XR programming, Big 5 personality traits, ethnicity, prior familiarity with dyad member, study purpose identification) and dyad-level (i.e., gender composition) differences were additionally tested as covariates. Although some covariates showed isolated significant associations with individual outcomes (*p*-values ranged from 0.00267 to 0.994), none consistently predicted outcomes across models and were therefore excluded for parsimony and to reduce Type I errors. Outliers (3SD above or below the mean) were removed from the analyses. Final models test the effect of *number of discrepancies* while controlling for the no-manipulation trial. For synchrony, a dyad-level outcome, a linear model without a random effect was used, as each dyad contributed a single observation. Two dyads were dropped from synchrony analysis due to technical difficulties.

All linear mixed-effects models were fit to the data in R using the *lmerTest* package, with maximum likelihood estimation, incomplete data treated as missing at random, and statistical significance evaluated at alpha = 0.05. Conditional and marginal R^2 were calculated using the *MuMIn* package. Confidence intervals (CI) were calculated using the *r2glmm* package. The linear model was fit using the *stats* package, with ordinary least squares estimation, incomplete data treated as missing, and statistical significance evaluated at alpha = 0.05. Partial eta-squared and CI were calculated using the *effectsize* package. Figures were generated using the *ggplot2* package.

6. Results

6.1. Synchrony

There was a main effect of the *number of discrepancies* on synchrony, such that people were less in sync with one another when there was a greater number of discrepancies, b = −0.0454, SE = 0.0175, t(46) = −2.60, $p = 0.0126$ ($\eta^2 = 0.14$, 95% CI [0.02, 1.00]). There was no effect of baseline synchrony, $b = 0.0847$, $SE = 0.164$, $t(46) = 0.515$, $p = 0.6076$ ($\eta^2 = 0.06$, 95% CI [0.00, 1.00]). Model comparison indicated that including baseline synchrony did not significantly improve model fit, $F(1, 46) = 2.90$, $p = 0.0955$. Results are presented in Fig. 5.

6.2. Change in response

There was a main effect of the *number of discrepancies* on the number of items a participant changed from their initial response of their individual recall to the collaborative response of their dyadic recall, such that participants changed their response more when there was a greater number of discrepancies, $b = 1.20$, $SE = 0.558$, $t(97.0) = 2.14$, $p = 0.0347$ ($R^2 = 0.045$, 95% CI [0.001, 0.155]) (Fig. 6a). In other words, they were less likely to defend the discrepancy. Additionally, participants who changed their response more during the no-manipulation trial changed their response more overall, $b = 0.3063$, $SE = 0.1408$, $t(97.0) = 2.18$, $p = 0.0321$ ($R^2 = 0.047$, 95% CI [0.001, 0.157]). Conditional and marginal R^2 for the model were $R^2c = 0.073$, $R^2m = 0.073$, 95% CI [0.013, 0.208]. Model comparison indicated that including baseline changed response significantly improved model fit, $\chi^2(1) = 4.76$, $p = 0.0291$.

6.3. Cognitive load

There was a main effect of the *number of discrepancies* on cognitive load, such that participants reported feeling greater cognitive load when there was a greater number of discrepancies, $b = 0.464$, $SE = 0.182$, $t(49.02) = 2.55$, $p = 0.0138$ ($R^2 = 0.069$, 95% CI [0.006, 0.189]) (Fig. 6b). Additionally, participants who experienced higher baseline cognitive load during the no-manipulation trial reported greater cognitive load overall, $b = 0.413$, $SE = 0.0877$, $t(92.6) = 4.706$, $p < 0.001$ ($R^2 = 0.189$, 95% CI [0.074, 0.331]). Conditional and marginal R^2 for the model were $R^2c = 0.312$, $R^2m = 0.21004$, 95% CI [0.101, 0.365]). Model comparison indicated that including baseline cognitive load significantly improved model fit, $\chi^2(1) = 20.078$, $p < 0.001$.

6.4. Willingness to submit collaborative response

There was a main effect of the *number of discrepancies* on willingness to submit the collaborative response, such that participants reported feeling less willing to submit the collaborative response when there was a greater number of discrepancies, $b = -0.721$, $SE = 0.292$, $t(45.7) = -2.47$, $p = 0.0171$ ($R^2 = 0.078$, 95% CI [0.008, 0.204]) (Fig. 6c). There was no effect of baseline willingness to submit the collaborative response, $b = 0.222$, $SE = 0.182$, $t(95.5) = 1.22$, $p = 0.225$ ($R^2 = 0.016$, 95% CI [0.000, 0.101]). Conditional and marginal R^2 for the model were $R^2c = 0.341$ and $R^2m = 0.0846$, 95% CI [0.019, 0.228]. Model comparison indicated that including baseline willingness to submit the collaborative response did not significantly improve model fit $\chi^2(1) = 1.54$, $p = 0.214$.

