MirrorBlender: Supporting Hybrid Meetings with a Malleable Video-Conferencing System

Preprint · February 2021

DOI: 10.1145/3411764.3445698

4 authors:

Jens Emil Grønbæk
University of Melbourne

Banu Saatçi
University of Milan

Carla Griggio
Aalborg University

Clemens Nylandsted Klokmose
Aarhus University

MirrorBlender: Supporting Hybrid Meetings with a Malleable Video-Conferencing System

Jens Emil Grønbæk
Computer Science, Aarhus University
Aarhus, Denmark
jensemil@cs.au.dk

Banu Saatçi
Digital Design and Information Studies, Aarhus University
Aarhus, Denmark
banu.saatci@cc.au.dk

Carla F. Griggio
Digital Design and Information Studies, Aarhus University
Aarhus, Denmark
carla@cc.au.dk

Clemens Nylandsted Klokmose
Digital Design and Information Studies, Aarhus University
Aarhus, Denmark
clemens@cavi.au.dk

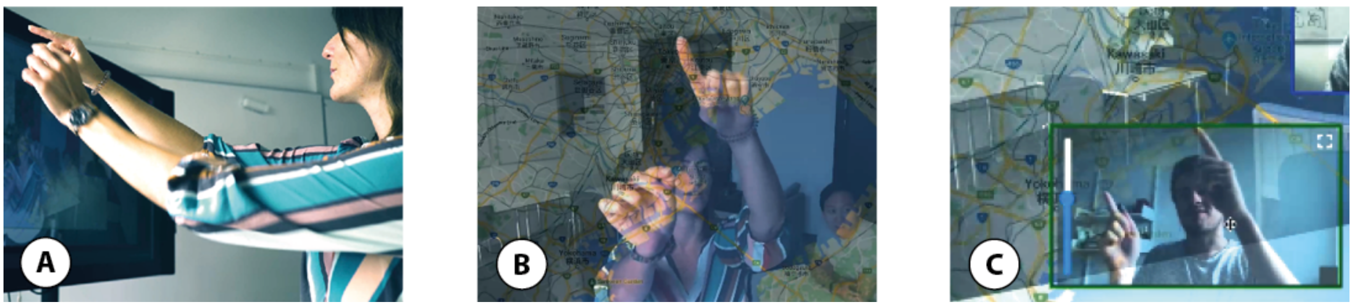

Figure 1: MirrorBlender is a malleable video-conferencing system that creates a WYSIWIS blended space for co-located and remote users. Deploying MirrorBlender in three hybrid meeting sessions, we studied A: deictic gestures, B: embodiment of mirror images, and C: how remote users are included and grab attention via their malleable representation in the interface.

ABSTRACT

In hybrid meetings, multiple co-located participants communicate with remote participants through video. But video communication inhibits non-verbal cues, and this often causes remote participants to feel excluded. To address this issue, we built MirrorBlender: a What-You-See-Is-What-I-See video-conferencing system for blending, repositioning, and resizing mirrors. Mirrors here denote shared video feeds of people and screens. In a qualitative study of MirrorBlender with three hybrid meeting sessions, we found that the shared control of mirrors supported users in negotiating a blended interpersonal space. Moreover, it enabled diverse acts of inclusion of remote participants. In particular, remote participants brought attention to themselves by manipulating the position, scale, and translucency of their camera and screen feeds. Participants also embodied and leveraged their mirror images for deictic gestures and playful interactions. Based on these findings, we discuss new opportunities for supporting video-mediated collaboration.

CHI '21, May 8–13, 2021, Yokohama, Japan

CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in collaborative and social computing**; **Collaborative and social computing systems and tools**.

KEYWORDS

Hybrid meetings, video conferencing, blended interaction spaces, shareable dynamic media, inclusive meetings

ACM Reference Format:
Jens Emil Grønbæk, Banu Saatçi, Carla F. Griggio, and Clemens Nylandsted Klokmose. 2021. MirrorBlender: Supporting Hybrid Meetings with a Malleable Video-Conferencing System. In *CHI Conference on Human Factors in Computing Systems (CHI '21), May 8–13, 2021, Yokohama, Japan.* ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3411764.3445698

1 INTRODUCTION

With remote work and video-mediated communication becoming increasingly common, we are seeing new challenges and opportunities emerge in how people conduct meetings. A popular meeting form is *hybrid meetings* [31, 32], which involve physically co-located and remote participants. Such meetings suffer from reduced non-verbal communication [35] due to a lack of spatial consistency between the physical spaces from where meeting participants connect [5, 29]. Unlike in face-to-face interaction, participants struggle to attract the attention of others through subtle glances and body gestures [14, 15]. This may lead to *primary room dominance* [18] where remote participants feel excluded from the room of co-located participants [31, 32]. To address these issues, Media Spaces (e.g., [17, 26, 29]) and Mixed Reality approaches (e.g., [5, 9, 28]) create a physically realistic frame of reference for a blended interpersonal space. This creates a sense of virtually "being there" together and enables deictic gestures with spatial referencing. However, these approaches dictate fixed spatial configurations of users and require elaborate (and costly) physical hardware setups.

We follow Hollan and Stornetta's suggestion to build telecommunication tools that go "beyond being there" [16] with the goal of developing new principles for supporting non-verbal communication and better inclusion of remote participants in hybrid meetings. Rather than establishing a *physical* frame of reference, we investigate the collaborative qualities of a malleable *virtual* frame of reference. Our work addresses the following research questions:

RQ1 How do participants leverage a malleable blended space for going beyond being there?
RQ2 How can participants use their camera feeds in a virtual frame of reference for deictic gestures?
RQ3 How can we better support inclusion of remote participants in hybrid meetings?

To this end, we built MirrorBlender – a prototype of a malleable video-conferencing system that enables remote and co-located participants to adjust the position, scale and translucency of video feeds in a shared 2D space. We refer to the video feeds as *mirrors*, which can either be *mirrored camera feeds* of participants or *screen mirroring* of shared content. MirrorBlender grants spatial consistency [29] by creating a shared 2D frame of reference in a What-You-See-Is-What-I-See (WYSIWIS) style [36] so that everyone can see the same layout of mirrors. To investigate our research questions, we deployed MirrorBlender in three hybrid meeting sessions each with three co-located and two remote participants.

The paper thus contributes the following:

- A novel approach to malleability in video communication through the manipulation of position, scale, and blending of camera and screen mirrors.
- Identification of opportunities and challenges of enabling embodiment of camera mirrors for deictic gestures.
- Insights into how hybrid meetings can be made more inclusive for remote participants.

2 RELATED WORK

We build upon recent work on hybrid meetings and motivate our study by reviewing research on media spaces and relevant commercial video-conferencing systems.

2.1 Hybrid Meetings

Hybrid meetings are a popular type of video-mediated meetings in business settings which consist of a primary room with many co-located participants and a display that connects to one or more remote endpoints [31, 32]. While this type of setup allows for flexible working conditions, research has identified several shortcomings. Hybrid meetings suffer from primary room dominance [18], which causes co-located participants to constantly interact with each other and forget about the remote participants [31]. Remote participants struggle with limited physical and social awareness and unequal meeting experiences to those who are co-located [40, 41]. Recent research shows that remote participants feel more disadvantaged and excluded in hybrid meetings than in fully virtual ones, in which every participant connects through their own device [32, 35]. Technical limitations of video-mediated communication are often blamed for the low engagement of remote participants and set the future expectations from remote participation in such meetings [24]. Unequal distribution of participants over different endpoints exacerbates the already existing attention problems in video conferencing and makes it difficult for remote participants to receive and pay attention throughout the meeting [23].

In Ens et al.'s recent survey of the landscape of collaborative mixed reality, they highlight hybrid collaboration as one of the central themes for future research [8, pp.90]:

> "[...] further research needs to explore how to support collaboration between team-members who are both co-located and remote, as the physical embodiment of collaborators affects how they can work with one another. A key challenge to address here is to afford all the benefits of collocation while similarly realizing the presence of remote collaborators in ways that all can participate effectively."

Neumayr et al. [27] present a descriptive framework for hybrid collaboration and coupling styles in partially distributed teams with a particular focus on transitions between synchronous and asynchronous work in hybrid settings. They emphasise that future hybrid meeting systems should support territoriality[1] and ways for remote participants to raise awareness. Lastly, Saatçi et al. propose the metaphor of "configuring the meeting" [31] to emphasize that meeting systems should support meeting moderator(s) and participants in setting the meeting needs and goals and adjusting the meeting flow before configuring the software. Their findings point to the need for "making the software adaptable to meeting needs" rather than the other way around. These insights motivate the exploration of ways to reduce primary room dominance and grant more equal control to remote and co-located participants over their hybrid meeting settings.

