Perspectives: Creating Inclusive and Equitable Hybrid Meeting Experiences

JOHN C. TANG, Microsoft Research, USA
KORI INKPEN, Microsoft Research, USA
SASA JUNUZOVIC, Microsoft Research, USA
KERI MALLARI, University of Washington & Microsoft Research, USA
ANDREW D. WILSON, Microsoft Research, USA
SEAN RINTEL, Microsoft Research, UK
SHIRAZ CUPALA, Microsoft, USA
TONY CARBARY, Microsoft Research, USA
ABIGAIL SELLEN, Microsoft Research, UK
WILLIAM A.S. BUXTON, Microsoft Research & Buxton Design, Canada

With the shift to hybrid meetings in work spaces, there is an increasing need to create a more inclusive hybrid meeting experience where people meeting together in a room interact with those joining remotely. This paper describes a design exploration, implementation, and evaluation of Perspectives, a novel hybrid meeting system that aimed to create an inclusive and equitable space for hybrid meetings. Perspectives digitally composites everyone into a virtual room so that each person has a unique but spatially consistent viewpoint into the meeting. The user study compared Perspectives with three commercially available UX designs for hybrid meetings: Gallery, Together Mode, and Front Row. Results from this study revealed key benefits of Perspectives, including supporting natural interactions, creating a strong sense of co-presence, and reducing cognitive load. Results from the study also helped iterate on the design principles of Perspectives, which offer important insights on supporting hybrid meetings.

CCS Concepts: • **Human-centered computing → Empirical studies in HCI**.

Additional Key Words and Phrases: hybrid; video meetings; collaboration; design; spatiality; perspective view; consistency; first-person view; self-view; inclusion effectiveness

ACM Reference Format:
John C. Tang, Kori Inkpen, Sasa Junuzovic, Keri Mallari, Andrew D. Wilson, Sean Rintel, Shiraz Cupala, Tony Carbary, Abigail Sellen, and William A.S. Buxton. 2023. Perspectives: Creating Inclusive and Equitable Hybrid Meeting Experiences. *Proc. ACM Hum.-Comput. Interact.* 7, CSCW2, Article 287.2356 (October 2023), 25 pages. https://doi.org/10.1145/3610200

1 INTRODUCTION

Hybrid work, in which employees are distributed between working remotely and in-facility, is here to stay [1]. Recent studies have shown both increased productivity and appreciation of work

Authors' addresses: John C. Tang, johntang@microsoft.com, Microsoft Research, Mountain View, CA, USA, 94043; Kori Inkpen, kori@microsoft.com, Microsoft Research, Redmond, WA, USA, 98052; Sasa Junuzovic, sasa.junuzovic@microsoft.com, Microsoft Research, Redmond, WA, USA, 98052; Keri Mallari, kmallari@uw.edu, University of Washington & Microsoft Research, Seattle, WA, USA, 98195; Andrew D. Wilson, awilson@Microsoft.com, Microsoft Research, Redmond, WA, USA, 98052; Sean Rintel, serintel@microsoft.com, Microsoft Research, Cambridge, UK, CB1 2FB; Shiraz Cupala, shirazc@microsoft.com, Microsoft, Redmond, WA, USA, 98052; Tony Carbary, tony@microsoft.com, Microsoft Research, Redmond, WA, USA, 98052; Abigail Sellen, asellen@microsoft.com, Microsoft Research, Cambridge, UK, CB1 2FB; William A.S. Buxton, bill@billbuxton.com, Microsoft Research & Buxton Design, Toronto, ON, Canada, M4E 1A1.

2573-0142/2023/10-ART287.2356 $15.00
https://doi.org/10.1145/3610200

Fig. 1. The Perspectives prototype UX for those gathered together in the meeting room

flexibility, albeit with some increased stress [3, 11]. Video meetings are a crucial component of the many technologies and practices needed to enable effective hybrid work. The COVID-19 pandemic prompted widespread adoption of fully distributed video meetings where everyone joined remotely, but that does not accommodate hybrid work. Although they give all participants reasonably equivalent experiences, they do not leverage the face-to-face interactions for people who are co-located. Hybrid meetings, where some people are physically together while others join remotely, introduce challenges to meeting dynamics, such as in-grouping by co-located participants that excludes remote participants [4, 38, 39].

In this paper, we outline three design principles that are crucial for enabling more inclusive hybrid meetings: *blending* rather than separating in-room and remote users; *consistency* in the spatial arrangement of people to give everyone a 'seat at the table'; and *first-person perspective* view that is unique to each participant and spatially consistent with the relative positioning of meeting participants. We describe related work that informed the design exploration that led to these principles and the development of an experimental hybrid meeting system called 'Perspectives' (Figure 1)[1]. We report findings of a user study comparing Perspectives to three other hybrid video meeting layouts available in commercial video meeting systems (Gallery, Together Mode, Front Row). We found that Perspectives was strongly preferred by users, helped support natural interactions, increased feelings of co-presence, and reduced users' cognitive load. We also discovered important differences between in-room and remote experiences in all conditions, suggesting that future designs should take users' location into account, even if this leads to asymmetries in local and remote user interfaces. Overall, the results from the user study validate our hybrid design principles, indicate areas for further iteration, and provide important insights for hybrid meeting systems.

2 RELATED WORK

Hybrid meetings have a long history of research. Here we review salient investigations into partially-distributed meetings, including prior work on supporting hybrid work through telepresence meeting

[1]Because we thought it was important to show people's faces and their gaze orientation towards each other in the prototypes, we obtained explicit permission from the participants to include their images presented in this paper.

rooms and robots. We focus on research that explored consistent spatial metaphors for meeting UXs and sharing a common background among participants in that space, which are two core design aspects of Perspectives.

2.1 Prior research on hybrid meetings

A range of research has looked at hybrid meetings, or partially distributed meetings. Bos et al. [4] documented how partially distributed groups formed in-groups, where those gathered together in the meeting room tended to ignore or isolate those who joined remotely. While there were no significant differences in performance on the task between those in the room and those joining remotely, they did find that those in the room rated the task as more fun than the remotes. Later they examined whether these differences existed in groups who already knew each other compared to strangers [5]. While they found that groups who were familiar with each other performed better than strangers, there were still inequities in that those in the room tended to set the strategy and do more of the work while the remotes went along with those plans. Recent field ethnographies have documented similar issues [38, 39]. These studies all show how the structural asymmetry of being in the room versus joining remotely creates differences in group dynamics that need to be addressed to enable equitable participation in the hybrid context.

Plotnick et al. [36] analyzed survey data from 243 professionals who regularly participated in partially distributed teams and documented how in-group dynamics that privilege the in-room participants have a strong negative relationship with perceived effectiveness. The study also provided evidence that training to reduce the negative effects of in-group dynamics actually can have the opposite effect. Their research demonstrates the pervasive nature of negative in-group dynamics in hybrid meetings, and suggests that designing a UX that implicitly reduces the asymmetry between in-room and remote participants could have advantages over training.

Neumayr et al. [33] provided a recent descriptive framework for activities that occur in hybrid collaboration. Their focus was on a general collaboration task that comprised a mix of group and individual work beyond interactive discussions in hybrid meetings. They found that territoriality, which was enabled by a shared spatial awareness, shaped the kind of collaboration activity that emerged over the course of the task, and that shared awareness, especially through audio, was key to staying aware of the group's activities.

Lacy et al. [24] used surveys and interviews to identify challenges in virtual, in-person, and hybrid meetings. This study was motivated in part by observing how virtual activities during the pandemic enabled new opportunities for participation for people with disabilities, and they wanted to understand the tradeoffs in inclusivity, accessibility, and experience among those settings. They focused on providing a "seat at the virtual table" as a way of affording inclusivity, especially for those joining remotely. Our research builds on their observations of improving inclusiveness and equality in online meetings.

2.2 Prior approaches to hybrid meeting systems

Studies of partially distributed meetings have largely shown the challenges of supporting them well [18] or identified approaches for going beyond just the meeting time to always-on video portals [21]. Most approaches to creating a meeting UX for partially distributed meetings focused on creating matching physical meeting spaces to connect through a video call (optimized for connecting meeting rooms together) or using telepresence robots to physically integrate the remote participant into a meeting room (optimized for individual remote people joining an in-person meeting).

Commercial telepresence rooms, such as Cisco TelePresence Systems [9] and HP Halo [14], created a unified partially distributed meeting UX by carefully crafting meeting rooms to look identical and configured to connect matching semi-circular shaped tables into a seemingly contiguous round table. These high-end systems required carefully controlled construction at each connection point with the layout of displays, tables, cameras, and lighting, and were very expensive. O'Hara et al. [34] critically examined the HP Halo system for its "blended system" properties, where the spatial geometries of each local space continued across the distributed boundary into the remote space, thus naturally affording spatial conversational and interactional cues. Their research led to the BISi (Blended Interaction Spaces) prototype that focused on enabling a shared workspace around a tabletop workspace for shared content among small teams. This prototype still required constructing matching setups for each connection point. In today's context of meetings where people are largely joining remote meetings from their homes, we cannot expect to be able to remodel each person's home to fit a common physical layout.

