Hybridge: Bridging Spatiality for Inclusive and Equitable Hybrid Meetings

PAYOD PANDA, Microsoft Research, UK
LEV TANKELEVITCH, Microsoft Research, UK
BECKY SPITTLE, Birmingham City University, UK
KORI INKPEN, Microsoft Research, USA
JOHN TANG, Microsoft Research, USA
SASA JUNUZOVIC, Microsoft Research, USA
QIANQIAN QI, Microsoft Research, USA
PAT SWEENEY, Microsoft Research, USA
ANDREW D. WILSON, Microsoft Research, USA
WILLIAM A.S. BUXTON, Microsoft Research, USA
ABIGAIL SELLEN, Microsoft Research, UK
SEAN RINTEL, Microsoft Research, UK

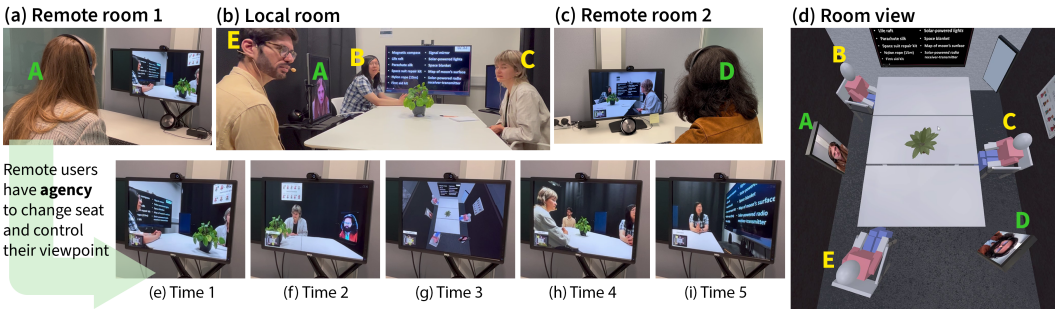


Fig. 1. A hybrid meeting using Hybridge. (a, c) show remote participants A and D, who may view a Hybridge room from any digital seat. (b) shows the local Hybridge room, with remote participants A and D distributed in-room. (d) show the meeting map of the Hybridge room digital twin with in-room (yellow) and remote (green) participants in their physical and digital seats. (e-i) Remote view choice: Remote participant A pans right to see Remote participant D (Time 1 - 2), switches to virtual view (Time 3), chooses a new digital seat (Time 4), pans right to see content (Time 5).

Hybrid meetings limit inclusion for remote participants. The Hybridge experimental system provides different interfaces for remote and room endpoints, focusing on improving inclusion via shared spatiality and remote

Authors' Contact Information: Payod Panda, payod.panda@microsoft.com, Microsoft Research, Cambridge, UK; Lev Tankelevitch, Microsoft Research, Cambridge, UK; Becky Spittle, Birmingham City University, Birmingham, UK; Kori Inkpen, Microsoft Research, Redmond, USA; John Tang, Microsoft Research, Redmond, USA; Sasa Junuzovic, Microsoft Research, Redmond, USA; Qianqian Qi, Microsoft Research, Redmond, USA; Pat Sweeney, Microsoft Research, Redmond, USA; Andrew D. Wilson, Microsoft Research, Redmond, USA; William A.S. Buxton, Microsoft Research, Redmond, USA; Abigail Sellen, Microsoft Research, Cambridge, UK; Sean Rintel, Microsoft Research, Cambridge, UK.

ACM 2573-0142/2024/11-ART501
https://doi.org/10.1145/3687040

agency. In-room participants see remotes on displays around a table, and remotes see video integrated into a digital twin. Remotes can choose where to appear and from where they view the room. We tested Hybridge in a within-subjects study of group survival tasks. An in-person condition was followed by a counterbalanced order of hybrid traditional videoconferencing ("Gallery") and Hybridge. We found that co-presence and agency differences between in-room and remotes were alleviated in Hybridge but remained in Gallery. Physical presence for remotes was higher in Hybridge than Gallery. Conversation flow was better in Hybridge than Gallery, but ease of awareness was not different. We argue that asymmetry should be embraced when designing hybrid meeting systems, with inclusivity achieved by tailoring features for the needs of different endpoints.

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

Additional Key Words and Phrases: videoconferencing; hybrid; meetings; collaboration; spatial; agency; asymmetry; inclusion; design; 3D; digital twin

ACM Reference Format:
Payod Panda, Lev Tankelevitch, Becky Spittle, Kori Inkpen, John Tang, Sasa Junuzovic, Qianqian Qi, Pat Sweeney, Andrew D. Wilson, William A.S. Buxton, Abigail Sellen, and Sean Rintel. 2024. Hybridge: Bridging Spatiality for Inclusive and Equitable Hybrid Meetings. *Proc. ACM Hum.-Comput. Interact.* 8, CSCW2, Article 501 (November 2024), 39 pages. https://doi.org/10.1145/3687040

1 Introduction

Hybrid video meetings, where some participants are together in a room while others join remotely, notoriously lack the inclusion afforded by physical presence and simultaneity in a shared physical space [66]. Inclusion, in this sense, has three elements: equitable opportunity of participation for remote and in-room participants, increasing remote participants' agency within the meeting, and reducing remote participants' feelings of being isolated or ignored. [1] Traditional videoconferencing shows people in a video grid (a "gallery" view), or a large content area andstrip of videos Both versions are unnatural compared to being in person, leading to both inclusion and fatigue problems [2, 62, 76]. In fully remote meetings, disadvantage is equally shared because each endpoint is fairly similar. However, in hybrid meetings,the number of people, physical configuration, and technology of endpoints are substantially different in-room compared to remote.

In canonical hybrid meetings, remote and in-room participants are divided. The gallery is scaled up on a wall display with a single camera nearby or on a table around which people in the room sit. Remote participants are disadvantaged in two key ways. First, in-room participants have spatial relationships with each other enabled by sharing the physical space, and also agency to navigate the physical space by looking around while seated, moving seats, or moving around the room. Videoconferencing constrains almost all these spatial relationships and agentic capabilities for remote participants. Workarounds such as remote participants asking in-room participants to move the camera add social friction and changes the view for all remotes. The technical solution of dynamic frame cropping to in-room participants disregards the room itself and reduces the agency of remotes by the very process of automation.

The traditional videoconferencing stage for hybrid meetings takes the form it does because it assumes a need to hew closely to symmetrical interfaces at all endpoints. This enables scaling to large numbers of people and meetings, reduces the complexity of setup and training, and assumes that 'What You See Is What I See' is the best way to enable reciprocity of perspectives [36, 78]. However, symmetry of interfaces ignores the asymmetries inherent in remote versus in-room endpoints, and heightens the conditions that exclude remote participants. The challenge of designing for hybrid meetings is how to attend to its inherent asymmetries of space and agency. As such, our driving

[1] Inclusion also involves broader issues relating to gender, ethnicity, ability, social class, etc. (e.g. [56, 69]), but these were out of scope for this research.

question is: how does an asymmetrical videoconferencing system, that emphasises shared spatiality and heightened agency, affect remote participants' inclusion in a hybrid meeting?

This paper reports our exploration of this research question using Hybridge, an experimental hybrid video meeting system which manifests a combined physical-virtual space shared by all participants on commodity hardware. We conducted a within-subject study with 48 participants in 12 groups, who completed survival tasks across three conditions:

(1) In-Person (acting as baseline measurement for collaborative meetings in knowledge work).
(2) Gallery (acting as baseline measurement for the commercial standard for *hybrid* meetings in knowledge work).
(3) Hybridge (the treatment of interest).

We begin by outlining prior research that motivates our work, followed by our principles for designing Hybridge. We then report on the methods (including hypotheses) and results for our study. In sum, Hybridge showed significantly better results compared to traditional Gallery views for reported measures of agency, co-presence, and physical presence for remote participants as well as conversation flow quality in the meeting space. However, there was no significant difference observed between the two systems for ease of awareness. Critically, we observed no difference for agency and a reduced difference for co-presence between in-room and remote participants in Hybridge, suggesting that it leveled the playing field between in-room and remote endpoints compared to the standard Gallery view. We discuss our results in terms of their wider implications for designing future hybrid-meeting systems, arguing that the goal of more inclusive meetings is best met by embracing asymmetry by tailoring features that best fit the resources available at each endpoint.

2 Related Work

2.1 Common interactional spaces and the challenge of video-mediated hybrid meetings

The basic state of conversation is inherently spatial and agentic. In this common interactional space [48], meeting in groups involves not simply gathering, but displaying mutual orientation in ways that distinguish focused from unfocused interaction [21]. Focus is displayed at bodily scale (e.g., F-formations [34]), or gaze, head, shoulder, and arm poses and gestures, and facial expressions, all of which also contribute to conversational flow [8, 22, 33]. Meetings have some specific forms of embodied action to bid for turns and holding the floor (e.g., nuanced ways of raising hands and highlighting attention [20, 47, 49, 50]. The physical configuration of spaces is also crucial. A room with a central meeting table surrounded by chairs (e.g. [20, 67]) provides different affordances for conversational agency than a large room with a stage and an audience facing it (e.g., [49]. When meeting in person, then, we have a holistic experience of interlocutors' embodied communication, their proximity, and the relative placement of personal and shared resources. Gathering on literal common ground helps establish social common ground [10].

Video meetings have always struggled to mimic the common interactional space of in-person meetings [2, 17, 28, 61]. This reduces their effectiveness [11] while increasing their potential for fatigue [15]. The overarching reason for this struggle is that video meetings rely on capture, transmission, and display technologies that fracture and fragment the ecology of communication, creating asymmetries [25, 26, 45] that disrupt signals of engagement [64, 68] and attention [38]. Further, it means that the person, reference, and task spaces are separated, attenuated, and/or potentially missing [9]. Common ground, in this case a shared frame of reference, is fragmented and harder to establish [82].

Hybrid video meetings, in which some participants are co-located and others are remote, exacerbate these asymmetries [16, 51, 66, 76]. Conversational flow is disrupted, especially aspects

such as bidding for the floor, responding in a timely manner, and knowing who is talking [65, 76]. In-grouping is common in hybrid meetings, with in-room participants disregarding or isolating remote participants [32, 65]. The favoring of in-room participants has a strong negative correlation with perceived effectiveness [60] and in-room participants may find tasks more enjoyable than remote participants [5]. These issues can be heightened for those with disabilities [1]. For both acquainted groups and strangers, in-room participants tend to be more active (e.g., setting strategy, doing their work and collaborating on more, and engaging in syntheses [6]), while remote participants tend to be more passive [6] and prey to social loafing [12]).

2.2 Space and agency in remote and hybrid meeting systems

Addressing these problems has a long history in Media Space [75] and related video-mediated communication research [19, 24, 53]. Many Media Space systems assumed that reciprocity of perspectives demands symmetrical interfaces [80]. In MatrixView [27], Multiview [54, 55], and Halo [58], multiple people sit at each endpoint at semi-circular desks looking at displays showing life-sized people. More complex hybrid interfaces bring remote participants off walls and into the physical space to heighten agency. MMSpace [59] enables two endpoints at which a pair of participants in a room meet with a pair of remote participants in another room. Each remote participant is shown on an individual display with motors driving the display physically through four degrees of freedom based on captured head pose. Here, agency takes the form of viewpoint choice of remote participants, and in-room participants seeing where remote participants have chosen to view. OmniGlobe [42] is less naturalistic. It uses 360° cameras on top of spherical globe displays, enabling people at each endpoint to move freely around their space and see the remote space, and also provided agency by using the globe via touch to bring people or objects into view. However, symmetrical interfaces are inherently inflexible, which tends to limit the amount of inclusion possible for remote participants, and does not fit modern conditions of work.

Leveraging asymmetries. Less attention has been paid to how asymmetries may be *leveraged* [46, 80] to balance out inequities between remote and in-room participants in hybrid meetings, but some examples do exist. In MirrorBlender [23], transparencies and background subtraction are leveraged to create a layered blend of person and task space at each endpoint. MirrorBlender provides agency in the the relative positioning of video on the mirrored canvas, but not agency in the 3D space. Gazelens [40] allows in-room participants to sit around a table. For remote participants, the in-room combination of people and artifacts are captured by cameras facing people and a top-down camera showing the collaboration surface. Remote participants see the in-room participants and other remote participants in a strip of videos of faces that surround the top-down view of the collaboration surface and artifacts, and have agency over their view with a control to zoom in and out of details of places on the collaboration surface. VROOM: Virtual Virtual Robot Overlay for Online Meetings [30] is an extremely asymmetrical system. In VROOM, a Beam telepresence robot is augmented by adding a life-size avatar overlay of the remote participant, viewed through a HoloLens worn by a local user, and giving the remote user an immersive 360° view of the local space, viewed through a VR headset. VROOM has very high agency for remote viewers, as they can physically move around the local activity space and gesture at local objects. That being said, VROOM is also a verycomplex system requiring significant resources and complex setup at both endpoints, making it hard to deploy at scale.

Perspectives [76] has a more subtle asymmetry and is more practically scaleable. Like MatrixView, MultiView, and Halo before it, a meeting room endpoint is set up with approximately life-size projection of remote participants, and individual cameras to capture each in-room participant. Remote participants are displayed on a wall or large screen as background-extracted video sitting

on one side of a virtual desk in a virtual room, facing the physical room participants. For in-room participants, all remote participants appear to be in the same virtual space, which is contiguous with the local space. For remote participants the experience is somewhat different. All other remote *and* in-room participants are captured as individual background-extracted video and displayed in a single virtual room. No physical rooms are shown. However, crucially, all participants occupy a seat at the virtual table in a spatially-consistent virtual room, which enables every participant to have a unique first-person view that is relative to all other participants. Perspectives is the closest prior system to our approach to enabling inclusion, but we also made different design decisions. As such, as we turn to describing Hybridge below, we will note some of its specific differences to Perspectives.

In sum, the struggle of videoconferencing, especially hybrid, to mimic the common interactional space of in-person meetings has largely been engaged with *overcoming* the asymmetries of end-points. However, if we take the focus away from reproducing the *material appearance* of in-person inclusivity, and instead focus on *what matters* about the shared spatiality and heightened agency of in-person inclusivity – agency, co-presence, physical presence, awareness, and conversational flow – then we might embrace asymmetrical features. And, further, we might do so to improve the inclusivity of remote participants, who are currently the most disadvantanged in hybrid meetings.

3 Designing Hybridge

3.1 Design principles

The driving motivation behind our iterative design and prototyping was to bridge the gap between remote and in-room participants in a hybrid meeting, and to level the playing field in terms of how much each participant felt part of the meeting. Our final prototype hinged upon two design principles: (1) distributed spatial presence of remote participants in the meeting room, and (2) heightened agency for remote participants to choose their viewpoint and where they appeared in the room.

Distributed spatial presence of remote participants in meeting room. Hybrid meetings create an imbalance due to the remote participants being in their own isolated physical space while the in-room participants share a physical space together. The standard solution (gallery view) displays all remote participants on a single screen in the meeting room. This is constraining due to the lack of individual presence that such an arrangement conveys, exacerbating the sense of remoteness for remote participants. This arrangement also makes it hard for an in-room participant to direct conversation at a specific remote participant. Additionally, since all remote participants share one camera view of the room, it is difficult for them to direct conversation to people in the room (and impossible to do so to other remotes). In order to mitigate this, we designed Hybridge fundamentally prioritizing the remote participants' spatial presence in the meeting room to create an equitable experience and a shared sense of spatiality between in-room and remote participants.