2.2 Media Spaces and Physical Frames of Reference

The challenges of undermined non-verbal communication occur not only in hybrid meetings, but more generally in meetings via media spaces: spaces of audio, video and computing technologies that connect distributed colleagues to support their work as well as more informal social interactions [6].

When communicating face-to-face, people rely on a shared frame of reference together with deictic expressions such as "here" and "there" when using their bodies to point and gesture around objects in the world [39]. In traditional video-conferencing systems such as Zoom[2] or Skype[3], there is no common frame of reference for pointing to objects or towards other people [29]. While techniques such as telepointers can help users point in remote collaboration [11], our bodies are much more expressive. We express non-verbal

[1] For a discussion of territoriality in HCI see, e.g., [33]
[2] Zoom: http://zoom.us/
[3] Skype: http://www.skype.com/

cues with our hands in the form of diverse gestures which can be deictic (e.g., pointing at an object), iconic (e.g., forming a circle with two fingers to indicate a round object), spatial (e.g., distance), and kinetic (e.g., a bouncing motion) gestures for non-verbal communication [3, 10]. Studies have shown that support for remote gestures improves remote collaborative activities [19, 20]. The challenge of deixis has also been a central issue in early groupware work based on WYSIWIS [36], e.g., for remote pointing [11]. Hybrid meetings complicate remote pointing as co-located participants may not have access to a pointing device but share a large display that may be out of immediate arm's reach [22].

A common approach to supporting gestures and other non-verbal cues is to achieve spatial consistency by creating a physically realistic 3D frame of reference. For example, Morikawa and Maesako create a shared workspace by blending camera feeds from two remote participants into one shared virtual space [26]. Camray connects two wall-sized displays in different rooms where the remote camera feed follows the remote user's position, giving users a sense of walking along the same wall [2]. Holoportation connects two users in separate rooms by reproducing 3D representations of each other in their head-mounted displays [28]. Previous work also explored how to help meeting participants embody the system's frame of reference, such as by aligning spatial referencing with gaze direction [1, 4], by overlapping semi-transparent video feeds [37, 38] or by "blending" different physical workspaces into Blended Interaction Spaces *"where the spatial geometries of the local space continue coherently across the distributed boundary into the remote site, providing the illusion of a single unified space"* [29].

Our work is mainly inspired by ClearBoard [17] and HyperMirror [26], two of the first-generation mixed reality systems. HyperMirror [26] merges video feeds from different physical spaces in a blended space, creating the illusion of a shared mirror. ClearBoard [17] enables face-to-face collaboration on the same shared digital content between two remote users by integrating a remote video feed with a shared drawing canvas, supporting the integration of interpersonal space and shared workspace for remote collaboration.

We believe that hybrid meetings can also benefit from a common frame of reference to better support gestures and other non-verbal cues. However, systems such as HyperMirror and ClearBoard rely on hardware that are costly and hard to adapt to different meeting settings (e.g., a colleague connecting from home or a café), which motivates us to explore cheaper, more flexible software alternatives.

2.3 Video Conferencing and Virtual Frames of Reference

Widely-adopted video communication platforms such as Skype, Zoom, BlueJeans, and Microsoft Teams offer cheap and simple solutions that support hybrid and fully-virtual meetings. Their interaction design is relatively conservative, primarily (and most importantly) focusing on performance, stability, and ease of use. However, the opportunities for using deictic gestures in the conversation are few in these platforms. For example, in an online presentation, the "talking heads" are dislocated from shared screen content and there is no spatial consistency across what two remote users see in their individual views of the meeting.

More recent features and software applications support a sense of "being there" by supporting a shared 2D virtual frame of reference. Microsoft Teams announced the *Together Mode*[4] that masks meeting participants out of their physical backgrounds and places them in a shared virtual background, e.g., in seats in an auditorium. MakeSpace[5] proposes a more malleable alternative, where users can move around avatars and camera feeds in a WYSIWIS 2D meeting canvas. Other solutions use mixed reality for mediating communication, such as placing participant videos on AR markers [5], or connecting through VR spaces, such as Facebook Spaces [9], Rec Room[6], and VRChat[7]. In VR, users are represented by avatars in a virtual world that offers diverse embodied affordances, such as shaking hands to "befriend" another user, sitting in round tables to face others while talking, or blocking others with a "talk to the hand" gesture [25]. Since these systems embody users in the virtual frame of reference where their avatars live, users' sense of location may shift from their real bodies to their representations on the screen [34], transforming "here" and "there" to virtual positions instead of their own surroundings. As these technologies become more widely adopted, we find it important to understand how the embodiment of users in virtual frames of reference supports or hinders social interaction.

We believe that allowing participants to "configure" their hybrid meetings requires their video-conferencing systems to provide means for adaptation and appropriation that allow them to *"make it theirs"* [13]. This motivates us to better understand how a malleable video-conferencing system (i.e., highly adaptable and appropriable) invites participants to build their own frame of reference, how it shapes their social interactions, and how it may support or hinder non-verbal cues such as deictic gestures.

3 MIRRORBLENDER

We developed MirrorBlender to explore how video-conferencing systems can better support non-verbal communication and inclusion of remote participants in hybrid meetings. MirrorBlender is based on the following three design principles: *a shared WYSIWIS frame of reference*, *malleable mirrors*, and *mirror blending*, which we will explicate in the following. The principles are derived from related work and a critique of state-of-the-art in contemporary video-conferencing systems.

3.1 Shared Frame of Reference: WYSIWIS

To enable any form of deixis in a collaborative system, it is necessary to establish a shared frame of reference. MirrorBlender builds on the principle of a WYSIWIS (see Figure 2) where everyone sees the same 2D view. The so-called *WYSIWIS canvas* contains mirror projections of people from webcams and from screen sharing (see Figure 2A) and the visual state of the canvas is shared and synchronized between all the devices in the meeting. These mirrors come from regular devices such as laptops, webcams, and large meeting room displays. Remote meeting participants have a dual-screen setup for viewing their screen content and the WYSIWIS canvas

[4] Microsoft Teams Together Mode: https://news.microsoft.com/innovation-stories/microsoft-teams-together-mode/
[5] MakeSpace: http://makespace.fun
[6] Rec Room: https://recroom.com/
[7] VRChat: https://hello.vrchat.com/

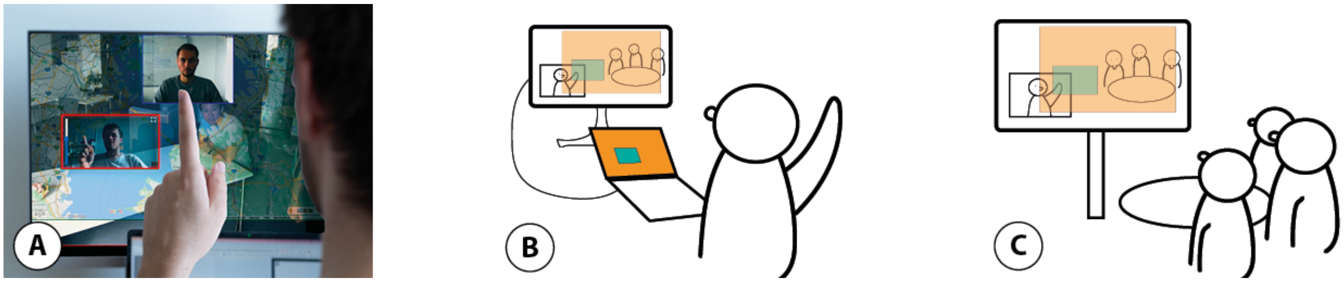

Figure 2: WYSIWIS frame of reference. A-B: Remote participants can point via their mirror image. C: Co-located participants in the primary room see the reference point.

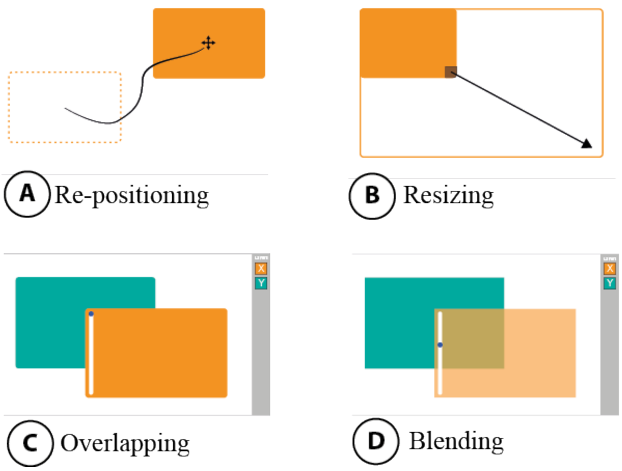

Figure 3: Malleable mirrors. A-B: Mirrors can be dynamically re-positioned and re-scaled. C-D: Mirrors can be overlapped and blended by changing their translucency.

simultaneously (see Figure 2B). They can view the WYSIWIS canvas on the external screen while controlling their own screen content on the laptop screen. Using the representation of their own body in the camera mirror, they can point and gesture with spatial references around the content of a screen mirrored into the canvas. Co-located participants at the other end see the spatial reference in the WYSIWIS canvas on a large touch display (see Figure 2C).