Another hybrid meeting arrangement that has been studied in prior work is a hub and satellite configuration where one or two people remotely connect into a meeting room where the rest of the participants have gathered [43]. This research focused on giving the remote participants a "seat at the physical table" through a physical representation in the meeting room. Commercial telepresence robots, such as the Beam from Suitable Technologies or the Double from Double Robotics, enabled remote people to drive mobile video conferencing terminals to meeting sites and interact through video, audio, and movement. Studies of these systems showed how the telepresence robots enabled remote participants to physically interact with colleagues located at the main office, and especially supported informal communication and connections [25] and the hallway conversations before and after meetings [42]. While telepresence robots are effective ways of including individual remote workers into an in-person meeting, this approach requires effort from the user to drive and become familiar with the physical workplace, and does not scale well to large numbers of remote workers in hybrid meetings. While telepresence robots gave remote participants a strong sense of presence to the participants in the room, they are not great at integrating with video conferencing systems to give a strong sense of presence to other remote people connecting through conventional video conferencing tools. To other remote participants, telepresence robot participants are captured through the camera and microphone in the room, which means that their representation in the video conferencing tool is reduced to the smaller than life-size video display and audio coming from the speaker of the robot. Both of these streams are degraded compared to joining the video conferencing tool directly.

2.3 Spatial cues

Much of the research on improving video conferencing UXs focuses on providing more natural spatial cues [40]. One problem with the typical matrix of video windows, presented in most video conferencing tools, is that it is impossible to tell who is looking at whom, especially since each person's layout of video windows is different. A core design concept of our prototype explores sharing spatial cues among in-person and remote collaborators in a hybrid meeting.

MAJIC, an early prototype for video calling, created a virtual spatial layout of video conference participants by projecting life-size video images onto a curved screen [35]. They prototyped an arrangement among three remote participants where each person's perspective enabled them to distinctly and consistently detect eye gaze direction among the group members. Participants liked being able to make eye contact and sharing a sense of togetherness, but found that the prototype took up a lot of space and effort to configure and that seams in the background of the video detracted from the feeling of co-presence. MAJIC represents a line of research prototypes that looked to create a virtual sense of shared space among remote collaborators.

HERMES was an early prototype for spatially situating remote and in-person participants in a hybrid meeting [17]. They interspersed video displays of remote people with people sitting in the room around a circular table. This layout provided a physical location for each participant, regardless of whether they were in the room or remote. The remote participant's view was switched to focus on the currently active speaker, since it was not possible to get a view that included everyone sitting around a circle. Compared to a conventional representation of remote video participants on a flat plane at the front of the room, HERMES provided a sense of eye gaze and the ability to direct speech to those in the room. However, because the remote people had a different UX, they did not perceive benefits from the HERMES layout, pointing out the need to create a hybrid meeting UX that works for both people in the room and remote.

Yamashita et al. [47] studied the importance of seating position in prototypes that create a virtually shared space. The study involved two pairs of participants—one remote and the other local, either sitting across from each other, or adjacent to the local participants. The results showed that when remotes were interspersed with locals, there was more even distribution of turn-taking between remotes and locals, and a higher sense of unity. This study documents the importance of spatial arrangement, even when that layout is synthesized through communication technology.

Inkpen et al. [16] explored adding spatialized audio to the video compared with mono audio or no video. They found that people preferred spatialized audio along with the video, especially for tracking who was talking and managing concurrent talk. Recently, mainstream video calling systems have started incorporating spatial audio as a feature (e.g., Spatial Audio in Apple FaceTime), and we included spatial audio in our prototype.

Kantonen et al. [20] developed a mixed reality teleconferencing prototype that leveraged the Second Life virtual world. Their prototype inserted avatars of remote participants into a physical meeting room, which could be perceived by wearing see-through, head-mounted display glasses. Meanwhile, remote participants entered a virtual representation of the conference room, re-created in Second Life, and interacted with avatars representing people from the meeting room as well as other remote participants. This prototype enabled both in-room and remote participants to interact with spatially consistent representations of each other, whether they were in the room or in Second Life. Participants in the room had to wear see-through augmented reality glasses, and the room relied on sensing equipment to generate avatars of the people in the room that could be displayed in Second Life. We wanted to create a similar effect using conventional video and audio technology, without having to rely on additional augmented reality devices.

Virtual Reality (VR) potentially offers more flexibility in synthesizing spatial arrangements for meetings [31]. Facebook Horizon Workrooms has demonstrated a VR meeting UX that requires wearing a VR headset and is therefore unsuited to the case where multiple people are meeting together in a room. A pilot study compared video calling in Zoom with immersive VR (wearing a VR headset) and a browser portal view into VR, using Mozilla Hubs [41]. They found a higher sense of presence in the immersive VR condition wearing a headset, while the non-immersive view of VR through a browser was rated lowest in terms of spatial presence, involvement, and realism. Their study did not address a hybrid meeting where some people met together in person while others joined in VR.

Jones et al. [19] explored combining physical and VR presence in a hybrid meeting. They created VROOM (Virtual Robot Overlay for Online Meetings), which super-imposed a VR avatar over a telepresence robot. VROOM allowed those wearing an augmented reality headset to see a full body avatar at the position of a telepresence robot that could gesture and point in a way that was not possible using the telepresence robot alone. Furthermore, by mounting a 360° camera on the robot, the remote user driving the robot could also experience an immersive view of the space through a VR headset. Their experimentation identified opportunities and complexities of this approach,

hypothesizing that it revealed a telepresence "uncanny valley" where the additional simulation from the VR overlay actually detracted from the sense of belonging compared to the telepresence robot or even video conferencing. Their experiences are a cautionary tale when designing hybrid meeting layouts that combine physical and virtual aspects.

Krogh et al. [22] extended the sense of space through a notion of socio-spatial literacy. They pointed out that beyond physical interpretations of space, spatial understanding can be socially interpreted. They use video conferencing as an example where our sense of proxemics (nearness) to the other participants is shaped more by social interpretation of the video UX beyond just the physical layout. They suggest that socio-spatial literacy can be used in the HCI design of video conferencing tools to create a sense of spatial closeness that goes beyond the literal spatial layout of the UX. Perhaps one example of this is by Iitsuka et al. [15], which provides gaze cues not through spatial layout but through arrows and modifying window size. In our prototype, we explore using visual and audio cues to create a shared space for hybrid meetings to evoke social interactions around sharing that space.

2.4 Unified Backgrounds

Another core design concept that has been used to integrate in-person and remote participants in a hybrid meeting is providing a common background among them, as mentioned with commercial telepresence rooms. Research prototypes have also explored digitally creating a common space among local and remote participants in video meetings.

HyperMirror was an early prototype that combined video call participants into a shared video wall [30]. Rather than a metaphor of looking across from each other face-to-face, HyperMirror positioned each person beside each other, as if looking at each other reflected in a mirror. This novel UX enabled people to use full-body gestures and interact with each other as if they were side-by-side.

An alternative UX approach replaced the remote video caller's video background with the actual background of the viewer's room [32]. By removing the remote participant's background (using a Kinect depth camera), and merging the remote person's video with a video capturing the background of what is behind the viewer's display screen, a video call UX is created that situates the remote person in front of the physical background of what is behind the viewer's display. This prototype generates a sense of the remote person being within the same environment as the viewer.

The Wish You Were Here research prototype explored inserting a live video feed into a shared video background [44]. A remote person's video was inserted into the video preview of the mobile user's camera view, so that the two people would appear visually together in the same environment. In this UX, they could feel like they are exploring a museum together, and even enable the remote person to "interact" with some of the museum displays. A related concept has appeared in commercial video calling products such as Together Mode in Microsoft Teams [28] or Immersive View in Zoom [48]. In these views, the videos of all participants (with backgrounds removed) appear in front of a common digital background, as if in the same space together. Everyone sees the identical layout, including themselves, wherever they are inserted into the scene.

MirrorBlender extends the visual sharing beyond the background to also include screen sharing. Each participant sees the same thing (as if looking at everyone reflected in a mirror), but there is flexible control over positioning and layering of each video window, shared screen, or other visual artifact shared in the video call [13]. Sharing a common view and allowing each person to control the position and translucency of their video window enabled deictic pointing, enlarging and moving your video window to get attention, and other novel forms of interaction. While these interactions required some effort to position and adjust the visibility of video windows, it offered a lot of flexibility in how to configure their shared video workspace.

Taken together, these research prototypes explore video call UXs where participants feel together in a shared background. In our prototype, instead of showing all participants the same view, we created a virtual shared space and placed each participant so that they could see the space from their unique point of view.

3 DESIGNING A NOVEL HYBRID MEETING SYSTEM

Informed by learnings from prior research, we conducted an iterative design and prototyping process to explore a novel hybrid meeting system to:

(1) provide an inclusive and equitable interaction space among in-person and remote attendees
(2) support spatial cues to enable natural use of gaze awareness and gestures
(3) offer an enhanced sense of co-presence
(4) reduce meeting participants' cognitive load

Our design exploration resulted in us focusing on three key principles: 1) everyone is composited into a common scene; 2) there is a consistent spatial layout giving everyone a seat at the table; and 3) everyone has their own unique but spatially consistent view (or "perspective") from their seat at the table. We then explored how these principles could be leveraged to achieve the above stated goals for a novel hybrid meeting system.

3.1 Blending in-room and remote participants with digital compositing

To help reduce remote users' feelings of being isolated or ignored, we attempted to blend the in-room and remote experiences by digitally compositing video streams into a virtual meeting room to create a more equitable experience for everyone. For users in the conference room, the virtual meeting room was an extension of the physical meeting room, projected on the wall or a large display, and remote attendees appear in this virtual extension. Utilizing a projector that creates a life-size representation of the remote attendees facilitates the illusion that members are all equally present in the physical meeting room. For remote users, the virtual meeting room is where the meeting occurs: in-person attendees and other remote attendees appear together in the virtual meeting room. In this case the illusion is that everyone is in the virtual meeting room, blurring the line between remote and in-person attendees. Digitally compositing everyone into a virtual room provides a unifying shared scene, similar to HyperMirror [30], Wish You Were Here [44] and even commercial offerings such as Together Mode in Microsoft Teams [28].

