ThingShare: Ad-Hoc Digital Copies of Physical Objects for Sharing Things in Video Meetings

Erzhen Hu
University of Virginia
Charlottesville, VA, USA
eh2qs@virginia.edu

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

Wen Ying
University of Virginia
Charlottesville, VA, USA
wy7yv@virginia.edu

Ruofei Du
Google Research
San Francisco, CA, USA
me@duruofei.com

Seongkook Heo
University of Virginia
Charlottesville, VA, USA
seongkook@virginia.edu

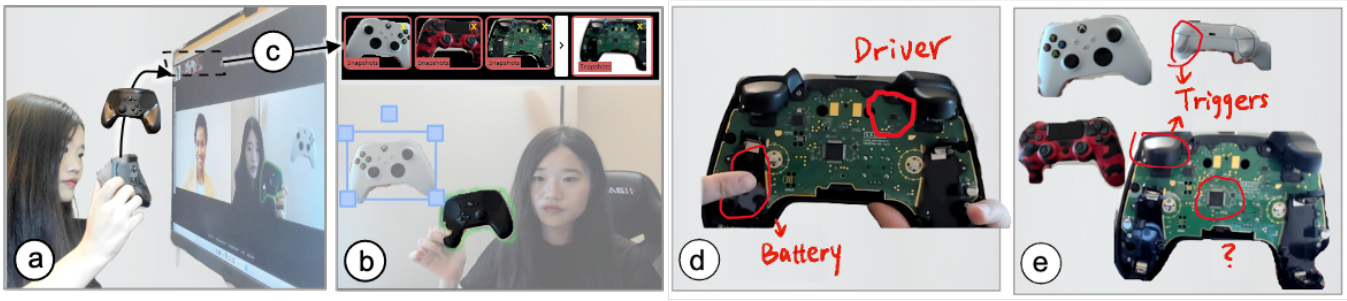

Figure 1: With ThingShare, (a) users can quickly and easily create digital copies of physical objects. These copies can be displayed (b) on the user's video, (c) stored in the object library for future use, or (d) displayed on a separate container called the Task View for more in-depth object-centric discussions. The Task View allows users to add annotations to the digital objects and (e) also supports displaying multiple copies for easy comparisons and multiple perspectives.

ABSTRACT

In video meetings, individuals may wish to share various physical objects with remote participants, such as physical documents, design prototypes, and personal belongings. However, our formative study discovered that this poses several challenges, including difficulties in referencing a remote user's physical objects, the limited visibility of the object, and the friction of properly framing and orienting an object to the camera. To address these challenges, we propose ThingShare, a video-conferencing system designed to facilitate the sharing of physical objects during remote meetings. With ThingShare, users can quickly create digital copies of physical objects in the video feeds, which can then be magnified on a separate panel for focused viewing, overlaid on the user's video feed for sharing in context, and stored in the object drawer for reviews. Our user study demonstrated that ThingShare made initiating object-centric conversations more efficient and provided a more stable and comprehensive view of shared objects.

CHI '23, April 23–28, 2023, Hamburg, Germany

ACM ISBN 978-1-4503-9421-5/23/04.
https://doi.org/10.1145/3544548.3581148

CCS CONCEPTS

• **Human-centered computing** → **Collaborative and social computing systems and tools**.

KEYWORDS

video-mediated communication, object-centered meetings, shared task space, collaborative work, augmented communication

ACM Reference Format:
Erzhen Hu, Jens Emil Grønbæk, Wen Ying, Ruofei Du, and Seongkook Heo. 2023. ThingShare: Ad-Hoc Digital Copies of Physical Objects for Sharing Things in Video Meetings. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23), April 23–28, 2023, Hamburg, Germany*. ACM, New York, NY, USA, 22 pages. https://doi.org/10.1145/3544548.3581148

1 INTRODUCTION

People increasingly use video-conferencing platforms for workplace meetings [108], education, entertainment [98], and social interaction with families and friends [79], accelerated by the need to interact remotely during the global COVID-19 pandemic. In video communication, people often need to show physical things (*e.g.*, documents and design artifacts) to remote users for work [58] or in social settings [61], just as they do in in-person social interactions [33].

However, showing physical objects in video meetings poses several challenges. Due to the constraints of screen space and the camera's field of view, users often have to choose between a close-up view of an object without showing themselves and the entire view that simultaneously shows both themselves and the object, at the expense of the object's details. Observing an object from different perspectives requires asking the local user to adjust the camera or the object since the camera only captures a 2D image of an object from a fixed point of view. Referencing objects often involves multiple rounds of verbal communication as the remote user does not have access to the object. Moreover, presenting digital objects, such as scanned 3D models, images, and video clips, can be an effective way to facilitate object-centered communication in meetings. However, preparing these objects prior to the meeting can be time-consuming and may not be adaptable to unplanned situations such as impromptu questions from other attendees.

To support remote collaboration, Buxton [6] identified three essential communication channels for video-mediated and remote communication: *person space* for delivering facial expressions, *task space* for establishing a shared workspace, and *reference space* for supporting shared gesturing in the task space. Prior work has explored *shared (first-person) view* and *desk view* in object-focused scenarios sharing and accessing remote spaces or specialized surfaces that separated the *task-reference space* from the *person space* and with extensive hardware configurations (*e.g.*, multiple cameras [77], projectors [52], robot telepresence [59, 87], shape-changing interfaces [23]). While hardware setups that enable stable and hands-free interactions [52, 79, 90, 103, 104] can be beneficial, they are complex, require preparation, and are not yet ubiquitous [58]. This constrains use cases to more special-purpose scenarios that are not yet common in everyday video-conferencing setup [58, 61].

In contrast to hardware solutions, we are interested in understanding and developing software solutions that allow people to share physical objects using everyday video-conferencing tools. Previous studies [22, 61, 73] have shown that the *face-to-face* configuration in conventional video-conferencing platforms can hinder the remote collaboration for tasks involving physical objects. However, there is a lack of understanding of how people overcome these challenges in their everyday use of video-mediated communication [7, 61]. To better understand these practices and the challenges people face, we conducted a formative study with 124 crowd workers. Participants reported encountering challenges such as showing detailed views of objects, maintaining objects within the camera view, referencing specific sections of an object, and displaying multiple objects or perspectives.

Informed by the findings from the formative study, we developed ThingShare, a videoconferencing system that supports users to fluidly share physical objects with a remote partner by creating digital copies. ThingShare provides user interface controls and interactions for users to easily create, manipulate, and reference digital copies for effective discussions around physical objects. Imagine the following usage scenario with ThingShare, where two hardware engineers, Alice and Charlie, are discussing the design of a game controller. Using ThingShare, Alice can create a digital copy of a game controller in her hand by a simple drag-and-drop (Figure 1a), which can then be placed on her video feed (Figure 1b). Since the digital copy is displayed on her video feed, she can use gestures to refer to the specific parts of the controller or hold another game controller in her hand. During their discussion on the placement of electrical components, Alice can open a Task View to display a close-up view of the controller, where they can also make annotations (Figure 1d). Charlie finds that the trigger button design of her controller differs from Alice's, and she creates a digital copy of her controller and adds it to the Task View for comparison (Figure 1e). In brief, these interactions show that users can have expressive conversations around physical objects through ad-hoc interaction with digital copies via ThingShare.

To evaluate the system, we conducted a user study with 16 participants. Our findings illustrate that capturing and storing physical objects enhances conversations around physical objects and improves the local user's ability to reference properties of objects and manage the privacy boundaries between sharing a view of their local environment vs. the objects within it. Supporting the remote user's access, control, and reference to digital copies helped them to better understand and contribute to the communication. We conclude that future designs should focus on providing an object-separated layer in front of users.

In summary, we contribute: 1) ThingShare, a video-conferencing system with commodity hardware designed for object-sharing practices. 2) Findings from a formative study on everyday object-sharing practices and challenges in remote meetings. 3) Insights from a user study on the use of ThingShare for sharing and discussing objects in remote meetings.

2 BACKGROUND AND RELATED WORK

In this section, we review prior literature on remote collaborative technologies that support shared task spaces. We situate our work in the broader context of remote technologies and conclude with a design space.

2.1 Challenges in Supporting Object-Focused Sharing and Remote Collaboration

2.1.1 Background of Systems and Interfaces for Object-Focused Collaboration. Prior work has studied remote object-focused collaboration [22, 73, 77, 83], with a focus on understanding users' perception of different perspectives on physical objects. Various research systems have been proposed to offer new ways of visually accessing remote spaces [26, 30, 68], such as remote camera control [83] or multiple viewpoint capture [59, 103], which use stationary or head-mounted cameras to expand the remote access to a physical space and its objects. Media space research has studied approaches to connect spaces across remote office sites, where there are challenges of spatial inconsistency [100], multiple views [82] (*e.g.*, desk view, birds-eye view) and affordances to control viewpoints [103]. They found the particular importance of accessing objects and remote environments and supporting collaboration in various work-related contexts [82, 100]. In contrast, other studies focused on families and friends who are geographically dispersed [53, 76, 106]. where interfaces were developed to enable *always-on* media space systems [50, 51, 81] to enhance family connection and awareness, and specialized shared table space for activities such as reading (*e.g.*, [81]) and playing (*e.g.*, [79, 96, 105]). In spite of these studies, many of the systems developed either do not focus on accessing physical objects (*e.g.*, [50, 51]) or require additional hardware (*e.g.*,

additional tabletop or projector [52, 79, 96, 105]). As a result, only a limited number of these systems can be used in everyday life, especially when people are working from home or connecting with remote friends and family, due to the specialized hardware required. In light of recent efforts to improve video-mediated communication [34, 41, 43, 75, 89] and the current trend of remote work [5, 8] and online social gatherings [25, 43], studying object-sharing practices might suggest new and alternative approaches to enhance remote meeting technologies.

In this paper, we consider details of how video-conferencing systems can be designed to accomplish everyday work-related or leisure-related activities with **a single display and camera**. This motivates us to design and investigate a system that tackles the challenges of sharing physical objects in the **conventional face-to-face, opposing view perspective of video chat systems**. We outline the following subsection to see how our work connects and extends prior work by supporting the *perspective* of space (2.1.2) and *remote referencing* (2.1.3) on the physical objects.

2.1.2 Challenges of Opposing View Perspective. The conventional face-to-face view can be challenging due to the opposing view perspective of video chat systems [22, 73], which flattens the view of physical objects [22]. Furthermore, the need for dynamic manipulation and re-orientation of handheld objects towards the front camera in a live video can be tedious, as users are responsible for presenting the handheld object to their remote partner. In particular, users may forget to present the object to their remote partners when their attention is on the physical material experience with the object, as found by Oehlberg *et al.* [73] in a scenario of critiquing design prototypes. To solve these challenges, ReMa [22] enabled a remote re-orientation of objects by supporting tele-manipulation with robot arms. CamBlend [77] utilized the wide FOV camera with multiple controllable, in-context views that enabled remote reference and captured snapshots. Licoppe *et al.* [61] explored how users orient objects with the support of a maneuverable device, the Kubi telepresence system. In contrast to these hardware-intensive solutions, our work focuses on developing a hardware-free approach to address object-sharing challenges when the user holds and orients objects towards the camera, with both dynamic live sharing and static non-live storage of physical objects in the virtual space.

2.1.3 Remote Gesture, Referencing, and Manipulation. The central theme in Computer-Supported Cooperative Work (CSCW) systems has been to facilitate remote collaboration by conveying the conduct of remote collaborators via remote referencing and gestures. Various methods have been explored in previous studies to facilitate remote gestures, including the use of cursor [77], laser pointers or tele-pointers [42, 56], as well as virtual overlays of bodies and arms [90] to provide supplementary non-verbal cues and subtleties. Furthermore, AR-based mobile collaboration supports a different perspective of space of what we aim to investigate. To solve ambiguous verbal referencing problems, this work focuses on supporting annotations and labeling to videos and helps users to move their mobile camera around the scene freely for physical tasks [10, 27, 29, 48, 74].

Moreover, prior work also supports remote gestures by enabling re-orientation and tele-manipulation of a physical proxy of the object, *e.g.*, utilizing robot arms [22] or shaping-changing actuation [22] to create physical affordance. However, such systems often require physical proxies of the same objects from both sides. Extending this work, we want to further investigate how enabling remote reference, access, and control to the parallel and shared digital copies of a physical object can help or hinder the remote user's understanding without physical proxies and hardware setups.

2.2 Integrating Physical Artifacts into Shared Surfaces and Spaces

Earlier tabletop research has explored the use of interactive tables to support and enhance creative group collaboration. Some of these systems focused on enabling the co-habitation of physical and digital artifacts on the tabletop, merging image capturing and display space [36, 39, 69, 80, 84, 85, 95]. With intuitive pen-based interfaces, the NiCE Discussion Room [37] incorporated paper artifacts and digital media to support sharing information in co-located group meetings. Transferring non-digital artifacts to digital output for sharing objects has also been explored in remote setups [38, 70]. Media space research has explored providing remote access to documents using multiple-camera setups, a shared desk area, and the life-size image of remote users in work [57, 67, 68] and social contexts [79, 103, 104]. To support sharing any surfaces, IllumiShare [52] enabled task and reference spaces with a peripheral device combined with a low-cost camera and projector. However, such metaphors usually require external hardware setups such as tabletop surfaces or projectors to capture the object and were better suited to capturing **flat artifacts** such as documents and books [54]. Such setups, thus, limit the flexibility of sharing three-dimensional physical objects [19, 22], leaving the reference to remote space disembodied and fractured [66, 99] with unsmooth transition between the person and task space [7].

In contrast to prior work that separated the working area (via capturing table surface) from the image of the remote user, CamBlend [77, 78] supported co-orienting and capturing larger objects such as information drawn on whiteboards and proposed a focus-in-context method in a panoramic video collaboration system. We aim to investigate the design possibilities embedded in the video mirror (face-to-face) space by integrating the physical artifacts in the shared virtual space, thus supporting fluid and spontaneous transitions between sharing objects and face-to-face communication during remote meetings. Unlike CamBlend [77], which supports spatially-anchored object windows as a focus in the environment for bi-directional object referencing in wider environments, our focus is on supporting the ad-hoc need to bring physical objects into the video window as digital copies that are entirely independent of the environment.