While most current video-conferencing systems lack a shared reference frame and prior research is focused on physically realistic frames, MirrorBlender deliberately enables the creation of unrealistic reference frames (see Figure 4).

3.2 Malleable Mirrors: Continuous Interactions

The principle of malleable mirrors means that mirrors can be manipulated in the WYSIWIS canvas to support non-verbal coordination and continuous negotiation of the shared 2D workspace. Figure 3A-B shows that mirrors can be re-positioned and resized in the canvas. Figure 3C further shows that mirrors can be overlapped, and Figure 3D shows that the degree of translucency can be manipulated with a slider to support adjustments of the blending of overlapping feeds, e.g., to control which mirror should be more prominent and in center of attention.

Current video-conferencing systems provide only limited support for fluidly configuring attention in meetings. Although current systems support turning on/off video or a "raise hand" button, the nature of grabbing attention is still very binary. In hybrid meetings, people in the primary room feed are often too small to allow for reading their body language [32]. The camera feeds of remote participants are often too big and if they turn off video, others lack visual confirmation that they are paying attention [31].

The ability to reconfigure position, scale, and translucency of mirrors freely in a 2D space enables deliberate creation of unrealistic spatial relations between mirrors (see Figure 4). For instance, in hybrid meetings, the camera feed from the primary room can be enlarged to cover most of the WYSIWIS canvas, so that the body language of co-located participants is more easily graspable for remote participants. When configured as the shared background, the remote participants can place their own mirror in it (see Figure 5) to either embed themselves into a shared screen or a physical location in the primary room. Furthermore, the ability to combine continuous input of the features enables rather complex gestures for attention grabbing and deixis. For instance, remote participants can "take the stage" by moving their mirrors to the screen content of discussion, scaling up/down and changing their degree of translucency to adjust the degree with which they take the stage.

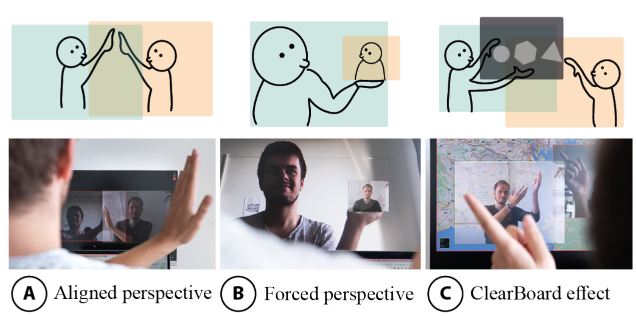

Figure 4: Examples of optical illusions. A: Similar physical appearance. B: Dissimilar physical appearance. C: A screen mirror as a "clearboard" blended with camera mirrors.

3.3 Mirror Blending: Deixis and Embodiment

Mirror blending is the principle of overlapping mirrors and controlling their translucency to create a blended image. Blending of mirrors in the 2D WYSIWIS canvas enables two important illusions:

- Blending of different physical contexts for optical illusions of shared interpersonal space (see Figure 4A-B).
- Embedding of a user's body into both digital and physical contexts for deictic referencing (see Figure 4C).

An example of configuring mirrors for a blended interpersonal space is shown in Figure 4B with the optical illusion of *forced perspective*. This is a well-known effect used in tourist photography, e.g., taking a photo with the Tower of Pisa in the background, where it looks like the tourist is holding the tower with the finger and stopping it from tipping over.

The malleable WYSIWIS canvas can further be used to create effects that support deixis in conversations between co-located and remote participants. In MirrorBlender, people's blended mirror images serve an important role in screen sharing sessions. By mirroring screen content into the canvas and blending the screen feed with the camera feeds, the shared screen can be treated as a transparent whiteboard. In line with Ishii et al.'s ClearBoard [17], meeting participants can now integrate a shared digital workspace with interpersonal space. We refer to this optical illusion as *the ClearBoard effect*. In Figure 4C, it is shown how users can do deictic gestures around shared content by using their own bodies in the mirror. Everyone viewing the WYSIWIS canvas can see the same spatial references as the person doing the gesture.

It is important to be able to view people's unmediated bodies (and especially people's hands) as they provide expressive non-verbal cues in face-to-face communication. Alternative approaches enable embodiment of avatars [9, 12, 34] or shared cursors in a WYSIWIS canvas [11]. However, with MirrorBlender, we want to investigate how optical effects with people's real faces, hands and bodies can support deixis in collaboration.

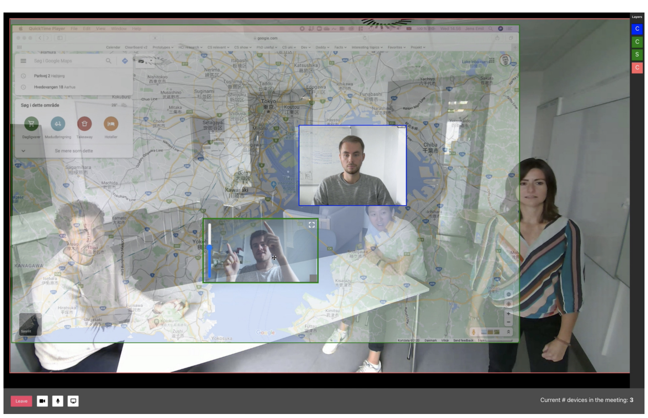

Figure 5: The MirrorBlender user interface. Mirrors from the same user will have a unique border color. The z-order of mirrors is controlled through the layer panel to the right.

4 PROTOTYPE IMPLEMENTATION

A meeting in MirrorBlender is represented by a web page with a unique URL. Meeting participants can join an ongoing meeting by navigating to the URL and sharing their audio and camera feed from within the browser. When sharing the camera feed, their mirror will appear on the WYSIWIS canvas and every participant can manipulate the position, scale and translucency of each mirror. Figure 5 shows the user interface of MirrorBlender. Pressing the share screen button will open a dialog for selecting either a screen or a window to share. The mirror of the screen will likewise appear in the WYSIWIS canvas and is manipulable in the same manner as camera mirrors. Layering of mirrors can be controlled using a layer panel in the right side of the screen (see Figure 6).

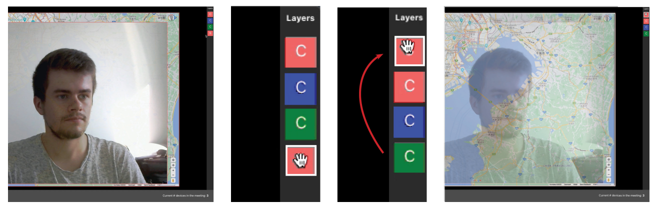

Figure 6: The layer panel: Users can re-order overlapping mirrors as layers via drag and drop. This affects how the mirrors are blended together.

The prototype of MirrorBlender is built with Webstrates [21] using its authoring tool, Codestrates [30]. Webstrates is a web-based software framework that allows for rapid development of real-time collaborative software. Video streaming is implemented using WebRTC[8] through an API provided by Webstrates. Because the Screen Capture API[9] is a working draft, the prototype only works fully in browsers where the API is supported. The design of the final prototype went through an iteration based on a pilot study. The pilot study was conducted with the same setup as in the final study (see Figure 8). However, we focused the debriefing interviews especially on how to improve the prototype design for usability. In the following, we provide detailed considerations about two central aspects of the prototype design.

4.1 WYSIWIS and Screen Mirroring

For a fully consistent WYSIWIS, it is necessary that the canvas looks exactly the same across different computer displays, up to a scale factor. Hence, it is important that the aspect ratio is the same. However, in our current implementation, we do not address this problem. The canvas elements (video feeds) have absolute pixel positions and pixel sizes. To avoid issues of aspect ratio, we controlled for it in the study setup by adjusting the size and aspect ratio of the browser windows on the external screens in the remote rooms and on the large touch display in the primary room.