3.2 Consistency of spatial layout

To support users' natural interactions and leverage spatial cues, we sought to leverage the consistent spatial arrangement demonstrated in MAJIC [35], HERMES [17], and even commercial telepresence rooms. We wanted to create a spatial arrangement digitally to avoid the effort and expense of having to carefully craft the physical environment at all sites. Within our virtual meeting room, all users have a distinct spatial location that is consistent for all participants. For example, if Alice is seated to the right of Bob, then everyone in the meeting sees Alice on Bob's right. Our virtual meeting room also uses artifacts such as plants and wall art to help reinforce distinct spatial cues.

3.3 First-person perspective

Rather than give everyone the same view into the meeting, we explored giving each person their own viewpoint into the common space, as they might have when sitting in a room together. In this way Alice views the room from her "seat" and Bob views the room from his "seat". While each person sees a different arrangement of their peers in the virtual room, it is spatially consistent. One

person might see Alice on the center of their screen, while another may see Alice on the far left or far right.

3.3.1 Video preview. Furthermore, as part of the first-person perspective, we also experimented with removing each user's video preview (also referred to as the self view or confidence monitor). Although users have become used to seeing a video preview, it is inconsistent with our interactions in the physical world where we do not also see of mirror of ourselves when talking with others. Recent studies of videoconferencing fatigue [8, 37] draw attention to how the video preview contributes to nonverbal overload [2] and, as an unnatural "constant mirror" [23], leads to excessive and often negative self-evaluation. Prior research has also explored how the video preview can hinder effective communication [6]. We hoped that the spatial cues provided by creating a common meeting space would restore some of the natural conversational and gaze cues that could obviate the need for a video preview and thereby help reduce the cognitive load of video calling.

4 PERSPECTIVES HYBRID MEETING PROTOTYPE

Based on our design principles, we developed a novel hybrid meeting system called Perspectives to support hybrid meetings where some meeting participants gather together in a meeting room and others join remotely.

4.1 Perspectives In-Room

Figure 1 shows the Perspectives in-room UX for a five-person meeting where two people are gathered in the room and three are joining remotely. The virtual room extends the appearance of the physical room to include the remote people within a shared common room. We included a virtual, semi-circular table that reflects the physical table that was in the room and the wall color of the virtual room echoed that of the physical room. The background was removed from the videos of the remote participants so that they could be placed in chairs around the virtual table giving each person a consistent space within the room. This interface was projected on the wall facing the participants in the room. Notice there is no video preview, so participants in the room do not appear in the UX. A Dolby Atmos sound system was used to provide spatialized sound for the voices of the remote participants correlated with their visual location.

To capture the isolated images and sound from the people in the room, we used a web camera dedicated to each participant and a head-worn microphone. This approach anticipates the availability of conference room cameras that can generate isolated video images of multiple people meeting together in a room, and microphones that similarly provide isolated audio of those people. Isolated audio and video streams enable flexible positioning of the people into the virtual room for the remote participants.

4.2 Perspectives Remote

Each remote participant receives a unique view of the virtual room from their perspective. Since all meeting participants are conceptually seated around a circular table, each remote person gets a specific view of the other participants and the surrounding room from their specific position around the table. Figure 2 shows the remote UX for the "middle" remote participant of the same five-person meeting, with the two in-room participants directly across and one remote person to the left and right of them. The virtual room background places all the participants in a common room with some "landmarks" such as wall art, potted plants, and doorways, that create a spatially consistent room. Not having a video preview helps preserve a sense of immersion in the virtual environment, similar to first-person perspective computer games. Headsets are used to deliver spatialized sound so that people's voices correlated with their visual location in the UX.

Fig. 2. Remote UX of the Perspectives prototype (on a 27" display)

5 USER STUDY

We designed a user study to compare the design features of our Perspectives prototype with current commercially available video conferencing layouts for hybrid meetings. We wanted to see if the Perspectives design provides a hybrid meeting experience which is more inclusive for remote and in-person attendees. We describe the layouts chosen to compare with Perspectives and the design of the study, including participant recruitment, tasks, and measures.

5.1 Comparing with three commercially existing video conferencing layouts

We wanted to compare Perspectives with people's current hybrid meeting experiences. We selected three modes currently offered in Microsoft Teams that are representative of hybrid meeting layouts in commercially available systems: Gallery, Together Mode, and Front Row. Table 1 summarizes the UX for each of the experimental conditions. Layout refers to how the video streams were composited in the scene (grid or linear). Common Background refers to whether or not users' backgrounds were removed from their videos and placed in a common scene. Video preview indicates whether or not a video preview was available. View coordination indicates how the views were coordinated (or not) across the different endpoints. Figure 3 shows the layouts for both in-room and remote participants for the four experimental conditions.

Condition	Layout	Common Background	Video Preview	View Coordination
Gallery	grid	no	yes	random
Together Mode	linear	yes	yes	identical
Front Row (in-room)	linear	yes	no	random
Front Row (remote)	grid	no	yes	random
Perspectives	linear	yes	no	consistent

Table 1. Comparison of the hybrid meeting UX for each experimental condition.

5.1.1 Gallery. Gallery is a conventional grid view of video streams that is used in many commercial videoconferencing systems (e.g., Microsoft Teams [29], Zoom [49], Google Meet [12], Cisco Webex [45]). Typically the grid of videos fills the display when there is no screen sharing. Each user has their own background, and everyone can see a video preview of themselves (or the conference room). In hybrid meeting situations, the conference room video is sometimes increased in size to make it easier to see the people in the conference room. Typically each person's grid view is

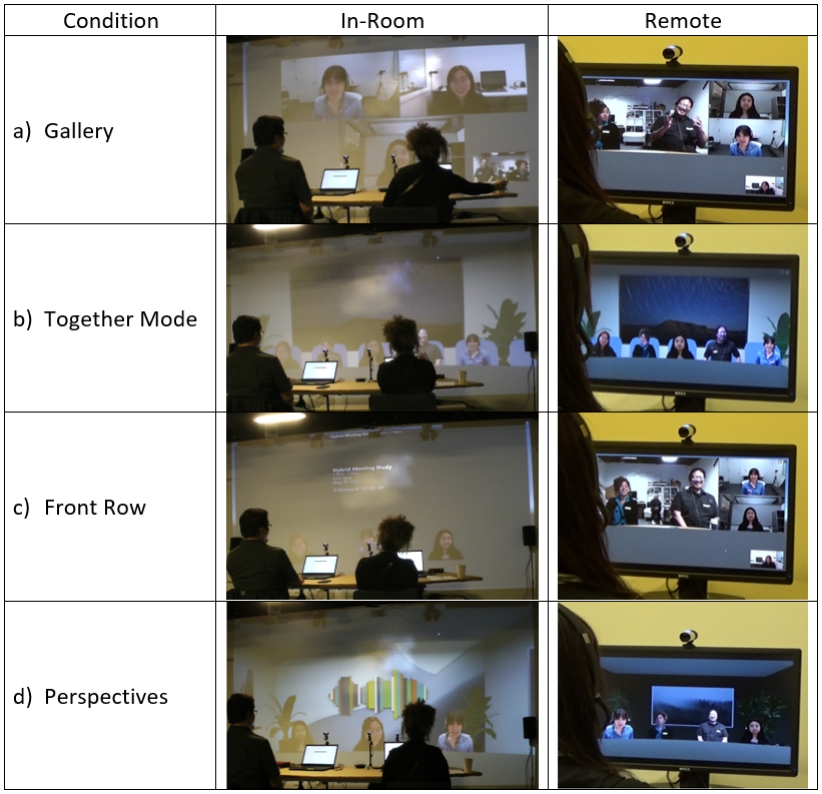

Fig. 3. Comparison of the UX displayed to people in the room and remotes for each condition.

different and the grid dynamically shuffles as people join, leave, or turn their video on or off. We included Gallery as it is the most commonly used layout in video calling systems.

Figure 3a shows our in-room implementation of Gallery which displayed the three remote video streams in a grid that filled the projected display. A video preview of the room was shown in the bottom right corner of the display. Our remote implementation displayed the room video stream, and the two additional remote participants in a grid view that filled the display. The grid layout was different for each remote user and the video stream from the room was doubled in width and height to better capture multiple people. A small video preview was shown in the bottom right corner. Audio was spatialized according to where the video streams appeared.

5.1.2 Together Mode. Microsoft Teams Together Mode [28] is an innovative UX that removes the background from all video streams and places the videos in a common, digital background, like a meeting room or lecture hall. The layout of video streams depends on the scene chosen. In Together Mode, the video streams are mirrored, which allows each person to naturally position their bodies in coordination with others. This approach builds on prior research such as HyperMirror [30] and MirrorBlender [13]. In Together Mode there is no separate video preview: everyone can see themselves within the Together Mode layout itself, although it is larger than the typical video preview and interspersed with the other participants. Additionally, all participants see the same view, thus giving everyone a common visual experience of the meeting. As people join or leave a call, or turn their video off or on, everyone's location in the interface dynamically shuffles. Although

everyone typically sees the same thing, there is no spatial correspondence as to where people's videos appear, especially for those meeting together in the room. We included Together Mode as a commercially available layout where everyone is placed in a common background.

Our implementation of Together Mode composited all five participants into a virtual scene so that they appeared to be sitting behind a table spanning the width of the display (Figure 3b). Since normal placement of users in Together Mode is random, in-room and remote participants were interspersed (i.e., not "seated" next to each other). For both in-room and remote participants, the Together Mode scene filled the display, and audio was spatialized according to where the video streams appeared.