2.3 Design Space: Supporting Shared Task Space for Remote Communication

Buxton proposed three communication channels for remote communication and differentiates the benefits and drawbacks of separating or integrating person and task-reference space. Utilizing Buxton's framework [7], we grouped prior literature that created systems to support access to remote shared physical task space into a Design Space (Figure 2).

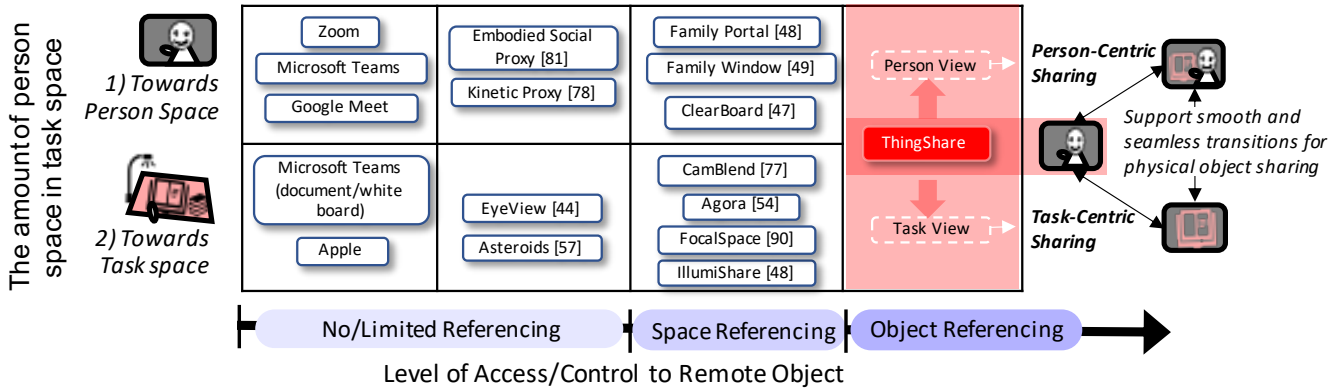

Figure 2: ThingShare Design Space of prior work that supports access to remote physical space or objects.

2.3.1 Towards Task Space. Current commercial tools, such as Microsoft Teams, use contour detection to capture the shared task space and support desk or whiteboard view. Apple has recently announced an approach that uses a single iPhone ultrawide angle lens camera as the webcam to quickly transition from face-to-face to desk view. FocalSpace [102] utilized the "focus+context" approach and captured both persons and objects as interactive elements using depth and shared image space in layer models, but its shared task space still primarily focuses on capturing specific surfaces, such as a whiteboard. CamBlend [77] supports task space using a different approach, as it integrates the person and task space yet blurs the human and unrelated task spaces, with a movable focus-in-context window deblurring certain objects for object-focused discussion.

2.3.2 Towards Person Space. Commercial video-conferencing platforms can support the quick showing of physical objects that form part of the task space via the *person space* video channel [7]. Media space and video-mediated collaboration research such as ClearBoard [46] primarily focuses on integrating the person and digital task space, limiting its ability to share physical objects through its person space channel like video-conferencing tools. ThingShare builds up on this metaphor of integrating the person and (physical) task space and focuses specifically on supporting the design of remote meetings for sharing physical objects with person-centric (Towards *Person Space*) and task-centric sharing (Towards *Task Space*).

2.3.3 Design Space. We devised a design space (Figure 2) to situate our techniques in relation to previous object-focused sharing system, suggest connections between techniques, and direct attention to relatively under-explored combinations.

On the left side of Figure 2, the rows indicate the communication configuration of either Towards *Person Space* (2.3.2) or Towards *Task Space* (2.3.1) of person and shared task space. We identify the task space here as the surrounding or wider physical space and environment rather than the digital shared task space.

The columns of the Design Space encompass our design of remote reference space into three levels of access to remote object referencing. We define this as a qualitative measure of how a given interaction technique limits the remote user's access to the physical objects of shared task space, *i.e.*, the degree to which a user provides the input (*i.e.*, remote reference) into the remote space, from no/limited referencing to object referencing. No Referencing refers to tools that do not support remote gestures and thus require a remote user to use ambiguous verbal referencing. Supporting tele-presence [59, 61, 88, 93] enables Limited Referencing , allowing a local user to be aware of where the remote user is looking at, and the remote user to access different viewpoints. However, only the direction of the space is noticed, and access to specific physical objects is limited. The use of ephemeral or persistent spatial-anchored remote gestures, such as arm shadows [90], cursor pointing, simple annotation [50, 51] and focus-in-context windows [77] can provide more remote control over Space Referencing to a specific object. By contrast, Object Referencing , which our work supports, allows the remote user to manipulate parallel and shared digital copies of objects. While many techniques for shared task space, including the work that supports entirely digital sharing or other display configurations such as VR/AR (*e.g.*, [91]) and hardware-intensive work that requires physical proxies (*e.g.*, [22, 23]), have been proposed, we have intentionally emphasized examples of work that support sharing physical environment and object-focused work with **2D screen or projected display**, since those are the most relevant to the ideas in this paper.

3 FORMATIVE STUDY OF OBJECT SHARING PRACTICES

Though there was a significant body of prior art, we found a dearth of knowledge regarding the current state of everyday object-sharing practices in video-conferencing tools and how their behavior and objectives vary depending on context. Therefore, we conducted an IRB-approved survey to help us understand how people currently show objects, to learn typical behaviors, and primary purposes for showing physical objects, and to gauge people's concerns and challenges to show things under different contexts, *i.e.*, work-related and leisure-related meetings or video calls.

3.1 Methods

We posted a survey on Mechanical Turk and recruited Mturk workers with a 98% approval rating and with more than 5000 HITs (*i.e.*, requests on Mturk) approved and paid each $3.00 for approximately 10-15 minutes. A screening question was included to filter out respondents who did not have regular video calls (less than

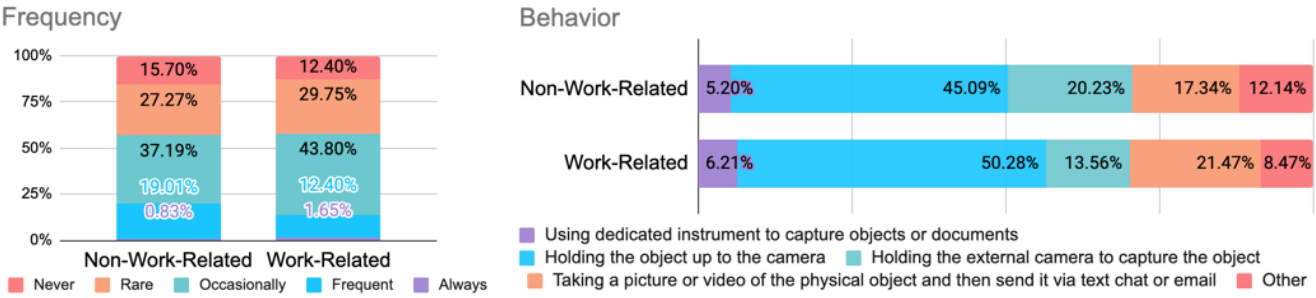

Figure 3: Frequency and behaviors of sharing physical objects in work-related and leisure-related contexts

once or twice a week) and hence could not provide insights for our research questions. We received 124 valid responses (74 male, 49 female, and 1 non-binary) from 156 respondents, which represent diverse occupations, including IT professionals, research lab managers, instructors, designers and creative professionals, salespeople, and homemakers. For respondents' frequency of having video calls in different contexts, 86.29% of respondents reported having work-related online meetings or video calls at least once or twice a week. 60.17% of respondents reported having leisure-related online meetings or video calls at least once or twice a week.

3.2 Findings

Here, we report the frequency, behaviors, purpose, and challenges of showing and sharing physical objects in work-related and leisure-related meetings. The complete summary of results is available in Appendix A.1. Overall, for users having work-related meetings at least once or twice a week, nearly half of respondents (72 respondents, 58.0%) reported showing objects at least occasionally (43.8% occasionally; 12.4 % frequent; 1.7% always). For users having leisure-related meetings at least once or twice a week, nearly half of respondents (71 respondents, 57.2%) reported showing objects at least occasionally (37.2% occasionally; 19.0% frequent; 0.8% always).

3.2.1 Behaviors of Sharing Physical Objects. Most respondents "Hold the object up to the camera" in work-related (50.3%) and leisure-related meetings (45.1%)[1] (See Figure 3 for other behaviors).

3.2.2 Use Cases of Sharing Physical Objects. We analyzed the variety of use cases for sharing physical objects in video meetings. Overall, the purposes in our sample include sharing objects for collaborative, instructional, or ad-hoc needs that vary across work-related and leisure-related meetings. We first categorized use cases found in work-related meetings (72 respondents) as follows.

(1) **Discussing details of products (31/72)** focuses on the detailed properties (size, color, design styles) of an object, such as the prototypes and products, *e.g.*, "*I work in a women's clothing wholesale. I showed color swatches or fabric samples as an example of the product. Images on paperwork were shown to see the item's style.*"

(2) **Remote instruction (11/72)** requires users to convey instructions for training or troubleshooting purposes, *e.g.*, to instruct others how to test and use a prototype: *"I showed my test device in a company training because I want to show others how I test the products"*, or a troubleshooting meeting that requires a user to show the broken parts to determine the root cause of a failure.

(3) **Ad-hoc demonstration and clarification (43/72)**, *e.g.*, "*I showed sample documents because I find it easier at the moment than sharing digital doc.*" The ad-hoc need for showing physical objects could also be to explain some complicated details with data graphs, sketches, visualizations, and calculations on paper or whiteboards (10/72) for clarification and explanation. For example, when discussing product modifications, a product analyst mentioned the need to show product changes or make calculations and notes on paper or tablets and show it to the camera for remote users to see, *e.g.*, "*Some of the company's products that are marketed to customers occasionally would refer to a wall chart with information; I might make notes or calculations spur of the moment and hold them up for others to see.*"

(4) **Showing personal items for small talk (17/72).** For example, some showed personal items when others saw things in the background and asked "*I showed a 3D printed item I made during a weekly meeting because people saw it in the background and were curious about it.*" " Some respondents showed items as part of the small talk, *e.g.*, "*I showed my dogs at various meetings to break the ice and garner some soft attention.*" or used objects to show reactions, *e.g.*, "*I showed a cup raising it as a toast to acknowledge a good idea.*"

For respondents who never (12.4%) or rarely (29.8%) showed objects, respondents generally indicated three reasons for not needing or wanting to share physical objects: 1) their work-related objects are digital, 2) the objects felt too personal and not professional for work-related reasons, or 3) it is challenging with current commercial tools.

During leisure-related meetings (71 respondents), respondents often share personal items in their conversations. Most respondents reported showing handheld and background items during social settings when the purpose could be ad-hoc and related to the conversation.

(1) **Sharing handheld Items.** More than half of them (37/71) showed items that they just bought, "*I showed outfits I purchased because my mother was curious about what I bought after my shopping trip to the mall.*" Moreover, respondents also show

[1]The list of behaviors included (1) Using a dedicated instrument to capture objects or documents; (2) Holding the object up to the camera; (3) Holding the external camera to capture the object (4) taking picture or video of the physical object and then send it via text chat or email; (5) other.

items to enjoy the moment with their family and friends (11/71), *e.g.*, having food and drinks together (8/71), and to play music remotely (3/71).

(2) **Sharing background items.** Respondents (15/71) showed background items, or how things fit in the space with remote families, *e.g.*, "*I was excited to show my new house to my mother, who lives far away...*" and "*I showed flowers in my garden in a conversation with my mother-in-law because we were talking about yard work and the season.*" Moreover, some of them provided how-to instructions (10/71) to remote families.

3.3 Challenges and Concerns of Showing Physical Objects

The survey findings verified our hypothesis that there was significant use of ad-hoc sharing and showing behavior of "holding a physical object up to the camera." The challenges and concerns we emphasized here occur more or less using all kinds of devices (desktops, laptops and mobile devices) yet slightly differ when using different behaviors (*e.g.*, capturing desk surfaces). Here we summarized the challenges and concerns of showing physical objects when using the behavior of "holding the objects up to the camera" that motivated our design.

- **C1: Difficult to show the finer detail or the actual size of the object.** The respondents found it hard to show the detail of an object and the actual size of an object, *e.g.*, *"I can not control the position quickly to show the right size."* P54 (Engineer) who holds the object to the camera found that it was hard to "*... show off some of the finer details of our prototype...*" P41 (Art Designer) usually shows some documents and paper artifacts with a combination of taking images and holding the paper up to the camera during work meetings, and she noted that "*The first I learned a lot about how to have a screenshot placement so they could see it..It was not right after I tried a few more times... I hold them up so they can see them.*"
- **C2: Unable to capture multiple pages of a document, different perspectives, or the entire image of a large item.** The respondents noted that "*it is hard to keep together and refer to the document when the document is of multiple pages.* The majority of them found it difficult to display multiple sides of the product, *e.g.*, "*Some idea is not easily copied on the phone or laptop, such as in a product design draft picture.*" and to show two items at once, *e.g.*, "*I can not show the object and my image at the same time.*" and items that cannot be picked up, *e.g.*, "*...trying to show my mom a vacuum cleaner that had a confusing part to it trying to get her to solve it and couldn't manage to get it into view.*"
- **C3: Limited view to reference and understand remote user's attention due to occlusion.** Respondents reported challenges to "*point to a specific section so that everyone can see exactly what I'm referring to.*" (P47, engineer, work meeting) and concerned whether the concept has been conveyed properly to everyone. "*It has to be clearly shown to others, and everyone has to notice what is being shown.*" (P66, production manager, work meeting)
- **C4: Laborious to repeatedly frame, coordinate, and adjust the position and angle of the camera or the object to show.** In both work-related and leisure-related meetings, respondents reported adjusting the angle to fit the physical objects in the video properly, and some mentioned feeling tired from doing so. This can be because "*the camera reverses images*" and also that the user wants to make sure to "*get the object in the field of vision*" and "*hold the camera or the thing steady.*"
- **C5: Incompatible with the virtual background in the current video chat or meeting platforms.** Respondents noted the challenges when showing objects with virtual background, *e.g.*, "*I was showing the notebook and phone; the blurring background obscured them and made it harder to decipher what they were at first glance.*" and they will need to breach their privacy to show objects, *e.g.*, "*I have to turn off the virtual background to show objects.*"

3.4 ThingShare Design Goals

The formative study findings demonstrated the need to show physical objects in both work and social contexts in everyday applications and the potential challenges of current showing practices. This finding reinforced our idea and hypothesis to support the ad-hoc and collaborative sharing of physical objects in a face-to-face metaphor. To facilitate this, we decided to support a heterogeneous set of physical objects, ranging from customer goods to work-related artifacts such as paper documents, sketchbooks, and hybrid objects like mobile phone screens with digital displays in physical formats. In light of these considerations, we formulated five objectives to direct the design of ThingShare.