Hybridge creates a shared space blending the physical and virtual for which the baseline truth is rooted in the **shared physical space**. Remote participants have a physical presence at seats in the physical room, but they are not anchored to those seats. The virtual room experienced by remote participants is a digital twin of the physical room, rather than a wholly virtual space. Wholly virtual spaces might make remote participants dissociate from the physical room, furthering the disconnect between the local and remote participants.

Heightened agency and control for remote participants. Another factor in imbalance is the agency and the amount of control that the remote and in-room participants have over the meeting space. The physical meeting room is given precedence over the physical spaces that the remote participants

join from, since the meeting room hosts multiple participants. Consequently, this leads to the mental model of the remotes 'dialing in' to the meeting room. Any adjustments in the remotes' view of the room is either impossible (if the meeting room hardware is immovable), or mediated through in-room participants (e.g., requesting a room participant to nudge the camera to bring something into view). This lack of control and reliance on in-room participants can lead to feeling like they can't participate in the meeting at the same level [6, 12]. We argue that to level the playing field, designing for hybrid meetings should provide remote participants with as much or even more agency than the in-room participants. Hybridge allows remote participants to choose (1) where they appear in the meeting room, and (2) where they view the meeting room from. Additionally, remote participants are able to view the meeting room from viewpoints that are not possible for in-room participants. We offer these controls through an interactive digital twin of the meeting room, which marks the occupancy status of all seats (for both in-room and remote participants), and the ability to navigate the digital twin. So, like Perspectives [76], we provide a first-person view, as if remote participants are sitting in a seat in the physical room. However, we take this further and enable remote participants to explore other perspectives, leveraging the virtuality of the digital twin. A remote participant in Hybridge can see a bird's eye view of the room, which is unnatural, but provides remote participants with agency not possible in most videoconferencing systems, and agency not available to in-room participants. We leverage this asymmetry to level the playing field between remote and in-room participants not in terms of literal capabilities, the rather extent of capabilities. We hypothesized that providing remote participants with heightened agency—over their viewpoint of, and position in, the room via spatially distributed seats, and via an interactive digital twin of the room—will lead to their feeling of being able to participate and engage in the meeting, and enhance their feeling of being present in the meeting room.

3.2 Hybridge hybrid meeting prototype

As a manifestation of our design principles, we created Hybridge, a meeting prototype system tailored for hybrid meetings. We debated between using a canvas-based 2D vs. a spatial 3D interface for the remote participant. As a first step, we studied the implications of the two interfaces which we report in a previous paper [72]. We learned that 3D outperformed 2D in the participants' perceived sense of awareness, sense of agency, and physical presence. The majority of participants in that study also subjectively preferred 3D over 2D. Thus, we used the 3D interface as a starting point for the present study, and incorporated further feedback from our participants from the first study in order to design the final version of Hybridge as studied in this paper. This section talks about designed features and UX—for technical implementation details, please see Appendix A.

3.2.1 In-room setup. In the meeting room, Hybridge provides smaller 27" displays in portrait orientation with a camera and speakers, positioned around a table among the physical seats (see subsection A.2 for implementation details). The size of these monitors was chosen to render the remote participant at approximately life-size. A separate front-of-room display (52") is provided in the meeting room for shared digital content (Figure 5 (c)). The smaller displays act as "digital seats" that can be occupied by a single remote participant, who can be seen and heard as if they were sitting at the table. This arrangement creates a dedicated space for each remote participant in the physical meeting room, and affords them a unique visual and aural perspective from this position. It also aids implicit non-verbal cues within conversations permitting natural gaze cues. In-room participants can address the remotes uniquely by turning their heads towards a particular display, helping in-room participants feel they are interacting directly with the physically distant participant. Simultaneously, it in turn helps the remote participant acknowledge that the in-room participant is talking to them, and to direct conversation to specific individuals—both in the room

and remote. This allows for increased sense of connection between participants, lending depth to dialogues between distant and in-room collaborators.

3.2.2 Remote setup and interface. Remote participants join the meeting from a satellite room located in the same building, equipped with a 27" monitor, keyboard, mouse, and headphones to render spatial audio (see subsection A.3 for implementation details). Participation in the meeting is mediated through the Hybridge software prototype, which is a digital twin of the meeting room. The digital twin blends the other remote and in-room participants through camera and virtual representations, with spatialized audio—the remote participant sees *and* hears others in a spatially consistent way. The digital twin broadly offers three types of views to the participant, with simple interaction features:

Room view. In the "*room view*", participants see a 3D replica of the meeting room (Figure 1 (d)) with a freely movable camera. The room view shows furniture and other landmarks (e.g., potted plant, painting) positioned in the meeting room, including the displays that act as potential seats for the remote participants. Occupied "digital seats" show other remote participants' video, and occupied chairs show in-room participants as static, generic avatars.

Seated view. The participant can select an empty digital seat to occupy by clicking on the display in the *room view*. Upon selecting a digital seat, the remote participant's view smoothly transitions from the *room view* to the *seated view*. This involves occupying the digital seat at the chosen display, and concurrently the view blending from the 3D representation to the webcam feed from that digital seat (Figure 2). In the *seated view*, the participant is able to control their viewpoint by panning the camera left and right (Figure 3 (1c,2c)). As the participant pans their camera and changes their attention target, everybody in the meeting gets feedback—the other participants see a billboard effect on this participant's video on the physical display this participant has occupied (Figure 3 (1b,2b)), and the participant themselves see their FOV cone on the mini-map (described below) rotate to reflect their current view direction (Figure 3 (1a,2a)).

Mini-map. A mini-map at the bottom-left corner of the screen shows a live top-down view of the room. The map shows the current location of the participant as a yellow circle, along with a field-of-vision (FOV) cone that indicates the participant's gaze direction (Figure 3 (1a,2a)). Seats occupied by others are also marked: in-room seats with a red square, and digital seats with a red circle. The name initials of all seat occupants are also shown. The occupancy state information from the 3D room view and the map can aid the remote participant decide which digital seat to occupy.

3.2.3 Meeting UX. With Hybridge, in-room participants maintain their spatial navigability and ability to change seats, while seeing remotes on individual displays positioned around a table sharing the same physical space. Remotes see video feeds from the room blended into a spatialized digital twin of the meeting room, which enables them to see and hear meeting participants in

Fig. 2. Seat selection: the participant smoothly transitions from the *room view* in the digital twin (a) to the *seated view* (e).

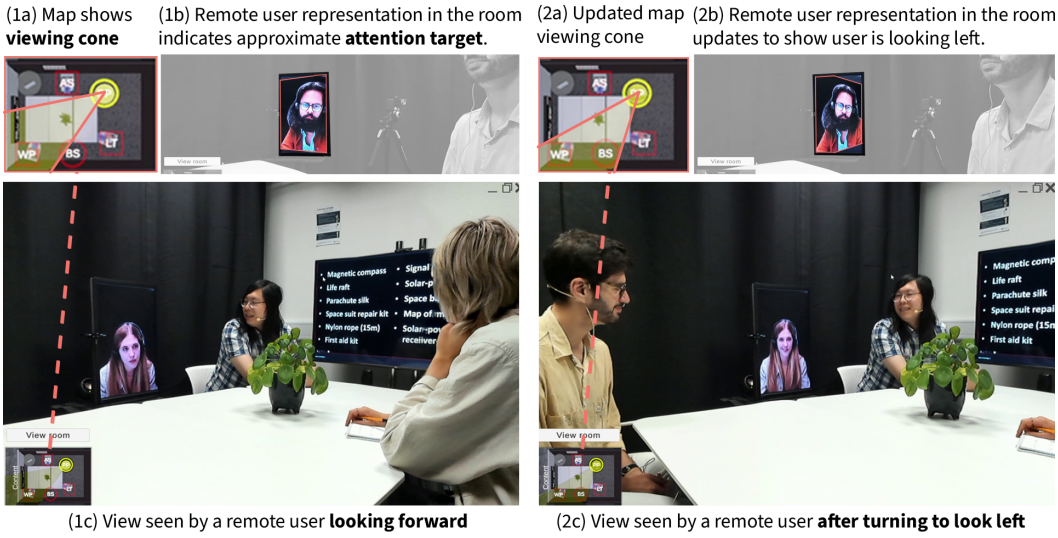


Fig. 3. Remote participants are able to control their viewpoint by panning their camera view left and right. The participant's viewpoints are shown in (1c) and (2c)—we can see that the participant turned their attention to the participant in the yellow shirt (in 2c). As the participant rotates their view, we see this reflected on the mini-map through their FOV cone (1a, 2a: marked in orange outline). The in-room as well as other remote participants can also see the video of the participant turn, creating a billboard effect, which gives an estimate of where the participant is currently paying attention (1b, 2b: video billboard marked in orange outline, background color-muted).

distinct locations in the meeting room. It also affords them the agency to choose where to appear in the room, and where to view the meeting room from. The combination of these features allows the meeting to progress in a more naturalistic manner. For instance, meeting participants are able to tell to a large extent where somebody else's attention target is. Specific remote and in-room participants can engage in conversations (Figure 4 (a)). Remote participants are able to acknowledge when in-room participants point towards and refer to content on a shared screen (Figure 4 (b))—which is impossible to do in a standard gallery view configuration since remote participants and shared content appear on the same digital display in the meeting room. If a remote participant

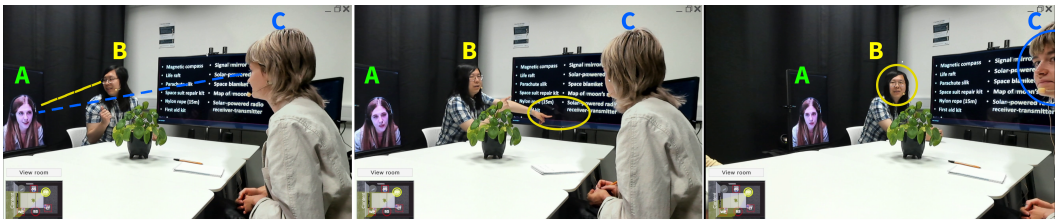


Fig. 4. Spatially distributing remote participants in the meeting room while keeping content separate in its own location allows remote participants to unambiguously identify attention targets. These images are shown from the viewpoint of a remote participant, from which one can gauge when (a) B and C talk to A, (b) B refers to the content being shared by pointing, and (c) B and C turn their attention to the participant.

contributes to the conversation, they can tell when other participants turn their attention towards them (Figure 4 (c))—which is also impossible to do in a standard gallery view with multiple remote participants due to shared presence in the meeting room.

4 Study and Methods

We conducted a study to compare Hybridge against other common ways of conducting a collaborative meeting—a fully in-person meeting, and a hybrid meeting using the currently widely used solution of a gallery view with a grid of videos. We were interested in seeing how our design principles for a hybrid meeting system levelled the playing field between remote and in-room participants. We introduce the three conditions we tested, followed by participant details, study procedure, and measures.

4.1 Comparing against common meeting practices

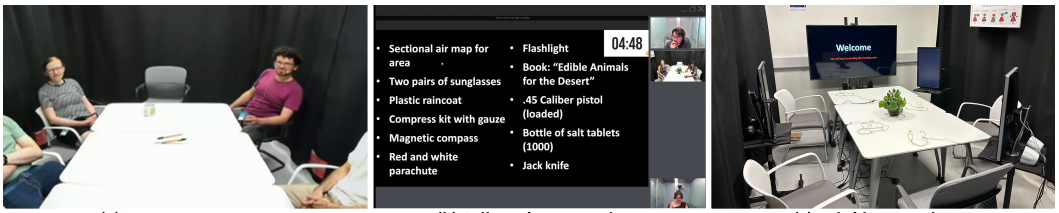
We decided to compare the Hybridge system to two other commonly used methods for a collaborative meeting—a fully in-person meeting, and a hybrid meeting using a standard Gallery view. We selected the in-person meeting as the baseline comparison for the "ideal" collaborative meeting—remote and hybrid meeting systems aspire to be as effective as a completely in-person meeting. We also chose to compare Hybridge against the standard video grid ("Gallery" view) since it is the modality that is most commonly used today for hybrid meetings. For consistency, both video-mediated conditions (Gallery and Hybridge) were implemented using the same prototype platform (a Unity application using Microsoft Teams NDI streaming capabilities). For technical implementation diagrams, see Appendix A.

A member of the research team acted as the moderator, sharing a slide deck with instructions for participants in all conditions. The shared screen showed instructions for the participants as well as a timer which showed the remaining time in that task. While the moderator could verbally coordinate with the study participants before and after each condition, they did not have any visual representation during the sessions. Other common video calling features such as text chat and reactions (e.g., hand raising, likes) were not included—the study focused only on the audio-visual layout for video calling.

In-person. For the in-person condition, all four participants met in a meeting room, sitting around a table. The participants were free to choose their seat in the room—there were five chairs for the four participants (Figure 5 (a)). A large screen at one end of the room showed the instructions for the participants, via a slideshow controlled by the meeting moderator.

Gallery. In this condition, the room had three physical chairs. Two participants participated from the room (choosing two of the three chairs), while two were escorted to individual on-premise satellite rooms where they joined the meeting remotely from. The large screen in the meeting room showed the instructions as well as the videos of the two remote participants, akin to a standard video call with a shared screen. The digital content on the shared screen occupied most of the display, and the videos from other endpoints appeared in a list on the right (Figure 5 (b)). The self-view webcam was shown on the bottom right. Also similar to a standard video call, this interface was mirrored on the remote endpoints. Neither remote nor in-room participants experienced spatialized audio from other participants.

Hybridge. In this condition, the room had three physical chairs and three smaller digital displays that served as the "digital seats" for the remote participants. Two participants stayed in the room (choosing two of the three chairs), while two each joined from on-premise satellite rooms. The two remote participants joined the meeting from the Hybridge remote interface (described in



(a) In-person setup (b) Gallery view screen layout (c) Hybridge room layout

Fig. 5. (a) The in-person setup consisted of five chairs around a conference table, and the four participants had the choice of their seats. This image is taken from the webcam of the large display at the front, which had the instructions for the task. (b) Our implementation of gallery view places the participants on a vertical strip on the right while the content occupies the majority of the space. There is a self-view at the bottom-right. (c) The Hybridge room setup for our study places small displays interspersed amongst room chairs around the conference table.

subsection 3.2), which allowed them seat selection and viewpoint agency through a digital twin. They received audio from all other participants (remote and in-person) correctly spatialized from their vantage point. Remotes went through a 2-minute training during which they learnt how to use the interface and made their initial seat choice (which they could change afterwards).

4.1.1 Hypotheses. Given these conditions, we developed our hypotheses focusing our interest on remote inclusivity in hybrid settings. Typical video meetings struggle to reproduce the common interactional space of in-person meetings [2, 17, 28, 61], and hybrid meetings in particular exacerbate the asymmetries between remote and in-person participation [16, 51, 66, 76]. This leads us to our first hypothesis:

> **H1:** Remote participants would report less inclusion in the Gallery condition than the In-Person condition.