5.1.3 Front Row. Microsoft Teams Front Row [27] is a UX explicitly designed to support hybrid meetings. It delivers a specialized UX for people in the meeting room only. Video streams of remote participants have their backgrounds removed and are spread horizontally along the bottom of a room display with a neutral background. The video camera is placed below the display to afford better eye gaze awareness. Remote participants have a standard Gallery view and there is no indication on either end that the UXs for in-person and remote participants are different. We included Front Row as a commercially available layout specifically designed for hybrid meetings.

Our in-room implementation of Front Row composited the three remote people along the bottom of the screen, with their background removed, placed on a plain, dark grey background (Figure 3c). No video preview was provided for the in-room participants. Our remote implementation of Front Row was a Gallery view with the videos shuffled so that they appeared in different locations than the Gallery condition. As in Gallery mode, a video preview was provided in the remote implementation. Audio was spatialized according to where the video streams appeared.

5.2 Study Implementation

For consistency, all four experimental conditions were implemented using the same prototype platform (a Unity application using Microsoft Teams NDI streaming capabilities), and focused only on the audio-visual layout for video calling. Other common video calling features such as screen sharing, text chat, or other reactions (e.g., hand raising, likes) were not included. While these important video calling features could be added to the overall layout, we wanted to focus the study on the impact of the differences of the visual layout. Additionally, the study only included five people who consistently stayed in the video call throughout their session. Therefore, we did not implement support for dynamic shuffling when people join, leave, or toggle their videos. Although spatialized audio was not available in Microsoft Teams at the time of the study, we expect it to be common in video conferencing soon, so we chose to include spatial audio in all four conditions.

5.3 Study Setup

5.3.1 Participant Recruitment. Recruitment of participants was done in compliance with our organization's IRB policies. We recruited 60 participants aged 18 years and older, with roughly even numbers of users who identified as man or woman (56.7% man and 43.3% woman), and the age distribution shown in Table 2. The participants included one person who is blind (who participated in the room) and one who is low vision (who participated as a remote). Inspired by prior research that demonstrated the value of evaluating teams of people who know each other [5], we recruited participants in groups of five people who were already familiar with each other. Given the COVID protocols in place at the time, eligibility to enter the building was limited so most participants were employees of our company. All participants signed a consent form to participate in the study and were given a $75 Tango gift card as a gratuity.

Age Range	Number of Participants
18-30	19
31-45	25
46-60	15
over 60	1

Table 2. Age Distribution of Participants

5.3.2 Hybrid Meeting Setup. We simulated a hybrid meeting by having all five participants come to the same building, and directing them to different rooms for the study. Two people were located in the same room outfitted as a conference room. The conference room had a curved table for the participants to sit at, laptops on the table to capture their videos, headworn microphones to capture each person's audio, and a ceiling-mounted projector to project a large display on the wall in front of the table. A Dolby Atmos 7.1 audio system was used to spatialize audio in the room, and each remote users' audio was spatialized according to their visual placement on the large display.

The other three participants were individually located in separate offices outfitted as home office workspaces. One workspace utilized a laptop for the remote endpoint with a headset for audio; one workspace utilized a standard desktop setup with a 27" monitor, and a headset for audio; and the third workspace utilized a desktop setup with a large wide-screen monitor, and a headset for audio. Video cameras were placed at the back of each room to record the sessions, and the conference room was used for an end-of-study discussion with the group.

5.4 Study Design

Our study utilized a mixed design with experimental condition as the within-subjects variable and room-versus-remote as the between subjects variable. This design enabled all groups to experience all four experimental conditions; however, each person in the group only experienced one endpoint. A basic Latin square design was used for counterbalancing the four experimental conditions (Gallery, Together Mode, Front Row, Perspectives), repeated three times, for a total of 12 groups (see Appendix A.1). At the beginning of the study, participants were told that they would participate in four videoconferencing setups, but were not given details on the different configurations.

5.4.1 Tasks. At the beginning of the study, participants completed an online survey comprising demographic questions and questions related to their prior experiences with video calling tools. At the start of each condition, participants completed two tasks to familiarize themselves with the condition. First, we asked them as a group to count off to 30, and the study moderator designated the first person to start. The group was not given time to discuss or plan how they would do this task, so they needed to organize themselves on-the-fly into some order to facilitate the counting. The groups were then asked to discuss an ice-breaker question (e.g., "What did you eat for breakfast today?") for two minutes making sure that each person had an opportunity to speak.

Upon completion of the warm-up tasks, the moderator gave the group a scenario to discuss and come to an agreement on as a group. For example, "plan a group dinner for your larger team, keeping in mind any dietary constraints such as vegetarian or gluten-free, and working out what the set menu would be for appetizers, entrees, desserts, etc." The group discussion tasks were designed to elicit opinions from all participants and negotiate a shared decision (see Appendix A.2). A timer was set for five minutes to prompt the group to finish up in that amount of time. After they completed the task, their display was turned off and the participants completed an online survey to collect their opinions about the condition they just finished. The participants were not allowed to

discuss with each other while completing the online survey. This cycle was repeated until all four conditions were completed.

Throughout the study, the moderator would join the meeting through the in-room endpoint and verbally give the participants instructions such as beginning a new prompt, wrapping up a conversation, or switching to the online survey. Upon completion of the final condition and online survey, the group met in the conference room and participated in a short, semi-structured feedback session with the moderator. The entire study took between 90 and 110 minutes.

5.5 Measures

We collected a variety of measures to compare among the conditions. The online survey after each condition asked a number of Likert scale questions about users' sense of co-presence, conversation flow, ease of turn-taking, amount of inappropriate interruptions, perceived comfort, preference, sense of group agreement and their opinion on the presence or absence of the video preview. Participants were also asked what they liked best about the condition, and what they felt should be improved. Over-the-shoulder video recordings were captured in each room, as well as isolated audio recordings of each participant.

We collected several measures to try to identify which conditions would cause the participants to expend less (or more) effort and assess participants understanding of the spatial layout of people in the meeting. The first measure was a self-report where participants rated their perception of effort compared to typical meetings, on a 5-point scale (1=much less effort to 5=much more effort). The second measure of effort / cognitive load involved a sketching exercise. After each condition (when the visual layout of the condition was no longer visible), one question in the online survey directed each participant to sketch the visual layout from memory, including who was located in each position. The sketches were evaluated for any errors, such as people missing, located in the incorrect position, or the layout being wrong. The number of people in each group with sketch errors was counted for each condition.

The third measure of effort / cognitive load was the warm-up counting task. As previously described, at the beginning of each condition, groups were asked to count off to 30. Some layouts, such as Perspectives and Together Mode, presented a visual order for counting off, whereas Gallery had no sense of order, and Front Row presented a visual order for those in the room but not for those joining remotely. Our hypothesis was that groups who could quickly form a mental model of the spatial layout of people sitting around the table, would be faster at the counting task. Counting to 30 meant that each group would have six rounds of counting through the five people in the group. We measured the time it took to complete each round of counting and included rounds 2-6 in our analyses. We chose to omit the first round when participants were figuring out an order. A round was defined as including the transition time from completing the previous round and starting a new round (i.e., from 5 to 6) up through completing the round (i.e., after counting to 10).

After all four conditions were complete, participants were asked to rank order the conditions on a scale of 1 to 7, where 1 was most preferred and 7 was least preferred. They were asked to use the entire scale, and were not allowed to give the same rating to more than one condition, as governed by the online survey. This measure provided their relative preferences among the conditions, including a sense of which conditions felt similar as well as those about which they had strongly different opinions. Following this, the group gathered together in the conference room for a brief semi-structured group interview to explore their preferences about the conditions and comments about the video preview together as a group. Thematic analysis was used to summarize the common themes that emerged in the interviews.

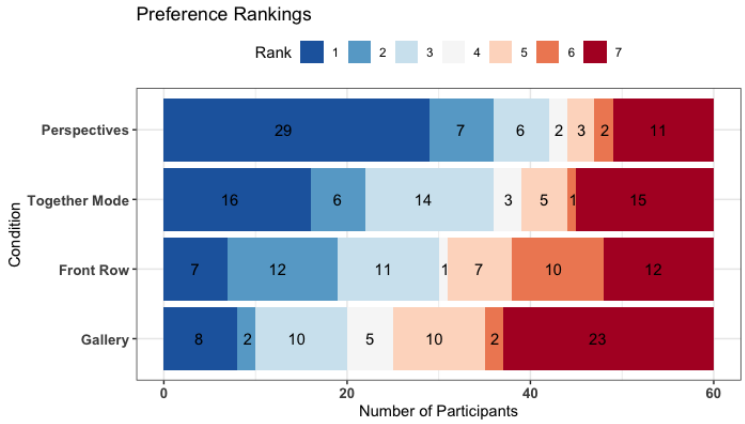

Fig. 4. Rank ordered preference of each condition from 1 (most preferred) to 7 (least preferred)

6 RESULTS

Throughout the paper, participants are referred to by their group number and letter designating their location: A, B were located in the conference room, and C, D, and E were the remote endpoints. Thus, P5B would be a participant in the fifth group sitting in the conference room, whereas P9D would be a participant in the ninth group joining remotely.

6.1 Preferred layout

Result 1a: Perspectives was ranked highest in terms of preference.
Result 1b: Whether users were in-room or remote affected their preferences.