For screen mirroring, we implemented two different interfaces; one for the remote and one for the co-located participants. Figure 2A shows that remote participants have the WYSIWIS canvas on an

[8]WebRTC: https://www.w3.org/TR/webrtc/
[9]Screen Capture API: https://developer.mozilla.org/en-US/docs/Web/API/Screen_Capture_API

Figure 7: The problem of parallax. A-B: Horizontal and vertical parallax. C-D: Perceiving one's body in real space during gesture vs. perceiving one's own body representation in the camera feedback.

external screen and their own screen content (to be mirrored) on the laptop screen. For the co-located participants, we implemented a Companion Mode for when they are screen sharing from their own laptop in the primary room. By choosing "Join without video", the laptop becomes a companion that neither produces nor consumes audio and video. Instead, the Companion Mode merely supports sharing and reconfiguring a screen mirror on the WYSIWIS canvas. During screen sharing, a small control panel enables the companion to remotely reconfigure the screen mirror in the WYSIWIS canvas on the large touch display. The need for this mode was discovered during the pilot study, where the sharing of several camera feeds together with screen feeds from the primary room caused confusion.

4.2 Blending and the Problem of Parallax

In order to point in screen content so that everyone can see it, the person gesturing must adopt the illusion of using their own representation in the camera mirror for pointing, rather than their real body for pointing on the physical display in the room. This is due to the problem of parallax [17] between the camera and screen. In ClearBoard [17], the hardware setup achieves the alignment that is sufficient for consistent spatial cues in a two-person configuration. However, for MirrorBlender, we want to support off-the-shelf hardware and do not rely on a hardware setup that can reduce the problem of parallax. In Figure 7, it is illustrated how the parallax affects the offset of the arm as it reaches towards the display for pointing. This is a limitation when attempting to do deictic gestures. The closer the user is to the display, the larger the parallax angle, and this results in a larger perceived offset. Therefore, in order to be inclusive of both remote and co-located participants, the user needs to get used to making deictic gestures by embodying their mirror image rather than their real body in space.

5 STUDY DESIGN

Our goal is to gain rich insights on how malleability supports communication dynamics such as deictic gestures, attention grabbing, and inclusion of remote participants. In order to qualitatively assess the effect of MirrorBlender, we designed a series of tasks to specifically invoke such dynamics. We conducted a qualitative observation study consisting of three hybrid meeting sessions with groups of five, each completing three collaborative tasks together.

5.1 Participants

We recruited 15 participants (5 women, 10 men, age range 23-31) from neighbouring research labs, through internal mailing lists, Slack and Facebook groups. We grouped the participants into three groups of five according to availability. Some participants were familiar with each other and some were not, but we made sure that we mixed the groups accordingly. We looked for participants who were familiar with current video-conferencing systems such as Skype, Zoom or Microsoft Teams, in order to contrast the everyday use of these to their experience with MirrorBlender. No compensation was offered, but we provided a free lunch.

5.2 Procedure

Three co-located participants including a moderator (*MC1*, *C2*, *C3*) were seated in the primary room with a Microsoft Surface Hub as well as their personal computers. Two remote participants (*R1*, *R2*) were seated in separate rooms (see Figure 8). In each group, we asked them to decide whether they wanted to be remote or co-located, and we asked one from the co-located participants to volunteer as meeting moderator (*MC1*). We designed this setup to align with the power dynamics in hybrid business meetings observed by Saatçi et al. [31]. When managers and/or meeting moderators are present, they often join the meeting from the primary room, which consists of more than two people, causing remote participants to feel ignored and excluded [31].

Each study session lasted for about two hours and consisted of a training session, three tasks and a focus group interview for debriefing. The three tasks are not intended for comparison, but rather as resources for eliciting diverse interactions among participants where different features of MirrorBlender may be relevant.

5.2.1 Training Session (20 minutes). A pre-recorded tutorial video demonstrated the key features of MirrorBlender and illustrated examples of use cases, such as pointing with a finger from a camera mirror to another participant's screen mirror, or re-positioning and scaling mirrors. We purposefully showed these examples to introduce participants to the idea of embodying themselves in their mirror images, since we are interested in learning about how they utilize this to interact with others rather than studying how they learn to use the system. Then we described the tasks and distributed the participants. Lastly, we shared a sandwich menu in MirrorBlender for ordering lunch and invited all participants to get familiar with MirrorBlender by activating their camera feeds and pointing via their mirror images at their preferred picks.

5.2.2 Turn Taking Task (5 minutes). The first task was a quick round of introductions. This task served as a warm up and mimicked a typical "status round" common in workplace meetings [31].

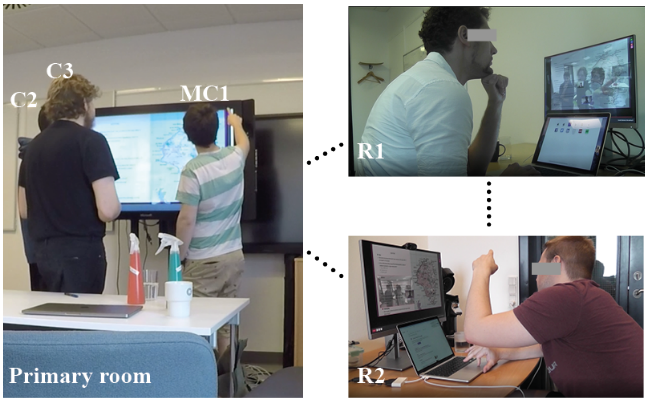

Figure 8: Meeting setup: A primary room of three co-located meeting participants and two separate rooms each with one remote participant.

5.2.3 Map Task (20 minutes). The second task was a puzzle challenge[10] where the group had to answer a list of questions about a map (e.g., how many places with the word "Cave" are on the map). The task requires that the group finds details on a tiled map while maintaining an overview of where they have already searched. We chose this task to resemble a sensemaking meeting where all participants have to interact with the same visual content. We expected this task to inspire diverse non-verbal expressions and deictic references on the map.

5.2.4 Planning Task (30 minutes). The third task was planning a conference trip to Yokohama, Japan. We chose this task to observe how participants adopted MirrorBlender in a context that elicited collaborative creativity (e.g., brainstorming of suggesting post-conference touristic activities) and decision making (e.g., from all the ideas, which one pleases the entire group). We expected this task to be more open ended and to display diverse dynamics around multiple camera and screen mirrors.

5.2.5 Debriefing Focus Group Interviews (30 Minutes). After the participants completed the three tasks, we conducted a focus group interview where we asked the participants to explain key episodes that we identified with the live observation notes, as well as any other example of a situation that felt particularly novel, fun or challenging during the meeting.

One of our co-authors stayed in the primary room to provide technical support and to guide the participants from one task to the other on time. Other co-authors monitored the meeting from a separate room by connecting to MirrorBlender from an extra computer without sharing our video or audio feed. From there, three researchers did live observations and took timed, detailed notes of episodes where MirrorBlender played an instrumental role in mediating interaction. We looked for episodes where participants reconfigured (i.e., adapted) the position, scale and translucency of their mirrors to support their communication with others; examples of overlaying and blending mirrors with others; instances of failed and successful deictic references to shared content; and generally surprising appropriations of MirrorBlender to meet communication needs between remote and co-located participants.

5.3 Data Collection

We video-recorded remote participants with a stationary camera in each room and the primary room with co-located participants was recorded with two separate cameras. We screen-recorded the MirrorBlender interface and group interviews were audio-recorded.

5.4 Analysis

Two researchers watched the recorded videos and took additional, detailed notes to those compiled during the study. Then, we conducted a top-down (deductive) thematic analysis [7] on all the notes: based on our observations, we first discussed preliminary themes that aligned with our research questions and then grouped notes into those themes. We used the content of the group interviews to support our interpretation of the video-recorded data. We iteratively discussed the cohesion of the themes until agreeing on the findings we present next.

6 RESULTS

The deployment of MirrorBlender showed diverse use of the system, including both playful and novel kinds of interpersonal exchanges facilitated by the prototype. The material is rich, yet for this paper we focus on a thematic analysis that addresses the research questions listed in the Introduction. The results are presented as example situations (incl. screenshots from video recordings[11]) under the following themes: *Malleable Blended Space* (6.1), *Deictic Gestures* (6.2), and *Inclusion and Attention Grabbing* (6.3).

For purposes of readability, we named the three groups as the *blue group*, the *red group* and the *green group*. Remote participants in each group are referred to as *R1* and *R2*; whereas the co-located participant, who is moderator, is referred to as *MC1* and other co-located participants as *C2* and *C3*.