6.5. Confidence in memory

Self. There was no main effect of the *number of discrepancies* on the confidence in the self's recall, $b = -0.338$, $SE = 0.196$, $t(101.0) = -1.73$, $p = 0.0872$ ($R^2 = 0.029$, 95% CI [0.000, 0.124]). Additionally, participants who had higher baseline confidence in the self's recall during the no-manipulation trial reported greater confidence in the self's recall, $b = 0.521$, $SE = 0.0916$, $t(101.0) = 5.69$, $p < 0.001$ ($R^2 = 0.244$, 95% CI [0.119, 0.387]). Conditional and marginal R^2 for the model were $R^2c = 0.253$, $R^2m = 0.253$, 95% CI [0.136, 0.406]. Model comparison indicated that including baseline confidence in the self's recall significantly improved model fit, $\chi^2(1) = 28.9$, $p < 0.001$.

Partner. There was no main effect of the *number of discrepancies* on the confidence in the partner's recall, $b = 0.0694$, $SE = 0.212$, $t(50.91) = 0.327$, $p = 0.745$ ($R^2 = 0.001$, 95% CI [0.000, 0.055]). Additionally, participants who had higher baseline confidence in the partner's recall during the no-manipulation trial reported greater confidence in the partner's recall, $b = 0.359$, $SE = 0.113$, $t(96.7) = 3.18$, $p = 0.002$ ($R^2 = 0.094$, 95% CI [0.015, 0.222]). Conditional and marginal R^2 for the model were $R^2c = 0.192$, $R^2m = 0.0971$, 95% CI [0.024, 0.238]. Model comparison indicated that including baseline trust within the technological medium significantly improved model fit, $\chi^2(1) = 9.92$, $p = 0.00164$.

Collaborative. There was a main effect of the *number of discrepancies* on the confidence in the collaborative recall, such that participants reported feeling less confident in the accuracy of their collaborative recall when there was a greater number of discrepancies, $b = -0.687$, $SE = 0.240$, $t(49.2) = 2.87$, $p = 0.0061$ ($R^2 = 0.084$, 95% CI [0.011, 0.208]) (Fig. 6d). There was no effect of baseline confidence in collaborative recall, $b = 0.0847$, $SE = 0.164$, $t(88.3) = 0.515$, $p = 0.6076$ ($R^2 = 0.003$, 95% CI [0.000, 0.061]). Conditional and marginal R^2 for the model were $R^2c = 0.1704$, $R^2m = 0.08102$, 95% CI [0.017, 0.217]. Model comparison indicated that including baseline confidence in collaborative recall

[3] Our other independent variable was initially included in our primary models but was removed from the final models as it did not significantly predict any outcome variables. We tested all possible interaction terms, including two-way and three-way interactions between the two independent variables and baseline outcome measures. These terms were excluded from the final models due to insignificance across most outcome variables.