Participants' preference rankings of all four conditions is shown in Figure 4. A non-parametric Aligned Rank Transform analysis [46] was run with *Condition* as the within-subjects variable and *Room-vs-Remote* as the between-subjects variable. The results revealed: 1) a significant main effect of *Condition* ($F_{3,174} = 5.87, p < .01$); 2) no significant main-effect of *Room-vs-Remote* ($F_{1,58} = 0.31, p = .58$); and 3) a significant interaction effect between *Condition* and *Room-vs-Remote* ($F_{3,174} = 3.69, p = .01$). Overall, Perspectives was rated highest, with 48% of participants rating it first or second; Gallery was rated lowest, with 38% of participants rating it last; and ratings for Together Mode and Front Row typically fell between Gallery and Perspectives.

6.1.1 In-room versus Remote. The significant interaction effect between *Condition* and *Room-vs-Remote* indicates that users' preferences were affected by whether they were sitting in the room or joined remotely. Figure 5 shows the in-room and remote preference rankings for each condition, with rankings grouped according to positive (1-2), neutral (3-5) and negative (6-7). Examination of the rank data combined with the qualitative data reveal several interesting trends which will be discussed next.

Perspectives. Overall, Perspectives was the highest ranked condition for both in-room and remote participants. The most common reasons given for preferring Perspectives were that 1) it felt the most natural; 2) it provided a consistent layout; and 3) it gave people a sense of being together in the same room (see Table 3). In-room participants ranked Perspectives significantly higher than Gallery ($p < .01$). Remote participants ranked Perspectives higher than both Gallery ($p = .09$) and

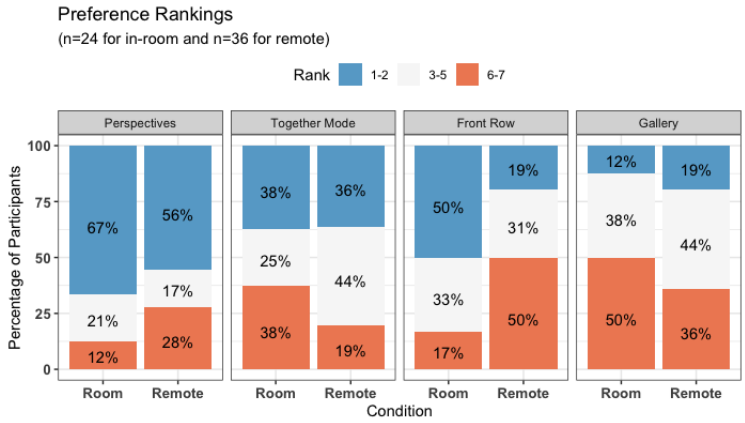

Fig. 5. Preference rankings for each condition according to being in the room or remote

Front Row ($p < .01$), which makes sense since the remote layout for both of those conditions is equivalent.

A larger percentage of remote participants (28%) rated Perspectives negatively compared to in-room participants (12%), which suggests that Perspectives was received differently in-room versus remotely. Comments from the surveys and interviews suggest that the lack of a video-preview in Perspectives contributed to some of the negative ratings, which was especially true for remote participants (*"Not seeing myself sucked"*, (P6E), and *"Perspectives was nice as well (better than classic meeting remote UI) but I was not able to see myself"*, (P7E)). There also seemed to be a difference in whether users felt that they were all "sitting around a table". In-room, the visual effect of extending the room onto the large display seemed to support the notion that all of the participants were sitting around the table. However, for remote participants, several expressed that they felt like they were on one side of the table, and the other participants were on the other side. As one participant commented: *"I did not like how I felt a little more separate from the group- almost like I was 'presenting' to them."* (P3C).

Together Mode. Overall, Together Mode was ranked lower in terms of preference than Perspectives, and higher than Gallery, however these differences were not statistically significant. The most common reasons given for preferring Together Mode were 1) it gave people a sense of being together in the same room; 2) it felt natural; and 3) it provided a video preview (see Table 3.) Interestingly, whether people were in-room or remote had little effect on the number of participants who ranked Together Mode positively; however, a higher percentage of participants ranked Together Mode negatively when they were in-room. Evidence from the qualitative data suggested that seeing themselves life-size on the screen when in the room contributed to this effect: *"I did not like seeing myself on the screen with everyone else"* (P2A). Additionally, in the room there were two representations of the in-room participants, their physical body, and a digital representation of them on the screen, which sometimes seemed to cause confusion for the in-room participants (especially during the counting task).

Front Row. The results for Front Row were strikingly different, depending on whether the participants were in-room or remote. This is not surprising since the participants in the room had a different layout compared to remote participants. In-room, the Front Row condition was ranked positively, was comparable to Perspectives and Together Mode, and was ranked higher than Gallery

Condition	Reason	Mentions	Participant Quotes
Perspectives	Natural	16	*"I enjoyed Perspective because it feels more natural and immersive." (P8E)*
	Consistent Layout	6	*"Perspectives and Together Mode made it very easy to understand where to look and what the experience for others was." (P6C)*
	Feeling Together	5	*"I liked the feeling of being physically in a room, it makes me feel more focused and engaged in the conversation and on others" (P8D)*
Together Mode	Feeling Together	9	*"Prefer everyone being together."(P6A) "I prefer the together mode as it felt like people were in the same room." (P7D)*
	Natural	5	*"Turning and looking toward the active talker was more natural in the linear layout (Together) mode."(10B)*
	Video Preview	5	*"Together mode is nice as it felt natural, and I could see myself." (P7E)*
Front Row (in-room)	Plain Background	2	*"I liked Front Row because everyone felt like they were in the same room with no background distractions." (P2B)*
Front Row (remote)	Video Sizes	2	*"I chose "Front Row" because it keeps larger videos for remote participants, but the in-room view is to scale" (P12C)*
Gallery	Familiar	6	*"The gallery is the one I'm the most comfortable with"* (P5B)

Table 3. The top reasons participants preferred each of the conditions, how many participants mentioned that reason, and an example quote.

($p = .09$). For remote participants, Front Row was ranked poorly. Seven people in the room rated Front Row as their first choice: their top reasons were 1) they felt the grey background was less distracting (in-room participants); and 2) they like the non-segmented videos as they were larger (remote participants), (see Table 3). Even though participants were not explicitly told the differences between the in-room and remote layouts, some participants mentioned noticing differences in their group's behavior: *"As a non-room participant you could tell that something was just a little different there and it distracted/etc. from the call."* (P6E)

Gallery. Gallery was ranked lowest, regardless of whether the participants were in-room or remote. This was especially true in-room where the participants did not like having the grid on the large display. The main reason given for why people preferred Gallery was because it was familiar (see Table 3.).

6.2 Spatial visual layouts and improved co-presence

Result 2: All three spatial layouts (Perspectives, Together Mode, Front Row (in-room)) significantly improved co-presence over Gallery.

Participants were asked to rate their feelings of co-presence during the study: "I felt like I was in the same space as the participants" on a 5-point Likert scale (1=strongly disagree to 5=strongly agree) (see Table 4). A non-parametric Aligned Rank Transform analysis with *Condition* as the within-subjects variable and *Room-vs-Remote* as the between-subjects variable revealed two significant main effects, and a significant interaction effect. *Condition* ($F_{3,174} = 15.99, p < .001$); *Room-vs-Remote* ($F_{1,58} = 4.64, p = .03$), and *Condition* X *Room-vs-Remote* = ($F_{3,174} = 10.78, p < .001$). Given the significant interaction effect, we examined post-hoc comparisons using pairwise wilcoxon rank sum test for in-room versus remote separately. For in-room participants, the three conditions that

	Perspectives	Together Mode	Front Row	Gallery
in-room	3.83	3.33	3.46	**2.21**
remote	3.58	3.72	**2.33**	**2.75**

Table 4. Mean co-presence ratings across conditions, for in-room and remote participants, on a 1 (strongly disagree) to 5 (strongly agree) scale for: "I felt like I was in the same space as the participants". Gallery as well as Front Row (remote) were rated significantly lower on co-presence.

	Perspectives	Together Mode	Front Row	Gallery
in-room	**2.08**	2.75	2.63	2.92
remote	**2.31**	**2.39**	2.81	2.67

Table 5. Perception of cognitive effort, for in-room and remote participants, compared to typical meeting on a 1 (much less effort) to 5 (much more effort) scale.

had spatial views (Perspectives, Together Mode and Front Row) were rated significantly higher than the Gallery condition ($p < .01$). For remote participants, Perspectives and Together Mode were rated significantly higher than both Gallery and Front Row ($p < .01$). Qualitative data from participants also supported these results with participants in the Perspectives, Together Mode, and Front Row (in-room) all commenting on feeling like they were together:

- [Perspectives] *"I liked the feeling of being physically in a room, it makes me feel more focused and engaged in the conversation and on others"* (P8D)
- [Front Row] *"It felt like everyone was in the same room as me and felt like a shared experience"* (P3B)
- [Together Mode] *"I liked when it seemed like we were all in the same room and the conversations were flowing better"* (P5C)

These results indicate that the spatial layouts evoked stronger feelings of co-presence than the Gallery layouts. This result is important given that Gallery is still the most common experience for video calls (hybrid or otherwise).

6.3 Perceived Effort and Performance

Result 3: Perspectives was rated easiest for in-room participants while Perspectives and Together Mode were rated easiest for remote participants. Participants also had fewer errors in the sketching task and completed the counting task faster using Perspectives, which suggests users had a stronger mental model using Perspectives which contributed to it being less effortful.