Past work has found positive effects of spatializing remote presence [27, 42, 54, 55, 58, 59] and additional agency for remotes [23, 29, 41], but focused on creating symmetry in video meetings with exceptions like Perspectives [76]. We designed Hybridge to overcome space and agency issues with interactional space while allowing for the inherent asymmetrical nature of a hybrid meeting (see section 3). Our hypotheses about inclusion for the Hybridge condition compared to the other two, then, were:

> **H2a:** Remote participants would report more inclusion in the Hybridge condition than the Gallery condition.
>
> **H2b:** Remote participants would report the same inclusion as in-room participants in the Hybridge condition.

4.2 Study design

The study employed a mixed design, where the *Meeting Type* (Fully in-person, Gallery, Hybridge) was defined as the within-subjects variable and participant *Location* (Local, Remote) as the between-subjects variable. Each session consisted of a meeting with four participants. All participants completed the In-Person condition as a baseline before being split into two Local and two Remote participants for the remaining conditions. While the first condition was always the in-person condition serving as the baseline, we alternated the order of the Gallery and Hybridge conditions for each group. This was done for two reasons: (1) first, because some of our instruments (e.g., physical presence) asked our participants to compare their experience of using a meeting system (Gallery or Hybridge) to meeting in person, and (2) second, for the pragmatic consideration of

logistical simplicity in running the study and escorting participants between rooms. This enabled all groups to experience all three conditions—two participants remaining in the room for all sessions, and two experiencing the remote endpoints for Hybridge and Gallery. A basic Latin square design was used for counterbalancing the video-mediated conditions.

4.2.1 Participants. The study involved a sample of 48 people (12 groups of 4—30 male, 18 female). Despite the added logistical complexity, we chose to recruit groups of people familiar with each other in order to better reflect knowledge work meetings. This could mitigate confounding the influence of our design principles with the novelty of both acquaintance *and* using a novel meeting interface [63]. We aimed to recruit 14 groups of participants, an even number to ensure balance in the video-mediated condition order. To capture relevant social and conversational dynamics and prevent group-related asymmetry in the interactions, local and remote endpoints needed at least 2 participants each, hence we opted for groups of four (e.g., as per [52]). Thus, our total target sample size approximated that of similar previous work using a within-subjects design [76]. Due to the challenges of recruiting groups of familiar participants, we were ultimately able to recruit 12 groups for the study.

All but two groups included both male and female participants. Participants were mostly aged 18-29 (n = 32) or 30-44 (n = 14), with two being 45-59. Participants were students and staff sourced from university mailing lists (n = 24) and industry professionals in research and computing (n = 24). Apart from one participant, who reported having no peripheral vision, participants had no uncorrected visual or auditory impairments.

All participants were experienced with remote meetings, with most attending them weekly (n = 25) or daily (n = 19). Three participants attended them monthly. Few participants spent more than 10 hours in video calls per week (n = 4), with most attending them for 6-10 (n = 12), 3-5 (n = 19) or 1-2 (n = 12) hours. One participant attended video calls for less than 1 hour per week. Participants had a range of experience with hybrid meetings, with most attending them at least once per week (n = 21), or once per month (n = 19). Fewer participants took part in hybrid meetings daily (n = 8), every 3 months (n = 5), or less than every three months (n = 2). Platforms used notably included Microsoft Teams (n = 43), Zoom (n = 38), Google Meet (n = 22) and Apple FaceTime (n = 12). Half (n = 24) participants had experienced virtual meetings on both desktop and mobile platforms, whereas many reported to have only used desktop (n = 23), and one had only used mobile.

4.3 Procedure

Each study session lasted 60-90 minutes, with the study protocol having previously received IRB approval. The protocol was broken down into the following steps:

4.3.1 Pre-test. Informed consent and demographic information, including experience, was attained from participants prior to each study session. After welcoming participants, the purpose of the study and test protocol was explained before they were escorted to the conference room to prepare for the baseline (all in-person) condition. A second researcher located in the conference room then allowed participants to get comfortable around the meeting table.

4.3.2 Training. No training was required for room participants. Similarly, no training was provided for Gallery since it used an interface familiar to participants. For Hybridge, remote participants were given standardized training. This involved participants watching a two-minute video (provided as supplementary material) and following interaction steps to become comfortable with the different controls and functions of the prototype. Tasks the participant was asked to follow included the navigating the 3D *room view*, selecting and switching seats, and controlling their *seated view* viewpoint (e.g. by looking at the shared content screen). This ensured all participants were familiar

with the different controls and functions associated with the Hybridge prototype. Participants were also provided with the opportunity to ask questions.

4.3.3 Test. At the start of each condition, the group was given discussion topics like "If you could have a superpower, what would it be?" as a way to get familiar with that prototype. Following this, the study moderator introduced one of three discussion tasks in counterbalanced order. These were presented as hypothetical scenarios: "Survival in the Desert", "Survival on the Moon", and "Spy Mission Task" (see Appendix B for details). Survival tasks were chosen as they have often been considered in research on collaborative work. They exemplify typical interactions in small group meetings, requiring effective communication, consensus building, and problem-solving skills [13, 35, 73]. The goal was for the group to select 3 items from a list of 11 to optimise their chances of success within a 5-minute time frame. This required the group to deliberate their options and consider different strategies for surviving the scenario.

After completing the baseline condition, 2 participants (selected at random) were escorted by the moderator to individual satellite rooms. Participants had a short break before repeating the process for the remaining conditions.

4.3.4 Post-test. After completing all three conditions, all participants reconvened in the conference room to participate in a semi-structured focus group facilitated by one moderator (see Appendix C for the focus group guide). Each participant was compensated with an Amazon voucher at the end of the session.

4.4 Measures

After each condition, we asked participants to complete a range of survey items about sense of agency, co-presence, physical presence, conversation flow, and ease of awareness, (see Appendix D for all the survey items).

We take *agency* to refer to a participant's sense of control over their engagement in the meeting. Agency in videoconferencing, both remote and hybrid, is typically considered in terms of agency over the presentation of self [37, 77]. In addition to (1) presentation of self, we are also considering a participant's control over (2) their view of others and (3) their position in the meeting with respect to others [23, 39, 76, 77]. We developed a 3-item questionnaire asking all participants about their sense of control over these three aspects. We also asked remote participants to answer these three items in relation to the local participants (see Table 2 for all items).

We were also interested in evaluating participants' sense of belonging together in the space [31]. For us, the most relevant aspects of *spatiality* are the psychological dimensions of presence [44, 57, 71]. This consists of *co-presence*, the sense of being together *with others* [4], and *physical presence*, the sense of being in a different environment [43, 74]. To measure co-presence, we adapted eight items from the Networked Minds questionnaire to our hybrid meeting context [3], asking both about people's sense of co-presence with those physically in the room, and those joining remotely. Items asking about remote participants were excluded for the fully in-person condition. We also adapted one item from [70] asking participants to compare their hybrid meeting experience to a fully in-person meeting. To measure physical presence (for remote participants only), we adapted 5 items from the Slater-Usoh-Steed (SUS) questionnaire [79]. We also adapted the one-item measure from [7]. While physical presence is traditionally measured for *virtual* environments, we were interested in the extent to which participants feel present in the *physical* meeting room—we adapted the items accordingly.

Given the turn-taking difficulties reported in hybrid meetings [68, 76], we sought to also measure the ease of turn-taking or 'conversation flow'. We used a 6-item questionnaire, adapted from [76]. Related to conversation flow is participants' ability to be aware of people and activities throughout

the meeting, and thus we developed a 4-item questionnaire asking about participants' ease of awareness of other participants and activity during the meeting.

After all three conditions, we asked participants to rank the conditions in terms of preference, and provide an explanation for their ranking.

Finally, each group completed a semi-structured focus group to better understand and contextualize their questionnaire responses regarding each condition. It also allowed participants to gain insights into and reflect upon the experiences from both local and remote endpoints (see Appendix C for the focus group guide).

We piloted incremental protocols with several convenience-sampled groups within our organization, and the final protocol with one group recruited through the same means as our study participants. Adapted survey items were also used in a previous study [72] and piloted to ensure comprehension and face validity among participants.

5 Results

Out of the 12 study sessions conducted, results were derived from 10 groups of participants, with 2 sessions excluded from the data analysis. These sessions were omitted due to technical issues: one because of a call dropout on the Microsoft Teams platform, and another due to the deactivation of spatial audio, which rendered the data incomparable. To enable a concise analysis of all quantitative survey measures and reduce the number of statistical comparisons, for each coherent set of items (e.g., 6 items asking about conversational flow) we computed a summary score averaging (or summing, for the SUS questionnaire [79]) each participant's responses across the items. Statistical analyses were conducted on these summary scores, and, where relevant, were complemented with descriptive reporting for individual items to aid in interpretation. Averages and standard deviations for each individual item are reported in Appendix D. We analyzed these data using a non-parametric aligned rank transform (ART) mixed-effects model [81], with *condition* (Gallery, Hybridge, fully in-person) as a within-subjects factor and participant *location* (local, remote) as a between-subjects factor, where relevant. Some survey measures were only relevant to the remote participants and these analyses therefore excluded participant location as a factor; similarly, some measures were only relevant to the hybrid meeting conditions (Gallery and Hybridge) and these analyses therefore excluded the fully in-person condition.

For the fully in-person condition, although all participants were (by definition) local, for the purpose of having a balanced participant location factor for analysis, we labelled the participants who were subsequently remote in the later conditions as also "remote" in the fully in-person condition. As such, in the fully in-person condition, we don't expect any differences between the "remote" and local participants in their survey responses (making this comparison function as an additional validity check for the survey data).

Significant main effects or interactions were followed up with pairwise contrast tests using the ART-C procedure [18]. When only main effects were found, pairwise contrast tests were conducted on data collapsed across the non-significant factor. Pairwise comparison p-values were adjusted to control for the false discovery rate using the Benjamini-Hochberg procedure.

To support the quantitative results, we also include quotes from the focus group, where relevant. Participant IDs are listed as P<session ID>-<**I**n-room or **R**emote> (e.g., P7-R corresponds to a "Remote" participant from session 7).

5.1 Sense of Agency

There were two distinct ways in which remote participants using Hybridge had control over their viewpoints. First, they could choose where they appeared in the room. Second, once seated, they could pan their camera view sideways to focus on a specific part of the room from their vantage

point. We first measured participants' sense of agency for all participants in all three conditions, using a 3-item scale that asked about their feeling of control over *their view* of the conference room, *other people's view* of them, and their *position* in the conference room. We found a significant main effect of condition ($F_{2,76}$ = 32.45, p < 0.001) and participant location ($F_{1,38}$ = 17.97, p < 0.001), as well as an interaction between the two ($F_{2,76}$ = 22.89, p < 0.001). In Gallery, remote participants rated their sense of control lower than local participants (p < 0.001; Figure 6a). In contrast, Hybridge showed no significant differences between remote and local participants (p = 0.56). Thus, the difference in the sense of control between the remote and local experience was practically mitigated in Hybridge but not in Gallery (t_{76} = 5.51, p < 0.001, for the difference in differences).

Among remote participants, Hybridge was scored substantially higher on all three items compared to Gallery (see Table 2). Participants appreciated being able to choose where they appeared in the room, even indicating that they felt more present in the room as a result:

> **P5-R** (in Hybridge) I felt like I had more agency to move about. And I felt like I had more presence in the room.

This additional control allowed them to decide where they sat in relation to other participants, as well as objects of interest in the room. This was a consideration for in-room participants as well:

> **P2-I** It was interesting to think about where to physically sit in relation to the monitors (in Hybridge).

Remote participants with poor vision particularly benefited by being able to sit closer to the in-room content screen:

> **P11-R** For me, the Hybridge was really good because I have really bad vision. And the first thing I said when I walked in here is I want to grab a seat that's close to the slides... So I felt more in control.

However, among local participants, differences were smaller between Hybridge and Gallery. In fact, Gallery scored slightly higher on average on the item concerning "other people's view of you in the physical conference room" (see Table 2). Unlike Gallery, Hybridge did not show a self-view for local participants, which may have contributed to this, as noted by one local participant:

> **P7-I** For the standard meetings (Gallery), I can see what it looks like for them. I can have the control. But for the prototype (Hybridge) I don't know.

Unsurprisingly, relative to the fully in-person condition, remote participants' ratings were lower in both Gallery (p < 0.001) and Hybridge (p = 0.001). Local participants' ratings were not different between Gallery and the fully in-person condition (p = 0.14), but, interestingly, were lower in Hybridge than in the fully in-person condition (just above the significance threshold; p = 0.06). This potential difference emerged in Hybridge but not Gallery because local participants scored the Hybridge condition lower on the items concerning "your view of the physical conference room" and "your position in the physical conference room" (see Table 2). This may be related to the presence of the remote displays around the conference room which necessarily restricted local participants' positions in (and views of) the room, and the displays' 2D nature, which limited local participants' view of remote participants due to the restricted viewing angles, as noted by some local participants:

> **P6-I** (in Hybridge) I had less control of how they saw me and also cos the rigs were quite big it was like quite imposing in the room... yes, they're more there, but also like it's less comfortable for me.

> **P10-I** In the prototype (Hybridge) if I want to look at <name redacted>, I had to kind of lean around and then I knew that my fat head would be obscuring the screen

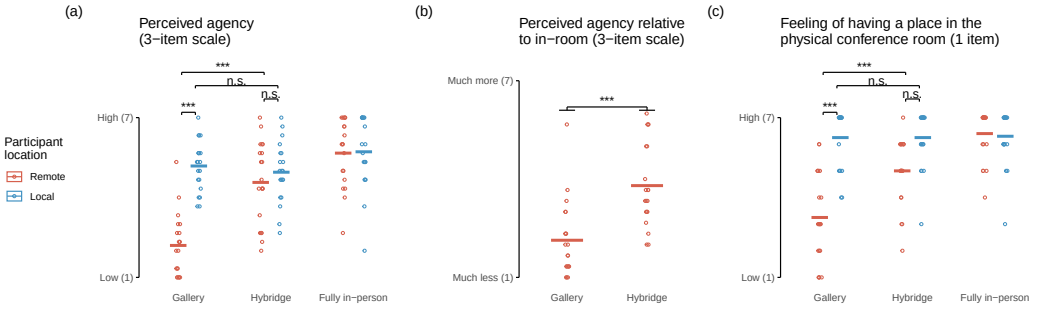


Fig. 6. (a) Perceived sense of agency as a function of condition and participant location. There was an interaction between condition and participant location, with ratings for remote participants (red) in Hybridge being higher than those in Gallery. Indeed, there was no significant difference in ratings between local and remote participants in Hybridge (compare blue and red bars), whereas there was a substantial difference in Gallery. Ratings for the fully in-person condition were highest of all. (b) Perceived sense of agency for remote *relative* to local participants for Gallery and Hybridge conditions, with Hybridge scoring higher. (c) Feeling of having a place in the room, as a function of condition and participant location. There was an interaction between condition and participant location, with ratings for remote participants (red) in Hybridge being higher than those in Gallery. There was no significant difference in ratings between local and remote participants in Hybridge (compare blue and red bars), whereas there was a difference in Gallery. Ratings for the fully in-person condition were highest of all. For all three panels, horizontal bars indicate averages, and circles indicate individual data points. To maintain figure readability, significant differences between the fully in-person condition and the others are not shown in the figure (they are reported in the main text). *** is $p < 0.001$, ** is $p < 0.01$, * is $p < 0.05$, n.s. is "not significant".