- **G1: Provide in-context and detailed views of sharing the physical objects.**
- **G2: Capture and store various perspectives of an object.**
- **G3: Support remote gestures for collaborative referencing.**
- **G4: Support an efficient hands-free and temporal manipulation of object size and position.**
- **G5: Support flexibly showing and hiding objects from the surrounding environment.**

4 THINGSHARE OVERVIEW

With the five design goals, we aim to leverage vision-based instance segmentation methods and develop the ThingShare system to enhance the ad-hoc and bi-directional object-sharing practice in video meetings. With ThingShare, we aim to answer the following two research questions. **RQ1: (Local User)** How can we support effective communication for the local user who owns the physical object to share the physical object? **RQ2: (Remote User)** How can we support the remote user to understand and perform remote gestures on the shared object?

4.1 Interaction Concepts, Workflow, and Interface Design

To support both the contextual and detailed views of physical objects **(G1)**, the system was designed with two types of video windows. Firstly, the users' Person View (Figure 4A-1) is used to show objects (with the corresponding person), which enables person-centric sharing. Secondly, the Task View (Figure 4B-2) displays a shared close-up view of digital copies, providing a staging area for collaborative interactions and discussions that support task-centric sharing. Moreover, in the Person View, the local user's video is mirrored to assist with physical references and gestures on the object and prevent left-right confusion. Conversely, in the Task

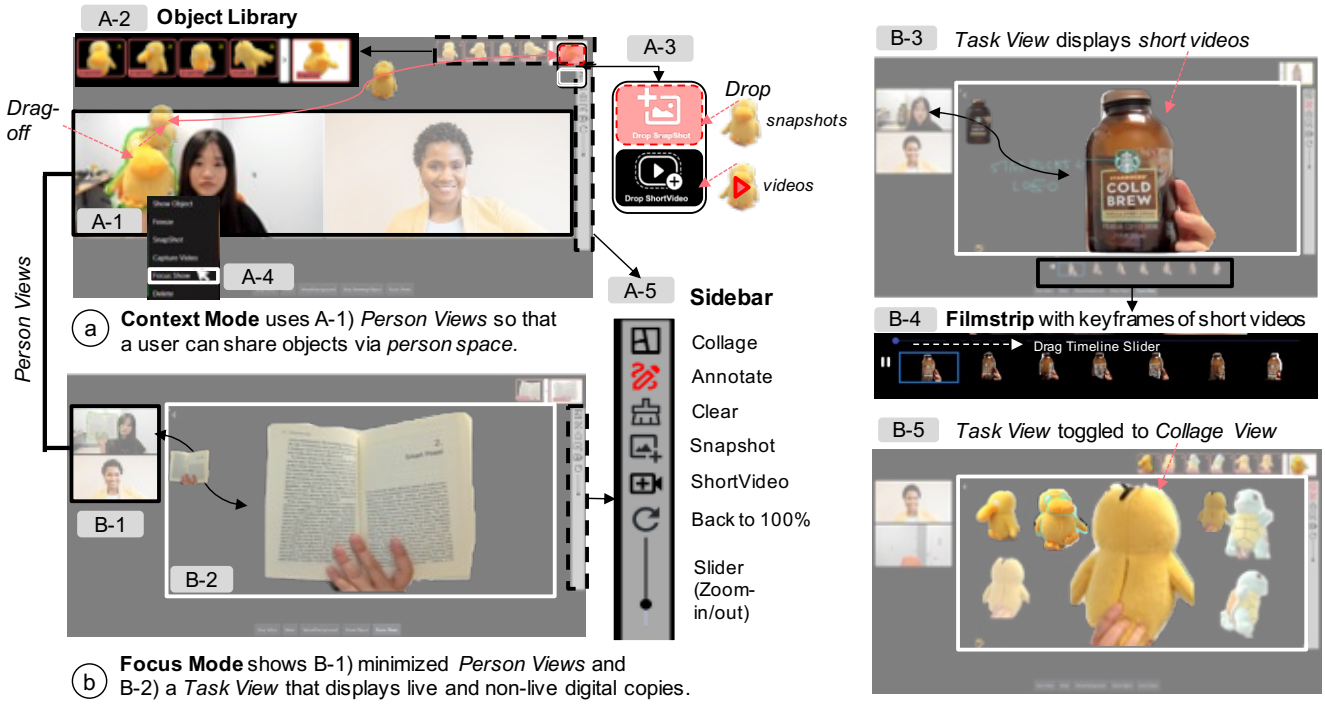

Figure 4: A walkthrough of the ThingShare user interface with two video layouts: a) Context Mode (default) and b) Focus Mode.

View, the close-up video of the object is not mirrored for both local and remote users, which allows them to focus their attention on the digital space and see the words correctly (unflipped) (*e.g.*, Figure 4b).

4.1.1 Person View. Based on the results of the formative study, it was discovered that people often share physical objects via their video feeds in an ad-hoc manner, such as causally sharing their personal belongings or prototypes without focusing on the fine details of the object. However, users may still encounter difficulties in sharing objects due to challenges in presenting the object from various perspectives **(C2)**, keeping it within the video frame **(C4)**, and avoiding occlusion by the virtual background **(C5)**. To address this disconnect and enable lightweight and expressive sharing of physical objects, we designed the *Person View*. Unlike traditional video containers, Person View serves as a temporal container for digital copies. Local users can create digital copies of physical objects in their Person View, which can then be manipulated and placed in the Person View, or moved to another user's Person View, Object Library, or the shared Task View. This allows users to easily create digital copies of a physical object from various views, combine them with their own video, and separate digital copies from the background.

4.1.2 Task View. Physical objects are brought to meetings not only for ad-hoc demonstrations but are also produced, modified, and reviewed afterwards, such as product reviews, hands-on training, and design sessions. Our formative study identified several challenges in showing objects in detail, especially during work-related meetings. Existing video meeting systems require users to hold objects up to the camera to show details **(C1)**, which can be difficult to keep steady and within the camera view **(C4)**. Additionally, referencing a specific section of an object can be challenging **(C3)** and effective work-related discussions may require sharing multiple objects or perspectives **(C2)**. Our study also found that sketches and paper artifacts were frequently shared and held up to the camera during discussions involving complex concepts, *e.g.*, showing quick drawing, scribbles of data visualization, or quick calculations. Furthermore, direct group attention to specific parts of artifacts can help. All of these use cases require thorough and close-up sharing beyond the Person View. To solve this, we designed a *Task View* (Figure 4b), a magnified object view, that covers the majority of the interface.

4.1.3 Video Layouts. Utilizing these two video windows support a balance between the focus and context, and enables both Towards *Person Space* and Towards *Task Space*, resulting in two video layouts. **Context Mode** (4.2) consists of the person views (Figure 4A-1) so that it shows physical objects via its person space channel, helping the user to drag-off a digital copy from the person view for a quick, ad-hoc, and expressive demonstration. **Focus Mode** (4.3) enabled two main UI changes: person view (Figure 4B-1) will be shrunk to the left side and a shared task view (Figure 4B-2) appears at the staging area, allowing the user to display different data formats of physical objects in more detail and collaborative manner.

Moreover, we support **Object Library** (Figure 4A-2) for both layouts as a temporal space to store captured perspectives of physical objects. In sum, the person view and task view video windows and Object Library served as *containers* for **Digital Copies of Physical Objects**. Any user can capture the items from any video window

and store them in the Object Library. They can also reuse the digital copies by bringing them back into any video window.

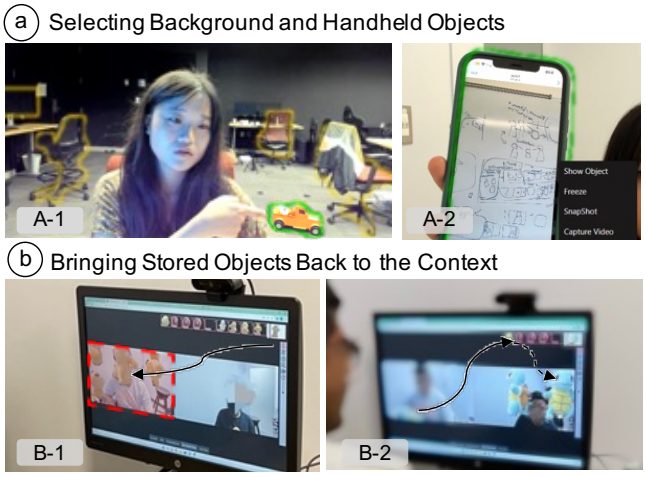

Figure 5: Interaction with the *Person View*: a) Contour; b) Bringing stored objects back to the person views.

4.2 Supporting Person-Centric Sharing of Physical Objects

Context Mode (Figure 4a) is the default layout for users, and it has a grid of equal-sized user person views (Figure 4A-1) in the center of the screen. To support person-centric sharing (Towards *Person Space*), users will be able to initiate sharing, freezing, and storage of physical objects by interacting with their own person views. When the user hovers over the person view, all detectable items will be highlighted in yellow contours (Figure 5A-1) to inform the local user that they are shareable. A green contour will appear around the user's selected object (Figure 5a).

4.2.1 Object Interactions. Right-clicking on the detected object triggers a Context Menu (Figure 4A-4) with six options: show/hide object, freeze, snapshots, short video, focus share, and delete.

- **Show or Hide Object:** To support flexible control of shared and private space with the virtual background on **(G5)**, a user can use the "Show/Hide Object" command to determine if a selected physical object is visible to a remote user.
- **Freeze and Interact with the Digital Copy:** By clicking on the "Freeze" command or double-clicking the targeted object, the system duplicates the selected physical object at the clicked position as a digital copy and enables further manipulation without storing it, such as dragging and resizing it inside the person view video. The physical object may be further dragged out of the video window into the Object Library to save as a snapshot.
- **Capture Snapshots:** When the user selects the "Capture Snapshots" command or drag-and-drops the targeted object into the Object Library with the "Drop Snapshot" drawer chosen (Figure 4A-3top), the captured item will be stored as a snapshot in the Object Library (Figure 4A-2).
- **Capture Short Videos:** If the user selects the "Capture Video" command or drag-and-drops the targeted object into the "Drop Shortvideo" drawer (Figure 4A-3bottom), the captured item will be initiated as a short video and stopped by the user at any time.
- **Focus Share:** Clicking on the "Focus Show" (Figure 4A-4) or dragging the item to the right side of the interface (near to Sidebar) starts the Task View, a shared staging area for discussing the details of an object. The system will then navigate to a detailed view of the selected object (See Section 4.3).
- **Delete:** To remove an object from the person view, the user can select the object and click "Delete".

4.2.2 Manipulating 2D Parallel Copies In and Across Video Windows. To enhance and augment the immediate and ad-hoc demonstration needs of physical objects and the efficient hands-free and temporal manipulation of objects **(G4)**, we enable the user to capture the static viewpoint of physical objects as a digital copy in the user video feeds (Figure 6).

Within the user's video streams, the replicated 2D digital copy can be dynamically re-positioned and re-scaled (Figure 6a). Another option is to record a short video of the chosen item and save it so that the user may play it back and forth with different viewpoints of the object and narrate it to display a dynamic activity with the object. Moreover, handheld items (Figure 6b), hybrid items such as contents on the mobile phone screen (Figure 5A-2), or background items (*e.g.*, Figure 6c-d) can be captured. To enable both local and remote reference on the captured object, all users can reuse the any copies by grabbing the stored parallel objects or frozen things into their own video feed to manipulate them *in parallel* (Figure 5B-2). In addition, the local user's cursor on a remote user's video feed will be tracked from the remote user's side, so that references to the remote user's physical objects in the virtual space are visible. As everyone can interact with the digital copies of a physical object, this provides distributed shared virtual items for everyone to view and engage with, as well as a quasi-shared view of two people looking at the same physical object as if they were side-by-side **(G3)**.

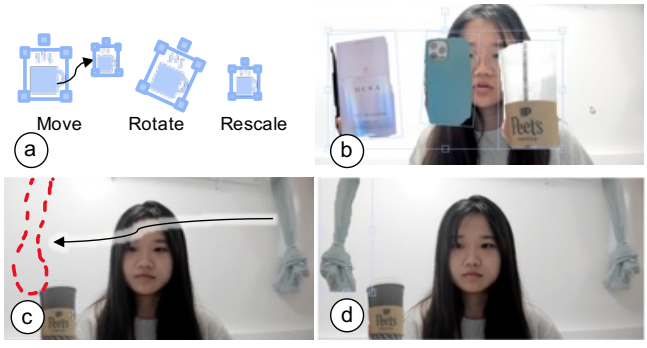

Figure 6: Manipulating 2D parallel digital copies in the video window: a) Rescale, Move, and Rotate. b) the user groups the frozen items; c) the user freezes the background curtain d) and locates it in a symmetric position.

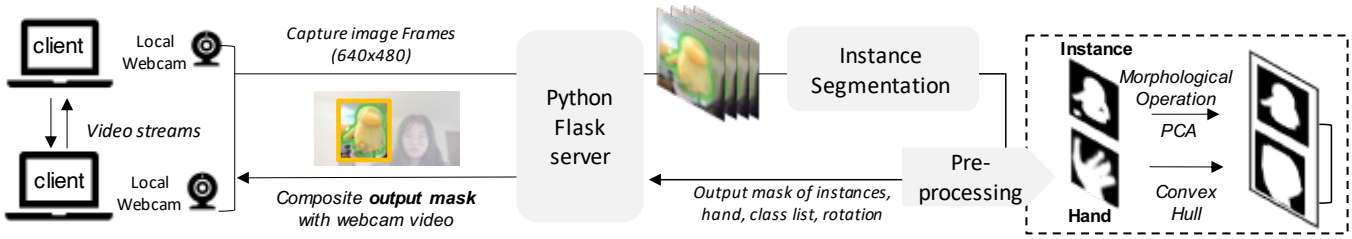

Figure 7: Workflow of ThingShare.