Examining participants' perception of effort compared to typical meetings, a non-parametric Aligned Rank Transform analysis with *Condition* as the within-subjects variable and *Room-vs-Remote* as the between-subjects variable revealed a significant main effect for *Condition* ($F_{3,174} = 5.00, p < .01$). No main effect was found for *Room-vs-Remote* ($F_{1,58} = 0.16, p = .68$) but a significant interaction effect was found between *Condition* X *Room-vs-Remote* = ($F_{3,174} = 3.53, p = .016$), indicating that the level of effort for each condition was impacted by whether or not the participants were in-room or remote. Although the post-hoc pairwise comparisons were not significant, we do see that on average, Perspectives and Front Row were easier in-room than remote; and that Together Mode and Gallery were easier when remote compared to in-room (see Table 5). Overall, in-room, we see that Perspectives was rated easiest while both Perspectives and Together Mode were rated easiest for remote participants.

	Perspectives (room/remote)	**Together mode (room/remote)**	**Front row (room/remote)**	**Gallery (room/remote)**
Number of Errors	**2 (0/2)**	13 (7/6)	9 (6/3)	10 (5/5)
Round time (mean)	**5.60s**	6.53s	5.95s	6.37s

Table 6. Measures of cognitive effort according to: a) number of errors in sketching the layout from memory after each condition, and b) average time needed to complete a round in counting off to 30.

Table 6 shows the number of errors per condition for the sketching exercise, split by whether they were made by in-room or remote participants. A repeated measures ANOVA revealed a significant main effect of condition ($F_{3,177} = 2.62, p = .05$). Post-hoc pairwise comparisons revealed that participants made significantly fewer sketch errors in the Perspectives condition, compared to all other conditions (Together Mode, Front Row, Gallery) ($p < .05$).

Table 6 shows the average round time for the counting task for each condition. The data were analyzed using a repeated measures ANOVA with *Round*, and *Condition* as the within-subjects variables. A significant main effect was found for *Round* ($F_{4,217}, 3.91, p < .01$) as groups typically increased in speed as the task progressed. A main effect of *Condition* was also found ($F_{3,217}, 5.16, p < .01$), but no interaction effect was found. Post-hoc pairwise analyses of *Condition* revealed that groups were significantly faster in the Perspectives condition than in Gallery (p=.02) or Together Mode (p<.01) which suggests that they had a stronger mental model of the group's spatial arrangement in the Perspectives condition.

The sketching and counting tasks provide evidence that participants were able to more readily develop a mental model in Perspectives compared to the other conditions, and that Perspectives may have a lower cognitive load than the other conditions. This may explain, in part, why users reported that they expended less cognitive effort in Perspectives.

6.4 Diverse opinions about video preview

Result 4: Preference for video previews was highly variable across participants and conditions.

The presence or absence of video preview evoked strong and diverse opinions, and was often a factor in which condition participants preferred. Some participants strongly preferred not having a preview, since they felt it enabled more natural interaction. Others missed the presence of the video preview, especially since they have become accustomed to always having a preview. The following quotes, with the condition they preferred [in brackets], illustrate the range of opinions observed in our data:

- [Perspectives] It felt the closest to reality, but I would still have liked to see a preview of myself (P11E)
- [Perspectives] I didn't like the one's [sic] where I see myself. Like in real life, I like the one's [sic] I only see others. Then I don't need to focus on myself. (P1A)
- [Perspectives] I liked not having the video preview. Made me look at other people vs. looking at myself. (P2C)
- [Together Mode] is nice as it felt natural, and I could see myself. Perspectives was nice as well… but I was not able to see myself. I think best option is Perspectives but with the ability to see video preview. (P7E)
- [Perspectives] I felt better about not seeing myself because it wasn't a preview that was distracting anywhere. Although I prefer to be able to view myself every once and a while to check. (P3C)

- [Gallery] I liked the prototype video arrangements in which I could see myself, because I think I've already been habituated to this: it's convention. (P5A)

One interesting pattern that emerged with Perspectives was that people who were in the room tended to not miss the preview as much as those joining remotely. Sharing the room with another person seemed to evoke a sense that they were being heard and seen properly, which may have lessened the need for a video preview. Conversely, those joining remotely needed more reassurance about how they were being portrayed. Eliminating the video preview is also a substantive change that may require more than one session to develop an informed reaction.

6.5 Role of spatial audio

Result 5: Spatial audio was highly valued for contributing to more natural conversation and making overlapping talk easier to interpret.

Spatial audio was integrated into all four conditions anticipating that this would soon become a standard feature in video calling. Spatial audio was often mentioned as a valued feature:

- Knowing where to look based off where sound was coming from (Perspectives P6C)
- The layout matched with audio was great. Easy to track and hear. No overlapping. (Together mode P1C)
- I liked that I was able to follow the conversation around very naturally just with the sound of voices (Front row P6E)
- The spatial audio makes it much easier for us to have a natural conversation. Multiple people were able to talk over each other and still be understandable (like a real conversation). (Gallery P1E)

While all the conditions offered spatial audio, the layouts that spread videos apart horizontally (i.e., Perspectives, Together Mode, and Front Row in the room) offered a more distinct spatial audio separation than layouts that involved stacking video streams vertically (Gallery). Yet, P1E's comment above, which was made when experiencing the Gallery layout, indicates how spatial audio still helped users understand overlapping talk more easily. Overall, participants appreciated the benefits of spatialized audio, especially when used in concert with the visual layout to create distinct spatial locations for each participant.

7 DISCUSSION AND REFLECTION ON DESIGN PRINCIPLES

The results of our user study validated the design principles featured in Perspectives, achieving many of our outlined goals. Overall, users in our study generally preferred Perspectives because they found that it felt more natural, enhanced their feeling of co-presence, and reduced cognitive load. Based on the results of our study, we reflect on our design principles and discuss other design suggestions.

7.1 Digital compositing to blend in-room and remote

Results from our study suggest that digitally compositing users into a virtual meeting scene was beneficial as it helped create a more inclusive and equitable meeting space, and enhanced users' sense of co-presence. However, many of the results were stronger for the conference room than for the remote endpoints. This suggests that specific implementations of this principle can impact its benefit, and tuning the visuals based on the endpoint may be required. For example, using different table shapes and sizes, more sophisticated 3D models of the room, and 3D avatars to match the orientations of peoples' representations around the table (e.g., frontal or at some degree of profile) could lessen the sense of facing a panel of examiners and create more of a sense of sitting together around a table.

Another suggestion is to take into account the placement of video cameras with respect to where people are displayed in the scene. Our design of Perspectives displayed people's videos along the bottom edge of the screen, situated along a table to capture important spatial cues. This design worked well (for gaze awareness) in the conference room when the video cameras capturing the viewers' images were located near the bottom of the screen. However, typical remote endpoints such as laptops and desktop setups have video cameras positioned at the top of the display. A better design for these remote endpoints may be to locate the video streams higher up on the display to enhance gaze awareness.

Compositing also needs to take into account other meeting content such as chat or shared content which is often used in meetings. Perspectives has space in the scene for this content (e.g., above the video streams) but it may require reducing the size or slightly repositioning users' video streams. These adjustments would enable all of the meeting information to be embedded into the virtual meeting room and is consistent with shifting the focus from the people to the shared content.

7.2 Consistency of Spatial Layout and First-Person Perspective View

Results from our study validated that providing a consistent spatial layout helped reduce users' cognitive load given that Perspectives, Together Mode, and Front Row (in-room) improved users' sense of co-presence. The way that Perspectives uniquely added the first-person perspective view enhanced the experience more, with users preferring this layout and feeling that it created a more natural experience. These two features also helped reduce users' cognitive load as evidenced by users' improved performance in the sketching and counting tasks. One observation however was that it sometimes took time for users to fully understand the spatial layout of "people around the table". Although not examined in our user study, the Perspectives design includes an interface for displaying an overhead view of the meeting room at the start of a meeting and allowing people to select a seat. An animated transition shifts from the aerial view of the room to the user's perspective view in the meeting from that selected seat. We believe this transition will help establish the spatial layout of people in the room, as well as users' perspective views, and help orient everyone relative to each other.

We believe that Perspectives leverages people's familiarity with the spatial layout of a room to evoke smooth coordination of interaction with each other in a socially translucent system [10]. The consistent spatial layout and virtual room environment makes the necessary social information visible to enable smooth turn taking and a robust sense of awareness of each other (evidenced in the accurate sketches in recall). While our study looked at groups who were already familiar with each other, these affordances might be even more important in groups that are just getting to know each other.

7.3 Video Preview

Our design choice to remove the video preview in Perspectives had mixed results. Some people loved it and felt that it was more natural and less distracting, and commented that it contributed to them choosing Perspectives as their favorite condition. Others however hated the lack of video preview, saying that they felt very self-conscious that they could not see how they were being portrayed in the scene, and that it contributed to them choosing a different layout as their favorite. These results demonstrate that choosing whether or not to have a video preview, and how to represent it is a complex problem. Given the highly variable results in our study, the current design guidance should be to give the user control over whether to show video previews. Since people have become accustomed to having a video preview, and many do not fully trust whether the camera is capturing them correctly, it suggests that some reassurance of how people are being captured is also needed, especially when joining alone remotely.

7.4 Other Design Learnings

The study also found design implications for the other commercially available videoconferencing systems. In almost all measures, participants ranked any of the conditions as significantly better than the default Gallery mode. Thus, setting any of these layouts as the default for hybrid meetings may generate a better experience. Additionally, the asymmetric layout of Front Row, where people in the room got a spatial layout in contrast to the default Gallery experience for remote participants, resulted in significantly higher ratings in the room compared to remote. This difference suggests an opportunity to improve the overall Front Row experience by revising the remote experience to leverage some of the benefits experienced by those in the room. Together Mode also enjoyed some of the benefits of a spatial layout, although that could be improved by more intentionally locating each person's position in the spatial layout, rather than randomly locating them in the space and allowing them to be dynamically moved around due to others' actions in the meeting.