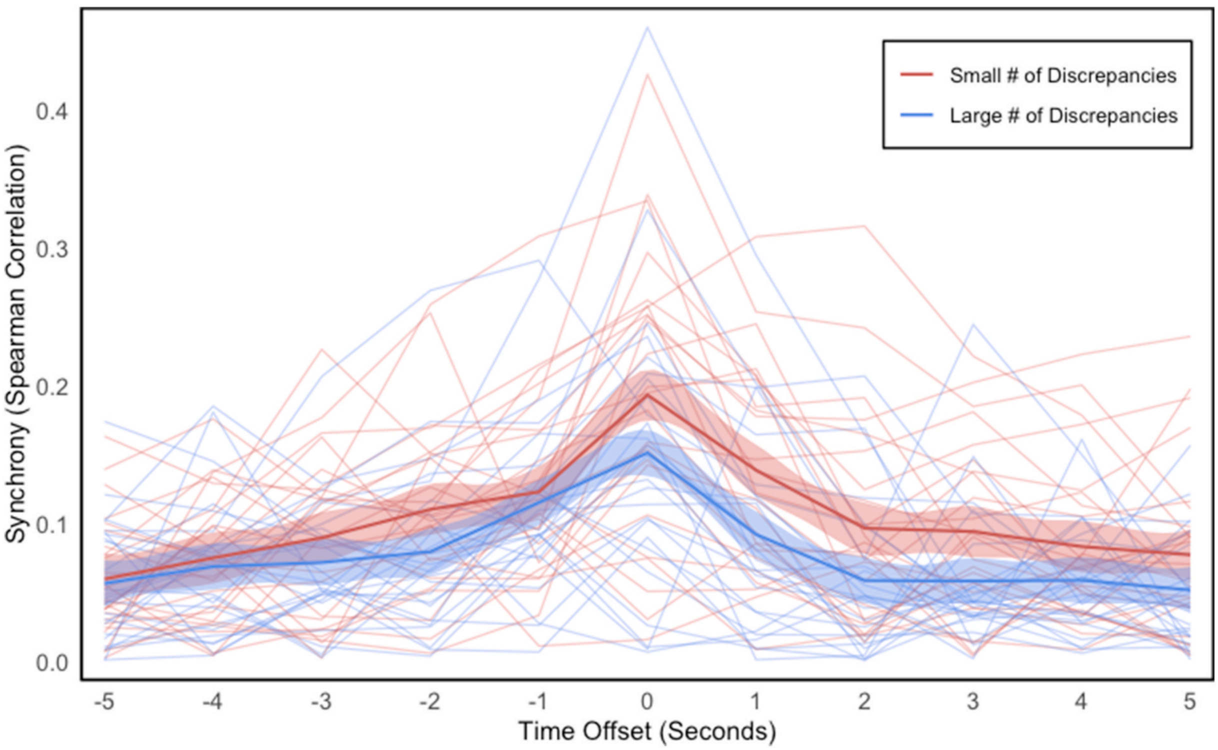

Fig. 5. Synchrony for each dyad in the manipulation trial is represented by transparent lines. The X-axis represents the time offset used to calculate Spearman correlations, created by shifting one participant's motion signal forward or backward in time relative to their partner's. As the offset moves further from zero (i.e., the motion signals are less temporally aligned), synchrony decreases. The averages of the two conditions at each offset are represented by the darker lines, with their ribbons indicating 95% confidence interval based on the underlying distributions. For the final analysis, the average values from these 11 correlations were taken.

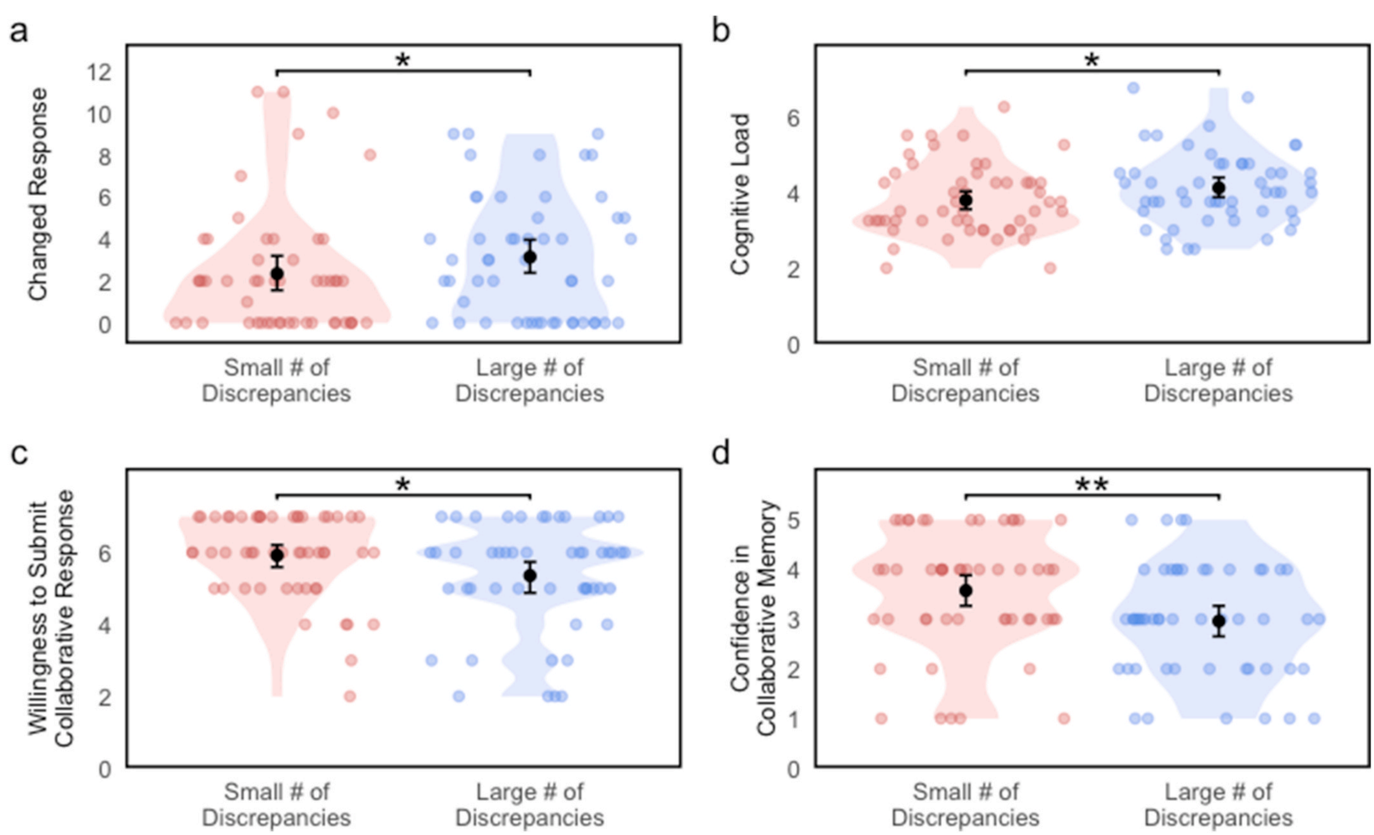

Fig. 6. Average outcome variables with 95% CI, by condition.

did not significantly improve model fit, $\chi^2(1) = 0.237$, $p = 0.627$.