4.3 Supporting Task-Centric Sharing of Physical Objects

To support task-centric discussion (Towards *Task Space*), the main part of the screen (Figure 4B-2) shows the Task View, in which the physical objects are shown in a shared staging area in details. The Task View is used to display different digital copies: Snapshots, Short Videos and Live Video. The Task View can also be toggled to Collage View (4.3.1) to display multiple digital copies at the same time.

Similar to the workflow of **Context Mode**, the user can double-click or drag-off items from their person views to freeze and capture snapshots or short videos. They can drag off selected physical objects from their person views into the shared Task View to show quick snapshots. When a user clicks on the thumbnails of snapshots or short videos in the Object Library (Figure 4A-2), the stored items will be displayed on the Task View (Figure 4B-1) and any user can click to toggle between different digital copies. If the user clicks on the "Short Video" icon of the Sidebar (Figure 4A-5), the interface will record a short video of the object and store it as a thumbnail in the Object Library. Clicking the short video thumbnail in the Object Library initiates a single playback of the video in the Task View and displays the Filmstrip (Figure 4B-4) with seven keyframes. The user can slide along the timeline, pause, or click on key frames of the sequence of the snapshots to display different keyframes in the task view, and annotate key frames of short videos. To keep the reference space in the digital copy, capturing snapshots and short videos using Sidebar, or sharing live videos in Task View will preserve the hand (Figure 4B-2 and B-3).

4.3.1 Collage View: Supporting Multiple Perspectives and Multi-User Sharing. Supporting the detailed view of snapshots, short videos, or live videos does not include viewing different perspectives of an item or two items at the same time **(G2)**. Collage View supports all users to capture their own objects from the physical environment or bring digital copies stored in the Object Library into the freeform collage in a What-You-See-Is-What-I-See (WYSIWIS) canvas. Hence, switching to the Collage View (Figure 4B-5) by toggling it in the Sidebar enables both local and remote users to drop stored items into the Task View for collaboration, thus supporting the discussion over multiple perspectives or different items on the Task View with the moving, rotating and rescaling interactions.

4.3.2 Annotation. Annotation (Figure 4B-3) on the static and key frames of a short video will be saved in the displayed thumbnails in real-time. Moreover, as the live video of objects is stabilized at the center of the Task View, any annotation in the system during live video sharing is object-anchored rather than spatially anchored. This indicates that when annotating on a physical object in a live video, the annotation will follow the physical objects rather than stay in mid-air or world-stabilized [28, 29, 48].

4.4 Object Library and Stored Digital Copies

The digital copies can be stored in non-live data formats in the Object Library. The box in the upper-right corner of Object Library displays the most recent acquired images and/or videos. Expanding or collapsing the Object Library reveals or conceals all historically captured copies. In sum, digital copies of physical objects can be shown in three formats: snapshots, short videos, and live videos of physical objects, thus allowing a captured object to exist simultaneously across different time (static, sequentially a sequence of static viewpoints, or live), contexts (drag from any container and drop into any container), and spaces (local and remote), *i.e.*, frozen in the user video feed, saved in the Object Library, displayed in the shared task view video window.

4.5 Applying Instance Segmentation in Video-Conferencing Interfaces

The front-end web interface was built using React, Javascript, HTML, and CSS. The backend authentication server was built using Flask to run the instance segmentation (Yolact [4]) and object tracking model (Deepsort [97]). The front-end and the back-ends communicate using a Flask1.1.2 and WebSocket to transmit data and HTTPS requests to access the APIs. To enable P2P WebRTC communication with more than one peer, the Simple-Peer library [3] was used.

4.5.1 Pipeline. The web client sends video stream from the user's webcam to a Flask back-end server using socketio as a feature input into the model. The model in the server then predicts the mask and class list of object detection before sending the output prediction back to the front-end. The default training set for the instance segmentation model Yolact [4] is the COCO dataset [1], which contains over 200,000 images and 80 item categories. Based on the best model performance given in [4], we utilize the pre-trained ResNet-101 [40] model with FPS 33.5 as the backbone.

4.5.2 Pre-Processing. Body segmentation was employed to segment persons out of the surrounding home environment in order to ensure privacy, yet instance segmentation approaches were not used for object-sharing practices in video meetings. Most errors are caused by mistakes in object detection such as misclassification, or box misalignment. There are also some issues caused by the mask generation algorithm [4] and the nature of classification tasks.

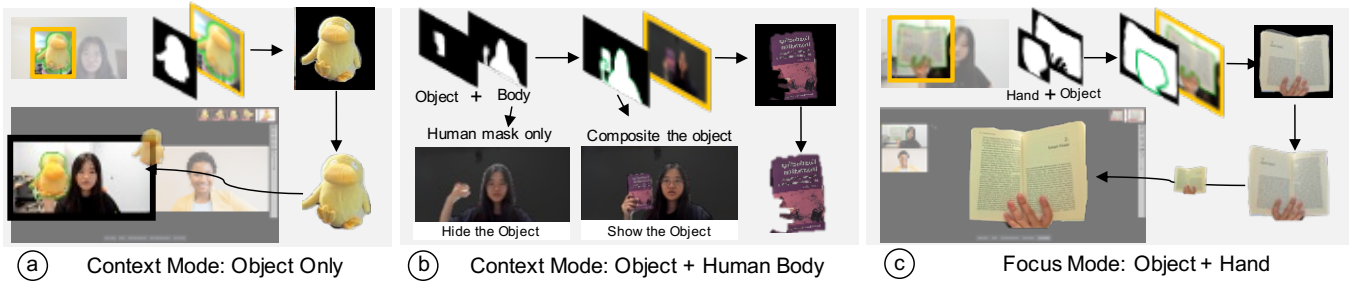

Figure 8: Compositing Operation: a) Context Mode (without hand); b) Context Mode with Virtual Background (reveal/diminish the object) c) Focus Mode (with hand).

Hence, prior to obtaining a suitable mask for the end-user experience in video meetings, there are several critical pre-processing steps: 1) **Eliminating Leakage Noises:** The Yolact approach crops the masks after assembly, and does not suppress the noises outside of the cropped region [4]. For example, Figure 17 in the Appendix exhibits several leakage issues. This indicates that whenever the predicted bounding box is imprecise or misaligned, the leakage noise may creep into the instance mask, generating some tiny artifacts or even merging the segmented mask of some other classes (e.g., human) with the predicted item. Hence, we first used a MediaPipe body segmentation model [2] to get a separate mask of the person to segment the whole virtual background. We then performed Morphological Operations for opening (*i.e.*, erosion followed by dilation) and closing (reverse of opening, dilation followed by erosion) to filter out the holes in the scene on the output masks (with binary color) (See Figure 7: Pre-Processing) and applied Gaussian Blur to smooth the boundary. 2) **Maintaining the Reference Space:** As the general instance segmentation approach is aimed for classification tasks that output a unique label to every instance, it directly detects and segments all the items apart out of the scene, including the hands (reference space) that are pointing towards an object (task space). However, to apply this in the remote meeting context and preserve the remote gestures in Task View for detailed sharing, we perform Compositing Operation (Section 4.5.3 and Figure 8) in the front-end to composite the masks of objects and human body or human hands with the video stream. Therefore, we segmented the hand and then utilized Convex Hull (See Figure 7: Pre-Processing) to expand the convex polygon of the hand mask so that the hand could be composited with the item in the foreground.

4.5.3 Compositing Operation. The output mask for an item and the remote gesture (*i.e.*, hand or human body) are key requirements for implementing instance segmentation in a video meeting environment. In the Context Mode, for the scenarios with a virtual background, we sequentially compose the person segmentation (from MediaPipe) with the output mask of handheld objects (Figure 8b). To temporarily hide the object, the composition only includes human segmentation and the video stream. The composition involves human segmentation, the object, and the video feed in order to reveal the item. When there is no virtual background, we composite the object mask with the video stream (Fig 8a).

Moreover, In the Focus Mode that aims to show a close-up view of the object, in order to get an ideal mask where only the remote gesture (i.e., hand) is included within the segmentation boundary (see Figure 8c), we compose the hand with the original mask of the physical object.

5 EXPLORATORY USER STUDY

We conducted an IRB-approved user study to understand how ThingShare improves communication over the physical object by 1) supporting local users (who own the physical object) to effectively share physical objects; 2) supporting remote audiences to understand and reference the remote physical object being shared.

5.1 Study Design

The study used a within-subjects design. The independent variable was the object-independent features out of the environment (*i.e.*, object-separated (ThingShare) or object-embedded (Baseline)). We chose Zoom as the baseline as it is one of the mostly used video meeting system. Since Zoom does not support any physical object sharing features, users need to rely on the live video feed mainly used for showing faces. For dependent variables, we measured users' attitudes towards communication, and the effectiveness of showing and remote referencing on physical objects to understand the effect of ThingShare on user experiences in shared task space using a questionnaire.

5.2 Participants

Sixteen participants (10 male and 6 female) were recruited as pairs from the university. Their familiarity with such tools was high (Median=6, IQR=1; 1-7 with 1 being no experience) and their frequency of sharing things was diverse (Median=3, IQR=2; 1-5 with 1 being Never). The study took around 90 minutes per deployment, and each participant was compensated with $20 USD.

5.3 Study Setup and Procedure

Pairs of participants were invited to a room divided into two spaces so that they could not see each other nor the physical objects. The server was deployed on a PC with an Intel Core i9-9900K CPU, 32GB RAM, and an NVIDIA GeForce RTX 2080 GPU. The input videos from both sides were captured by an external Logitech BRIO webcam. Upon the completion of the consent, the participants were given a demonstration of ThingShare features by the researchers and briefly used the system to get familiar with ThingShare. Then, the participants were introduced to three tasks. Before each task, the researchers explained the purpose and procedure to participants for 10 minutes. Following the instructions, participants engaged in

the baseline and ThingShare conditions with different sets of objects (Figure 9). The condition orders were counterbalanced across deployments. After completing each condition and context, participants responded to an intermediate questionnaire about the drawbacks and benefits of the tried system and the effectiveness of the system in completing the tasks using a 5-point Likert scale and provided comments. After completing all tasks, a final questionnaire collected user preferences over different conditions, subjective feedback, and willingness to use in real-world applications, followed by an interview.

5.4 Tasks

The tasks selected for this study were based on video meeting scenarios that involve showing objects, as identified through prior research and our own formative study.

Task 1: Ad-hoc and Casual Sharing Task *(10min).* The first task was designed to emulate the ad-hoc and informal object-sharing scenarios that we observed in our formative study, both in work-related and leisure-related meetings. This task was relatively simple, serving as a warm-up with a brief round of introductions. Participants were asked to choose two objects, one from handheld objects (Figure 9A) and another from background objects (Figure 9B) prepared in the room, and show and explain the objects to each other using ThingShare and baseline. The participants completed the task in the following two contexts: 1) Work context: the participants were asked to turn on the virtual background as if they were having a work meeting from home; 2) Social context: the participants were instructed to assume that they were friends on a casual video call and not use the virtual background.

Figure 9: Objects used in the study: a) Task 1: handheld and b) background objects; c) Task 2: VR and game controllers; d) Task 3: floor plans and furniture brochures.

Task 2: Single User Sharing Task (20min). This task mimics the remote assistance meeting where a local user with a device needs the help of a remote user to operate a device, which was seen in previous studies (*e.g.*, [58]) and the formative study responses. In this task, the participants took the roles of a local user needing assistance and a remote expert who helps the local user. The local user was given either a VR controller or a game controller (Figure 9C) and the remote expert was given an instruction document of the controller that explained how to perform two specific tasks (each for ThingShare and baseline) using the controller. The objective of the task was for the local user to learn the procedure for using the controller from the remote expert, and the task was completed when the local user successfully restated the procedure. After completing the task, the participants switched roles and completed the same task with a different controller that uses different instructions. The complete controller instructions are available in Appendix A.2.

Task 3: Collaborative Sharing (20min). This task was designed to mimic a scenario where two people need to collaborate remotely on a task that involves showing and referencing objects in their environments. In this task, we told the participants to assume that they would be sharing an apartment as roommates. The participants were asked to decide on the room allocation and find furniture to decorate their rooms. The same sets of floor plans and furniture brochures were provided to both participants (Figure 9D). The brochures had page numbers and item numbers that the participants could use to refer to a specific item if they wanted. The two sets of floor plans and furniture brochures were prepared to differentiate tasks between ThingShare and baseline. The task was completed when they finished allocating rooms and finding the furniture.

Interview Session (10min). After the completion of all tasks, we conducted a semi-structured interview to ask the participants how they used ThingShare to interact with each other, how ThingShare influenced the way they share objects, their positive and negative experiences with ThingShare, how they would use ThingShare in future applications, and other thoughts on the system. We also asked them to explain key sharing behaviors that we identified with the live observation notes.

5.5 Interview Data Analysis

We used open, axial, and selective coding [16] to analyze the interview data into higher-level categories. The researcher first read over the transcripts and developed and broke them into discrete codes. Our axial codes included the categorizations of the open codes. As the analysis progressed, we began to observe recurring themes, such as the choices of various digital copies, exhibiting and reducing digital copies in person views, and utilizing task views to construct shared perspectives. The researcher then established connections between these discrete codes, which were then organized and grouped into broader categories (axial). Finally, the categories were brought together to form overarching themes.

6 RESULTS

A one-way ANOVA (α = 0.05) was used to analyze the general and task-related questionnaire questions. The results of the questionnaire are presented in Table 1 and the visualizations can be found in Appendix A.2.1. For overall preference, 11 out of 16 participants preferred ThingShare to the Baseline during Task 1. For Task 2, 15 participants preferred ThingShare to the Baseline. For Task 3, 11 participants preferred ThingShare over the Baseline.

We present our qualitative findings with the following themes: reflecting on usability (6.1), presence and absence of digital copies (6.2), factors affecting choices of digital copies (6.3), and constructing shared perspectives (6.4).