Our study also identified how being in a room with others, contrasted with joining remotely alone, affected how the different visual layouts were perceived. For example, when using Perspectives, in-room participants were less concerned about the lack of video preview than remote users, suggesting that sitting beside someone helped reassure them that they were being captured appropriately or that they felt that they were part of the meeting without additional video cues. Additionally, in Together Mode, users struggled more when they were in the room with someone else, as it created two different representations of the other person (a physical one and a digital one). This duplication caused confusion and made interactions more awkward. For example, people had to decide whether to look at the person seated beside them in the room or at the digital representation of them on the screen. These observations raise a more general point that the reactions to the conditions did not depend only on the layouts, but also the context in which they were being viewed. Layout designs may need to account for contextual differences such as participating in a room together with other people or joining remotely as an individual. While video-mediated collaboration research has historically aimed to rectify the fragmented and asymmetrical problems of video meetings[26], we should not be afraid to explore asymmetrical layouts and features for local and remote participants if they provide clear inclusivity benefits while evoking consistent, relative spatial awareness.

8 LIMITATIONS

While our study showed the promise of the Perspectives prototype in supporting hybrid meetings, it is still early in both the development of Perspectives and the evolution of hybrid meeting practices. We want to acknowledge several limitations in this early study.

Our study did not include a comparison of screen sharing, text chat, reactions, or other meeting tools which have become important features in commercial videoconferencing systems. We wanted to first test the visual layout of Perspectives as its most distinctive feature to discern whether the Perspectives approach was promising enough to merit further development. While digital screen sharing could be readily accomplished by reserving some space in the UX to display shared content, integrating digital and physical shared objects is a more ambitious challenge for the future. Our study also did not exercise some of the advantages of Perspectives around consistent spatial positioning of each participant. Unlike the other conditions, which would have shuffled or changed the layout of participants' as they joined, left, or toggled their videos on/off, people's position in the Perspectives layout remains constant once they have joined a meeting. We have not exercised how this stability could increase the benefit of Perspectives' consistent spatial layout in real meetings.

The meeting size (five participants), configuration (2 in-room, 3 remote) and type (brainstorming) used in our study were chosen to represent common scenarios in real work environments but more

research is needed to understand how well the features in Perspectives support different kinds of meetings as well as different sizes and group configurations. However, we do feel that our results are fairly representative for a broad class of meetings as data show that the vast majority of video conferencing meetings involve six or fewer people[7], and a brainstorming meeting like the one used in our study represents a common meeting type where anyone can contribute to the meeting. Future work should explore larger meetings, where everyone cannot be represented on the display but selective viewing can be leveraged, just as in large in-person meetings today. Meetings with clearly defined roles, such as classrooms, would also suggest a different hybrid UX. Additionally, it will be beneficial to understand how users' opinions of Perspectives change as they become more comfortable with the system over time. Our future work will also explore other measures of participation, such as how much time each person spoke, to see how these quantitative measures complement the qualitative survey responses reported in this paper.

COVID-19 challenges affected many aspects of running the study including limitations on recruitment to people within our company. Although this may have introduced bias into our sampling, we believe that our results are still valid and can be generalized to information workers. Furthermore, since we brought the entire group into the same site for the purposes of the study, we did not experience other boundary crossing which can occur in hybrid meetings, such as including people from different sites or even cultures. More research is needed to explore how the shared space metaphors of Front Row, Together Mode, and Perspectives interact with crossing site and cultural boundaries.

9 CONCLUSIONS

As we continue to experience changes in our work patterns brought on by the pandemic, the rise in hybrid meetings seems to be one practice that could be long-lived. The Perspectives prototype is an early attempt to create a more inclusive and equitable hybrid meeting experience among in-room and remote participants. Our study shows advantages of using digital compositing, consistent spatial positioning, and first-person perspective views to create a better hybrid meeting experience, as well as ways of improving on the concept. While we are still early in developing work practices around hybrid meetings, we have an opportunity to modify the designs of our current video meeting tools to create a more satisfying hybrid meeting experience.

Given the growing flexibility in work environments, we need a much richer and flexible array of options to fully embrace an inclusive, hybrid future. Our results suggest that implementations should be developed taking into account the context of use, so an important next step is to examine a broader spectrum of endpoints. This should include endpoints that are both highly constrained (e.g., mobile devices) as well as those with rich capabilities (e.g., 3D virtual reality environments). Understanding how to envision the design principles in these environments, and exploring integration across a diversity of endpoints will yield more inclusive solutions. Broadening the use of multimodal approaches to support our design principles could be especially promising for constrained environments and for making hybrid meetings accessible to people of all abiltiies.

ACKNOWLEDGMENTS

Extreme thanks for assistance with this research goes to: Robin Moeur, Pat Sweeney, Payod Panda, Jason Falcon, Aaron Garner, Bala Kumaravel, Pete Ansell, Richard Hughes, Greg Baribault, Sam Albert, Mike Messer, Mar Gonzalez-Franco, Eyal Ofek, Jenny Dileo, Marcus Ash, Kunal Gupta, the MSR AV/Media team, Rico Malvar, and Ece Kamar. We greatly appreciate the anonymous participants who came in for the study under the uncertain COVID conditions at the time.

REFERENCES

[1] Cevat Giray Aksoy, Jose Maria Barrero, Nicholas Bloom, Steven J. Davis, Mathias Dolls, and Pablo Zarate. 2022. Working from Home Around the World. https://doi.org/10.3386/w30446

[2] Jeremy N Bailenson. 2021. Nonverbal overload: A theoretical argument for the causes of Zoom fatigue. *Technology, Mind, and Behavior* 2, 1 (2021), 1–6.

[3] Nicholas Bloom, Ruobing Han, and James Liang. 2022. How Hybrid Working From Home Works Out. https://doi.org/10.3386/w30292

[4] Nathan Bos, N Sadat Shami, Judith S Olson, Arik Cheshin, and Ning Nan. 2004. In-group/out-group effects in distributed teams: an experimental simulation. In *Proceedings of the 2004 ACM conference on Computer supported cooperative work*. 429–436.

[5] Nathan D Bos, Ayse Buyuktur, Judith S Olson, Gary M Olson, and Amy Voida. 2010. Shared identity helps partially distributed teams, but distance still matters. In *Proceedings of the 16th ACM international conference on Supporting group work*. 89–96.

[6] Jose Eurico de Vasconcelos Filho, Kori M Inkpen, and Mary Czerwinski. 2009. Image, appearance and vanity in the use of media spaces and video conference systems. In *Proceedings of the ACM 2009 international conference on Supporting group work*. 253–262.

[7] Debbie Chew, Mahsa Azizi. 2022. The state of video conferencing 2022. https://www.dialpad.com/blog/video-conferencing-report/.

[8] Nicola Döring, Katrien De Moor, Markus Fiedler, Katrin Schoenenberg, and Alexander Raake. 2022. Videoconference Fatigue: A Conceptual Analysis. *International Journal of Environmental Research and Public Health* 19, 4 (Feb. 2022), 2061. https://doi.org/10.3390/ijerph19042061

[9] engadget. 2006. Cisco's TelePresence Meeting does video meetings in ultra-HD. https://www.engadget.com/2006-10-23-ciscos-telepresence-meeting-does-video-meetings-in-ultra-hd.html/.

[10] Thomas Erickson and Wendy A Kellogg. 2000. Social translucence: an approach to designing systems that support social processes. *ACM transactions on computer-human interaction (TOCHI)* 7, 1 (2000), 59–83.

[11] Denae Ford, Margaret-Anne Storey, Thomas Zimmermann, Christian Bird, Sonia Jaffe, Chandra Maddila, Jenna L Butler, Brian Houck, and Nachiappan Nagappan. 2021. A tale of two cities: Software developers working from home during the covid-19 pandemic. *ACM Transactions on Software Engineering and Methodology (TOSEM)* 31, 2 (2021), 1–37.

[12] Google Meet Help. 2022. How to view people in Google Meet. https://support.google.com/meet/answer/9292748.

[13] Jens Emil Grønbæk, Banu Saatçi, Carla F Griggio, and Clemens Nylandsted Klokmose. 2021. MirrorBlender: Supporting Hybrid Meetings with a Malleable Video-Conferencing System. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–13.

[14] Hewlett Packet Development Company. 2008. HP Halo Collaboration Meeting Room. https://www.hp.com/hpinfo/newsroom/press_kits/2008/halo/ds_halo-meetingroom.pdf/.

[15] Rikuto Iitsuka, Ikkaku Kawaguchi, Buntarou Shizuki, and Shin Takahashi. 2021. Multi-party Video Conferencing System with Gaze Cues Representation for Turn-Taking. In *International Conference on Collaboration Technologies and Social Computing*. Springer, 101–108.

[16] Kori Inkpen, Rajesh Hegde, Mary Czerwinski, and Zhengyou Zhang. 2010. Exploring spatialized audio & video for distributed conversations. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work*. 95–98.

[17] Tomoo Inoue, Ken-ichi Okada, and Yutaka Matsushita. 1997. Integration of face-to-face and video-mediated meetings: HERMES. In *Proceedings of the international ACM SIGGROUP conference on Supporting group work: the integration challenge*. 405–414.

[18] J Johnson. 1999. A field study of partially distributed group support. In *Proceedings of the 32nd Annual Hawaii International Conference on Systems Sciences. 1999. HICSS-32. Abstracts and CD-ROM of Full Papers*. IEEE, 9–pp.