6.6. Trust

Dyadic Trust. There was no main effect of the *number of discrepancies* on trust within the dyad partner, $b = -0.0185$, $SE = 0.10016$, $t(48.4) = -0.185$, $p = 0.854$ ($R^2 = 0.00$, 95% CI [0.00, 0.052]). There was no effect of baseline trust within the dyad partner, $b = 0.132$, $SE = 0.07508$, $t(95.6) = 1.76$, $p = 0.0816$ ($R^2 = 0.031$, 95% CI [0.00, 0.129]. Conditional and marginal R^2 for the model were $R^2c = 0.131$, $R^2m = 0.03047$, 95% CI [0.002, 0.142]. Model comparison indicated that including baseline trust within the dyad partner did not significantly improve model fit, $\chi^2(1) = 2.99$, $p = 0.0839$.

Technology Trust. There was no main effect of the *number of discrepancies* on trust within the technological medium, $b = 0.235$, $SE = 0.137$, $t(44.01) = 1.72$, $p = 0.0924$ ($R^2 = 0.029$, 95% CI [0.000, 0.125]). Additionally, participants who experienced higher baseline trust within the technological medium during the no-manipulation trial reported greater trust overall, $b = 0.888$, $SE = 0.09049$, $t(99.1) = 9.808$, $p < 0.001$ ($R^2 = 0.493$, 95% CI [0.368, 0.610]). Conditional and marginal R^2 for the model were $R^2c = 0.511$, $R^2m = 0.51002$, 95% CI [0.399, 0.632]. Model comparison indicated that including baseline trust within the technological medium significantly improved model fit, $\chi^2(1) = 69.3$, $p < 0.001$.

7. Discussion

This study investigates the following question: when perceptual conflicts arise in participants' virtual experience, violating expectations, how will they shape the communication process? Participants shared the same physical space but experienced perceptually conflicting stimuli in MR. Unlike conflicts arising from individual perspectives or self-presentation, this study focuses on an external, visible referent as the source of conflict.

The effects of perceptual conflict were not uniformly reflected across the different levels of measurement, spanning self-report and behavioral, as well as interpersonal and intrapersonal outcomes. Notably, the magnitude of the discrepancy did not appear to alter how participants viewed themselves or their partners individually, such that confidence in self or partner recall were not affected by the manipulation, nor trust in the partner or technology. However, participants felt less confident in their collaborative recall and were less willing to submit joint responses when this discrepancy was greater. The effects were particularly evident at the behavioral level: participants were less synchronized with their partner, made more changes in their recall during the collaborative task, and reported feeling greater cognitive load. On a higher level, this suggests that while perceptual conflicts may not fundamentally change how oneself or their communication partner, the greater the discrepancy, the more it undermines the *collective* process – the shared foundation that joint outcomes depend on.

In relation to past literature, the finding that dyad members were less in sync when there was a greater discrepancy aligns with past research demonstrating that synchrony is closely tied to social outcomes, such that it arises naturally during active joint activities (Gordon et al., 2020)). Here, we contribute to the line of work on interpersonal synchrony as not only a marker of successful joint activity but also a potential signal of when expectations have been violated. Although the goal of the activity was to collaborate with one another, the lack of joint action introduced by the active disruptions to the dyad members' shared reality, along with the degree of the severity of this disruption, may have contributed to the downstream effect of decreased synchrony.

Results also showed that participants changed their responses from individual recall during the collaborative phase when discrepancies were greater, suggesting some social influence. Past research shows individuals may adjust responses under majority influence, even when it contradicts their own perception (Mori & Arai, 2010). Dyads were required to reach consensus and received a monetary incentive to either defend their response or collaborate, even if it conflicted with their judgement. By design, if one member held firm to their original response, the other showed greater change, so these changes cannot be solely interpreted as conformity. The findings from the willingness to submit and confidence in memory may shed some light on this finding. The number of discrepancies did not significantly influence participants' confidence in their own memory or in their dyad partner's memory. However, it did influence their collaborative memory and their willingness to submit the final response. This suggests that while their perceptions of individual and dyad's performance remained stable, the discrepancy may have impacted their trust in the joint decision-making process. Moreover, individual differences played a role, similar to findings of gender by Mori and Arai (2010), with those who were more likely to change their responses at baseline also showing greater susceptibility to change.

Furthermore, the number of discrepancies did not reduce trust in either the partner or technology, consistent with Seuren and colleagues (2021), who found that latency had procedural consequences, such that interactants treated these silences and overlaps as turn-taking problems, not latency problems. Similarly, the discrepancies here may not have been seen as significant enough to alter trust but did affect the collaborative process. This is reflected in reduced synchrony, lower willingness to submit, and decreased confidence in collaborative memory, suggesting that while trust in partner and technology remained stable, the shared process was affected.