Thus, Hybridge may have slightly lowered local participants' sense of agency, but ultimately fostered equitable experiences between remote and local participants.

Additionally, we asked remote participants to rate their sense of agency (for the three items above) *relative* to the local participants. There was a significant effect of condition ($F_{1,19} = 30.5$, $p < 0.001$), with participants rating their sense of agency higher in Hybridge than in Gallery (Figure 6b); substantial differences were observed for all three items (see Table 2).

Finally, we asked all participants to rate the extent to which they felt they had a place in the conference room. As above, we found a significant main effect of condition ($F_{2,76} = 26.79$, $p < 0.001$) and participant location ($F_{1,38} = 36.40$, $p < 0.001$), as well as an interaction between the two ($F_{2,76} = 24.31$, $p < 0.001$; Figure 6c). Again, the difference between the remote and local experience was practically mitigated in Hybridge but not in Gallery ($t_{76} = 4.27$ $p < 0.001$, for the difference in differences). Comments from remote participants expressed the relationship between their spatial presence in the room and their increased agency in conversation:

P11-R (in Hybridge) I felt more in control and like if I'm at the screen sitting next to someone, I would grab peoples attention.

While Hybridge was rated highly for agency, not every participant used the ability to change seats *during* the meeting.

P13-R we could move around the room although I didn't because it felt weird. It's like, obviously, if you're in a meeting, like, I'm not just gonna stand up.

Nevertheless, our telemetry shows that remote participants in Hybridge condition used the ability to change their seats on average 3.45 times per session. Additionally, the participants viewed the *room view* on average 5.25 times per session (including the starting scene). Remote participants

also panned their viewpoint in order to attend to elements of interest. The *mean angle* in Table 1 indicates the average angle that participants chose to view from that respective seat, which tended to be either the content screen or other participants (for a per-participant breakdown, see Figure 7). The *standard deviation* reported in Table 1 indicates the extent to which participants panned their view per seat–a higher deviation indicates panning further left and right. The participants thus used the panning feature less in the *room view* than while viewing the room from the *seated view*–which is unsurprising, since the *room view* provided a bird's eye view of the whole digital twin from a single vantage point.

Table 1. **Camera panning usage**: Mean and standard deviation, in degrees. Mean indicates the region of highest interest for each seat, and the standard deviation indicates the extent to which participants panned their view (higher deviation indicates more panning). Telemetry indicates that participants used the camera panning feature more from seats (Seat 1-3) than in *room view*.

Seat	Mean	Std. Deviation
Room view	0.83°	9.29°
Seat 1	11.29°	26.94°
Seat 2	-15.66°	31.75°
Seat 3	4.96°	26.25°

However, even while acknowledging the higher level of agency and control from Hybridge (relative to Gallery), some participants still expressed a feeling of helplessness compared to the experience of being physically together (P5-R), and desired even more control while seated in Hybridge, such as being able to translate the physical display on which they appear in the room (P6-R).

> **P5-R** I feel completely helpless because you, you're in the room. Yeah, you can control my screen... I can like pan with a mouse, whatever. But I can't see behind me... If you were to make a sign or something, while I'm panned that way, I can still see you in my peripheral vision (when physically together), whereas (in Hybridge) I can't.

> **P6-R** ... great to be able to control the rotation, but also it would be better if I can control translation and the position of the chair ... physically in the room I can look, for example do this (leans forward) as part of the conversation ... was like I don't have full control of moving myself.

5.2 Co-presence

For perceived co-presence, we first compared scores on a subset of four items that were asked of participants in all three conditions. We found a significant main effect of condition ($F_{2,76}$ = 41.85, p < 0.001) and participant location ($F_{1,38}$ = 20.66, p < 0.001), as well as an interaction between the two ($F_{2,76}$ = 17.48, p < 0.001). In both Gallery and Hybridge, remote participants rated co-presence as being lower than local participants (p < 0.001 for both; Figure 8a). However, ratings for remote participants were higher in Hybridge than in Gallery (p < 0.001; with a particularly large difference for the item "I often felt as if the people physically in the room and I were together"; see Table 3). Indeed, the difference between remote and local participants was smaller in Hybridge than in Gallery (t_{76} = 2.55, p = 0.013, for the difference in differences), suggesting that Hybridge at least somewhat mitigated the difference in experiences between local and remote participants.

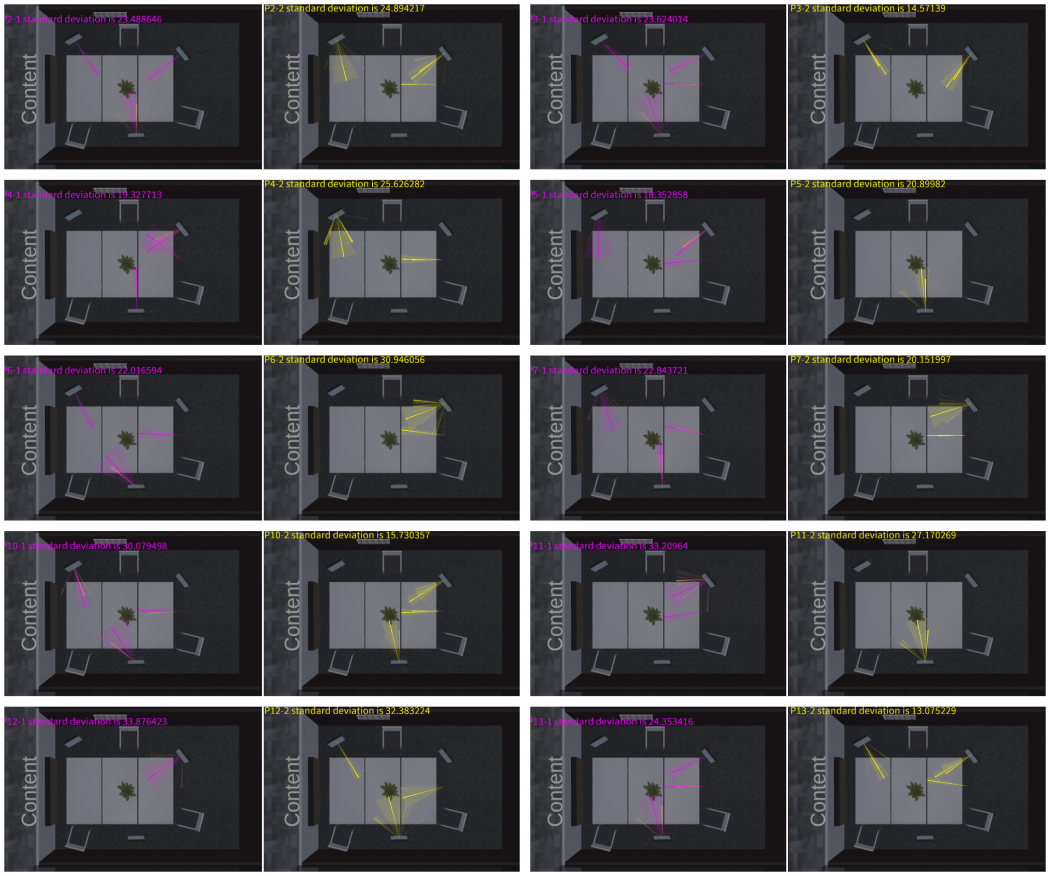


Fig. 7. This figure shows usage data for remote participants using Hybridge collected via telemetry visualized on the mini-map. The ten rectangles each represent a session, and the magenta and yellow colors represent the two remote participants for that session. Each map shows a "heatmap" of the participant's FOV cone—so, in aggregate, the visualization gives an indication of how much each participant used the panning and seat change features. The solid lines on each map show that participant's average viewing angle for that seat. The translucent triangles represent that participant's standard deviation for that seat—how much did that participant deviate from their average viewing angle for that seat? The figure shows that participants tended to focus their attention either on other seats or on the content screen at the front of the room.

Unsurprisingly, remote participant ratings in both Gallery and Hybridge were lower than in the fully in-person condition ($p < 0.001$ for both). Interestingly, in Gallery, even *local* participant ratings were lower than the fully in-person condition ($p = 0.02$), a difference that was not observed between Hybridge and the fully in-person condition ($p = 0.28$). This experience in Gallery was at least partly reflected in low ratings on the item "I often felt as if the people physically in the room and I were together" (see Table 3).

Next, we compared scores on a longer set of eight items that were asked of all participants in the Hybridge and Gallery conditions only. We found a significant main effect of condition ($F_{1,38} = 19.61$, $p < 0.001$) and participant location ($F_{1,38} = 10.22$, $p = 0.003$), but no significant interaction ($p = 0.59$). Perceived co-presence ratings were higher overall in Hybridge than in Gallery, and higher

for local than remote participants (Figure 8b). Hybridge was rated higher than Gallery across all eight items. Among remote participants, there was a substantial difference for the item "I often felt as if the people physically in the room and I were together", whereas among local participants, there was a substantial difference for the item "I often felt as if the people who joined remotely and I were together" (see Table 3). This could be explained by remote participants feeling disconnected from local participants during a discussion in the gallery condition:

P11-R I felt more out of the loop (in gallery)Especially when you guys (in-room) were talking to each other.

Interestingly, this participant felt a sense of camaraderie with the other remote participant, leading to a feeling of sitting together:

P11-R knowing that <name redacted> is experiencing what I'm experiencing. So if I couldn't hear them, he also couldn't hear them. So I felt like I was sitting with him (gallery).

Lastly, we asked all participants in the Hybridge and Gallery conditions to rate how similar their sense of being together was to a fully in-person meeting. We found only a main effect of condition ($F_{1,38}$ = 27.52, p < 0.001), with Hybridge rated as more similar to a fully in-person meeting than Gallery (Figure 8c). For instance, participants often remarked the benefit of creating eye contact and how Hybridge enabled this for them, from both a remote and in-room perspective:

P11-R (during Hybridge) when you shifted to look at me it really looked like we were looking in the eye...

P6-I I think like the eye contact thing is pretty good. Like if you're having like a brainstorming type thing.

However, while Hybridge made it easier for remote participants to connect with in-room participants, some participants expressed concerns about not being able to see the other remote participants very well, including their facial expressions:

P7-R It was just less from you (other remote participant) like, getting the facial expressions.

5.3 Physical Presence

We measured perceived physical presence only for remote participants in the Hybridge and Gallery conditions. We found a significant effect of condition ($F_{1,19}$ = 13.99, p = 0.001), with ratings being higher in Hybridge than Gallery (Figure 10a); this was the case across all 6 items (see Table 4). This was supported by remote participants' comments:

P2-R (Hybridge) Was more of like in the physical space. It really brought my attention back into the room

P4-R I like it (hybridge) more than the video one (gallery) because I felt like I was actually in the room

As P2's comment suggests, a sense of physical presence can increase remote participants' attention towards in-room activity. Likewise, Hybridge's design affords remote participants with *literal* physical presence via the distribution of dedicated A/V displays throughout the conference room. This in turn drew the attention of in-room participants who positively acknowledged remote participants' increased physical presence:

P10-I It's nice having a chair or place in the room, It's nice having that dedicated space for you.

P7-I I thought it was not as good as physical because of those little nuances. However, you did feel like people had a position in the room and you could address them and look at them.

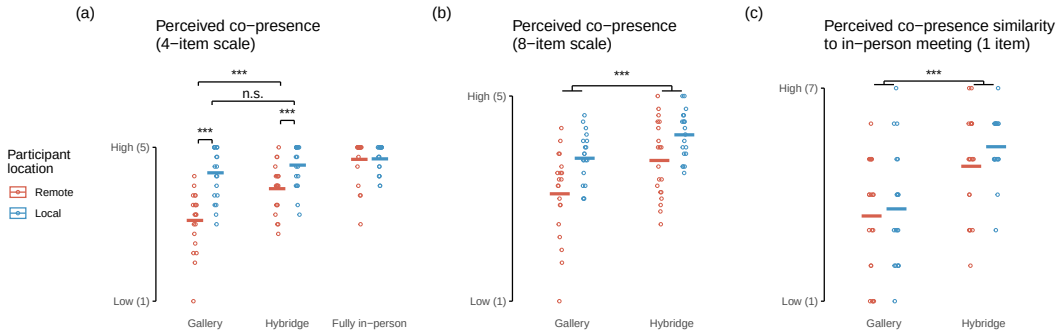


Fig. 8. (a) Perceived co-presence for a subset of four items in all three conditions. There was an interaction between condition and participant location, with ratings for remote participants in Hybridge being higher than those in Gallery. Indeed, the difference in ratings between local and remote participants was smaller in Hybridge than in Gallery (compare blue and red bars for each). Ratings for the fully in-person condition were highest of all. (b) Perceived co-presence for the full set of eight items in Hybridge and Gallery. For both local and remote participants, ratings were higher in Hybridge than in Gallery. (c) Perceived similarity of co-presence relative to being fully in-person, in Hybridge and Gallery. For both local and remote participants, ratings were higher in Hybridge than in Gallery. For all three panels, horizontal bars indicate averages, and circles indicate individual data points. To maintain figure readability, significant differences between the fully in-person condition and the others are not shown in the figure (they are reported in the main text). *** is $p < 0.001$, ** is $p < 0.01$, * is $p < 0.05$, n.s. is "not significant".

5.4 Conversation Flow

Hybridge's distributed spatial presence and heightened agency for remote participants should in turn support improved turn-taking and conversational flow between all participants. For perceived conversation flow, we found a significant main effect of condition ($F_{2,76} = 77.8$, $p < 0.001$). Participants rated the fully in-person condition as higher than both Hybridge and Gallery ($p < 0.001$ for both). However, Hybridge was rated higher than Gallery ($p = 0.001$) (Figure 9a). The spatial audio in Hybridge was crucial in helping participants distinguish between multiple people talking at the same time.

> **P13-I** it makes a difference that there's like different audio sources so that we could hear you talking over Austin or like interrupting him.

For both local and remote participants, Hybridge scored higher than Gallery on all conversational flow items, with a particularly large difference (i.e., over 1 scale point) among remote participants for the item "I could easily refer to or direct comments to others in the conversation" (see Table 5). This went both ways—some participants felt that others could refer to and acknowledge them in the Hybridge condition:

> **P7-R** having someone look at you, you kind of tacitly, like, get their approval to be like ohh you have the floor now you have the microphone now.