Question	*Task 1*			*Task 2*			*Task 3*			*General*		
	p-value	TS	BS	p-value	TS	BS	p-value	TS	BS	p-value	TS	BS
It was easy to complete the task with the interface.	0.349	4.5, 1	4, 1.25	0.004**	5, 1	3, 2.25	0.071	4.5, 1.25	3. 2.25	0.005**	13, 2	10, 3.25
It was easy to communicate with my partner.	0.176	5, 1	4.5, 1.25	0.010**	5, 1	4, 2	0.085	5, 1	4, 1.25	0.002**	14, 2	11, 2
I was able to effectively show the physical object.	0.141	4, 1	3.5, 1.5	0.009**	5, 1	4, 2.25	0.011*	4, 0.5	2, 2	0.001**	12.5, 1.25	10, 3.25
I was able to understand what my partner was talking about the physical object.	0.688	4, 1	4, 1	0.111	5, 1	4, 2	0.040*	4.5, 1	4, 1	0.099	13, 2.25	12, 2.25
I was able to effectively reference the object.	0.227	4.5, 1	4, 1	0.019*	5, 1	3.5, 1.5	0.013*	4, 1	3, 1.25	0.007**	13, 2.5	10, 3
Task-Related Questions	p-value	TS	BS	p-value	TS	BS	p-value	TS	BS	p-value	TS	BS
I was able to show physical objects without compromising privacy.	0.006**	4.5, 1.25	3, 2									
I was able to effectively show and discuss different perspectives of a physical object.				0.001**	5, 1	2, 2.25						
I was able to effectively show and discuss specific details of a physical object.							0.004**	4, 1	2, 1.25			

Table 1: Summary of significant survey results of ThingShare (TS) vs. Baseline (BS) p-value with (Median and IQR). $(*)$ indicates significant difference by Wilcoxon signed-rank test $(p < 0.05)$. (1: Disagree - 5: Agree)

6.1 Reflecting on Usability of Capturing and Storing Digital Copies

Overall, the participants were positive about the user interface and the interactions of ThingShare. Many participants (11/16) found the contour highlighting helpful for identifying and selecting detected objects, and (9/16) found the drag-and-drop interaction to be intuitive and easy to use. Most participants (15/16) mentioned that they found resizing and repositioning the snapshots useful. On the other hand, some participants (4/16) suggested including the user's hand in the captured snapshots of Context Mode, as the hand serves as a reference for perceiving the size of the object. For example, P16 stated, "*...I prefer having the hands with captured objects because I can let the remote partner understand how large the item is...*" The inability to directly place objects into another user's person view video was also an occasional source of difficulty, as participants mentioned that directly placing objects would be more efficient (P11) and help gauge attention (P15) in some situations. This feature was intentionally blocked as we designed one's person view video as their own private space that others cannot manipulate. Though this was not explicitly implemented, a download button in the Object Library beside each snapshot or short video was asked for by several users (4/16) so that they could archive and take notes of the item being discussed "*Also, making a collage and saving it for later was so cool, but I would like to download it immediately.*" and they sometimes forgot the context of some snapshots being captured. Additionally, P11 and P15 suggested having some auto or semi-auto features for object selection (*i.e.*, auto selection or suggestions for object selection) with less cursor-based interaction. As P11 noted, "*quickly capture snapshots without mouse interaction.* " These suggestions may be valuable for future design considerations.

6.2 Presence and Absence of Digital Copies

Our results show that users appropriated the combination of interactions with digital copies (*i.e.*, freeze, store, diminish and reveal the digital copies within their person views) for a variety of social interactions. These interactions not only augmented object-sharing behaviors, but also allowed users to reveal objects from virtual backgrounds.

Blending Digital Copies and Person View Window. We observed that participants frequently used digital copies of physical objects to blend with their person view videos for casual and playful conversations. For example, P9 froze an apple in the *person view* and pretended to eat it in the video, saying "*See, I am eating this*" (Figure 10b). Another participant (P10) blended digital copies of her psyduck toy and her partner's apple to create an interesting scene, saying, "*my duck just stole your apple.*" P7 created multiple copies of her partner's object to decorate and tile her virtual background. Participants also used their body projections to perform embodied interactions with the captured objects, as if they were interacting with a physical object (*e.g.*, see Figure 10c). Moreover, similar to findings in [73], participants occasionally forgot to present the item to the remote audience when looking at their objects in the baseline. However, after creating digital copies with ThingShare, participants rarely showed objects live to their remote partners via the person view. Instead, some of them would freeze an object in the person view and then shift between looking at the physical item and pointing to the digital copy in their video window using their embodied projections (see Figure 10d). This demonstrated the efficiency of the blended effect in facilitating the transition from examining the physical properties of an object to physically referencing it. Overall, ThingShare provided an effective embodied affordance for more efficient sharing.

Remote Access to Digital Copies Enables Parallel Exploration. Participants (16/16) appreciated that they were able to interact with the digital copies of their partner's objects despite being remote from the actual object. From the remote user's perspective, we also observed them dragging off digital copies of physical objects from the local user's person view to their own view to magnify and see details during a discussion. For instance, when a local user was promoting an apple variety in Task 1, the remote user dragged a stored apple into his own person view and zoomed in on it, saying "*but this apple has a hole on it.*" In another example, a user was describing the details on the Starbucks cup sleeve (Figure 10e), while the remote user viewed it in his own person view and asked questions about specific details. This enabled parallel exploration

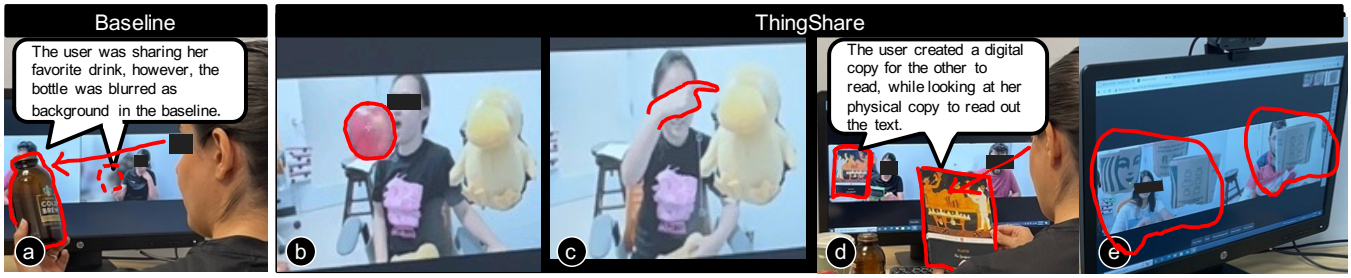

Figure 10: This sequence in Task 1 shows a) how the baseline condition failed to support the flexibility between shared and private space; and how ThingShare enabled b) blending illusions; c) embodied gestures; d-e) augmented discussion.

and provided a more efficient way to reference specific details of an object for instant questions.

Maintaining Privacy While Sharing Physical Objects. According to the questionnaire results, participants were able to show objects without comprising their privacy with ThingShare (ThingShare: Median = 4.5, IQR = 1; Baseline: Median = 3, IQR = 1). Furthermore, participants (14/16) appreciated the ability to maintain privacy while selectively expanding the shared space to include physical objects, while all participants found it difficult to show objects with *baseline* with the virtual background (*e.g.*,"*the object gets blurred if the object is placed outside the human body.*", P4). This was particularly the case for background objects that could not be held up to the camera. Some participants appreciated the controllability of the baseline condition, as they found that the physical object could be shown when the handheld object was within the boundary of the human body, and hidden when moved out of the human boundary (*i.e.*, P1 noted "*I can control whether I want the audience to see the object or not using the background.*"). However, all participants found the amount of shared area enabled by the body area in the baseline condition to be limited (See Figure 10a).

6.3 Factors Affecting Choices of Digital Copies

We sum up factors that lead to the use of different formats of digital copies, *i.e.*, live videos, snapshots, and short videos.

Handsfree Benefits of Sharing Non-Live Digital Copies over Live ones. Participants (13/16) preferred non-live copies (*i.e.*, snapshots and short videos) to live videos as non-live copies allow them to free their hands from the object and efficiently discuss the objects (*e.g.*, "*I do not need to hold the item all the time, and also I can zoom in to let the picture become bigger*, P12). In particular, some participants (4/16) expressed concern about the difficulty of digesting multiple frames of information and remembering details in live videos, *e.g.*,"*I need to digest multiple details of live video over time and remember the details I saw versus what I am seeing at the moment*" (P11). Three participants also commented on the need to capture copies for local users and revisit past activities in the case of live sharing only.

The Potential Value of Combining Live and Non-Live Digital Copies. While non-live copies were useful in many situations, we also observed instances that showed the potential value of combining live and non-live digital copies for increased awareness and collaboration. In Task 2, we observed that some local users would occasionally show live videos of themselves holding the controller through their person view window in order to ask the remote user if they were holding or using it correctly while both users were viewing the non-live copy of the controller in the task view. Some participants suggested combining live and non-live copies to provide more detailed information in certain situations with a trade-off between live and non-live views, *e.g.*, P13 commented -"*I hope in Focus Mode, there could be a way to expand my video [person views] when I wish to signpost the non-live copies but also show it lively through my video.*"

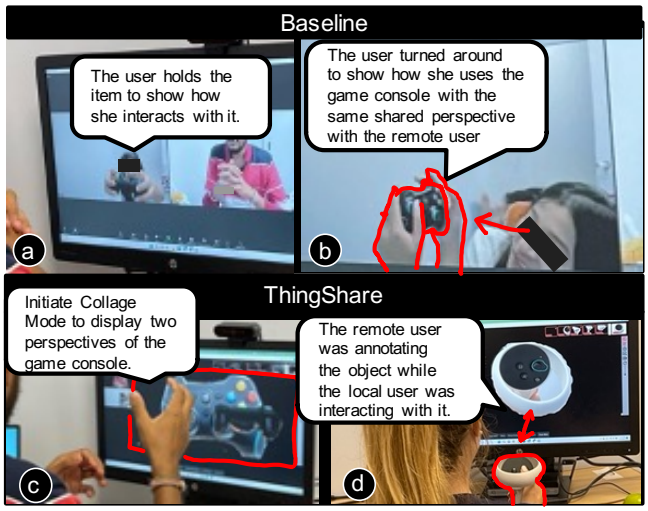

Figure 11: This sequence in Task 2 illustrates how ThingShare helped remote assistance and constructed share perspectives of objects.

Use Cases Affecting the Sharing of Non-Live Digital Copies. Participants have divided preferences over capturing the snapshots and short videos. Some participants (11/16) liked that short videos can mitigate the tediousness of capturing snapshots multiple times, especially when the captured snapshot is not optimal. They preferred the short videos when showing objects with multiple facets (P11, P13, P16), such as a cube, or capturing the motion of an object as tutorials (P9, P12) such as demonstrating switching light bulbs. In addition, some participants (4/16) did not favor short videos as the short videos can be unnecessarily long with transitioning scenes,

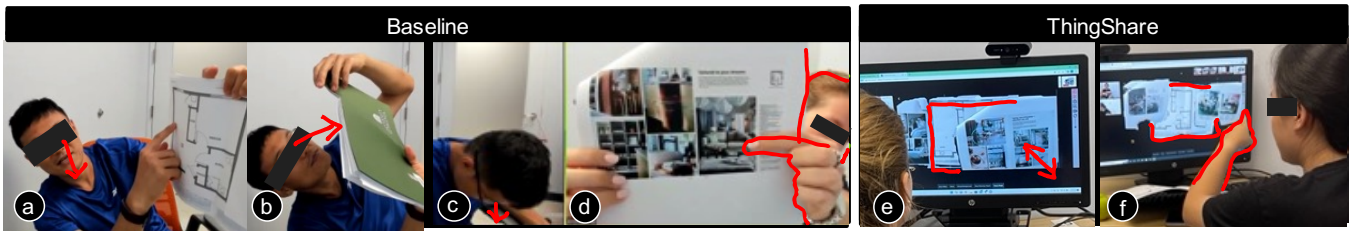

Figure 12: This sequence in Task 3 exhibits a-b) how a user frequently shifted from showing the room plan in the opposing view to looking at the physical document in baseline condition; c-d) the user was showing the object, but another user was looking at his physical space; e) with ThingShare, the user enlarged and annotated the item to demonstrate the specifics and how the furniture fits into the room plan; f) The user pointed to the virtual location using physical reference.

e.g., P7 commented - "*it is hard to find the best framing of the short video so that I need to coordinate the framing again and again and make a longer video than expected.*"

6.4 Constructing Shared Perspectives

The Focus Mode reconfigured the layouts with a *Task View* that enacts shared perspectives for diverse uses in both asymmetric sharing tasks (one user shares objects) and symmetric sharing tasks (both users have the same set of objects).

Digital Copies Enable a Shared Perspective for Asymmetric Tasks. For Task 2 (Asymmetric Task), participants generally used two object-sharing behaviors in the baseline condition: 1) the remote user(s) instructed the local user(s) to show different perspectives of the object, and the local user(s) rotated the object in their hand to interact with it before rotating it back towards the webcam for communication; 2) the local user(s) turned themselves towards the object to share the same perspective with the webcam as if they were looking at the object from the same side (See Figure 11a-b). With ThingShare, however, local users rarely turned themselves or moved objects back-and-forth for interaction, instead utilizing the snapshots and Task View (See Figure 11d). Participants appreciated the Collage View, which allowed for the display of objects from multiple perspectives at once (See Figure 11c), reducing the need for inefficient verbal communications requesting for perspective changes. Participants (15/16) also reported that ThingShare helped construct a detailed shared view with the remote user reference, allowing for live annotation on the focus-shared object and the ability to interact with the object based on remote instructions without the need for switching views (Figure 11d).