6.1 Malleable Blended Space

In the following, we describe instances of how participants used the malleable blended space in MirrorBlender to support their deictic gestures and engage others in playful interactions. In general, our results show that users creatively appropriated the combination of interaction techniques with the mirrors (i.e., changing scale, position and translucency in a shared virtual frame of reference) for diverse purposes of social interaction.

6.1.1 Optical Illusions of Interpersonal Space. The groups were experimenting with shared optical illusions for blending the remote participants into the primary room mirror. Throughout the different tasks, especially remote participants were aware of their own presence in the blended interpersonal space, as camera and screen mirrors were moved around. Remote participants reconfigured their mirrors not only for the purpose of having fun, but also to achieve certain goals in the communication. An interesting example of this happened in the status rounds in the red group, where remote participants chose to create a forced perspective by

[10] We used a puzzle from the British Ordenance Survey's Puzzle Book: https://getoutside.ordnancesurvey.co.uk/guides/os-map-puzzles-to-challenge-your-skills/

[11] The accompanying video shows clips of several instances described in the paper.

Figure 9: Playful optical illusions. A: Remote participants place themselves on physical objects in the primary room; e.g., a chair and a large screen. B: Placing remote participants on the faces of co-located participants. C: Blending a remote participant with a co-located participant so that the remote participant "wears" the co-located participant's torso.

co-locating their camera mirrors with physical objects and people in the primary room. Right before the turn taking task started, R1 and R2 placed their mirrors on two empty chairs between C3 and C2. However, to our surprise, the group ignored the round-table arrangement in the blended space and did not let the remote participants speak before all co-located participants were done. Because the participants made the effort of arranging virtually in the real seats, we expected their arrangement to govern the order of their status round. However, this shows that the virtual sense of "being there" is clearly something else than real co-location.

Another example is when one of the facilitators entered the primary room to present the tasks. She stood exactly where R1 had placed his mirror. This caused her face to be occluded behind R1's mirror. After a few seconds, R1 moved himself to another empty chair between C2 and C3 to make the facilitator visible to everyone again. But his mirror was placed this time above the chair to match the head level of the co-located participants.

6.1.2 Embodiment and Playfulness in the Blended Space. Overall, our observations show that the participants felt a sense of embodiment in a blended interpersonal space; they would sometimes forget about the location of their real body and they were able to effectively create a large variety of shared optical illusions.

For instance, in order to immerse themselves in the blended illusion, the co-located participants needed to focus on their mirrored representation in the WYSIWIS canvas, rather than the physical space with their co-located peers. This user experience is described by two co-located participants in the green group during the debriefing interviews. C2 stated that even as a co-located participant, it felt like being in a different space than others sometimes: "It felt like we were all remote". MC1: "I felt disembodied". This was followed by an anecdote that MC1 was nearly poking C2 in the eye; the moment which is shown in Figure 10. The problem of parallax with the MirrorBlender study setup causes a shift in the visual perspective from the cameras mounted with a horizontal or vertical offset to the center of the WYSIWIS screens. The instance in Figure 10 indicates a shifted sense of self-location [34], which means MC1 was temporarily immersed in embodying his mirror image to the degree that he was less aware of his real body.

Our observations further indicate that the participants were relying on the illusions for communicating more effectively. This

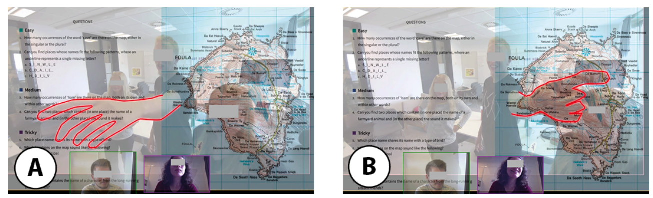

Figure 10: Embodiment and self-calibration. A-B: MC1 self-calibrates by adjusting his arm pose to embody his own mirror image. B: MC1 nearly pokes C2's face by accident.

was especially clear during technical breakdowns. For instance, in the blue group, when R2 and MC1 realized that MC1 did not see correctly where R2 was pointing in her mirror (due to a breakdown in the synchronization of the WYSIWIS canvas), MC1 concluded: "Okay, so What You See Is *Not* What I See then." This instance shows that they started to rely on a shared illusion for their collaboration.

Beyond the goal-oriented use, we further observed that all groups used the malleable mirrors and equal access to control for engaging in playful illusions by manipulating the position and scale of both their own and each other's mirrors. See Figure 9 for several examples of playful optical illusions. For instance, in the red group, C2 made circular movements with R2's and R1's mirrors during the training session and R1 immediately adapted to his manipulated presence by making circular movements with his head, acting dizzy.

6.2 Deictic Gestures

The principles of WYSIWIS and mirror blending enabled both co-located and remote participants to use their body and hands for deictic gestures when referring to shared screen content (see Figures 11-12). Moreover, the reconfiguration of mirrors supported the enactment of very diverse and expressive gestures.

6.2.1 Embodied Pointing and Self-calibration. Every participant needed to make sense of their own body projections in order to have agency of their mirrored representation. This was a clear pattern in the users' behavior for coping with the problem of parallax. Due to the horizontal parallax arising from the Surface Hub's camera location (see Figure 7B and 7C-D), co-located participants showed two different modes of embodied pointing: using their body to point

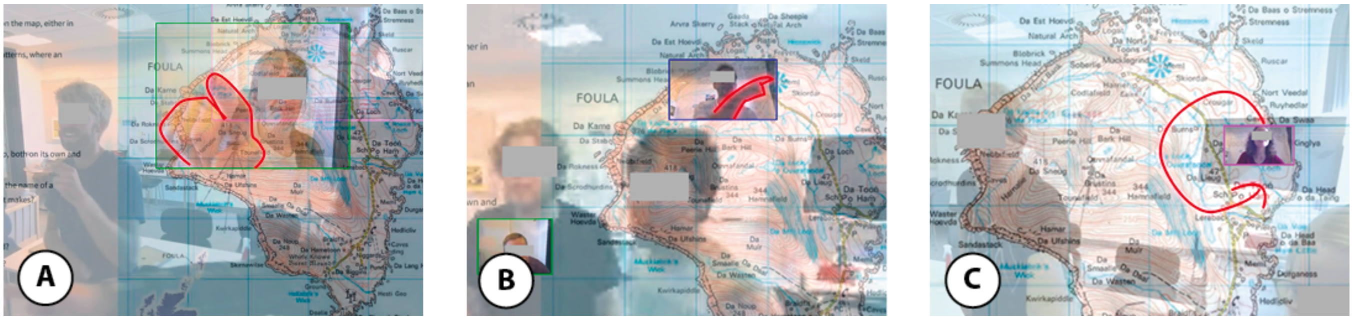

Figure 11: Remote deictic gestures. A: Bimanual pointing in the blue group. B: Bimanual pointing in the red group. C: Highlighting area by circling with a tiny feed as a cursor in the green group.

Figure 12: Co-located deictic gestures. A: A two-hand gesture to divide the labor of searching the map into groups of tiles for different participants. B: A pointing finger as an indicator travelling over the map in the reading direction. C: A two-hand gesture for indicating (and recalling) several name labels of interest.

in the real space and using their mirrors to point in the virtual frame of reference. When co-located participants addressed each other, they pointed with their finger on the screen, excluding the remote participants from the interaction. The unconscious use of pointing in real space was seen frequently among co-located participants in the primary room to rapidly grab the attention of co-located peers. In contrast, when they addressed the whole group, they pointed with their finger in the primary room mirror to include the remote participants. Pointing via mirrors requires practice and additional mental effort, especially for the co-located participants that often had to consciously switch from one mode of pointing to another. For example, participants tended to first notice their own hand and index finger in the mirror and then adjust its position to the desired target. We refer to this practice as *self-calibration* (see Figure 10).

6.2.2 Bimanual Pointing Gestures. During the map task, remote participants frequently pointed at labels by combining a deictic gesture with dragging their camera feeds along the screen. While they used one hand to drag their camera feed next to a label, they used the index finger of their other hand to point precisely at it. We refer to this as bimanual pointing gestures (see Figure 11A-B for examples).

Co-located participants in the primary room engaged in collaborative bimanual pointing to help each other. In the red group (see Figure 12B), MC1 was using his finger as a moving indicator to track their progress top-down. However, the bottom part of the map was not overlapping with the primary room mirror. Without being asked, C2 helped MC1 by dragging their mirror down.

6.2.3 Two-Hand Gestures. During the map task, participants discussed the most efficient strategy for accomplishing the task as well as how to divide the labor. Co-located participants would sometimes use confirming gestures to make sure that all participants had an equal understanding of the division of labor. For example, C3 from the blue group tried to distribute responsibilities among participants and divided the map into four areas by using both hands to indicate a vertical and a horizontal axis (see Figure 12A). Two-hand gestures were also utilized to point at multiple places at the same time, e.g., for keeping track of which labels the group had already found on the map (see Figure 12C).