5.5 Ease of Awareness

In tandem with improved conversational flow, we expected Hybridge to improve participants' awareness of others and of activity throughout the meeting. For perceived ease of awareness, we found a significant main effect of condition ($F_{2,76} = 45.28$, $p < 0.001$). Participants rated the fully in-person condition as higher than both Hybridge and Gallery ($p < 0.001$ for both); this was the

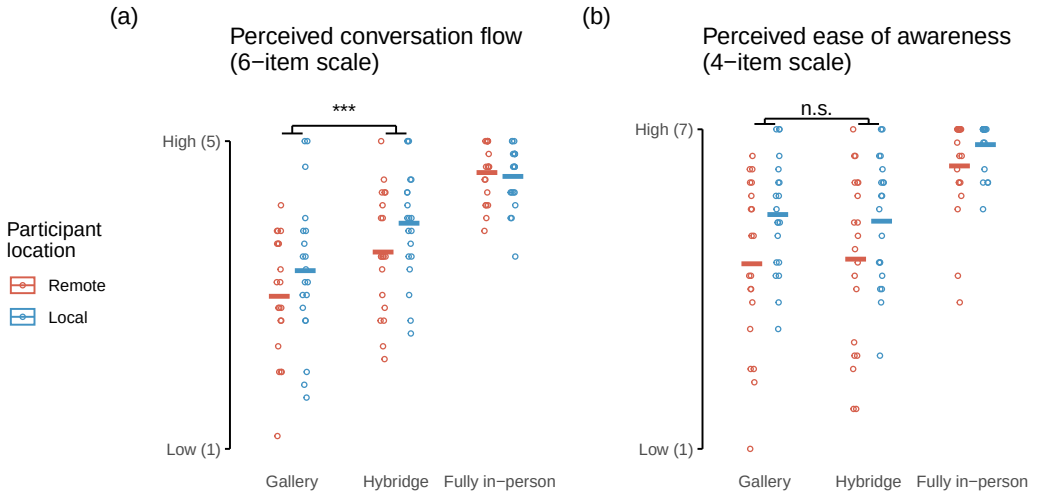


Fig. 9. (a) Perceived conversation flow as a function of condition and participant location. Both local (blue) and remote (red) participants reported higher conversation flow in Hybridge compared to Gallery, and in the fully in-person condition compared to both Hybridge and Gallery. (b) Ease of awareness as a function of condition and participant location. Among both local and remote participants, there was no difference between Hybridge and Gallery, whereas the fully in-person condition scored higher than both Hybridge and Gallery. For both panels, horizontal bars indicate averages, and circles indicate individual data points. To maintain figure readability, significant differences between the fully in-person condition and the others are not shown in the figure (they are reported in the main text). *** is $p < 0.001$, ** is $p < 0.01$, * is $p < 0.05$, n.s. is "not significant".

case for all four items, among both local and remote participants (see Table 6). However, there were no differences between Hybridge and Gallery ($p = 0.89$; Figure 9b).

Directional differences in ratings for individual items were inconsistent. For example, among remote participants, ratings were slightly higher for Gallery for "see everything going on in the meeting all the time", but higher for Hybridge for "stay oriented on what was happening where throughout the meeting". There was no difference for "see everyone you needed to throughout the meeting" and "see everything you needed to track what was going on throughout the meeting" (similar inconsistencies were observed among local participants; see Table 6) Some participants found that the spatial distribution inherent to Hybridge required both local and remote participants to actively look around to track the meeting's activity, whereas Gallery presented both people and content in one convenient location:

P3-R (Hybridge) doesn't really give you enough control to find a comfortable place to view everything.

P2-I It (Gallery) has the most clarity, I suppose, because you don't move your head ever

Lastly, there was also a main effect of participant location ($F_{1,38} = 4.94$, $p = 0.03$), with remote participants providing lower ratings than local participants.

5.6 Participant Preferences

Finally, we asked participants to rank their preference (first, second, third choice) of meeting mode for future meetings. Preference rankings were very similar between remote (Figure 10b) and local

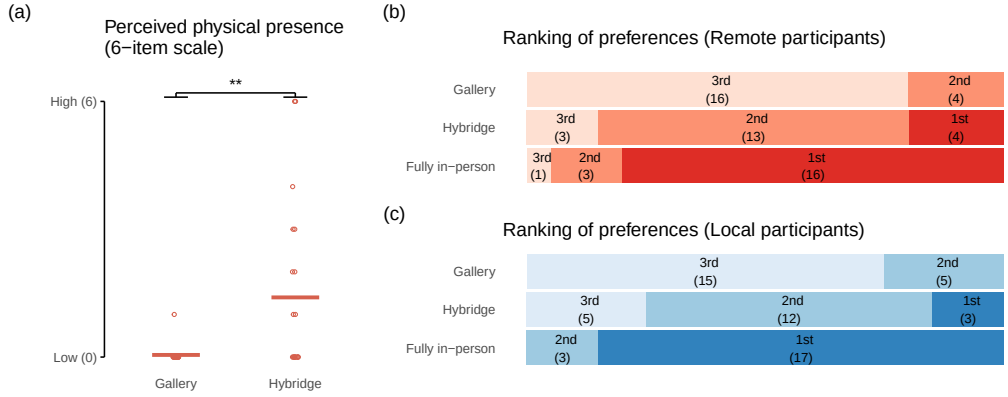


Fig. 10. (a) Perceived physical presence among remote participants in Hybridge and Gallery conditions, with Hybridge scoring higher. Horizontal bars indicate averages, and circles indicate individual data points. *** is $p < 0.001$, ** is $p < 0.01$, * is $p < 0.05$. (b) Participant rankings of preference (i.e., 1st, 2nd, 3rd) between Hybridge, Gallery, and fully in-person for future meetings, for remote (red) and local (blue) participants. Numbers in parentheses indicate the number of participants choosing this rank for a given condition (e.g., among local participants (blue), 17 ranked the fully in-person condition as their first preference, three ranked it as their second preference, and none ranked it as their third preference.

(Figure 10c) participants, with respectively 13 and 12 participants (out of 20) ranking Hybridge as their second choice after a fully in-person meeting. Interestingly, respectively 4 and 3 participants ranked Hybridge as their first choice (none ranked Gallery as their first choice). Thus, 17 out of 20 remote (85%) and 15 out of 20 local participants (75%) overall preferred Hybridge over Gallery. In the previous sections, we report how Hybridge affords an improved sense of agency, co-presence, physical presence, and conversational flow. Here, we further contextualize participants' preferences with other themes from the focus group.

When comparing to Gallery, remote participants found Hybridge to be *"more comfortable"* (P12-R) and *"more convenient"* (P2-R). Participants' comparisons of Hybridge relative to the fully in-person experience varied. Some local participants found it *"closer to reality than [they] thought"* (P2-I), and *"kind of a natural meeting"* (P12-I). Others found it *"really uncomfortable"* (P5-I), instead preferring the familiarity of Gallery (P6-I).

Participants appreciated Hybridge's potential to balance the opportunity for participation across local and remote endpoints. This was attributed to the dedicated spatial presence that Hybridge afforded remote participants.

> **P7-I** I think the worst kind of meetings, are the ones where [...] the remote people are definitely second class to the physical, whereas the Hybridge prototype equalised that imbalance because everyone had a seat

Remote participants echoed the impact of this spatial presence on their participation, noting that Hybridge makes them *"more confident to express [their] ideas"* (P12-R), or *"more entitled to [...] jump into the conversation"* (P2-R), and *"grab people's attention"* (P11-R). A related phenomenon was the increased meeting engagement in Hybridge observed by some participants. Though it may be a novelty effect, one participant noted *"It's naturally engaging to like turn to people and make my contact"* (P5-I). Similarly, *"midway through the meeting (Gallery) you just casually looked at your*

phones, whereas in (Hybridge and in-person) that didn't happen" (P5-R). This may be a product of Hybridge's spatial distribution and the presence and conversational flow that it affords.

Perception of self and others' view of oneself arose as a key issue in Hybridge. Local participants noted the lack of information in Hybridge on how they appeared to remote participants, particularly given the latter's ability to switch positions and viewing angles.

P10-I "I can't see how I appear to you, whereas with (Gallery) I can see what you're looking at."

However, as noted by some participants, this need for information may partially stem from their (understandable) lack of experience with the remote endpoint in Hybridge.

P6-I "if we've been on the other end (of Hybridge)... it would make it easier to use."

Remote participants also emphasized the value of *"seeing yourself and having a sense of what others see of you"* (P13-R), which is available in Gallery. For some, this information was worth the trade-off in terms of spatial presence: *"you can feel like you're some alien overlord (in Gallery), but you still know where you are in the room"* (P3-R). However, at least one participant noted that the presence that Hybridge affords to remote participants reduced their need for self-view.

P13-R "like when it's this sort of view where it feels like you're that person. It feels like the expectations are a bit different."

One design issue, noted by local participants in Hybridge, pertained to remote participants' sudden movements across displays, which local participants found "disorientating" (P2-I) or "creepy" (P5-I). This may also have contributed to local participants' slightly lower ratings for perceived agency for Hybridge relative to the in-person condition (see subsection 5.1). Other issues noted by participants pertained to Hybridge's functional implementation as a prototype. Some participants found the physical setup "very complicated" (P2-I) or resource-intensive (P11-I) to scale to larger meetings. Accordingly, it was suggested to be most useful for smaller meetings that may be social or otherwise dynamic, *"like smaller teams where its more fun and chill"* (P10-R), or *"when working on a project together. And so it's not a presentation style. It's like a group thinking sort of exercise"* (P11-I).

6 Discussion and Design Implications

The results from our study highlighted the impact of designing a hybrid meeting system that provides distributed presence and heightened agency for remote participants. Notably, these design principles helped bridge the gap between remote and in-person participants by enabling the remote participants to feel more in control of their presence in and view of the room, leading to a feeling of being present with others, feeling present in the meeting room, and an increased sense of agency more comparable to that felt by those in the room (providing partial support for H3). While it didn't impact the sense of awareness of others, Hybridge also substantially improved the conversational flow—alongside co-presence, physical presence, and agency—compared to the standard grid-based video system (providing support for H2). Lastly, as expected from prior work [2, 76], remote participants in the standard grid-based system felt a reduced sense of inclusion across all measures relative to the fully in-person condition (providing support for H1). In this section, we discuss some emergent design themes and future implications.

6.1 Rethinking symmetry in hybrid meeting systems

Many past meeting systems have focused on solving the issues of reciprocity of perspectives through deploying symmetrical UIs, i.e. systems where each endpoint is functionally equivalent [80]. However, while asymmetry was not one of our design goals, our designed system created asymmetry in several ways, the most important of which was the modality of engaging with the meeting. The in-person participants engaged primarily through the physical meeting room, where the remote

participants were "physicalized" in the physical space by having a dedicated space and seat at the table. On the other hand, remote participants engaged primarily through a digital twin of the meeting room, where the in-room participants (along with other remote participants) manifested digitally.

However, both these endpoints were connected by the spatial understanding they created for its participants—both in-room and remote participants could develop a sense of how the meeting room was laid out, and how the participants were seated, creating opportunities for non-verbal social cues not typically possible with standard grid-based video meeting systems. For instance, an in-person participant could direct conversation at a specific remote participant by looking towards them and speaking, similar to interacting with another in-person participant. Since the remote participant's viewpoint of the room is from that display's location, they are able to tell that they are being referred to directly.

Another way in which we leveraged the asymmetry of digital and the physical was to provide the ability for remote participants to view a digital twin of the meeting room (*room view*), intended to help the participants create a better mental model of the room and layout. Our telemetry shows that participants used this feature sparingly compared to viewing the meeting from the *seated view*. Thus, the current study provides limited evidence for the importance of this feature. It's plausible that the limited use of the *room view* in the current study stems from the fact that it did not show a high-fidelity view of the meeting room (i.e., the webcam), and showed the in-room participants as low-fidelity avatars. Moreover, participants did not get referential cues in the *room view* about when others were looking at or referring to them. Nevertheless, although our current study did not provide evidence for *room view*'s importance, we obtained some evidence for its role from our comparison of 2D and 3D interfaces of Hybridge in a previous study [72], which found that the 3D interface enhanced spatial awareness, *"Participants reported experiencing enhanced spatial awareness in H3D (n = 9), with the representation of the physical meeting room and the ability to pan the camera view allowing for more intuitive navigation of the meeting space when compared to H2D"*. We leave further exploration of this feature and its role in asymmetrical contexts for future work.

We propose that catering to the individual needs of each endpoint in an inherently asymmetrical setup would naturally lead to a meeting system that embodies one or more asymmetries. However, we are not suggesting that asymmetry should be the *primary* design principle or goal. The design principles should align with the design goals, for instance distributed spatial presence and heightened agency (principles) for a more equitable and inclusive meeting (goal). Rather, we suggest that the need for uniform systems across endpoints should be relaxed in favor of the needs for inclusion, and if the result is a system that has asymmetries then that is not something to shy away from *a priori*.

6.2 Equal agency ≠ equitable experience

Some video meeting systems try to bridge the gap between remote and in-person by trying to copy capabilities of being physically together on the remote end. However, the two endpoints are *inherently* different—even with spatial immersive technologies the experience can't be exactly replicated. Instead, we should aim to make the best use of the medium on each side. For a hybrid meeting, this means ensuring utilizing affordances of shared physical space for in-room endpoint, and a digital space for remote endpoints.

This could take the form of a novel feature or capability. In our implementation, for instance, the digital remote interface allows the participant to view the meeting room from a bird's eye view. Our specific implementation had this as a virtual camera, but one might imagine placing an RGB webcam on the ceiling, looking down, which gives a real-time view of the room from an elevated

viewpoint. This is a perspective that is not possible for an in-room participant to have, but is made possible for the remote participant due to the virtual nature of the interface.

Sometimes, even attempting to replicate a capability from one modality to the other might manifest in different behaviour between the endpoints, owing, for instance, to convention. For instance, in-room participants in our system (and other physical rooms with at least another vacant chair) have the capability of vacating their chair and occupy another. However, in practice, this behaviour is not a common occurrence due to two potential reasons:

(1) first, because it might not be needed—in-room participants being present with their physical bodies in the room can navigate the room visually with substantial freedom, including looking around, turning their bodies left and right, leaning on the desk, nudging their seats forward and back etc. which reduces their need for a completely new vantage point in the room, and
(2) second, because vacating their chair, moving in the room, and taking another chair could cause a disruption in the meeting flow and might be frowned upon—it is the convention in meetings to usually stick with the seat chosen by the participant at the beginning of a meeting.

While we implemented much the same capability on the remote end—i.e., the capability for the remote participant to "vacate" their seat, scout for a different seat, and occupy that seat instead—this feature was used by several remote participants to change their seats during the meeting. We ascribe this behaviour to the *need* for a different vantage point—being in the remote seat doesn't afford similar navigational capabilities to the remote participant while staying in the same seat. However, another interesting contributing factor could be that using the digital interface freed the remote participants of the convention of sticking with one's seat throughout the meeting. Our telemetry shows that several remote participants not only changed their viewport while in a particular seat, but also actively switched between seats in order to change their range of possible viewpoints (e.g. moving from a seat on the side to a seat opposite the content screen in order to get a better angle on the content while being able to rotate to view other participants).

While we didn't create *equal* experiences for remote and in-room participants, we succeeded in creating a more *equitable* experience. For Hybridge, our data doesn't show a difference in "Sense of Agency" between in-person and remote participants, i.e. *all* participants felt similarly about how much agency they had in controlling their view of the meeting room and others' view of them in the room. However, we see a significant difference in agency between in-person and remote participants for Gallery condition, as well as the agency for the remote participants between Hybridge and Gallery conditions (i.e. sense of agency was significantly higher in Hybridge than in Gallery).

6.3 Considerations for hybrid meeting system design

Tailored interfaces for each endpoint. One of the central implicit design tenets that we adopted in designing Hybridge was considering the needs of meeting participants at each endpoint individually, and designing an appropriate interface for them to engage with the meeting and other participants. While this allowed us to create tailored interfaces that created an equitable experience for all participants, it took substantial work to implement the technology (despite the relatively straightforward UX features). It is understandable why commercial solutions emphasize symmetrical/uniform interfaces: they streamline development, enable deployment at scale (both on a large number of different devices and in large quantities of participants), reduce system training/learning time, and reduce likelihood of burden on IT services. While deploying a commercially viable system at scale that is tailored to each endpoint will require a change in mindset and substantial work,

the benefits of inclusive hybrid meetings that engage more participants might outweigh the initial engineering time cost.