[19] Brennan Jones, Yaying Zhang, Priscilla NY Wong, and Sean Rintel. 2021. Belonging there: VROOM-ing into the uncanny valley of XR telepresence. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (2021), 1–31.

[20] Tuomas Kantonen, Charles Woodward, and Neil Katz. 2010. Mixed reality in virtual world teleconferencing. In *2010 IEEE Virtual Reality Conference (VR)*. IEEE, 179–182.

[21] Demetrios Karis, Daniel Wildman, and Amir Mané. 2016. Improving remote collaboration with video conferencing and video portals. *Human–Computer Interaction* 31, 1 (2016), 1–58.

[22] Peter Gall Krogh, Marianne Graves Petersen, Kenton O'Hara, and Jens Emil Grønbæk. 2017. Sensitizing concepts for socio-spatial literacy in HCI. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. 6449–6460.

[23] Kristine M. Kuhn. 2022. The constant mirror: Self-view and attitudes to virtual meetings. *Computers in Human Behavior* 128 (March 2022), 107110. https://doi.org/10.1016/j.chb.2021.107110

[24] Amanda Lacy, Seth Polsley, Samantha Ray, and Tracy Hammond. 2022. A Seat at the Virtual Table: Emergent Inclusion in Remote Meetings. *Proc. ACM Hum.-Comput. Interact.* 6, CSCW2, Article 426 (nov 2022), 20 pages.

[25] Min Kyung Lee and Leila Takayama. 2011. " Now, i have a body" uses and social norms for mobile remote presence in the workplace. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 33–42.

[26] Paul Luff, Hideaki Kuzuoka, Christian Heath, Keiichi Yamazaki, and Jun Yamashita. 2009. Creating Assemblies in Media Space: Recent Developments in Enhancing Access to Workspaces. In *Media Space 20 + Years of Mediated Life*, Steve Harrison (Ed.). Springer London, London, 27–55. https://doi.org/10.1007/978-1-84882-483-6_4

[27] Mary Jo Foley. 2022. Microsoft starts rolling out Teams 'Front Row' view for better hybrid meetings. https://www.zdnet.com/article/microsoft-starts-rolling-out-teams-front-row-view-for-better-hybrid-meetings/.

[28] Microsoft 365 Blog. 2020. Reimagining virtual collaboration for the future of work and learning. https://www.microsoft.com/en-us/microsoft-365/blog/2020/07/08/reimagining-virtual-collaboration-future-work-learning/.

[29] Microsoft 365 Support. 2022. Customize your meeting view. https://support.microsoft.com/en-us/office/customize-your-meeting-view-95aaeaf8-0f22-46cf-a6f9-34ca9b04a1b2.

[30] Osamu Morikawa and Takanori Maesako. 1998. HyperMirror: toward pleasant-to-use video mediated communication system. In *Proceedings of the 1998 ACM conference on Computer supported cooperative work*. 149–158.

[31] Hidenobu Nagata, Dan Mikami, Hiromu Miyashita, Keigo Wakayama, and Hideaki Takada. 2017. Virtual reality technologies in telecommunication services. *Journal of Information Processing* 25 (2017), 142–152.

[32] Mamoun Nawahdah and Tomoo Inoue. 2012. Building a high realistic media space by superimposing a remote person's figure on the local view. In *Proceedings of the 2012 IEEE 16th International Conference on Computer Supported Cooperative Work in Design (CSCWD)*. IEEE, 416–422.

[33] Thomas Neumayr, Hans-Christian Jetter, Mirjam Augstein, Judith Friedl, and Thomas Luger. 2018. Domino: A descriptive framework for hybrid collaboration and coupling styles in partially distributed teams. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 1–24.

[34] Kenton O'hara, Jesper Kjeldskov, and Jeni Paay. 2011. Blended interaction spaces for distributed team collaboration. *ACM Transactions on Computer-Human Interaction (TOCHI)* 18, 1 (2011), 1–28.

[35] Ken-Ichi Okada, Fumihiko Maeda, Yusuke Ichikawaa, and Yutaka Matsushita. 1994. Multiparty videoconferencing at virtual social distance: MAJIC design. In *Proceedings of the 1994 ACM conference on Computer supported cooperative work*. 385–393.

[36] Linda Plotnick, Starr Roxanne Hiltz, and Robin Privman. 2016. Ingroup dynamics and perceived effectiveness of partially distributed teams. *IEEE Transactions on Professional Communication* 59, 3 (2016), 203–229.

[37] Alexander Raake, Markus Fiedler, Katrin Schoenenberg, Katrien De Moor, and Nicola Döring. 2022. Technological Factors Influencing Videoconferencing and Zoom Fatigue. *arXiv:2202.01740 [cs]* (Feb. 2022). http://arxiv.org/abs/2202.01740 arXiv: 2202.01740.

[38] Banu Saatçi, Kaya Akyüz, Sean Rintel, and Clemens Nylandsted Klokmose. 2020. (Re) Configuring Hybrid Meetings: Moving from User-Centered Design to Meeting-Centered Design. *Computer Supported Cooperative Work (CSCW)* 29, 6 (2020), 769–794.

[39] Banu Saatçi, Roman Rädle, Sean Rintel, Kenton O'Hara, and Clemens Nylandsted Klokmose. 2019. Hybrid Meetings in the Modern Workplace: Stories of Success and Failure. In *International Conference on Collaboration and Technology*. Springer, 45–61.

[40] Abigail Sellen, Bill Buxton, and John Arnott. 1992. Using spatial cues to improve videoconferencing. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 651–652.

[41] Frank Steinicke, Nale Lehmann-Willenbrock, and Annika Luisa Meinecke. 2020. A first pilot study to compare virtual group meetings using video conferences and (immersive) virtual reality. In *Symposium on Spatial User Interaction*. 1–2.

[42] Katherine M Tsui, Munjal Desai, Holly A Yanco, and Chris Uhlik. 2011. Exploring use cases for telepresence robots. In *2011 6th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 11–18.

[43] Gina Venolia, John Tang, Ruy Cervantes, Sara Bly, George Robertson, Bongshin Lee, and Kori Inkpen. 2010. Embodied social proxy: mediating interpersonal connection in hub-and-satellite teams. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1049–1058.

[44] Gina Venolia, John C Tang, Kori Inkpen, and Baris Unver. 2018. Wish you were here: being together through composite video and digital keepsakes. In *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services*. 1–11.

[45] Webex Help Center. 2022. Switch your view in Webex Meetings, Webex Webinars, and Webex Events (classic). https://help.webex.com/en-us/article/dy3xzq/Switch-your-view-in-Webex-Meetings,-Webex-Webinars,-and-Webex-Events-(classic).

[46] Jacob O Wobbrock, Leah Findlater, Darren Gergle, and James J Higgins. 2011. The aligned rank transform for nonparametric factorial analyses using only anova procedures. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 143–146.

[47] Naomi Yamashita, Keiji Hirata, Shigemi Aoyagi, Hideaki Kuzuoka, and Yasunori Harada. 2008. Impact of seating positions on group video communication. In *Proceedings of the 2008 ACM conference on Computer supported cooperative*

work. 177–186.
[48] Zoom Blog. 2021. Introducing Immersive View, A Fun New Way to Meet. https://blog.zoom.us/introducing-zoom-immersive-view/.
[49] Zoom Support. 2022. Adjusting your video layout during a virtual meeting. https://support.zoom.us/hc/en-us/articles/201362323-Adjusting-your-video-layout-during-a-virtual-meeting.

A APPENDIX

A.1 Order of Condition per Group

Given the initial order of Gallery to Together to Front Row to Perspectives, for every subsequent group we shifted the nth position of a condition to n-1 modulo 4. The table below lists all the group scenarios across the conditions, where the value inside each box indicates at what order the condition was presented to a given group.

		Groups											
		1	2	3	4	5	6	7	8	9	10	11	12
Condition	Gallery	1st	4th	3rd	2nd	1st	4th	3rd	2nd	1st	4th	3rd	2nd
	Together	2nd	1st	4th	3rd	2nd	1st	4th	3rd	2nd	1st	4th	3rd
	Front Row	3rd	2nd	1st	4th	3rd	2nd	1st	4th	3rd	2nd	1st	4th
	Perspectives	4th	3rd	2nd	1st	4th	3rd	2nd	1st	4th	3rd	2nd	1st

A.2 Study Tasks

In this study, participants were tasked to discuss an ice-breaker question, and a scenario to come to an agreement as a group. The participants had 2 minutes to discuss the ice-breaker question, and had five minutes for the scenario discussion. Given that our study had 4 conditions, we had 4 sets of discussion tasks as listed below.

(1) Task One
   (a) Ice-breaker Question: What did you eat for breakfast?
   (b) Scenario Agreement Discussion: Where would you like to go for an organizational team meal together? (add vegetarian or seafood allergy contraints, agree on a set menu for more complexity).
(2) Task Two
   (a) Ice-breaker Question: What is the favorite car that you've driven?
   (b) Scenario Agreement Discussion: What kind of team-building activity would you like to do for your organization? (add constraints that wheelchair or blind user will join for more complexity).
(3) Task Three
   (a) Ice-breaker Question: What is the favorite city you've lived in?
   (b) Scenario Agreement Discussion: What organizational family social gathering over the summer would you like to suggest to be planned? (get specific about location and activities, add constraints that wheelchair or blind user will join for more complexity).
(4) Task Four
   (a) Ice-breaker Question: What is your favorite kind of pet?
   (b) Scenario Agreement Discussion: If you were planning a group reward celebration trip, where would you go? (What group activities would be planned? Add constraints that wheelchair or blind user will join for more complexity).

Received January 2023; revised April 2023; accepted May 2023