Finally, the number of discrepancies was found to influence cognitive load, suggesting these transformations may come at a cognitive cost by increasing the mental effort required to reconcile differing realities. The findings suggest that while transformed cues can serve strategic purposes, they may also lead to additional cognitive demands on individuals working together in shared virtual environments.

7.1. Limitations and future directions

In the present study, dyads completed two trials: the first had no manipulations and the second did. While the first trial familiarized participants with the task and each other, and established baseline measures (e.g., trust), it may have influenced expectations for the second trial. The absence of discrepancies initially likely fostered a sense of consistency and trust, making later discrepancies more salient and drawing disproportionate attention. Furthermore, the measure of dyadic trust showed poor reliability, limiting the ability to detect effects on interpersonal trust. Future studies should examine discrepancies without a preceding trial and use more targeted trust measures.

In this vein, while the first trial served as a control, it may have been more effective as a separate condition. This raises new questions about how these dynamics evolve over time. Although some dyads varied in prior familiarity (friends, strangers, partners, family), the study did not track changes across repeated interactions. Future work should investigate whether participants learn to navigate persistent conflicts and whether decreased willingness to submit, lower confidence, and reduced synchrony have lasting consequences.

Third, this study focused on dyads, but larger groups introduce new dynamics. Majorities and minorities may emerge. If there is a greater number of people arguing against an individual's perceptual experience, will that individual defend the discrepancy and voice their disagreement, or will they attribute the error to their own perception or to the fault of the technology? Furthermore, while we did not find ethnicity-related differences, cultural attitudes and beliefs, such as variations in how members reconcile disagreements or collectivist versus individualist orientations, were not measured in this study but may play a role. Given that recent research has demonstrated cultural differences in XR technology adoption (Monteiro et al., 2024), future work should explore how other factors, such as group size and culture, influence these dynamics.

Finally, the memory task allowed for clear right or wrong answers and was designed to be challenging but not dismissible as simple memory errors. Through piloting, a version that was challenging enough to engage participants but not difficult to the point that disagreements could be dismissed as simple memory failures rather than meaningful perceptual conflicts. However, this task also had its limitations. While the present task had measures to avoid obvious configurations (e.g., three red spheres in a row), there may have been other patterns that were salient to individuals, influencing how virtual objects were processed. Future studies should explore alternative tasks that better isolate collaborative processes, such as tasks that focus on joint problem-solving or shared decision-making.

7.2. Implications

7.2.1. For practice

One way to understand perceptual conflicts is through technological artefacts that alter the shared environment. Similar to latency in videoconferencing, MR systems can introduce distortions, such as object drift. In such cases, one MR user may point to an object that appears misaligned for the other, disrupting grounding. Perceptual conflicts may also emerge through personalization. Personalization is increasingly prevalent in digital life, where research on social media shows that no two users have the same experience on a platform (Kalyanaraman & Sundar, 2006). Similarly, virtual environments can be adapted for each user in real-time based on information such as object metadata and user feedback (e.g., Valmorisco et al., 2024). This, in conjunction with the affordances of the medium, such as realism and immersion, could lead people to not simply play, but live in alternate realities. To illustrate an example, consider the following scenario: for one MR user, the system implements diminished reality, a set of methodologies to eliminate objects in a perceived environment (Cheng et al., 2022). Meanwhile, their communication partner sees a different filtering of the same environment. As the two collaborate within the same space, one user might reference an object that, to the other, does not exist. This type of perceptual conflict highlights how an example of personalized MR experiences can complicate collaborative tasks.

It is important to note, however, the present study was conducted in a controlled laboratory setting, which allowed us to systematically create and manipulate perceptual conflicts while holding constant other variables that could introduce noise. It is important to acknowledge that in collaborative settings outside of the laboratory, numerous additional factors can, and have been shown to, cause similar conflicts beyond the manipulations examined here. Factors such as users' differences in physical location (Bailenson et al., 2008; Hasenbein et al., 2022), asymmetric access to different mediums (e.g., Drey et al., 2022; Grandi et al., 2019; Jeong et al., 2019; Numan & Steed, 2022), and variations in viewing angles or perspectives (Hoppe et al., 2021; Zhou & Won, 2024) may inevitably arise, potentially leading to both immediate disruptions such as those found in the present study, and/or long-term consequences for team coordination that may go unnoticed or unaddressed. Moreover, as users grow accustomed to these technologies and come to expect these conflicts, they may adjust their expectations of both the technology and their partner, raising additional questions.

7.2.2. For theory

From a theoretical perspective, these findings extend our understanding of EVT, as well as broader principles in communication that depend on establishing a sense of common ground for successful joint activity, by examining interactions in an MR-mediated context. Our results suggest that while perceptual conflicts during shared experiences may not immediately undermine trust in one's communication partner or the medium itself, nor diminish confidence in oneself or the partner, they have costs to nonverbal communication and decision-making. This raises new questions about whether interpersonal interactions in immersive environments require additional mechanisms to manage expectancy violations when a stable, shared reality cannot be assumed.