Digital Copies Help Gauge Attention in Symmetrical Tasks. As both users had the same set of objects in Task 3 (Symmetric Task), we observed that participants using the baseline condition struggled to synchronize their views. When discussing furniture, most participants (13/16) were looking at their physical artifacts instead of holding up the object to create a shared digital view. This made it difficult to confirm whether they were looking at the same page ("*In baseline, I didn't know whether the remote user was looking at the same page with me*", P9) and to gauge the other user's attention ("*when I was showing it [to the camera] to confirm, my peer is still looking at his book [brochure]*", P12). Figure 12c-d shows that a user did not pay attention to the screen when his partner(s) were holding up to show something. As a result, participants preferred to communicate using verbal descriptions, such as page numbers and locations on the page. In contrast, participants using ThingShare were able to use the Collage View to place and discuss multiple snapshots of the objects. Some participants also used the snapshots as a "cursor" by minimizing the snapshots to indicate the position they wanted to place furniture. However, we observed that two participants used finger gestures (Figure 12f) to point to shared items on the screen, which was not usable with ThingShare.

All participants mentioned that they liked that the shared object automatically filled the task view window without showing unrelated objects and details in ThingShare. In baseline, the participants had to physically move an object close to the webcam to show the detailed view, which was also difficult to maintain as the object could ward off their view (see Figure 12d). However, some participants (7/16) still brought objects closer to the webcam for better resolution, which led to challenges with object tracking. It often led to losing track of an object as the object tracking model could not detect the object boundaries, which may need to be addressed. Some participants suggested a post-cropping feature to further eliminate unnecessary information on snapshots.

7 DISCUSSION

We discuss the novelty of ThingShare in relation to prior work, the two RQs, and the five challenges with the design and study of ThingShare with design implication, applications, and limitations.

7.1 Supporting Ad-Hoc Copies of Objects from the Physical to the Digital Realm

We have investigated how features around digital copies support local users to share physical objects (RQ1) and enable Towards *Person Space* and Towards *Task Space* to address **C1**.

7.1.1 Towards Person Space. Our findings showed how the physical belongings of a user could be blended into the remote user's *person space* to enhance social experiences. Users reconfigured physical objects to create blending illusions with stored digital copies and remote users' video feeds. This enabled playfulness in their social interactions, embodied referencing, and privacy protection. The blending of spaces showed similar benefits to those of ClearBoard [46] and MirrorBlender [34]. In terms of the enhanced social

experience, previous work has explored the photorealistic aspect of digital copies, but it has primarily focused on the person space rather than objects, as evidenced by the work of Venolia *et al.* [94] to support the co-presence of distributed users in group photos, or Follmer *et al.* [24] to place human video streams in a storybook for playful interactions and emotional togetherness. Thus, ThingShare extends these ideas to the sharing of physical things.

Furthermore, prior work has explored how the choice of virtual background in video chats can affect the remote user's perception of the local user's personality traits and exhibit self-presentation [45], which leads to future investigation of the implications of incorporating personal belongings into remote shared environments. With ThingShare, we have shown how users can use digital copies in their person view to customize their virtual backgrounds, *e.g.*, tiling shared personal belongings from a remote partner's video into one's own person space. Moreover, ThingShare also addresses the challenge of objects being hidden by the virtual background **(C5)**. By selectively allowing objects to be visible regardless of the presence of a virtual background, users have flexibility to show objects in their local environment without compromising privacy. While ideas such as diminished reality are not new (*e.g.*, [11]), most previous work has not discussed or designed remote collaborative settings that do not require complicated setups [62]. Prior work has explored privacy use cases such as diminishing human activities from the video (*e.g.*, by hiding the temporary absence of activities such as looking at mobile phone [17] and freezing the user's video feed in a private conversation [43]). To complement these, we have studied the compelling use case of supporting object-sharing while the virtual background is being used.

An interesting future implication is to investigate the potential of using additional modalities to control or automate the boundary of diminished reality with speech, gesture, or proximity cues [11, 17, 102]. Designers could consider some semi-autonomous or gestural detection to initiate the control of an object being diminished or shown that can immediately respond to their behavior, *e.g.*, detecting movement of tangible objects to reveal/hide it to enable touchless and remote pointing like [14, 15].

7.1.2 Towards Task Space. ThingShare offers three options for sharing objects: live video, snapshots, and short videos. These options are available in both the Context and Focus Mode. We have discussed that interacting with these digital copies in a person view provides augmented context information that includes the surrounding home or workspace environment, the person, and the reference space.

In contrast, the use of digital copies for task-centric sharing enables the user to organize detailed information about the physical object with a shared perspective. This allows for more direct interaction with the object itself, and particularly raises two interesting dimensions to consider: 1) Live versus non-live formats of sharing and their combination, and 2) Images and videos with additional annotations as different types of captured data.

First, live videos were intended to support real-time interaction and manipulation of physical objects. Our study discovered that users appreciated that the Task View further diminished background details into a detailed object sharing. As expected, the non-live capturing of physical objects (*i.e.*, snapshots and short videos) also mitigated the onerous coordination efforts **(C4)**. Moreover, participants preferred the non-live options that freed their hands, as opposed to live video sharing which requires holding items. This aligns with the findings by Fakourfar et al. [21] that showed the effectiveness of freeze-frame in making annotations. Aside from temporally freezing digital copies in video window, a more frequent way to share object with the Task View is to use the combination of live and non-live format of copies to share physical objects. A local user may have the non-live copies (snapshots or short videos) displayed in the task view, while transitioning between looking at the object and showing, orienting, and physically referencing finer details of the item to the remote partner via his or her person view. An interesting implication here is that future designers may consider supporting a balance between the size of the Task View and the user's person views.

Second, our study found that capturing non-live data (i.e., snapshots and short videos) in addition to live videos improved the ability to show multiple aspects of objects to remote users **(C2)**. While short videos were designed to provide a more efficient way of taking multiple sequential snapshots at once, users were concerned about the potential inefficiency of making longer videos due to frequent coordination, similar to the live video.

In addition to the ad-hoc use of digital copies *during* conversations, participants requested the feature to download captured objects for later use *after* the conversation. Though this was not explicitly implemented, they still opened these copies in another window to privately check or attempted to right-click to "download" for later use. This points to an interesting new direction to support different kinds of recording formats, e.g., recording bodily movements [9], for later knowledge sharing [55] or note-taking [72], meeting summarization [64, 107], or for post-hoc learning [12, 20, 32]. Future designers may then need to consider additional curation for the large number of captured non-live copies.

7.2 Enabling the Remote User's Agency and Control over Remote Physical Objects

We have studied the remote referencing space over digital copies in both parallel and shared perspectives (RQ2).

The main benefit of ThingShare in Context Mode is that the remote user can have fine-grained control over the stored or frozen digital copies by bringing a parallel copy back to their video windows for independent and free manipulation. This was found to be very efficient for remote users, as they could easily browse and zoom in on the parallel object in detail. They appreciated that they could quickly ask questions about the digital copy and point to the specific part for quick reference, as their activities within the video feed are fully translucent to another user. Participants also liked the free control and access to collaboratively manipulate digital copies in the collage and focus mode. The remote access and control to the digital copies of physical objects thus addressed the challenge of remote referencing **(C3)**.

Moreover, participants suggested the capability to drop items into the remote user's video window with a directional handover of a physical object, which was not implemented for considerations of violating personal spaces. However, this feature could be potentially useful for some leisure-related scenarios, such as remote

child-adult interactions [47] where objects are used to create memorable experiences, as found by [25]. It could also be efficient for current video meeting setups with pairs and lead to further discussion of integrating such a feature in more scalable multi-party or one-to-many meetings. Future object-centered meetings could consider designing aggregated interaction feedback of multiple remote users over the physical object from multiple remote users who interact with their parallel digital copies to enhance the local user's understanding via more flexible and expressive remote input, similar to [13, 31].

7.3 Reflection on Promising Scenarios and Applications

ThingShare can be used in a diverse set of applications to improve the experiences of the local and distributed users.

Ad-hoc and Improvised Sharing of Everyday Physical Objects. ThingShare can be used to support the ad-hoc and improvised showing of physical objects, while also protecting privacy in both social and professional contexts. The user can also augment the part they are pointing to remote partners by hovering over the object so that the remote user can see exactly where the user is pointing to in their person space. Moreover, as found in the study, users can use ThingShare in remote social gathering for playfulness and background decoration [45]. ThingShare can also be useful for creating memorable experiences between child-parent pairs or between grandparents and grandchildren [25].

Collaborative Sharing and Brainstorming. ThingShare could be also useful to support a quick transition from ad-hoc conversation to teamwork. A user can share a screen together with the shared document with an object. Additionally, ThingShare's Task View feature allows users to capture digital copies of physical sketches or scribbles, making it easier to answer remote user's questions and clarify complicated calculations and visualizations in unplanned manner during synchronous meetings. Distributed users can also use ThingShare as a virtual Miro Board, where they can quickly jot down ideas on physical sticky notes and gather them in the Collage View for brainstorming sessions. ThingShare can also be a valuable tool for designers who rely on everyday objects to inform their design process.

Support Real-Time Remote Learner Inputs and Engagements for Hands-on Activities. ThingShare can also be helpful for classroom presentation with prototypes or collaborative learning for hands-on activities. In such scenarios, both Person and Task View can be used. Prior work in video-conferencing [31] and live streaming [13] have investigated to support viewer/remote user input systems with chat interface [101], aggregated [13, 31] responses with different levels of streamer adaptability and viewer expressiveness. ThingShare can support a more expressive way of interacting with physical objects in reality realm that differentiate with prior work that supports only digital visual inputs on visuals of physical objects. Moreover, ThingShare features suggested a new modality for systems supporting breakout room like FluidMeet [43] to support the awareness of activity and provides permeable, flexible spaces for hands-on activities [58].

7.4 Limitations and Future Directions

Need for More Robust Instance Segmentation. We currently use the Yolact instance segmentation model trained with COCO data set with additional image processing steps. Though the existing implementation was sufficient for evaluating the concept, the model's performance depends highly on the training data set. The study found that the model may also fail to detect objects that are partially visible in the camera view. A robust instance segmentation method is essential because improper object detection can compromise user privacy. We believe that the image segmentation methods will only become better in the future. Additionally, incorporating 6DoF ad-hoc object tracking [18] may further improve the robustness of the instance segmentation. However, we should also consider methods to prevent potential privacy leaks in case of model failures. We also encourage future work to enrich data sets for physical objects in home/office environments.

Explore Automated versus User Control. All interactions with the interface are limited to the 2D screen and the user's manipulation of the object. There were other lines of research that investigated the overlay of gestural menus or chironomia with tangible feedback when interacting with objects or data [35], speech-based interfaces [60], and visual captions [63]. These can be interesting future extensions of the current study but not under the scope of this study for communication.

Explore other Points of View (POVs) of the Space. The main scope of this work is to explore the transition between the person and task space using a single video channel, thus designed for the single-camera setup. However, the potential of the ThingShare system could be further extended to other perspectives and camera setups [71] in future research, including contexts such as fabrication workshops [86] that require the use of multiple cameras. In these circumstances, balancing simplicity and flexibility is important as earlier research has found that the variety of cameras and adjustable viewpoints could produce distractions and a lack of mutual orientation and coordination [86, 92]. Considering the semi-fixed cameras and mobile camera collaboration use cases, the mobility of camera setups and viewpoints can cause distractions and difficulties in mutual orientation and coordination, and the rapidly shifting environment in mobile camera collaborations [49] should be take into account for future design of digital copies.

Study in Real-World Contexts. The scope of our study is limited in terms of its scale and lab setup. Thus, further in-depth and long-term evaluations in real-world contexts would be necessary to gain a better understanding of its impact in such settings. For future work, deploying the ThingShare tool in work meetings, classrooms, and informal social gatherings could provide valuable insights. However, studying ThingShare in different contexts may require different design considerations. Currently, ThingShare only examines object-sharing experiences between pairs, so supporting one-to-many scenarios beyond pairs will necessitate additional considerations relating to how to aggregate multiple remote users' inputs and references [13, 31] in an expressive and effective way [65], as well as how to effectively manage multiple snapshots. Additionally, the integration of digital copies with digital task spaces (e.g., shared

screen) while maintaining the benefits of what-you-see-is-what-you-get (WYSIWIS) [44] will need to be taken into account when introducing ThingShare in real contexts.

8 CONCLUSION

We presented ThingShare, a novel video-conferencing system that allows remote participants to interact and manipulate physical objects in real time. By segmenting the physical object from the user's video window, both local and remote users can reference and manipulate the object without the need for additional hardware. We described the system design and implementation of ThingShare, with a particular focus on the face-to-face perspective. Our study found that ThingShare provides efficient transition between the person space and the object being discussed, as well as a better control over privacy by allowing users to selectively show objects with a virtual background. We believe that ThingShare will open up new opportunities for object-oriented sharing in video meetings and inspire further exploration of such design principles in the HCI community, leveraging the capabilities of real-time computer vision and instance segmentation models for object sharing.

ACKNOWLEDGMENTS

This work was supported by the Commonwealth Cyber Initiative and the European Research Council (ERC) under the European Union Horizon 2020 research and innovation programme (grant agreement No. 740548).