6.2.4 Malleable Mirrors as Cursors and Pins. A common limitation of screen sharing in video-conferencing systems is that only one person's cursor is visible at a time. With MirrorBlender, both co-located and remote participants discovered alternative ways of using their malleable mirrors as multiple simultaneous cursors for deictic gestures around screen content. The following examples reveal a legacy bias towards pointing with a mouse cursor, but also point to new opportunities from using malleable mirrors.

In the green group, during the map task (see Figure 11C), R1 had just found a label on the map. To draw the others' attention to it, she made her mirror tiny and used it as a cursor for making a circular movement around the area. During the planning task,

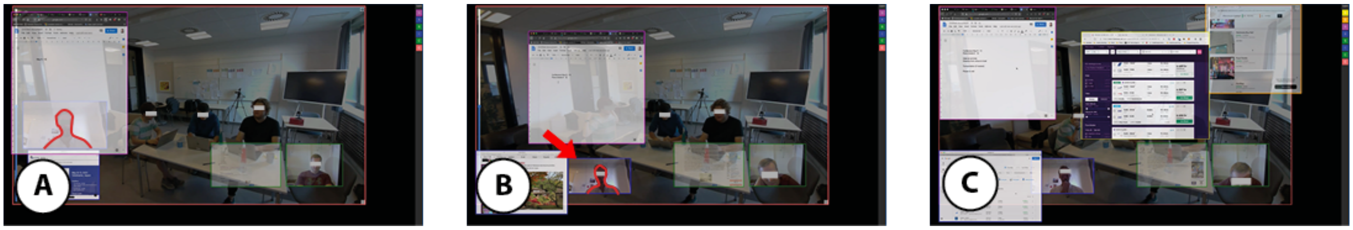

Figure 13: Territoriality and attention grabbing. Participants in the red group keep their own screen mirrors and camera mirrors together, self-organizing in personal territories. A: R1's and MC1's territories interfere as MC1 shares his screen. B: R1 politely moves out of MC1's territory. C: C3's screen is placed in the center, the group territory, and made opaque for everyone to collectively plan the flight itinerary.

MC1 from the blue group discovered that he could blend his screen mirror with another screen mirror to use his own mouse cursor for pointing to content on the other screen.

During the map task, participants often needed to make persistent marks on the screen. In the green group, they first used both their hands as pins on name labels (Figure 12C). Later, the group switched to using a painted overlay to more easily make pin on previously checked labels: they came up with the idea that R2 could share a mirror of a full-screen painting tool. He blended this mirror with the map and created an annotation overlay from which he could draw pins and mark the parts that they had already checked.

In the blue and the red group, we observed territorial behavior with the mirrors of the remote participants: They used their mirrors to pin themselves to a group of tiles in the map in order to claim the territory. This was negotiated in the very beginning of the map task, when dividing the labor. For instance, C3 in the blue group, assigned tasks to the remote participants with a two-hand gesture (Figure 12A). As R1 had already positioned her mirror in the top-right corner of the map, C3 assigned her the responsibility of that area. R1 remained pinned to this territory. The red group similarly divided the map into territories of responsibility (in this case, top, middle, and bottom) by pinning one camera mirror to each.

6.3 Inclusion and Attention Grabbing

MirrorBlender supported diverse ways for remote participants to grab attention and build a sense of being included, not only through the pointing via embodied mirrors but also by supporting territorial behavior [33] in the WYSIWIS canvas.

6.3.1 Reciprocal Acts of Inclusion. Co-located participants tried to include remote participants in their collaborative activities and remote participants also tried to include themselves by grabbing attention through the features of MirrorBlender. A common act of inclusion towards remote participants was when co-located participants switched from pointing with their real body to pointing with its representation in the mirror (see Figure 10). Our group interviews support that in the map task, embodied pointing with mirrors was used a lot, which made remote participants feel more included. R1 from the red group further explained that when the co-located participants "did it wrong" (referring to pointing on the actual screen), it made him feel excluded. Moreover, all groups agreed that there were less acts of inclusion in the planning task, where participants relied more on screen mirrors than camera mirrors.

We observed a balanced distribution of acts of inclusion that reduced the primary room dominance typical of hybrid meetings [18]. Both remote and co-located ends tried to reciprocally include each other in the social interaction. Remote participants proactively requested that co-located participants point in inclusive ways: in the blue group, C3 was pointing on the real screen, and due to the parallax, his index finger was outside the field-of-view. R1 objected: "I cannot see where you are pointing". MC1 corrected the pointing by using the primary room mirror to point so that the remote participants are included in the current co-located discussion of a name label on the map. Remote participants also used their embodied mirrors to grab their attention, relieving co-located participants from a one-sided responsibility of keeping them included. For example, as the WYSIWIS canvas tended to be crowded with several overlapping mirrors, remote participants often wiggled their own mirrors (a form of embodied kinetic gestures [10]) while adding verbal cues such as "Here I am" or "I want this". Especially in the planning task, when co-located participants were not standing in front of the Surface Hub, remote participants helped them by adjusting the appearance of the camera and screen mirrors based on their requests. These hybrid collaborations can be considered a direct outcome of supporting equal access to control of mirrors in the WYSIWIS canvas.

6.3.2 Negotiating Attention in Personal and Group Territories. We saw evidence of a sense of inclusion in how remote and co-located participants negotiated the terrotoriality of the WYSIWIS canvas in equal terms. Some of the examples we described in Section 6.1 show how remote participants positioned their mirrors to be next to physical objects and co-located participants (see Figure 9), presenting themselves as part of the primary room. Furthermore, remote participants often followed each other to stay close. C2 from the green group mentioned that "usually in hybrid meetings, remote people are large and end up taking up too much space in the meeting. In this system [MirrorBlender], people can be more on the same level".

We noticed the formation of *group* and *personal territories* [33], especially in the planning task (see Figure 13). The activities in the planning task required more diverse use of content (internet searches, notes, etc.), which resulted in the screen mirrors receiving more attention than the camera mirrors. Screen mirrors that were relevant to everyone were large, opaque and in the center of the WYSIWIS canvas, indicating a space for group interactions. Additionally, both co-located and remote participants tended to keep their screen mirrors within proximity of their mirror image in the periphery, denoting personal territories.

The distribution of personal and group territories supported different strategies for grabbing attention. During the map task, remote participants used the act of moving their camera mirrors into the map area as means for jumping into the conversation (see examples in Figure 11). Co-located participants, on the other hand, would walk closer to the display and camera of the Surface Hub to appear larger. Despite using different methods for grabbing the attention, they were able to participate more equally. In fact, remote participants were often more mobile than the co-located ones since moving the mirror is quicker than moving one's real body. During the planning task, the pattern was reused. This time, however, by all participants with their individual *screen mirrors* (see Figure 13). They often dragged their screen mirrors to the center to report their task progress or ask for opinions, then back into the periphery.

Remote participants also leveraged the periphery of the screen to *avoid* attention. After connecting to the meeting, they tended to adjust and move their camera mirrors to the most available part of screen, which usually was the empty corners. R2 from the green group explained his logic behind this behavior: "I wanted to not disturb the other parts. This was the only corner that was free". We see such situations as more evidence of how remote participants felt included in the meeting, since they had to take concrete actions to avoid attention when they did not need it.

Comparing across groups, they developed different organizational patterns for solving the tasks, which shaped their territorial behaviours. Both the blue and the red group used a divide-and-conquer strategy, dividing the map space into territories and distributing independent tasks for planning the trip. However, the green group developed a tighter coupling style in their collaboration with less explicit territoriality. This group came up with the idea of blending the map with a painting tool, and their work during the planning task happened more in plenary with MC1 taking a prominent coordinator role.

These insights show how important it is that our tools for video communication are malleable and flexible for different activities and group dynamics.

7 DISCUSSION

In the following, we discuss the three research questions we have posed in the Introduction with the design and study of MirrorBlender and put our results in the context of prior work. We further outline the implications of our work and how it points to future directions for supporting video-based collaboration.

7.1 Beyond Being There with MirrorBlender

We have investigated how our concept of malleable mirrors enable users to create blended spaces that go beyond being there (*RQ1*).

Several optical illusions added value to the social interactions between the meeting participants. They created several playful illusions such as stitching one participant's head to the body of another (see Figure 9C). Some effects even occurred across different groups, such as pinning remote participants to physical objects in the primary room (see Figure 9A). These illusions were instrumental in creating an informal atmosphere where participants across the distributed spaces had fun together and teased each other.