Application of Emerging Technology. Emerging technologies like AI and spatial computing offer opportunities to enhance the scalability and extensibility of Hybridge-like meeting systems. For instance, AI-driven solutions can enable the creation of dynamic UIs that adapt to the unique requirements of different meeting endpoints. The use of AI not only aids in customizing the participant experience based on specific participant profiles and environmental factors but also contributes to the system's overall efficiency and effectiveness in real-time adjustments—ultimately extending the possibilities of the kind of system we have shown in this paper by allowing even each remote endpoint to be different. Work on adaptive meeting UIs is already ongoing, and can be applied to hybrid meeting systems as well.

Additionally, advanced technology such as Virtual Reality (VR) and Augmented Reality (AR) can enrich the immersive aspect of hybrid meetings. For instance, VR can offer remote participants a more tangible and interactive experience of the physical meeting space, transcending traditional video conferencing limitations. This not only enhances the sense of presence for remote participants but also fosters a more inclusive and engaging meeting environment.

7 Limitations and Future Investigations

Our study demonstrated the potential of the Hybridge design principles for creating equitable hybrid meetings. However, as this is an early stage in the development of these concepts and technologies, we acknowledge several limitations in this preliminary investigation.

Collaborative work material. Many meetings in knowledge work rely upon working with shared content. While our study involved a screen shared by the moderator, our procedure did not involve any participant sharing their screen. Additionally, the interaction with the shared screen was passive—meeting participants only consumed the information, rather than create new artefacts together, such as working on a document. Anecdotally, this covers the most frequent use of shared content, where the organizer shares their screen with others and goes through the content. Future work should investigate multi-person content sharing from both remote and in-room participants in ways that provide agency for each endpoint to see shared content as best fits their needs. This should also include integrating agency around sharing of physical objects from each endpoint. How could we enable participants to smoothly transition from collaborating on a document to talking with each other? How could we maintain spatial relationships between meeting participants, lost in typical collaborative editing software? One approach could be exploring a "*collaboration view*" in addition to the *seated* and *room view*.

Meeting dynamics. Our 4-participant sessions used a specific configuration (2 in-room and 2 remote) and focused on discussion sessions around an assigned task. Further, each group only used Hybridge for a single session, whereas they were much more familiar with in-person or grid-based video meetings. We feel that our results are representative for a broad class of meetings—majority of video conferencing meetings involve six or fewer people[14], and an open discussion format (theoretically) allows any participant to contribute. However, Hybridge requires further research to assess how Hybridge affects meeting dynamics and level of participation in larger meetings. Additionally, longitudinal evaluation of how Hybridge might impact meeting practices and conventions will be particularly valuable.

Physical space. While our remote interface could be used on any computer in any space, our in-room setup was tied to the size and configuration of the meeting room. We fit three physical displays among the chairs around the table, and had another large display on one end of the table.

Different sized rooms, furniture, and displays might change the way that meeting participants experience Hybridge—just as it changes the dynamics of fully in-person meetings. Further work is needed to determine the appropriate balance of scale and agency—for instance, accommodating more remote people on a single larger screen in the physical space would allow more remote participants, while giving up some individual agency and a unique location in the meeting room. How does having larger screens with multiple people in a grid layout impact the remotes' agency and the local participant's feeling of the remotes being with them?

Participant diversity. The study was limited to participants from a university (n=24) and from a large tech organization (n=28). Additionally, we had 34 males and 18 females in our participant pool. Although we believe our findings are broadly applicable to information workers, more diverse sampling and inclusion of participants from varied cultural and organizational backgrounds as well as higher gender diversity would enhance the generalizability of our results.

8 Conclusion

Hybrid meetings are here to stay, and only getting increasingly popular with the rise of work-from-home. The Hybridge system is an early attempt at addressing the problems of (1) a lack of shared spatiality and (2) and imbalance in the level of agency between remote and in-room participants in a hybrid meeting. Hybridge was designed to bridge the gap between remote and in-room meeting participants with two primary design principles: distributed spatial presence and heightened agency for remote participants. Hybridge showed significantly better results compared to traditional Gallery views for reported measures of agency, co-presence, and physical presence for remote participants as well as conversation flow quality in the meeting space. However, there was no significant difference observed between the two systems for ease of awareness. Critically, we observed no difference for agency and a reduced difference for co-presence between in-room and remote participants in Hybridge, suggesting that it leveled the playing field between in-room and remote endpoints compared to the standard Gallery view.

Given the constantly changing technological landscape, we need a system that adapts to the varying needs of different endpoints, including devices, spaces, and people. Our study demonstrated a more equitable experience between remote and in-room participants, challenging the conventional symmetry and uniformity in meeting systems. Tailored interfaces for the remote and room endpoints played a crucial role in creating equitable experiences, albeit with the trade-off of increased implementation complexity. However this complexity could be lessened with the use of new technology like AI that could adapt the meeting interface to in-the-moment needs in real-time.

In conclusion, the Hybridge system shows a significant advancement in hybrid meeting technology, offering a more inclusive and engaging experience for both remote and in-person participants. The insights gained from this study can guide future innovations in the field, ultimately contributing to more effective and equitable hybrid collaboration environments.

Acknowledgments

Huge thanks for assistance with this research goes to the MSR Cambridge tech support team who enabled the complex system used for this study: Nathan Jones, Joshua Oliver, and Ijaz Mohammed. We also thank Robin Moeur, Richard Hughes, Pete Ansell, Nic Marquardt, and Bala Kumaravel for advice on project direction, research approach, and technical implementation. We thank our IRB team for reviewing our protocol and ensuring best practices. We also greatly appreciate the colleagues who helped pilot our system and study protocol, and the anonymous participants who came to our premises in Cambridge, UK for participating in the study.

References

[1] Rahaf Alharbi, John Tang, and Karl Henderson. 2023. Accessibility Barriers, Conflicts, and Repairs: Understanding the Experience of Professionals with Disabilities in Hybrid Meetings. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 605, 15 pages. https://doi.org/10.1145/3544548.3581541

[2] Rachel Bergmann, Sean Rintel, Nancy Baym, Advait Sarkar, Damian Borowiec, Priscilla Wong, and Abigail Sellen. 2022. Meeting (the) Pandemic: Videoconferencing Fatigue and Evolving Tensions of Sociality in Enterprise Video Meetings During COVID-19. *Computer Supported Cooperative Work (CSCW)* (Nov. 2022). https://doi.org/10.1007/s10606-022-09451-6

[3] Frank Biocca and Chad Harms. 2003. Guide to the networked minds social presence inventory v. 1.2. (2003).

[4] Frank Biocca, Chad Harms, and Judee K. Burgoon. 2003. Toward a More Robust Theory and Measure of Social Presence: Review and Suggested Criteria. *Presence* 12, 5 (2003), 456–480. https://doi.org/10.1162/105474603322761270

[5] 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.

[6] 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.

[7] Stéphane Bouchard, Geneviéve Robillard, Julie St-Jacques, Stéphanie Dumoulin, Marie-Josée Patry, and Patrice Renaud. 2004. Reliability and validity of a single-item measure of presence in VR. In *The 3rd IEEE international workshop on haptic, audio and visual environments and their applications*. IEEE, 59–61.

[8] Judee K Burgoon, Valerie Manusov, and Laura K Guerrero. 2021. *Nonverbal communication*. Routledge.

[9] Bill Buxton. 2009. Mediaspace – Meaningspace – Meetingspace. In *Media Space 20 + Years of Mediated Life*, Steve Harrison (Ed.). Springer, London, 217–231. https://doi.org/10.1007/978-1-84882-483-6_13

[10] Herbert H. Clark and Susan E. Brennan. 1991. Grounding in communication. In *Perspectives on socially shared cognition*. American Psychological Association, Washington, DC, US, 127–149. https://doi.org/10.1037/10096-006

[11] Ross Cutler, Yasaman Hosseinkashi, Jamie Pool, Senja Filipi, Robert Aichner, Yuan Tu, and Johannes Gehrke. 2021. Meeting Effectiveness and Inclusiveness in Remote Collaboration. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (April 2021), 1–29. https://doi.org/10.1145/3449247

[12] Mark H.D. Danton and Ian Bushnell. 2023. Zoom and its Discontents: Group Decision Making in Pediatric Cardiology in the Time of COVID (and Beyond). *Journal of Medical Systems* 47, 1 (May 2023), 59. https://doi.org/10.1007/s10916-023-01944-1

[13] Celso M. de Melo, Kangsoo Kim, Nahal Norouzi, Gerd Bruder, and Gregory Welch. 2020. Reducing Cognitive Load and Improving Warfighter Problem Solving With Intelligent Virtual Assistants. *Frontiers in Psychology* 11 (11 2020). https://doi.org/10.3389/fpsyg.2020.554706

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

[15] 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

[16] Melanie Duckert, Louise Barkhuus, and Pernille Bjørn. 2023. Collocated Distance: A Fundamental Challenge for the Design of Hybrid Work Technologies. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 612, 16 pages. https://doi.org/10.1145/3544548.3580899

[17] Carmen Egido. 1988. Video conferencing as a technology to support group work: a review of its failures. In *Proceedings of the 1988 ACM conference on Computer-supported cooperative work (CSCW '88)*. Association for Computing Machinery, New York, NY, USA, 13–24. https://doi.org/10.1145/62266.62268

[18] Lisa A Elkin, Matthew Kay, James J Higgins, and Jacob O Wobbrock. 2021. An aligned rank transform procedure for multifactor contrast tests. In *The 34th annual ACM symposium on user interface software and technology*. 754–768.

[19] Kathleen E. Finn, Abigail J. Sellen, and Sylvia B. Wilbur (Eds.). 1997. *Video-Mediated Communication*. L. Erlbaum Associates Inc., Mahwah NJ, USA.

[20] Ceclia E. Ford. 2008. *Women Speaking Up: Getting and Using Turns in Workplace Meetings*. Palgrave Macmillan, New York.

[21] Erving Goffman. 1963. *Behavior in public places: Notes on the social organization of gatherings*. Free Press.

[22] Charles Goodwin. 2002. Time in Action. *Current Anthropology* 43, S4 (Aug. 2002), S19–S35. https://doi.org/10.1086/339566 Publisher: The University of Chicago Press.

[23] 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* (<conf-loc>, <city>Yokohama</city>, <country>Japan</country>, </conf-loc>) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 451, 13 pages. https://doi.org/10.1145/3411764.3445698

[24] Steve Harrison (Ed.). 2009. *Media Space 20+ Years of Mediated Life*. Springer-Verlag, London. https://doi.org/10.1007/978-1-84882-483-6

[25] Christian Heath and Paul Luff. 1992. Media Space and Communicative Asymmetries: Preliminary Observations of Video-Mediated Interaction. *Human–Computer Interaction* 7, 3 (1992), 315–346. https://doi.org/10.1207/s15327051hci0703_3

[26] Jon Hindmarsh, Mike Fraser, Christian Heath, Steve Benford, and Chris Greenhalgh. 1998. Fragmented Interaction: Establishing Mutual Orientation in Virtual Environments. In *Proceedings of the 1998 ACM Conference on Computer Supported Cooperative Work* (Seattle, Washington, USA) *(CSCW '98)*. Association for Computing Machinery, New York, NY, USA, 217–226. https://doi.org/10.1145/289444.289496

[27] Ramon Hofer, Christoph Ganser, and Andreas Kunz. 2006. MatrixView: extending immersion in video conferencing. (2006), 3 p. https://doi.org/10.3929/ETHZ-A-005713962 Artwork Size: 3 p. Medium: application/pdf Publisher: ETH Zurich.

[28] Ellen A. Isaacs and John C. Tang. 1994. What video can and cannot do for collaboration: A case study. *Multimedia Systems* 2, 2 (Aug. 1994), 63–73. https://doi.org/10.1007/BF01274181

[29] Brennan Jones, Yaying Zhang, Priscilla N. Y. Wong, and Sean Rintel. 2020. VROOM: Virtual Robot Overlay for Online Meetings. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems* (<conf-loc>, <city>Honolulu</city>, <state>HI</state>, <country>USA</country>, </conf-loc>) *(CHI EA '20)*. Association for Computing Machinery, New York, NY, USA, 1–10. https://doi.org/10.1145/3334480.3382820

[30] Brennan Jones, Yaying Zhang, Priscilla N. Y. Wong, and Sean Rintel. 2021. Belonging There: VROOM-Ing into the Uncanny Valley of XR Telepresence. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW1, Article 59 (apr 2021), 31 pages. https://doi.org/10.1145/3449133

[31] Brennan Jones, Yaying Zhang, Priscilla N. Y. Wong, and Sean Rintel. 2021. Belonging There: VROOM-ing into the Uncanny Valley of XR Telepresence. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW1, Article 59 (apr 2021), 31 pages. https://doi.org/10.1145/3449133

[32] Demetrios Karis, Daniel Wildman, and Amir Mané. 2016. Improving Remote Collaboration With Video Conferencing and Video Portals. *Human–Computer Interaction* 31, 1 (Jan. 2016), 1–58. https://doi.org/10.1080/07370024.2014.921506 Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/07370024.2014.921506.

[33] Adam Kendon. 1967. Some functions of gaze-direction in social interaction. *Acta Psychologica* 26 (Jan. 1967), 22–63. https://doi.org/10.1016/0001-6918(67)90005-4

[34] Adam Kendon. 2010. Spacing and Orientation in Co-present Interaction. In *Development of Multimodal Interfaces: Active Listening and Synchrony: Second COST 2102 International Training School, Dublin, Ireland, March 23-27, 2009, Revised Selected Papers*, Anna Esposito, Nick Campbell, Carl Vogel, Amir Hussain, and Anton Nijholt (Eds.). Springer, Berlin, Heidelberg, 1–15. https://doi.org/10.1007/978-3-642-12397-9_1

[35] Kangsoo Kim, Celso M. de Melo, Nahal Norouzi, Gerd Bruder, and Gregory F. Welch. 2020. Reducing Task Load with an Embodied Intelligent Virtual Assistant for Improved Performance in Collaborative Decision Making. *IEEE VR Conference Proceedings* (03 2020). https://doi.org/10.1109/vr46266.2020.00074

[36] Hubert Knoblauch. 2013. Alfred Schutz' Theory of Communicative Action. *Human Studies* 36, 3 (Aug. 2013), 323–337. https://doi.org/10.1007/s10746-013-9278-9

[37] 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

[38] Anastasia Kuzminykh and Sean Rintel. 2020. Classification of Functional Attention in Video Meetings. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376546

[39] 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.

[40] Khanh-Duy Le, Ignacio Avellino, Cédric Fleury, Morten Fjeld, and Andreas Kunz. 2019. GazeLens: Guiding Attention to Improve Gaze Interpretation in Hub-Satellite Collaboration. In *Human-Computer Interaction – INTERACT 2019*, David Lamas, Fernando Loizides, Lennart Nacke, Helen Petrie, Marco Winckler, and Panayiotis Zaphiris (Eds.). Springer International Publishing, Cham, 282–303.