REFERENCES

[1] 2022. COCO Dataset. https://cocodataset.org/#home. Accessed: 2022-09-09.
[2] 2022. MediaPipe. https://google.github.io/mediapipe/solutions/selfie_segmentation.html. Accessed: 2022-9-15.
[3] 2022. Simple-Peer. https://github.com/feross/simple-peer. Accessed: 2022-09-15.
[4] Daniel Bolya, Chong Zhou, Fanyi Xiao, and Yong Jae Lee. 2019. Yolact: Real-Time Instance Segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 9157–9166. https://doi.org/10.1109/ICCV.2019.00925
[5] Erik Brynjolfsson, John J Horton, Adam Ozimek, Daniel Rock, Garima Sharma, and Hong-Yi TuYe. 2020. *COVID-19 and Remote Work: An Early Look at US Data*. Technical Report. National Bureau of Economic Research.
[6] Bill Buxton. 2009. Mediaspace-Meaningspace-Meetingspace. In *Media Space 20+ Years of Mediated Life*. Springer, 217–231. https://doi.org/10.1145/3170427.3173033
[7] William Buxton. 1992. Telepresence: Integrating Shared Task and Person Spaces. In *Proceedings of Graphics Interface*, Vol. 92. Canadian Information Processing Society Toronto, Canada, Canadian Information Processing Society Toronto, Canada, 123–129. https://doi.org/10.5555/155294.155309
[8] Hancheng Cao, Chia-Jung Lee, Shamsi Iqbal, Mary Czerwinski, Priscilla N Y Wong, Sean Rintel, Brent Hecht, Jaime Teevan, and Longqi Yang. 2021. Large Scale Analysis of Multitasking Behavior During Remote Meetings. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 448, 13 pages. https://doi.org/10.1145/3411764.3445243
[9] Scott Carter, Pernilla Qvarfordt, Matthew Cooper, and Ville Mäkelä. 2015. Creating Tutorials with Web-Based Authoring and Heads-Up Capture. *IEEE Pervasive Computing* 14, 3 (2015), 44–52. https://doi.org/10.1109/MPRV.2015.59
[10] Yuan-Chia Chang, Hao-Chuan Wang, Hung-kuo Chu, Shung-Ying Lin, and Shuo-Ping Wang. 2017. AlphaRead: Support Unambiguous Referencing in Remote Collaboration With Readable Object Annotation. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing* (Portland, Oregon, USA) *(CSCW '17)*. Association for Computing Machinery, New York, NY, USA, 2246–2259. https://doi.org/10.1145/2998181.2998258
[11] Yi Fei Cheng, Hang Yin, Yukang Yan, Jan Gugenheimer, and David Lindlbauer. 2022. Towards Understanding Diminished Reality. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 549, 16 pages. https://doi.org/10.1145/3491102.3517452
[12] Pei-Yu Chi, Joyce Liu, Jason Linder, Mira Dontcheva, Wilmot Li, and Bjoern Hartmann. 2013. DemoCut: Generating Concise Instructional Videos for Physical Demonstrations. In *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology* (St. Andrews, Scotland, United Kingdom) *(UIST '13)*. Association for Computing Machinery, New York, NY, USA, 141–150. https://doi.org/10.1145/2501988.2502052
[13] John Joon Young Chung, Hijung Valentina Shin, Haijun Xia, Li-yi Wei, and Rubaiat Habib Kazi. 2021. Beyond Show of Hands: Engaging Viewers Via Expressive and Scalable Visual Communication in Live Streaming. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 109, 14 pages. https://doi.org/10.1145/3411764.3445419
[14] Christopher Clarke, Alessio Bellino, Augusto Esteves, and Hans Gellersen. 2017. Remote Control by Body Movement in Synchrony With Orbiting Widgets: An Evaluation of TraceMatch. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol* 1, 3, Article 45 (sep 2017), 22 pages. https://doi.org/10.1145/3130910
[15] Christopher Clarke and Hans Gellersen. 2017. MatchPoint: Spontaneous Spatial Coupling of Body Movement for Touchless Pointing. 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, 179–192. https://doi.org/10.1145/3126594.3126626
[16] Juliet Corbin and Anselm Strauss. 2014. *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*. Sage publications.
[17] Anthony DeVincenzi, Lining Yao, Hiroshi Ishii, and Ramesh Raskar. 2011. Kinected Conference: Augmenting Video Imaging With Calibrated Depth and Audio. In *Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work*. 621–624. https://doi.org/10.1145/1958824.1958929
[18] Ruofei Du, Alex Olwal, Mathieu Goc, Shengzhi Wu, Danhang Tang, Yinda Zhang, Jun Zhang, David Tan, Federico Tombari, and David Kim. 2022. Opportunistic Interfaces for Augmented Reality: Transforming Everyday Objects Into Tangible 6DoF Interfaces Using Ad Hoc UI. In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems (CHI)*. ACM. https://doi.org/10.1145/3491101.3519911
[19] Carmine Elvezio, Mengu Sukan, Ohan Oda, Steven Feiner, and Barbara Tversky. 2017. Remote Collaboration in AR and VR Using Virtual Replicas. In *ACM SIGGRAPH 2017 VR Village* (Los Angeles, California) *(SIGGRAPH '17)*. Association for Computing Machinery, New York, NY, USA, Article 13, 2 pages. https://doi.org/10.1145/3089269.3089281
[20] Omid Ettehadi, Fraser Anderson, Adam Tindale, and Sowmya Somanath. 2021. Documented: Embedding Information Onto and Retrieving Information From 3D Printed Objects. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 424, 11 pages. https://doi.org/10.1145/3411764.3445551
[21] Omid Fakourfar, Kevin Ta, Richard Tang, Scott Bateman, and Anthony Tang. 2016. Stabilized Annotations for Mobile Remote Assistance. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 1548–1560. https://doi.org/10.1145/2858036.2858171
[22] Martin Feick, Terrance Mok, Anthony Tang, Lora Oehlberg, and Ehud Sharlin. 2018. Perspective on and Re-Orientation of Physical Proxies in Object-Focused Remote Collaboration. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–13. https://doi.org/10.1145/3173574.3173855
[23] Sean Follmer, Daniel Leithinger, Alex Olwal, Akimitsu Hogge, and Hiroshi Ishii. 2013. InFORM: Dynamic Physical Affordances and Constraints Through Shape and Object Actuation. In *Uist*, Vol. 13. 2501–988. https://doi.org/10.1145/2501988.2502032
[24] Sean Follmer, Hayes Raffle, Janet Go, Rafael Ballagas, and Hiroshi Ishii. 2010. Video Play: Playful Interactions in Video Conferencing for Long-Distance Families With Young Children. In *Proceedings of the 9th International Conference on Interaction Design and Children* (Barcelona, Spain) *(IDC '10)*. Association for Computing Machinery, New York, NY, USA, 49–58. https://doi.org/10.1145/1810543.1810550
[25] Verena Fuchsberger, Janne Mascha Beuthel, Philippe Bentegeac, and Manfred Tscheligi. 2021. Grandparents and Grandchildren Meeting Online: The Role of Material Things in Remote Settings. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–14. https://doi.org/10.1145/3411764.3445191
[26] Susan R Fussell, Robert E Kraut, and Jane Siegel. 2000. Coordination of Communication: Effects of Shared Visual Context on Collaborative Work. In *Proceedings of the 2000 ACM Conference on Computer Supported Cooperative Work*. 21–30. https://doi.org/10.1145/358916.358947
[27] Steffen Gauglitz, Cha Lee, Matthew Turk, and Tobias Höllerer. 2012. Integrating the Physical Environment Into Mobile Remote Collaboration. In *Proceedings of the 14th International Conference on Human-Computer Interaction With Mobile Devices and Services* (San Francisco, California, USA) *(MobileHCI '12)*. Association for Computing Machinery, New York, NY, USA, 241–250. https://doi.org/10.1145/2371574.2371610

[28] Steffen Gauglitz, Benjamin Nuernberger, Matthew Turk, and Tobias Höllerer. 2014. In Touch With the Remote World: Remote Collaboration With Augmented Reality Drawings and Virtual Navigation. In *Proceedings of the 20th ACM Symposium on Virtual Reality Software and Technology* (Edinburgh, Scotland) *(VRST '14)*. Association for Computing Machinery, New York, NY, USA, 197–205. https://doi.org/10.1145/2671015.2671016

[29] Steffen Gauglitz, Benjamin Nuernberger, Matthew Turk, and Tobias Höllerer. 2014. World-Stabilized Annotations and Virtual Scene Navigation for Remote Collaboration. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology* (Honolulu, Hawaii, USA) *(UIST '14)*. Association for Computing Machinery, New York, NY, USA, 449–459. https://doi.org/10.1145/2642918.2647372

[30] William W Gaver, Abigail Sellen, Christian Heath, and Paul Luff. 1993. One Is Not Enough: Multiple Views in a Media Space. In *Proceedings of the INTERACT'93 and CHI'93 Conference on Human Factors in Computing Systems*. 335–341. https://doi.org/10.1145/169059.169268

[31] Elena L. Glassman, Juho Kim, Andrés Monroy-Hernández, and Meredith Ringel Morris. 2015. Mudslide: A Spatially Anchored Census of Student Confusion for Online Lecture Videos. 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, 1555–1564. https://doi.org/10.1145/2702123.2702304

[32] Pauline Gourlet, Sarah Garcin, Louis Eveillard, and Ferdinand Dervieux. 2016. DoDoc: A Composite Interface That Supports Reflection-in-Action. In *Proceedings of the TEI '16: Tenth International Conference on Tangible, Embedded, and Embodied Interaction* (Eindhoven, Netherlands) *(TEI '16)*. Association for Computing Machinery, New York, NY, USA, 316–323. https://doi.org/10.1145/2839462.2839506

[33] Jens Emil Grønbæk, Mille Skovhus Knudsen, Kenton O'Hara, Peter Gall Krogh, Jo Vermeulen, and Marianne Graves Petersen. 2020. Proxemics Beyond Proximity: Designing for Flexible Social Interaction Through Cross-Device Interaction. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–14. https://doi.org/10.1145/3313831.3376379

[34] Jens Emil Grønbæk, Banu Saatçi, Carla F Griggio, and Clemens Nylandsted Klokmose. 2021. MirrorBlender: Supporting Hybrid Meetings With a Malleable Video-Conferencing System. (2021). https://doi.org/10.1145/3411764.3445698

[35] Brian D. Hall, Lyn Bartram, and Matthew Brehmer. 2022. Augmented Chironomia for Presenting Data to Remote Audiences. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology* (Bend, OR, USA) *(UIST '22)*. Association for Computing Machinery, New York, NY, USA, Article 18, 14 pages. https://doi.org/10.1145/3526113.3545614

[36] Michael Haller, Peter Brandl, Daniel Leithinger, Jakob Leitner, Thomas Seifried, and Mark Billinghurst. 2006. Shared Design Space: Sketching Ideas Using Digital Pens and a Large Augmented Tabletop Setup. In *International Conference on Artificial Reality and Telexistence*. Springer, Springer, 185–196. https://doi.org/10.1007/1194135_20

[37] Michael Haller, Jakob Leitner, Thomas Seifried, James R Wallace, Stacey D Scott, Christoph Richter, Peter Brandl, Adam Gokcezade, and Seth Hunter. 2010. The Nice Discussion Room: Integrating Paper and Digital Media to Support Co-Located Group Meetings. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 609–618. https://doi.org/10.1145/1753326.1753418

[38] Feng Han, Yifei Cheng, Megan Strachan, and Xiaojuan Ma. 2021. Hybrid Paper-Digital Interfaces: A Systematic Literature Review. In *Designing Interactive Systems Conference 2021*. 1087–1100. https://doi.org/10.1145/3461778.3462059

[39] Björn Hartmann, Meredith Ringel Morris, Hrvoje Benko, and Andrew D. Wilson. 2010. Pictionaire: Supporting Collaborative Design Work by Integrating Physical and Digital Artifacts. In *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work* (Savannah, Georgia, USA) *(CSCW '10)*. Association for Computing Machinery, New York, NY, USA, 421–424. https://doi.org/10.1145/1718918.1718989

[40] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep Residual Learning for Image Recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 770–778. https://doi.org/10.1109/CVPR.2016.90

[41] Zhenyi He, Keru Wang, Brandon Yushan Feng, Ruofei Du, and Ken Perlin. 2021. GazeChat: Enhancing Virtual Conferences with Gaze-Aware 3D Photos. In *The 34th Annual ACM Symposium on User Interface Software and Technology* (Virtual Event, USA) *(UIST '21)*. Association for Computing Machinery, New York, NY, USA, 769–782. https://doi.org/10.1145/3472749.3474785

[42] 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

[43] Erzhen Hu, Md Aashikur Rahman Azim, and Seongkook Heo. 2022. FluidMeet: Enabling Frictionless Transitions Between In-Group, Between-Group, and Private Conversations During Virtual Breakout Meetings. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 511, 17 pages. https://doi.org/10.1145/3491102.3517558

[44] Erzhen Hu, Jens Emil Grønbæk, Austin Houck, and Seongkook Heo. 2023. OpenMic: Utilizing Proxemic Metaphors for Conversational Floor Transitions in Multiparty Video Meetings. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3544548.3581013

[45] Angel Hsing-Chi Hwang, Cheng Yao Wang, Yao-Yuan Yang, and Andrea Stevenson Won. 2021. Hide and Seek: Choices of Virtual Backgrounds in Video Chats and Their Effects on Perception. *Proc. ACM Hum.-Comput. Interact* 5, CSCW2, Article 303 (oct 2021), 30 pages. https://doi.org/10.1145/3476044

[46] Hiroshi Ishii and Minoru Kobayashi. 1992. Clearboard: A Seamless Medium for Shared Drawing and Conversation With Eye Contact. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 525–532. https://doi.org/10.1145/142750.142977

[47] Qiao Jin, Ye Yuan, and Svetlana Yarosh. 2023. Socio-technical Opportunities in Long-Distance Communication Between Siblings with a Large Age Difference.. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI)*. ACM. https://doi.org/10.1145/3544548.3580720

[48] Hyungeun Jo and Sungjae Hwang. 2013. Chili: Viewpoint Control and On-Video Drawing for Mobile Video Calls. In *CHI'13 Extended Abstracts on Human Factors in Computing Systems*. 1425–1430. https://doi.org/10.1145/2468356.2468610

[49] Brennan Jones, Anna Witcraft, Scott Bateman, Carman Neustaedter, and Anthony Tang. 2015. Mechanics of Camera Work in Mobile Video Collaboration. 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, 957–966. https://doi.org/10.1145/2702123.2702345

[50] Tejinder K. Judge, Carman Neustaedter, Steve Harrison, and Andrew Blose. 2011. Family Portals: Connecting Families Through a Multifamily Media Space. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vancouver, BC, Canada) *(CHI '11)*. Association for Computing Machinery, New York, NY, USA, 1205–1214. https://doi.org/10.1145/1978942.1979122

[51] Tejinder K. Judge, Carman Neustaedter, and Andrew F. Kurtz. 2010. The Family Window: The Design and Evaluation of a Domestic Media Space. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Atlanta, Georgia, USA) *(CHI '10)*. Association for Computing Machinery, New York, NY, USA, 2361–2370. https://doi.org/10.1145/1753326.1753682

[52] Sasa Junuzovic, Kori Inkpen, Tom Blank, and Anoop Gupta. 2012. IllumiShare: Sharing Any Surface. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1919–1928. https://doi.org/10.1145/2207676.2208333