The territorial behavior (using the center and periphery of the WYSIWIS canvas) shows that they effectively adopted the ClearBoard effect for conversations and coordinated actions around the same content. The effect goes beyond being there in several ways. It supports being face-to-face while simultaneously looking at the same content. This is not typical of co-located collaboration (unless you actually stand on either side of a clear-glass whiteboard). Additionally, remote participants are more mobile and in control in the virtual reference frame than the co-located participants. The increased mobility of remote participants compensates for their inherent disadvantages (e.g., audio delay and limited video field-of-view). This shows the value of going beyond the approach of physical realism; doing so can shift the traditional power dynamics in hybrid meetings [18, 31] towards more equal participation.

These findings point to future work that investigates how different blended interpersonal spaces may change the power dynamics of meetings. MirrorBlender could further be used for deriving patterns in malleable blended spaces. Such patterns could be used to inform the design of (intelligent) presets for spatial structures of mirrors, so that it is less effort for the users to reconfigure the meeting layout for different meeting activities.

7.2 Embodying Mirrors for Deictic Gestures

We have investigated the feasibility of using mirror images for deictic gestures in a virtual frame of reference (*RQ2*). Our results suggest that users can embody their mirror image for effective deictic gesturing. However, they also reveal that there is mental effort associated with switching between modes of embodied pointing due to the problem of parallax (elaborated in Section 7.4).

We found that every group independently discovered the technique of self-calibration when attempting to do deictic gestures via their mirror. Moreover, while co-located participants would forget to point via their mirror, they still exhibited very rapid self-calibration when needing to include remote participants. The deictic gestures were also more expressive than what can be achieved with a mouse cursor. For instance, several participants were using their body for expressing complicated gestures like using both arms simultaneously to divide the map into separate areas.

Additionally, while mostly expecting the use of mirrors for deictic gestures, our results further show how participants discovered other kinds of gestures. For instance, kinetic gestures like wiggling a mirror for grabbing attention, or spatial gestures like moving and pinning mirrors to personal and group territories in the WYSIWIS canvas. In conclusion, while deictic gestures via mirrors require mental effort, they provide value in terms of including other people through video, providing equal opportunities for deictic expression, and enabling more expressive use of body language.

7.3 Inclusion of Remote Participants

Finally, we have investigated how the design principles of MirrorBlender support inclusion of remote participants (*RQ3*).

Quotes from the debriefing interviews suggest that the malleable mirrors enable the remote participants to include themselves more in the primary room and feel more equal to them. The provided example situations have further shown that they have several new mechanisms for grabbing the attention of the group.

It is important to stress that MirrorBlender is intended to support users in *coping* with the asymmetry of group interactions in hybrid meetings, rather than *erasing* it. I.e., we focus on designing new interaction mechanisms for inclusion rather realistically aligning remote spaces. For instance, the users can deliberately manipulate unrealistic scales of participants. According to Ens et al. [8], this is an under-explored concept for collaborative mixed reality. Our results show that the ability to manipulate the scale of meeting participants and screen content in MirrorBlender supported the dynamic adaptation of the WYSIWIS canvas for different configurations of attention. For example, co-located participants in the primary room mirror were persistently large and remaining in the background of all other mirrors. This made is easier for remote participants to read the body language of co-located participants. Aside from scale, remote participants would mainly adjust their translucency and position in the canvas to configure how much they were in focus. Thus, both position, scale, and translucency showed to be important for including remote participants.

Equally important is the nature of how users can manipulate these properties. In contrast to most current video-conferencing systems, MirrorBlender supports continuous gestures for configuring the degree to which users grab others' attention: at one end of the extreme, a camera mirror can be shortly moved into the center of attention and back out (without resizing or adjusting the translucency). At the other end of the extreme, a camera mirror can be scaled up, made opaque and more persistently claim the center of attention for a period of time. With this additional control of their own presence, remote participants would often configure the attention of the group by gradually making themselves more prominent in the meeting.

These novel gestures show the potential of continuous input to provide remote participants with mechanisms for feeling more included by being able to conduct themselves more appropriately.

7.4 Limitations and Future Work

While the tasks invoked gestures that leveraged MirrorBlender for deixis and inclusion, our results are also limited to the specifics of the tasks and setup. Future work is required to better understand how frequent and relevant such deictic gestures are in real meetings, how well the WYSIWIS concept scales to larger groups, and how the power dynamics in a real team would hinder or emphasize the acts of inclusion that we observed in our study.

Additionally, a clear limitation of the deictic gestures with MirrorBlender is the parallax problem. Especially, the co-located participants had to spend significant cognitive effort on self-calibrating to compensate for the parallax. They struggled more than the remote participants because the horizontal parallax of the Surface Hub is less common in everyday computers, such as laptops. The horizontal parallax was further amplified when being close to the Surface Hub to be able to view its content. However, thanks to the parallax issue, we were able to identify concrete efforts of inclusion that may have been taken for granted otherwise.

Studying the effect of the parallax problem for video-mediated deictic gestures under more controlled conditions would be an interesting avenue for future work. We saw hints of a training effect where the participants would be increasingly comfortable with pointing through mirrors, but a more controlled study would be required to document this effect. Furthermore, the problem of parallax is also the case for direct eye contact, which is not supported by MirrorBlender (as opposed to, e.g., ClearBoard [17]). This limitation is shared with other systems that rely on off-the-shelf technologies.

Another limitation is that video feeds are sometimes hard to blend, where two overlapping feeds easily render the content unintelligible. Simple visual optimizations would be to remove the background of remote participants (as possible in Zoom or Skype) to reduce visual clutter, or to have more fine-grained mechanisms for controlling color blending.

Finally, in the planning task, we observed challenges in transitions between closely coupled and loosely coupled collaboration where the primary room would accidentally exclude the remote participants (e.g., occluding the primary room mirror). Future work could include analysis of our data with a focus on collaborative coupling styles, e.g., using Neumayr et al.'s [27] framework.

8 CONCLUSION

We have built MirrorBlender to study the potential of malleable video-conferencing systems in supporting the complex dynamics of hybrid meetings. Our results show that the ability to manipulate the position, scale and translucency of video feeds in a WYSIWIS canvas supports users in negotiating a blended interpersonal space for collaboration. The support for deictic gestures and continuous control of mirrors in MirrorBlender shows potential for facilitating more equal participation in hybrid meetings. While users sometimes struggled or forgot to point to shared content in an inclusive way, our data exposed numerous instances where they embodied and leveraged their mirror images for deictic gestures and playful interactions. Our work lays out exciting new directions for supporting non-verbal communication and more inclusive participation in hybrid forms of video-mediated collaboration.

ACKNOWLEDGMENTS

Jens Emil Grønbæk is supported by the Center for Computational Thinking and Design. Banu Saatçi's work has been supported by Microsoft Research through its PhD Scholarship Programme in EMEA. Carla Griggio is supported by the Carlsberg Foundation. Special thanks to Sebastian Krog Knudsen for making the video. Furthermore, thanks to the participants for agreeing to take part in our user study in the midst of the COVID-19 pandemic.

REFERENCES

[1] Ignacio Avellino, Cédric Fleury, and Michel Beaudouin-Lafon. 2015. Accuracy of Deictic Gestures to Support Telepresence on Wall-Sized Displays. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15)*. Association for Computing Machinery, New York, NY, USA, 2393–2396. https://doi.org/10.1145/2702123.2702448

[2] Ignacio Avellino, Cédric Fleury, Wendy E. Mackay, and Michel Beaudouin-Lafon. 2017. CamRay: Camera Arrays Support Remote Collaboration on Wall-Sized Displays. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 6718–6729. https://doi.org/10.1145/3025453.3025604

[3] Mathilde M. Bekker, Judith S. Olson, and Gary M. Olson. 1995. Analysis of Gestures in Face-to-Face Design Teams Provides Guidance for How to Use Groupware in Design. In *Proceedings of the 1st Conference on Designing Interactive Systems: Processes, Practices, Methods, & Techniques* (Ann Arbor, Michigan, USA) *(DIS '95)*. Association for Computing Machinery, New York, NY, USA, 157–166. https://doi.org/10.1145/225434.225452

[4] Steve Benford, Chris Greenhalgh, Gail Reynard, Chris Brown, and Boriana Koleva. 1998. Understanding and constructing shared spaces with mixed-reality boundaries. *ACM Transactions on computer-human interaction (TOCHI)* 5, 3 (1998), 185–223.