Beyond the dyadic level examined in our study, these findings open new avenues for theory building in intergroup contexts. Research on the affordances of the sociomaterial environment suggests that people from different backgrounds perceive the affordances of their surrounding environment differently, shaped by their cultural practices, norms, and material surroundings. In other words, individuals from different backgrounds may fundamentally see and coordinate with the world in different ways (Van Dijk & Rietveld, 2017). These differences may complicate interactions with members of different social groups, particularly during communication processes where interactants must establish mutual beliefs and shared understanding for successful joint action (Clark & Brennan, 1991). Our experimental design, which systematically manipulates the affordances people see through MR, could therefore serve as a research tool to study and potentially improve intergroup dynamics through the lens of sociomaterial affordances and perspective-taking. This study introduces the potential to use tools like MR to deliberately manipulate immediate environmental affordances to scaffold shared experiences and facilitate grounding across group boundaries.

That being said, the implications of such perceptual conflicts, arising from technological artefacts or intentional personalization or diminishment, remain understudied. This present study contributes to the growing literature of how individual changes influence group-based outcomes. Although more research is needed to examine how these effects evolve over time as these perceptual conflicts become more common, and consequently, expected, the current findings underscore that these conflicts can introduce disruptions to the communication process.

CRediT authorship contribution statement

Eugy Han: Writing – review & editing, Writing – original draft, Visualization, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Portia Wang:** Writing – review & editing, Methodology, Data curation, Conceptualization. **Monique Santoso:** Writing – review & editing, Methodology, Data curation. **Keshav Rastogi:** Writing – review & editing, Investigation, Formal analysis, Data curation. **Jeremy N. Bailenson:** Writing – review & editing, Supervision, Formal analysis, Conceptualization.

Consent to participate

Written informed consent was obtained from all participants prior to their involvement in the study.

Ethical considerations

The recruitment and experiment process were approved by the university's IRB (ID: 74058).

Funding

This project was supported by the Sloan Foundation and the Data Sciences Institute at the University of Toronto with funding through grant number DSI-QRSY3R1P03.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors would like to thank James Brown, Roy Pea, and Jaylen Pittman for their support during the development of the study. The authors would also like to thank Emmanuel Angel Corona, Ali Muhammad, and Sam Shors for their support running the study.

Data availability

Data will be made available on request.

References

Abdelazeem, B., Abbas, K. S., Amin, M. A., El-Shahat, N. A., Malik, B., Kalantary, A., & Eltobgy, M. (2022). The effectiveness of incentives for research participation: A systematic review and meta-analysis of randomized controlled trials. *PLoS One, 17*(4), Article e0267534.

Asch, S. E. (1951). Effects of group pressure upon the modification and distortion of judgments. In *Documents of gestalt psychology* (pp. 222–236). University of California Press.

Bailenson, J. N., Yee, N., Blascovich, J., Beall, A. C., Lundblad, N., & Jin, M. (2008). The use of immersive virtual reality in the learning sciences: Digital transformations of teachers, students, and social context. *The Journal of the Learning Sciences, 17*(1), 102–141.

Barreda-Ángeles, M., & Hartmann, T. (2022). Psychological benefits of using social virtual reality platforms during the covid-19 pandemic: The role of social and spatial presence. *Computers in Human Behavior, 127*, Article 107047.

Bréhault, V., Dubois, E., Prouzeau, A., & Serrano, M. (2025). A systematic literature review to characterize asymmetric interaction in collaborative systems. In *Proceedings of the 2025 CHI conference on human factors in computing systems* (pp. 1–19).

Burgoon, J. K. (1993). Interpersonal expectations, expectancy violations, and emotional communication. *Journal of Language and Social Psychology*. https://doi.org/10.1177/0261927X93121003

Burgoon, J. K., Floyd, K., & Guerrero, L. K. (2010). Nonverbal communication theories of interaction adaptation. In *The handbook of communication science* (pp. 93–108). SAGE Publications, Inc.

Burgoon, J. K., Stern, L. A., & Dillman, L. (1995). *Interpersonal adaptation: Dyadic interaction patterns*. Cambridge University Press.

Chen, W., Clavel, C., Férey, N., & Bourdot, P. (2014). Perceptual conflicts in a multi-stereoscopic immersive virtual environment: Case study on face-to-face interaction through an avatar. *Presence (Cambridge, Mass.), 23*(4), 410–429.

Cheng, Y. F., Yin, H., Yan, Y., Gugenheimer, J., & Lindlbauer, D. (2022). Towards understanding diminished reality. In *CHI conference on human factors in computing systems. CHI '22: CHI conference on human factors in computing systems, New Orleans LA USA*. https://doi.org/10.1145/3491102.3517452

Cheong, J. H., Jolly, E., Xie, T., Byrne, S., Kenney, M., & Chang, L. J. (2021). Py-feat: Python facial expression analysis toolbox. http://arxiv.org/abs/2104.03509.

Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In *Perspectives on socially shared cognition* (pp. 127–149). American Psychological Association.

Condon, W. S., & Ogston, W. D. (1966). Sound film analysis of normal and pathological behavior patterns. *The Journal of Nervous and Mental Disease, 143*(4), 338–347.

Drey, T., Albus, P., der Kinderen, S., Milo, M., Segschneider, T., Chanzab, L., Rietzler, M., Seufert, T., & Rukzio, E. (2022). Towards collaborative learning in virtual reality: A comparison of co-located symmetric and asymmetric pair-learning. In *CHI conference on human factors in computing systems. CHI '22: CHI conference on human factors in computing systems, New Orleans LA USA*. https://doi.org/10.1145/3491102.3517641

Gergle, D., Kraut, R. E., & Fussell, S. R. (2012). Using visual information for grounding and awareness in collaborative tasks. *Human-Computer Interaction*, 1–39.

Gevins, A., Smith, M. E., McEvoy, L. K., Ilan, A. B., Chan, C. S., Jiang, A., Sam-Vargas, L., & Abraham, G. (2011). A cognitive and neurophysiological test of change from an individual's baseline. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology, 122*(1), 114–120.

Gordon, I., Gilboa, A., Cohen, S., Milstein, N., Haimovich, N., Pinhasi, S., & Siegman, S. (2020). Physiological and behavioral synchrony predict group cohesion and performance. *Scientific reports, 10*(1), 8484.

Grandi, J. G., Debarba, H. G., & Maciel, A. (2019). Characterizing asymmetric collaborative interactions in virtual and augmented realities. In *2019 IEEE conference on virtual reality and 3D user interfaces (VR). 2019 IEEE conference on virtual reality and 3D user interfaces (VR), Osaka, Japan*. https://doi.org/10.1109/vr.2019.8798080

Han, E., Miller, M. R., DeVeaux, C., Jun, H., Nowak, K. L., Hancock, J. T., Ram, N., & Bailenson, J. N. (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: JCMC, 28*(2). https://doi.org/10.1093/jcmc/zmac031

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (task load index): Results of empirical and theoretical research. In *Advances in psychology* (pp. 139–183). Elsevier.

Hasenbein, L., Stark, P., Trautwein, U., Queiroz, A. C. M., Bailenson, J., Hahn, J.-U., & Göllner, R. (2022). Learning with simulated virtual classmates: Effects of social-related configurations on students' visual attention and learning experiences in an immersive virtual reality classroom. *Computers in Human Behavior, 133*(107282), Article 107282.

Hoppe, A. H., van de Camp, F., & Stiefelhagen, R. (2021). ShiSha. *Proceedings of the ACM on Human-Computer Interaction, 4*(CSCW3), 1–22.

Jeong, K., Kim, J., Kim, M., Lee, J., & Kim, C. (2019). Asymmetric interface: User interface of asymmetric virtual reality for new presence and experience. *Symmetry, 12*(1), 53.

Kalyanaraman, S., & Sundar, S. S. (2006). The psychological appeal of personalized content in web portals: Does customization affect attitudes and behavior? *Journal of Communication, 56*(1), 110–132.

Kerr, N. L., & Levine, J. M. (2008). The detection of social exclusion: Evolution and beyond. *Group Dynamics: Theory, Research, and Practice, 12*(1), 39.

Koschmann, T., & LeBaron, C. D. (2003). Reconsidering common ground. In *ECSCW 2003* (pp. 81–98). Netherlands: Springer.

Krämer, N. C., Neubaum, G., Winter, S., Schaewitz, L., Eimler, S., & Oliver, M. B. (2021). I feel what they say: The effect of social media comments on viewers' affective reactions toward elevating online videos. *Media Psychology, 24*(3), 332–358.

Larzelere, R. E., & Huston, T. L. (1980). The dyadic trust scale: Toward understanding interpersonal trust in close relationships. *Journal of Marriage and Family, 42*(3), 595.

Leary, M. R., & Baumeister, R. F. (2000). The nature and function of self-esteem: Sociometer theory. *Advances in Experimental Social Psychology, 32*, 1–62. Academic Press.

Lugaresi, C., Tang, J., Nash, H., McClanahan, C., Uboweja, E., Hays, M., Zhang, F., Chang, C.-L., Yong, M. G., Lee, J., Chang, W.-T., Hua, W., Georg, M., & Grundmann, M. (2019). MediaPipe: A framework for building perception pipelines. http://arxiv.org/abs/1906.08172.

Mcknight, D. H., Carter, M., Thatcher, J. B., & Clay, P. F. (2011). Trust in a specific technology. *ACM Transactions on Management Information Systems, 2*(2), 1–25.

Miler, T. S., & Steinfeld, N. (2025). Interaction and collaborative intergroup dynamics in social virtual reality (SVR). In *2025 11th international conference on virtual reality (ICVR)* (pp. 88–97). IEEE.