[41] Khanh-Duy Le, Ignacio Avellino, Cédric Fleury, Morten Fjeld, and Andreas M Kunz. 2019. GazeLens: Guiding Attention to Improve Gaze Interpretation in Hub-Satellite Collaboration. *Lecture Notes in Computer Science* (01 2019), 282–303. https://doi.org/10.1007/978-3-030-29384-0_18

[42] Zhengqing Li, Shio Miyafuji, Erwin Wu, Hideaki Kuzuoka, Naomi Yamashita, and Hideki Koike. 2019. OmniGlobe: An Interactive I/O System For Symmetric 360-Degree Video Communication. In *Proceedings of the 2019 on Designing Interactive Systems Conference* (San Diego, CA, USA) *(DIS '19)*. Association for Computing Machinery, New York, NY, USA, 1427–1438. https://doi.org/10.1145/3322276.3322314

[43] Matthew Lombard, Frank Biocca, Jonathan Freeman, Wijnand IJsselsteijn, and Rachel J Schaevitz. 2015. *Immersed in media: Telepresence theory, measurement & technology*. Springer.

[44] Matthew Lombard and Theresa Ditton. 1997. At the Heart of It All: The Concept of Presence. *Journal of Computer-Mediated Communication* 3, 2 (09 1997), JCMC321. https://doi.org/10.1111/j.1083-6101.1997.tb00072.x

[45] Paul Luff, Christian Heath, Hideaki Kuzuoka, Jon Hindmarsh, Keiichi Yamazaki, and Shinya Oyama. 2003. Fractured Ecologies: Creating Environments for Collaboration. *Human–Computer Interaction* 18, 1-2 (2003), 51–84. https://doi.org/10.1207/S15327051HCI1812_3

[46] Paul K. Luff, Naomi Yamashita, Hideaki Kuzuoka, and Christian Heath. 2015. Flexible Ecologies And Incongruent Locations. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. Association for Computing Machinery, New York, NY, USA, 877–886. https://doi.org/10.1145/2702123.2702286

[47] Vassiliki Markaki and Lorenza Mondada. 2012. Embodied orientations towards co-participants in multinational meetings. *Discourse Studies* 14, 1 (Feb. 2012), 31–52. https://doi.org/10.1177/1461445611427210

[48] Lorenza Mondada. 2009. Emergent focused interactions in public places: A systematic analysis of the multimodal achievement of a common interactional space. *Journal of pragmatics* 41, 10 (2009), 1977–1997.

[49] Lorenza Mondada. 2011. The interactional production of multiple spatialities within a participatory democracy meeting. *Social Semiotics* 21, 2 (April 2011), 289–316. https://doi.org/10.1080/10350330.2011.548650 Publisher: Routledge _eprint: https://doi.org/10.1080/10350330.2011.548650.

[50] Lorenza Mondada. 2013. Embodied and spatial resources for turn-taking in institutional multi-party interactions: Participatory democracy debates. *Journal of Pragmatics* 46, 1 (Jan. 2013), 39–68. https://doi.org/10.1016/j.pragma.2012.03.010

[51] Thomas Neumayr, Mirjam Augstein, and Bettina Kubicek. 2022. Territoriality in Hybrid Collaboration. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2, Article 332 (nov 2022), 37 pages. https://doi.org/10.1145/3555224

[52] 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. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW (Nov. 2018), 128:1–128:24. https://doi.org/10.1145/3274397

[53] Carman Neustaedter, Steve Harrison, and Abigail Sellen (Eds.). 2013. *Connecting Families: The Impact of New Communication Technologies on Domestic Life*. Springer, London. https://www.microsoft.com/en-us/research/publication/connecting-families-the-impact-of-new-communication-technologies-on-domestic-life/

[54] David Nguyen and John Canny. 2005. MultiView: spatially faithful group video conferencing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '05)*. Association for Computing Machinery, New York, NY, USA, 799–808. https://doi.org/10.1145/1054972.1055084

[55] David T. Nguyen and John Canny. 2007. Multiview: improving trust in group video conferencing through spatial faithfulness. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '07)*. Association for Computing Machinery, New York, NY, USA, 1465–1474. https://doi.org/10.1145/1240624.1240846

[56] Loi Anh Nguyen, Rebecca Evan, Sanghamitra Chaudhuri, Marcia Hagen, and Denise Williams. 2023. Inclusion in the workplace: an integrative literature review. *European Journal of Training and Development* ahead-of-print, ahead-of-print (Jan. 2023). https://doi.org/10.1108/EJTD-10-2022-0104

[57] Catherine S. Oh, Jeremy N. Bailenson, and Gregory F. Welch. 2018. A Systematic Review of Social Presence: Definition, Antecedents, and Implications. *Frontiers in Robotics and AI* 5 (2018), 114. https://doi.org/10.3389/frobt.2018.00114

[58] Kenton O'hara, Jesper Kjeldskov, and Jeni Paay. 2011. Blended Interaction Spaces for Distributed Team Collaboration. *ACM Trans. Comput.-Hum. Interact.* 18, 1 (May 2011), 3:1–3:28. https://doi.org/10.1145/1959022.1959025

[59] Kazuhiro Otsuka. 2016. MMSpace: Kinetically-augmented telepresence for small group-to-group conversations, In 2016 IEEE Virtual Reality (VR). *IEEE Virtual Reality (VR)*, 19–28. https://doi.org/10.1109/VR.2016.7504684

[60] 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.

[61] Roger Pye and Ederyn Williams. 1977. Teleconferencing: is video valuable or is audio adequate? *Telecommunications Policy* 1, 3 (jun 1977), 230–241. https://doi.org/10.1016/0308-5961(77)90027-1

[62] 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.

[63] Sean Rintel. 2007. Maximizing environmental validity: remote recording of desktop videoconferencing. Springer-Verlag, 1772593, 911–920.

[64] Karen Ruhleder and Brigitte Jordan. 2001. Co-Constructing Non-Mutual Realities: Delay-Generated Trouble in Distributed Interaction. 10, 1 (jan 2001), 113–138. https://doi.org/10.1023/A:1011243905593

[65] 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.

[66] 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.

[67] Helen B Schwartzman. 1989. *The meeting: Gatherings in organizations and communities*. Springer Science & Business Media.

[68] Lucas M. Seuren, Joseph Wherton, Trisha Greenhalgh, and Sara E. Shaw. 2021. Whose turn is it anyway? Latency and the organization of turn-taking in video-mediated interaction. *Journal of Pragmatics* 172 (Jan. 2021), 63–78. https://doi.org/10.1016/j.pragma.2020.11.005

[69] Lynn M. Shore, Jeanette N. Cleveland, and Diana Sanchez. 2018. Inclusive workplaces: A review and model. *Human Resource Management Review* 28, 2 (June 2018), 176–189. https://doi.org/10.1016/j.hrmr.2017.07.003

[70] Mel Slater, Amela Sadagic, Martin Usoh, and Ralph Schroeder. 2000. Small-group behavior in a virtual and real environment: A comparative study. *Presence* 9, 1 (2000), 37–51.

[71] Mel Slater and Sylvia Wilbur. 1997. A Framework for Immersive Virtual Environments (FIVE): Speculations on the Role of Presence in Virtual Environments. *Presence: Teleoperators and Virtual Environments* 6, 6 (12 1997), 603–616. https://doi.org/10.1162/pres.1997.6.6.603

[72] Becky Spittle, Payod Panda, Lev Tankelevitch, Kori Inkpen, John Tang, Sasa Junuzovic, Qianqian Qi, Patrick Sweeney, Andrew D. Wilson, William Buxton, Abigail Sellen, and Sean Rintel. 2024. Comparing the Agency of Hybrid Meeting Remote Users in 2D and 3D Interfaces of the Hybridge System. In *Extended Abstracts of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA, May 11–16, 2024) *(CHI EA '24)*. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3613905.3651103

[73] D. Sandy Staples and Lina Zhao. 2006. The Effects of Cultural Diversity in Virtual Teams Versus Face-to-Face Teams. *Group Decision and Negotiation* 15 (07 2006), 389–406. https://doi.org/10.1007/s10726-006-9042-x

[74] Jonathan Steuer. 2006. Defining Virtual Reality: Dimensions Determining Telepresence. *Journal of Communication* 42, 4 (02 2006), 73–93. https://doi.org/10.1111/j.1460-2466.1992.tb00812.x

[75] Robert Stults. 1986. *Media Space*. Technical Report. Xerox PARC. https://www.academia.edu/44010741/Media%5FSpace%5FXerox%5FPARC%5F1986?auto=download

[76] 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 351 (oct 2023), 25 pages. https://doi.org/10.1145/3610200

[77] Cameron Teoh, Holger Regenbrecht, and David O'Hare. 2012. How the Other Sees Us: Perceptions and Control in Videoconferencing. In *Proceedings of the 24th Australian Computer-Human Interaction Conference* (Melbourne, Australia) *(OzCHI '12)*. Association for Computing Machinery, New York, NY, USA, 572–578. https://doi.org/10.1145/2414536.2414624

[78] Jason Turowetz, Matthew M. Hollander, and Douglas W. Maynard. 2016. Ethnomethodology and Social Phenomenology. In *Handbook of Contemporary Sociological Theory*, Seth Abrutyn (Ed.). Springer International Publishing, Cham, 387–410. https://doi.org/10.1007/978-3-319-32250-6_19

[79] Martin Usoh, Ernest Catena, Sima Arman, and Mel Slater. 2000. Using presence questionnaires in reality. *Presence* 9, 5 (2000), 497–503.

[80] Amy Voida, Stephen Voida, Saul Greenberg, and Helen Ai He. 2008. Asymmetry in media spaces. In *Proceedings of the 2008 ACM conference on Computer supported cooperative work*. ACM, San Diego CA USA. https://doi.org/10.1145/1460563.1460615

[81] 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.

[82] Nelson Wong and Carl Gutwin. 2014. Support for Deictic Pointing in CVEs: Still Fragmented after All These Years'. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing* (Baltimore, Maryland, USA) *(CSCW '14)*. Association for Computing Machinery, New York, NY, USA, 1377–1387. https://doi.org/10.1145/2531602.2531691

A Prototype System Implementations

The Hybridge prototype was a complex system involving many moving parts. In this section, we first introduce the high-level Hybridge system ecosystem, followed by the detailed apparatus used for the in-room and remote endpoints to realize the ecosystem.

A.1 The Hybridge software ecosystem

Refer to Figure 11 for a diagram of the various system-level components in the Hybridge set up. In order to enable the kinds of interactions that we wanted to implement, we needed a way to create our own meeting roster that could be kept up-to-date across all machines. To this end, underneath the interactive software we have a Seating Server that maintains an updated meeting roster and seat assignments in the meeting.

Broadly, each "visible" machine (i.e. seen / interacted with by a meeting participant—the two remote machines, and the three digital seats in the meeting room) in the Hybridge setup requires a similar set of software:

(1) **Microsoft Teams**: Rather than build our own A/V stack, we rely upon an off-the-shelf software, Microsoft Teams[2], to manage the transmission of audio-video content. This has the advantage of us not having to solve hard video communication challenges like compression, managing sync issues between audio and video etc. We take advantage of the NDI transport offered by Teams—all machines join the same Teams call and broadcast their A/V locally via NDI. These streams are picked up by the Hybridge front-end software (described below) and rendered appropriately.

(2) **Hybridge prototype**: The Hybridge front-end prototype is built with the Unity3D game engine. On a system level, this software serves two roles on both the remote and in-room machines (in addition to enabling all the interactions described in subsection 3.2): (1) first, it communicates to the Hybridge seating server to determine the current roster, and updates the roster when there is a seat change event, and (2) second, it renders the appropriate

[2]Microsoft Teams (Last accessed 22nd April 2024) - https://www.microsoft.com/en-gb/microsoft-teams/group-chat-software

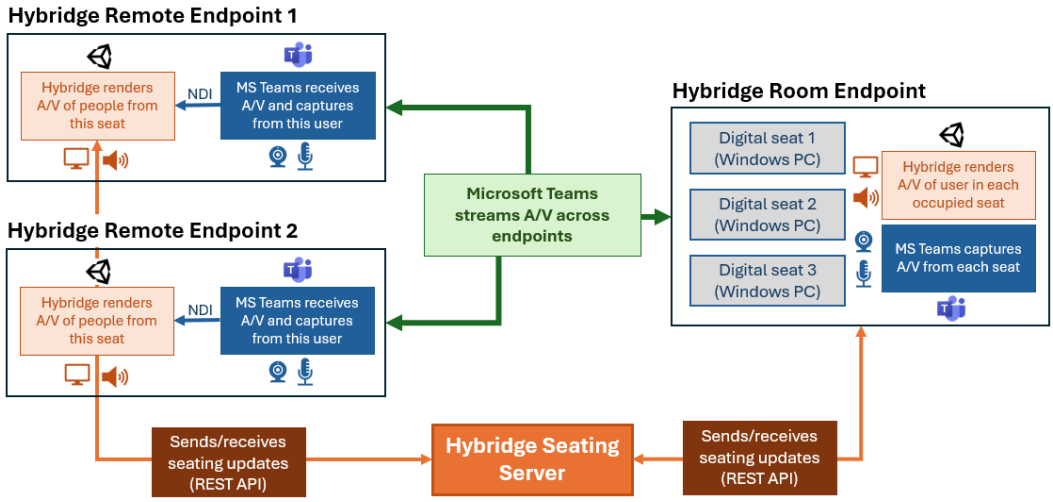


Fig. 11. System diagram for the Hybridge prototype showing connections between various parts of the system.

A/V streams on the audio-visual display connected to the machine. This involves selecting the current set of streams to render, and managing the camera pan angles for the remote participants.

A.2 Local meeting room apparatus

A.2.1 Basic setup. The local meeting room had a conference table in the center, with a large screen at one end for the content. This content screen setup included:

- Large content screen (52" monitor),
- stereo speakers to render moderator audio, and
- 180°wide-angle webcam (Jabra PanaCast) to capture the room (for gallery condition).

We provided one more chair around the table than the number of in-room participants in each condition—so, five chairs for in-person condition (for four in-room participants), and three chairs each for gallery and Hybridge conditions (for two in-room participants). These chairs were relatively evenly spaced around the available space around the table.

A.2.2 Digital seats. In addition to the above basic setup, the Hybridge condition also had three digital seats (for two remote participants). See Figure 1(d) for the Hybridge room layout in the digital twin of the physical room. Each of the digital seats comprised of:

(1) a Windows PC (Intel i7-12700K, 16GB RAM, NVIDIA GeForce RTX 3060 Ti), running:
   - Hybridge prototype,
   - Microsoft Teams,
   - software to enable spatial audio (see below), and
   - software for recording (OBS Studio [3] and Streamer.bot [4]).
(2) 27" display in portrait orientation (Dell U27), to render remote participant's camera,
(3) stereo speakers to render remote participant's audio, and
(4) 180°wide-angle webcam (Jabra PanaCast) to capture the view of the room from this seat.

A.2.3 Spatial audio. In order to enable spatial audio in the remote Hybridge endpoint, we needed to record isolated audio from in-room participants. To this end, each of the in-room participants was recorded through a unidirectional microphone (JK MIC-J 069). Each audio feed was sent to a separate machine (e.g., audio from Physical Seat 1 was sent to the Windows PC used for Digital Seat 1), and was processed through additional software. This path diagram is captured in Figure 12.