[5] Mark Billinghurst and Hirokazu Kato. 1999. Real World Teleconferencing. In *CHI '99 Extended Abstracts on Human Factors in Computing Systems* (Pittsburgh, Pennsylvania) *(CHI EA '99)*. Association for Computing Machinery, New York, NY, USA, 194–195. https://doi.org/10.1145/632716.632838

[6] Sara A. Bly, Steve R. Harrison, and Susan Irwin. 1993. Media Spaces: Bringing People Together in a Video, Audio, and Computing Environment. *Commun. ACM* 36, 1 (Jan. 1993), 28–46. https://doi.org/10.1145/151233.151235

[7] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101.

[8] Barrett Ens, Joel Lanir, Anthony Tang, Scott Bateman, Gun Lee, Thammathip Piumsomboon, and Mark Billinghurst. 2019. Revisiting collaboration through mixed reality: The evolution of groupware. *International Journal of Human-Computer Studies* 131 (2019), 81–98.

[9] Rachel Franklin. 2017. Facebook spaces: A new way to connect with friends in vr. https://about.fb.com/news/2017/04/facebook-spaces/.

[10] Susan R. Fussell, Leslie D. Setlock, Jie Yang, Jiazhi Ou, Elizabeth Mauer, and Adam D. I. Kramer. 2004. Gestures over Video Streams to Support Remote Collaboration on Physical Tasks. *Hum.-Comput. Interact.* 19, 3 (Sept. 2004), 273–309. https://doi.org/10.1207/s15327051hci1903_3

[11] Saul Greenberg, Carl Gutwin, and Mark Roseman. 1996. Semantic telepointers for groupware. In *Proceedings Sixth Australian Conference on Computer-Human Interaction*. IEEE, 54–61.

[12] Simon N. B. Gunkel, Hans M. Stokking, Martin J. Prins, Nanda van der Stap, Frank B. ter Haar, and Omar A. Niamut. 2018. Virtual Reality Conferencing: Multi-User Immersive VR Experiences on the Web. In *Proceedings of the 9th ACM Multimedia Systems Conference* (Amsterdam, Netherlands) *(MMSys '18)*. Association for Computing Machinery, New York, NY, USA, 498–501. https://doi.org/10.1145/3204949.3208115

[13] Steve Harrison and Paul Dourish. 1996. Re-Place-Ing Space: The Roles of Place and Space in Collaborative Systems. In *Proceedings of the 1996 ACM Conference on Computer Supported Cooperative Work* (Boston, Massachusetts, USA) *(CSCW '96)*. Association for Computing Machinery, New York, NY, USA, 67–76. https://doi.org/10.1145/240080.240193

[14] 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.

[15] Christian Heath and Paul Luff. 1993. Disembodied conduct: Interactional asymmetries in video-mediated communication. *Technology in working order: Studies of work, interaction, and technology* (1993), 35–54.

[16] Jim Hollan and Scott Stornetta. 1992. Beyond Being There. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Monterey, California, USA) *(CHI '92)*. Association for Computing Machinery, New York, NY, USA, 119–125. https://doi.org/10.1145/142750.142769

[17] Hiroshi Ishii, Minoru Kobayashi, and Jonathan Grudin. 1993. Integration of Interpersonal Space and Shared Workspace: ClearBoard Design and Experiments. *ACM Trans. Inf. Syst.* 11, 4 (Oct. 1993), 349–375. https://doi.org/10.1145/159764.159762

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

[19] David Kirk, Tom Rodden, and Danaë Stanton Fraser. 2007. Turn it this way: grounding collaborative action with remote gestures. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. 1039–1048.

[20] David Kirk and Danae Stanton Fraser. 2006. Comparing remote gesture technologies for supporting collaborative physical tasks. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*. 1191–1200.

[21] Clemens N. Klokmose, James R. Eagan, Siemen Baader, Wendy Mackay, and Michel Beaudouin-Lafon. 2015. <i>Webstrates</i>: Shareable Dynamic Media. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software and Technology* (Charlotte, NC, USA) *(UIST '15)*. Association for Computing Machinery, New York, NY, USA, 280–290. https://doi.org/10.1145/2807442.2807446

[22] Peter Gall Krogh, Marianne Graves Petersen, Kenton O'Hara, and Jens Emil Grønbæk. 2017. Sensitizing Concepts for Socio-Spatial Literacy in HCI. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 6449–6460. https://doi.org/10.1145/3025453.3025756

[23] 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*. 1–13.

[24] Anastasia Kuzminykh and Sean Rintel. 2020. Low Engagement As a Deliberate Practice of Remote Participants in Video Meetings. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–9.

[25] Joshua McVeigh-Schultz, Anya Kolesnichenko, and Katherine Isbister. 2019. Shaping Pro-Social Interaction in VR: An Emerging Design Framework. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3290605.3300794

[26] Osamu Morikawa and Takanori Maesako. 1998. HyperMirror: Toward Pleasant-to-Use Video Mediated Communication System. In *Proceedings of the 1998 ACM Conference on Computer Supported Cooperative Work* (Seattle, Washington, USA) *(CSCW '98)*. Association for Computing Machinery, New York, NY, USA, 149–158. https://doi.org/10.1145/289444.289489

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

[28] Sergio Orts-Escolano, Christoph Rhemann, Sean Fanello, Wayne Chang, Adarsh Kowdle, Yury Degtyarev, David Kim, Philip L. Davidson, Sameh Khamis, Mingsong Dou, Vladimir Tankovich, Charles Loop, Qin Cai, Philip A. Chou, Sarah Mennicken, Julien Valentin, Vivek Pradeep, Shenlong Wang, Sing Bing Kang, Pushmeet Kohli, Yuliya Lutchyn, Cem Keskin, and Shahram Izadi. 2016. Holoportation: Virtual 3D Teleportation in Real-Time. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology* (Tokyo, Japan) *(UIST '16)*. Association for Computing Machinery, New York, NY, USA, 741–754. https://doi.org/10.1145/2984511.2984517

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

[30] Roman Rädle, Midas Nouwens, Kristian Antonsen, James R. Eagan, and Clemens N. Klokmose. 2017. Codestrates: Literate Computing with Webstrates. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology* (Québec City, QC, Canada) *(UIST '17)*. Association for Computing Machinery, New York, NY, USA, 715–725. https://doi.org/10.1145/3126594.3126642

[31] 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 (Nov 2020), 769–794. https://doi.org/10.1007/s10606-020-09385-x

[32] 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.

[33] Stacey D Scott, M Sheelagh T Carpendale, and Kori M Inkpen. 2004. Territoriality in collaborative tabletop workspaces. In *Proceedings of the 2004 ACM conference on Computer supported cooperative work*. 294–303.

[34] Sofia Seinfeld, Tiare Feuchtner, Antonella Maselli, and Jörg Müller. 2020. User Representations in Human-Computer Interaction. *Human–Computer Interaction* 0, 0 (2020), 1–39. https://doi.org/10.1080/07370024.2020.1724790 arXiv:https://doi.org/10.1080/07370024.2020.1724790

[35] Jared Spataro. 2020. The future of work—the good, the challenging & the unknown. https://www.microsoft.com/en-us/microsoft-365/blog/2020/07/08/future-work-good-challenging-unknown/. (Accessed on 14/08/2020).

[36] M. Stefik, D. G. Bobrow, G. Foster, S. Lanning, and D. Tatar. 1987. WYSIWIS Revised: Early Experiences with Multiuser Interfaces. *ACM Trans. Inf. Syst.* 5, 2 (April 1987), 147–167. https://doi.org/10.1145/27636.28056

[37] Anthony Tang, Michael Boyle, and Saul Greenberg. 2003. Display and presence disparity in Mixed Presence Groupware. (2003).

[38] Aiden Wickey and Leila Alem. 2007. Analysis of Hand Gestures in Remote Collaboration: Some Design Recommendations. In *Proceedings of the 19th Australasian Conference on Computer-Human Interaction: Entertaining User Interfaces* (Adelaide, Australia) *(OZCHI '07)*. Association for Computing Machinery, New York, NY, USA, 87–93. https://doi.org/10.1145/1324892.1324909

[39] 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

[40] Nicole Yankelovich, Nigel Simpson, Jonathan Kaplan, and Joe Provino. 2007. Porta-person: telepresence for the connected conference room. In *CHI'07 extended abstracts on Human factors in computing systems*. 2789–2794.

[41] Nicole Yankelovich, William Walker, Patricia Roberts, Mike Wessler, Jonathan Kaplan, and Joe Provino. 2004. Meeting central: making distributed meetings more effective. In *Proceedings of the 2004 ACM conference on Computer supported cooperative work*. 419–428.