Miller, M. R. (2023). Virtual reality tracking data: Insights. *Risks, Opportunities*.

Monteiro, D., Ma, T., Li, Y., Pan, Z., & Liang, H. N. (2024). Cross-cultural factors influencing the adoption of virtual reality for practical learning. *Universal Access in the Information Society, 23*(3), 1203–1216.

Mori, K., & Arai, M. (2010). No need to fake it: Reproduction of the Asch experiment without confederates. *International Journal of Psychology: Journal International de Psychologie, 45*(5), 390–397.

New Meta ray-ban AI-Powered display glasses & neural band. (2025). Meta Store. https://www.meta.com/ai-glasses/meta-ray-ban-display/?srsltid=AfmBOooGgTGah7LFk11KiDwIVXQN7fj2037QVj2Lqy4T5yZks7aZCWyr.

Numan, N., & Steed, A. (2022). Exploring user behaviour in asymmetric collaborative mixed reality. In *Proceedings of the 28th ACM symposium on virtual reality software and technology. VRST '22: 28th ACM symposium on virtual reality software and technology, Tsukuba Japan*. https://doi.org/10.1145/3562939.3565630

Orion AR glasses. (2025). Meta. https://www.meta.com/emerging-tech/orion/?srsltid=AfmBOop6dU72HzXcgbqnxxWYb2HxmTMKRD3LSMGB58tofddbsYt4vDc1.

Ramseyer, F., & Tschacher, W. (2014). Nonverbal synchrony of head- and body-movement in psychotherapy: Different signals have different associations with outcome. *Frontiers in Psychology, 5*, 979.

Santoso, M., & Bailenson, J. (2024). How video passthrough headsets influence perception of self and others. *Cyberpsychology, Behavior, and Social Networking, 27*(11), 798–806. https://doi.org/10.1089/cyber.2024.0398

Seuren, L. M., Wherton, J., Greenhalgh, T., & Shaw, S. E. (2021). Whose turn is it anyway? Latency and the organization of turn-taking in video-mediated interaction. *Journal of Pragmatics, 172*, 63–78.

Sps 2024. (2024). Introducing new spectacles and snap OS: The next frontier of AR glasses. https://newsroom.snap.com/sps-2024-spectacles-snapos.

Stefik, M., Bobrow, D. G., Foster, G., Lanning, S., & Tatar, D. (1987). WYSIWIS revised: Early experiences with multiuser interfaces. *ACM Transactions on Information Systems, 5*(2), 147–167.

Tassinari, M., Aulbach, M. B., & Jasinskaja-Lahti, I. (2022). The use of virtual reality in studying prejudice and its reduction: A systematic review. *PLoS One, 17*(7), Article e0270748.

Tassinari, M., Marinucci, M., & Hasler, B. S. (2026). Immersive technologies and intergroup relations: A critical review of theoretical questions and methodological challenges. *Computers in Human Behavior*, Article 108908.

Tschacher, W., Rees, G. M., & Ramseyer, F. (2014). Nonverbal synchrony and affect in dyadic interactions. *Frontiers in Psychology, 5*, 1323.

Tsuchiya, A., Ora, H., Hao, Q., Ono, Y., Sato, H., Kameda, K., & Miyake, Y. (2020). Body movement synchrony predicts degrees of information exchange in a natural conversation. *Frontiers in Psychology, 11*, 817.

Valmorisco, S., Raya, L., & Sanchez, A. (2024). Enabling personalized VR experiences: A framework for real-time adaptation and recommendations in VR environments. *Virtual Reality, 28*(3). https://doi.org/10.1007/s10055-024-01020-0

Van Dijk, L., & Rietveld, E. (2017). Foregrounding sociomaterial practice in our understanding of affordances: The skilled intentionality framework. *Frontiers in Psychology, 7*, 1969.

Voida, A., Voida, S., Greenberg, S., & He, H. A. (2008). Asymmetry in media spaces. In *Proceedings of the 2008 ACM conference on computer supported cooperative work* (pp. 313–322).

Winter, S., Brückner, C., & Krämer, N. C. (2015). They came, they liked, they commented: Social influence on Facebook news channels. *Cyberpsychology, Behavior, and Social Networking, 18*(8), 431–436.

Zhang, X., & Furnas, G. W. (2002). Social interactions in multiscale CVEs. In *Proceedings of the 4th international conference on collaborative virtual Environments. CVE02:*

Collaborative virtual environments, Bonn Germany. https://doi.org/10.1145/571878.571884

Zhao, S., Tan, F., & Fennedy, K. (2023). Heads-up computing moving beyond the device-centered paradigm. *Communications of the ACM, 66*(9), 56–63. https://doi.org/10.1145/3571722

Zhou, S., & Won, A. S. (2024). Actor–observer asymmetry in virtual reality. *Journal of Media Psychology, 36*(1), 27–44.