- We used OBS Studio to take the raw audio input from the mic, and route it through an OBS audio filter plugin (gate) to remove background noise and cross-talk.
- This gated audio from OBS was fed into a virtual microphone, created using VB-CABLE Virtual Audio Device [5].
- The virtual microphone device from VB-CABLE was broadcast on the network using NDI [6]. These individual audio streams could then be captured by the remote endpoints, paired to their respective locations in the virtual room, hence creating spatial audio for the remote Hybridge participants.

[3]OBS Studio (Open Broadcaster Software)(Last accessed 22nd April 2024) - https://obsproject.com/
[4]Streamer.bot (Last accessed 22nd April 2024) - https://streamer.bot
[5]VB-CABLE Virtual Audio Device (Last accessed 22nd April 2024) - https://vb-audio.com/Cable/
[6]NDI (Last accessed 22nd April 2024) - https://ndi.video/tools/ndi-core-suite/

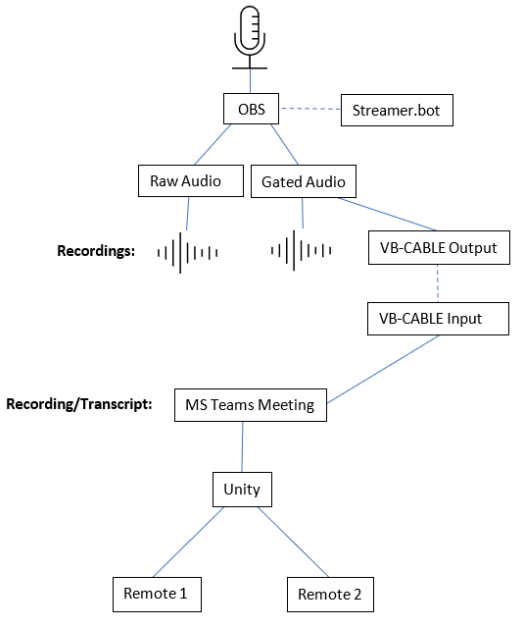


Fig. 12. Path to provide spatial audio in Hybridge.

A.3 Remote / satellite room apparatus

The two remote rooms allowed a single participant each to connect to the meeting room through the Hybridge remote interface. The hardware and software setup was similar to the digital seats in the local meeting room. Both remote rooms had:

- a Windows PC (Intel i7-12700K, 16GB RAM, NVIDIA GeForce RTX 3060 Ti), running:
  - Hybridge prototype,
  - Microsoft Teams, and
  - software for recording (OBS Studio and Streamer.bot).
- 27 inch monitor in landscape orientation (Dell U27),
- a unified speaker and microphone (Jabra Speak 710)–only the mic was used here, and
- headphones (Microsoft Teams certified headphones) to render spatial audio.

A.4 Moderator apparatus

The meeting moderator, part of the research team, had their own machine that allowed them to share their screen with the meeting participants. This allowed them to control the contents on the shared screen, as well as communicate the instructions to the participants for each task. Additional software (Streamer.bot) allowed the moderator to remotely control the A/V recordings from the remote satellite rooms and the digital seats in the local meeting room. As illustrated in Figure 12, Streamer.bot was used to remotely start and stop audio and video recordings being captured by all machines in OBS (using Websockets). This was to ease the logistics of running a distributed study. A gate audio effect was also applied in OBS on each of the computers within the meeting room. Gate settings could be updated on the fly using Streamer.bot, and tweaks could be easily made for individual participants. This helped to deliver effective spatial audio, as the ideal settings needed to isolate the speech from background noise are dependent on the nature of the voice input.

The virtual cable allowed for gated audio to be routed from OBS into the Microsoft Teams meeting, before being sent alongside visuals to the Hybridge application via NDI, as shown in Figure 11.

B Discussion tasks

For the three conditions, the following discussion topics were presented during each session in a counter-balanced order.

B.1 Survival on the moon

You are part of a 4-member team traveling to the station on the moon. Something is wrong with your navigation system, so you land safely, but 80km away from the station. Your survival depends on reaching the station, protecting yourself until someone finds you, or meeting a rescue party on the way to the station.

The moon has no atmosphere and no magnetosphere. Gravity is only 1/6 as strong as Earth's. The soil is a mixture that includes sharp, glassy particles. More than 80% of the moon is made of heavily cratered highlands. Temperatures vary widely, from -193°C to 111°C depending on time and location.

The next slide contains 11 items available. Rank the top 3 items in order of importance for the survival of you and your crew. You have 5 minutes to decide and write down your choices as a group.

- Magnetic compass
- Life raft
- Parachute silk
- Space suit repair kit
- Nylon rope (15m)
- First aid kit
- Signal mirror
- Solar-powered lights
- Space blanket
- Map of moon's surface
- Solar-powered radio receiver-transmitter

B.2 Survival in the desert

Your airplane has crash-landed in the desert in southwestern US. You and 3 others have survived unharmed. Before you crashed, you heard the pilot say that you are 110km away from the nearest town. Your survival depends on reaching the town or protecting yourself until someone finds you.

The immediate area is quite flat and appears barren, except for occasional cacti. The last weather report you heard said temperatures would reach 54°C. You are dressed in light-weight clothing—short-sleeved shirt, trousers, socks, and shoes.

The next slide contains 11 items available to you. Rank the top 3 items in order of importance to you and the others to give you the best chance of survival. You have 5 minutes to decide and write down your choices as a group.

- Sectional air map for area
- Two pairs of sunglasses
- Plastic raincoat
- Compress kit with gauze
- Magnetic compass
- Red and white parachute
- Flashlight
- Book: "Edible Animals for the Desert"
- .45 Caliber pistol (loaded)
- Bottle of salt tablets (1000)
- Jack knife

B.3 Spy mission task

Your team has been assigned a mission to acquire a classified document from a high-security building on a small remote island. The document contains vital information that could put lives at risk if it falls into the wrong hands.

You have been dropped off in a covert location near the target building. The building is protected by trained guards and equipped with state-of-the-art security systems, including motion sensors and infrared cameras. Time is of the essence, but if you get caught you will fail the mission.

The next slide lists 11 items that are available. Rank the top 3 items in order of importance to give you the best chance of completing the mission. You have 5 minutes to decide and write down your choices as a group.

- EMP device
- Lockpicking set
- Camping tent
- Nano-drones
- Infrared stealth suit
- Remote hacking device
- Binoculars
- Two grappling hooks and climbing gear
- Night vision goggles
- Decoy explosives
- Concealed body armour

C Focus Group Guide

Participants were invited to share their views and engage in discussions on the following points:

(1) What did you like best about each meeting type?
(2) What needs the most improvement in each meeting type?
(3) In what order did everyone rank the meeting types?
(4) Would you find the spatial arrangement of remote participants in the Hybridge prototype useful for other hybrid meetings you attend? If so, which ones, why? If not, why not?

D Study Instruments

After each condition, participants completed survey items about their sense of agency, co-presence, physical presence, conversational flow, and ease-of-awareness. By *agency*, we mean a participant's sense of control over their engagement in the meeting, including presentation of self [37, 77], one's view of others, and one's position in the meeting with respect to others [23, 39, 76, 77] (see Table 2 for all items). To measure co-presence, we adapted eight items from Networked Minds questionnaire [3] and one item from [70] (see Table 3). To measure physical presence (for remote participants only), we adapted 5 item from the Slater-Usoh-Steed (SUS) questionnaire [79], and the one-item measure from [7], adapting them to refer to presence in the *physical* meeting room (see Table 4). We measured the ease of turn-taking or 'conversation flow' using a 6-item questionnaire, adapted from [76] (see Table 5). Relatedly, we also measured participants' ability to be aware of people and activities throughout the meeting using a 4-item questionnaire (see Table 6). Finally, to measure physical presence, we used 5 items from the Slater-Usoh-Steed (SUS) questionnaire [79] After both conditions, we also asked participants which one they preferred and why.

Received January 2024; revised April 2024; accepted May 2024

Table 2. **Sense of agency**: survey items and summary scores indicating *mean (standard deviation)*

	Gallery		Hybridge		Fully in-person	
	Remote	Local	Remote	Local	Remote	Local
How much in control did you feel of *your view of the physical conference room?*	2.25 (1.33)	5.75 (1.12)	5 (1.49)	5.55 (1.43)	6.05 (1.23)	5.8 (1.51)
How much in control did you feel of *other people's view of you in the physical conference room?*	2.7 (1.95)	4.1 (1.48)	4.15 (1.73)	4.25 (1.52)	5.4 (1.67)	5.65 (1.6)
How much in control did you feel of *your position in the physical conference room?*	1.65 (0.99)	5.7 (1.34)	4.55 (1.57)	5.05 (1.47)	5.55 (1.54)	5.7 (1.38)
Compared to the people physically in the room, did you feel you had less or more control over *your view of the physical conference room?*	1.75 (1.21)		3.6 (1.57)			
Compared to the people physically in the room, did you feel you had less or more control over *others' view of you in the physical conference room?*	2.75 (1.48)		3.8 (1.24)			
Compared to the people physically in the room, did you feel you had less or more control over *your position in the physical conference room?*	1.9 (1.29)		4 (1.59)			
To what extent did you feel you had a place in the physical conference room?	3.25 (1.68)	6.25 (1.12)	5 (1.26)	6.25 (1.07)	6.4 (0.94)	6.3 (1.03)

Items 1-3 and 8 used a seven-point response scale from 'Not at all' to 'Very much so'. Items 4-7 used a seven-point response scale from 'Much less' to 'Much more'.

Table 3. **Co-presence**: survey items and summary scores indicating *mean (standard deviation)*

	Gallery		Hybridge		Fully in-person	
	Remote	Local	Remote	Local	Remote	Local
I often felt as if the people physically in the room and I were together.	2.4 (0.99)	4.1 (1.21)	3.6 (1.05)	4.45 (0.76)	4.75 (0.55)	4.75 (0.44)
I often felt as if the people who joined remotely and I were together.	2.65 (1.23)	2.5 (1.05)	3.35 (1.31)	4.15 (0.67)		
I think the people physically in the room often felt as if we were together.	2.95 (1.15)	4.35 (0.75)	3.65 (1.04)	4.4 (0.94)	4.7 (0.66)	4.7 (0.47)
I think the people who joined remotely often felt as if we were together.	2.65 (1.23)	2.7 (1.03)	3.45 (1.43)	3.5 (1)		
I was often aware of the people physically in the room during the meeting.	3.55 (1.19)	4.6 (0.6)	4.25 (0.55)	4.65 (0.59)	4.75 (0.44)	4.85 (0.37)
I was often aware of the people who joined remotely during the meeting.	3.7 (1.08)	4 (0.92)	3.8 (0.83)	4.25 (0.55)		
The people physically in the room were often aware of me during the meeting.	3.5 (0.76)	4.3 (0.8)	4.2 (0.62)	4.65 (0.59)	4.55 (0.83)	4.5 (0.76)
The people who joined remotely were often aware of me during the meeting.	3.35 (1.18)	3.75 (0.91)	3.65 (1.04)	3.9 (0.91)		
Please rate how closely your sense of being together with others in a fully in-person meeting resembles your sense of being with them in this hybrid meeting.	3.4 (1.6)	3.6 (1.6)	4.8 (1.44)	5.35 (0.88)		

Items 1-8 used a five-point response scale from 'Strongly Disagree' to 'Strongly Agree'. Item 9 used a seven-point response scale from 'Not at all' to 'Very much so'.

Table 4. **Physical presence**: survey items and summary scores indicating *mean (standard deviation)*

	Gallery	Hybridge
Please rate your sense of "being there" in the physical conference room, on a scale of 1 to 7, where 7 represents your normal experience of being in a place. I had a sense of being there in the physical conference room... *(7-point scale: 'Not at all' to 'Very much')*	0 (0)	0.4 (0.5)
To what extent were there times during the meeting when you felt that you were in the physical conference room? There were times during the experience when the physical conference room was the reality for me... *(7-point scale: 'At no time' to 'Almost all the time')*	0 (0)	0.25 (0.44)
When you think back to the meeting, do you think of the physical conference room more as images that you saw or more as somewhere that you were visiting? The physical conference room seems to me to be more like... *(7-point scale: 'Images that I saw' to 'Somewhere that I visited')*	0 (0)	0.2 (0.41)
During the meeting, which was the strongest on the whole: your sense of being in the physical conference room or of being elsewhere? I had a stronger sense of... *(7-point scale: 'Being elsewhere' to 'Being in the physical conference room')*	0.05 (0.22)	0.25 (0.44)
During the meeting, did you often think to yourself that you were actually in the physical meeting room? During the meeting, I often thought that I was really in the physical conference room... *(7-point scale: 'Not very often' to 'Very much so')*	0 (0)	0.15 (0.37)
To what extent did you feel present in the physical conference room, as if you were really there? *(7-point scale: 'Not at all' to 'Very much')*	0 (0)	0.15 (0.37)

Table 5. **Conversational flow**: survey items and summary scores indicating *mean (standard deviation)*

	Gallery		Hybridge		Fully in-person	
	Remote	Local	Remote	Local	Remote	Local
It was easy to tell when other people wanted to speak.	3.2 (1.01)	3 (1.21)	3.6 (1.05)	3.75 (0.91)	4.7 (0.57)	4.45 (0.6)
It was easy to jump in when I wanted to say something.	2.65 (0.93)	3.6 (1.1)	3.5 (1.05)	4.1 (0.79)	4.75 (0.55)	4.6 (0.5)
There were many inappropriate interruptions.	2.9 (0.97)	3.3 (1.08)	3.15 (1.14)	3.95 (1)	4.55 (0.76)	4.35 (0.93)
The flow of the conversation felt natural.	3 (1.17)	3.25 (1.25)	3.55 (1.23)	3.95 (0.94)	4.6 (0.6)	4.9 (0.31)
There were not many uncomfortable pauses.	3.35 (1.23)	3.1 (1.29)	3.7 (0.98)	3.95 (1.15)	4.45 (0.83)	4.35 (0.93)
I could easily refer to or direct comments to others in the conversation.	2.8 (1.2)	3.65 (1.23)	3.85 (0.93)	3.9 (1.02)	4.5 (0.61)	4.6 (0.68)

Items used a five-point response scale from 'Strongly Disagree' to 'Strongly Agree'

Table 6. **Ease of awareness**: survey items and summary scores indicating *mean (standard deviation)*

	Gallery		Hybridge		Fully in-person	
	Remote	Local	Remote	Local	Remote	Local
How easy was it to *see everyone you needed to throughout the meeting?*	4.8 (2.04)	5.7 (1.13)	4.8 (1.74)	5 (1.49)	6.35 (1.14)	6.65 (0.67)
How easy was it to *see everything going on in the meeting all the time?*	4.4 (1.93)	5.25 (1.37)	4.2 (1.82)	4.9 (1.45)	6.35 (1.18)	6.75 (0.55)
How easy was it to *stay oriented on what was happening where throughout the meeting?*	4.3 (1.53)	5.15 (1.5)	4.85 (1.69)	5.7 (1.34)	6.4 (0.75)	6.75 (0.55)
How easy was it to *see everything you needed to track what was going on throughout the meeting?*	4.4 (1.67)	5.5 (1.19)	4.4 (1.88)	5.5 (1.19)	6.15 (1.04)	6.7 (0.47)

Items used a seven-point response scale from 'Not at all easy' to 'Very easy'.