[53] David S. Kirk, Abigail Sellen, and Xiang Cao. 2010. Home Video Communication: Mediating 'Closeness'. In *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work* (Savannah, Georgia, USA) *(CSCW '10)*. Association for Computing Machinery, New York, NY, USA, 135–144. https://doi.org/10.1145/1718918.1718945

[54] Russell Kruger, Sheelagh Carpendale, Stacey D Scott, and Saul Greenberg. 2004. Roles of Orientation in Tabletop Collaboration: Comprehension, Coordination and Communication. *Computer Supported Cooperative Work (CSCW)* 13, 5 (2004), 501–537. https://doi.org/10.1007/s10606

[55] Stacey Kuznetsov and Eric Paulos. 2010. Rise of the Expert Amateur: DIY Projects, Communities, and Cultures. In *Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries* (Reykjavik, Iceland) *(NordiCHI '10)*. Association for Computing Machinery, New York, NY, USA, 295–304. https://doi.org/10.1145/1868914.1868950

[56] Hideaki Kuzuoka, Jun'ichi Kosaka, Keiichi Yamazaki, Yasuko Suga, Akiko Yamazaki, Paul Luff, and Christian Heath. 2004. Mediating Dual Ecologies. In *Proceedings of the 2004 ACM Conference on Computer Supported Cooperative Work*. 477–486. https://doi.org/10.1145/1031607.1031686

[57] Hideaki Kuzuoka, Jun Yamashita, Keiichi Yamazaki, and Akiko Yamazaki. 1999. Agora: A Remote Collaboration System That Enables Mutual Monitoring. *CHI '99 Extended Abstracts on Human Factors in Computing Systems* (1999). https://doi.org/10.1145/632716.632836

[58] Audrey Labrie, Terrance Mok, Anthony Tang, Michelle Lui, Lora Oehlberg, and Lev Poretski. 2022. Toward Video-Conferencing Tools for Hands-On Activities in Online Teaching. *Proc. ACM Hum.-Comput. Interact* 6, GROUP, Article 10 (jan 2022), 22 pages. https://doi.org/10.1145/3492829

[59] Jiannan Li, Maurício Sousa, Chu Li, Jessie Liu, Yan Chen, Ravin Balakrishnan, and Tovi Grossman. 2022. ASTEROIDS: Exploring Swarms of Mini-Telepresence Robots for Physical Skill Demonstration. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. https://doi.org/fullHtml/10.1145/3491102.3501927

[60] Jian Liao, Adnan Karim, Shivesh Singh Jadon, Rubaiat Habib Kazi, and Ryo Suzuki. 2022. RealityTalk: Real-Time Speech-Driven Augmented Presentation for AR Live Storytelling. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology* (Bend, OR, USA) *(UIST '22)*. Association for Computing Machinery, New York, NY, USA, Article 17, 12 pages. https://doi.org/10.1145/3526113.3545702

[61] Christian Licoppe, Paul K. Luff, Christian Heath, Hideaki Kuzuoka, Naomi Yamashita, and Sylvaine Tuncer. 2017. *Showing Objects: Holding and Manipulating Artefacts in Video-Mediated Collaborative Settings.* Association for Computing Machinery, New York, NY, USA, 5295–5306. https://doi.org/10.1145/3025453.3025848

[62] David Lindlbauer and Andy D. Wilson. 2018. Remixed Reality: Manipulating Space and Time in Augmented Reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3173574.3173703

[63] Xingyu Liu, Vladimir Kirilyuk, Xiuxiu Yuan, Peggy Chi, Xiang Chen, Alex Olwal, and Ruofei Du. 2023. Visual Captions: Augmenting Verbal Communication With On-the-Fly Visuals. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI)*. ACM. https://doi.org/10.1145/3544548.3581566

[64] Zhicong Lu, Seongkook Heo, and Daniel J. Wigdor. 2018. StreamWiki: Enabling Viewers of Knowledge Sharing Live Streams to Collaboratively Generate Archival Documentation for Effective In-Stream and Post Hoc Learning. *Proc. ACM Hum.-Comput. Interact* 2, CSCW, Article 112 (nov 2018), 26 pages. https://doi.org/10.1145/3274381

[65] Zhicong Lu, Rubaiat Habib Kazi, Li-yi Wei, Mira Dontcheva, and Karrie Karahalios. 2021. StreamSketch: Exploring Multi-Modal Interactions in Creative Live Streams. *Proc. ACM Hum.-Comput. Interact* 5, CSCW1, Article 58 (apr 2021), 26 pages. https://doi.org/10.1145/3449132

[66] 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/S15327051HCI181_3

[67] Paul Luff, Christian Heath, Hideaki Kuzuoka, Keiichi Yamazaki, and Jun Yamashita. 2006. Handling Documents and Discriminating Objects in Hybrid Spaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Montréal, Québec, Canada) *(CHI '06)*. Association for Computing Machinery, New York, NY, USA, 561–570. https://doi.org/10.1145/1124772.1124858

[68] Paul Luff, Naomi Yamashita, Hideaki Kuzuoka, and Christian Heath. 2011. Hands on Hitchcock: Embodied Reference to a Moving Scene. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 43–52. https://doi.org/10.1145/1978942.1978951

[69] Hanuma Teja Maddali and Amanda Lazar. 2023. Understanding Context to Capture when Reconstructing Meaningful Spaces for Remote Instruction and Connecting in XR. In *CHI '23: CHI Conference on Human Factors in Computing Systems*. ACM. https://doi.org/10.1145/3544548.3581243

[70] Jennifer Marlow, Scott Carter, Nathaniel Good, and Jung-Wei Chen. 2016. Beyond Talking Heads: Multimedia Artifact Creation, Use, and Sharing in Distributed Meetings. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*. 1703–1715. https://doi.org/10.1145/2818048.2819958

[71] Nicolai Marquardt, Nathalie Henry Riche, Christian Holz, Hugo Romat, Michel Pahud, Frederik Brudy, David Ledo, Chunjong Park, Molly Jane Nicholas, Teddy Seyed, et al. 2021. AirConstellations: In-Air Device Formations for Cross-Device Interaction Via Multiple Spatially-Aware Armatures. In *The 34th Annual ACM Symposium on User Interface Software and Technology*. 1252–1268. https://doi.org/10.1145/3472749.3474820

[72] Xiaojun Meng, Shengdong Zhao, and Darren Edge. 2016. HyNote: Integrated Concept Mapping and Notetaking. In *Proceedings of the International Working Conference on Advanced Visual Interfaces*. 236–239. https://doi.org/10.1145/2909132.2909277

[73] Terrance Mok and Lora Oehlberg. 2017. Critiquing Physical Prototypes for a Remote Audience. In *Proceedings of the 2017 Conference on Designing Interactive Systems*. 1295–1307. https://doi.org/10.1145/3064663.3064722

[74] Jens Müller, Roman Rädle, and Harald Reiterer. 2016. Virtual Objects As Spatial Cues in Collaborative Mixed Reality Environments: How They Shape Communication Behavior and User Task Load. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 1245–1249. https://doi.org/10.1145/2858036.2858043

[75] Archana Narayanan, Erzhen Hu, and Seongkook Heo. 2022. Enabling Remote Hand Guidance in Video Calls Using Directional Force Illusion. In *Companion Publication of the 2022 Conference on Computer Supported Cooperative Work and Social Computing* (Virtual Event, Taiwan) *(CSCW'22 Companion)*. Association for Computing Machinery, New York, NY, USA, 135–139. https://doi.org/10.1145/3500868.3559470

[76] Carman Neustaedter, Steve Harrison, and Abigail Sellen. 2013. Connecting Families. *The Impact of New* (2013). https://doi.org/10.1109/ISTAS.2008.4559768

[77] James Norris, Holger Schnädelbach, and Guoping Qiu. 2012. CamBlend: An Object Focused Collaboration Tool. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Austin, Texas, USA) *(CHI '12)*. Association for Computing Machinery, New York, NY, USA, 627–636. https://doi.org/10.1145/2207676.2207765

[78] James Norris, Holger M Schnädelbach, and Paul K Luff. 2013. Putting Things in Focus: Establishing Co-Orientation Through Video in Context. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1329–1338. https://doi.org/10.1145/2470654.2466174

[79] Jasmin Odenwald, Sven Bertel, and Florian Echtler. 2020. Tabletop Teleporter: Evaluating the Immersiveness of Remote Board Gaming. In *Proceedings of the 9TH ACM International Symposium on Pervasive Displays* (Manchester, United Kingdom) *(PerDis '20)*. Association for Computing Machinery, New York, NY, USA, 79–86. https://doi.org/10.1145/3393712.3395337

[80] Roman Rädle, Hans-Christian Jetter, Nicolai Marquardt, Harald Reiterer, and Yvonne Rogers. 2014. HuddleLamp: Spatially-Aware Mobile Displays for Ad-Hoc Around-the-Table Collaboration. In *Proceedings of the Ninth ACM International Conference on Interactive Tabletops and Surfaces* (Dresden, Germany) *(ITS '14)*. Association for Computing Machinery, New York, NY, USA, 45–54. https://doi.org/10.1145/2669485.2669500

[81] Hayes Raffle, Rafael Ballagas, Glenda Revelle, Hiroshi Horii, Sean Follmer, Janet Go, Emily Reardon, Koichi Mori, Joseph Kaye, and Mirjana Spasojevic. 2010. Family Story Play: Reading With Young Children (and Elmo) Over a Distance. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Atlanta, Georgia, USA) *(CHI '10)*. Association for Computing Machinery, New York, NY, USA, 1583–1592. https://doi.org/10.1145/1753326.1753563

[82] Abhishek Ranjan, Jeremy P. Birnholtz, and Ravin Balakrishnan. 2006. An Exploratory Analysis of Partner Action and Camera Control in a Video-Mediated Collaborative Task. In *Proceedings of the 2006 20th Anniversary Conference on Computer Supported Cooperative Work* (Banff, Alberta, Canada) *(CSCW '06)*. Association for Computing Machinery, New York, NY, USA, 403–412. https://doi.org/10.1145/1180875.1180936

[83] Abhishek Ranjan, Jeremy P. Birnholtz, and Ravin Balakrishnan. 2007. Dynamic Shared Visual Spaces: Experimenting With Automatic Camera Control in a Remote Repair Task. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '07)*. Association for Computing Machinery, New York, NY, USA, 1177–1186. https://doi.org/10.1145/1240624.1240802

[84] Jun Rekimoto and Masanori Saitoh. 1999. Augmented Surfaces: A Spatially Continuous Work Space for Hybrid Computing Environments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Pittsburgh, Pennsylvania, USA) *(CHI '99)*. Association for Computing Machinery, New York, NY, USA, 378–385. https://doi.org/10.1145/302979.303113

[85] Jun Rekimoto, Brygg Ullmer, and Haruo Oba. 2001. DataTiles: A Modular Platform for Mixed Physical and Graphical Interactions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 269–276. https://doi.org/10.1145/365024.365115

[86] Clara Rigaud, Gilles Bailly, Ignacio Avellino, and Yvonne Jansen. 2022. Exploring Capturing Approaches in Shared Fabrication Workshops: Current Practice and Opportunities. *Proc. ACM Hum.-Comput. Interact* 6, CSCW2, Article 391 (nov 2022), 33 pages. https://doi.org/10.1145/3555116

[87] Mose Sakashita, E. Andy Ricci, Jatin Arora, and François Guimbretière. 2022. RemoteCoDe: Robotic Embodiment for Enhancing Peripheral Awareness in Remote Collaboration Tasks. *Proc. ACM Hum.-Comput. Interact* 6, CSCW1, Article 63 (apr 2022), 22 pages. https://doi.org/10.1145/3512910

[88] David Sirkin, Gina Venolia, John Tang, George Robertson, Taemie Kim, Kori Inkpen, Mara Sedlins, Bongshin Lee, and Mike Sinclair. 2011. Motion and Attention in a Kinetic Videoconferencing Proxy. In *IFIP Conference on Human-Computer Interaction*. Springer, Springer, 162–180. https://doi.org/10.1007/978-3-642-23774-_16

[89] Jaeyoon Song, Christoph Riedl, and Thomas W. Malone. 2021. Online Mingling: Supporting Ad Hoc, Private Conversations at Virtual Conferences. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–12. https://doi.org/10.1145/3411764.3445776

[90] Anthony Tang, Carman Neustaedter, and Saul Greenberg. 2007. *VideoArms: Embodiments for Mixed Presence Groupware*. 85–102. https://doi.org/10.1007/978-1-84628-664-_8

[91] Balasaravanan Thoravi Kumaravel, Fraser Anderson, George Fitzmaurice, Bjoern Hartmann, and Tovi Grossman. 2019. Loki: Facilitating Remote Instruction of Physical Tasks Using Bi-Directional Mixed-Reality Telepresence. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*. 161–174. https://doi.org/10.1145/3332165.3347872

[92] Baris Unver, Sarah D'Angelo, Matthew Miller, John C. Tang, Gina Venolia, and Kori Inkpen. 2018. Hands-Free Remote Collaboration Over Video: Exploring Viewer and Streamer Reactions. In *Proceedings of the 2018 ACM International Conference on Interactive Surfaces and Spaces* (Tokyo, Japan) *(ISS '18)*. Association for Computing Machinery, New York, NY, USA, 85–95. https://doi.org/10.1145/3279778.3279803

[93] Gina Venolia, John Tang, Ruy Cervantes, Sara Bly, George Robertson, Bongshin Lee, Kori Inkpen, and Steven Drucker. 2010. Embodied Social Proxy: Connecting Hub-and-Satellite Teams. *Proc. Comput. Supported Cooperative Work* (2010). https://doi.org/10.1145/1753326.1753482

[94] Gina Venolia, John C. Tang, Kori Inkpen, and Baris Unver. 2018. Wish You Were Here: Being Together Through Composite Video and Digital Keepsakes. In *Proceedings of the 20th International Conference on Human-Computer Interaction With Mobile Devices and Services* (Barcelona, Spain) *(MobileHCI '18)*. Association for Computing Machinery, New York, NY, USA, Article 17, 11 pages. https://doi.org/10.1145/3229434.3229476

[95] Pierre Wellner. 1993. Interacting With Paper on the DigitalDesk. *Commun. ACM* 36, 7 (jul 1993), 87–96. https://doi.org/10.1145/159544.159630

[96] Andrew D Wilson and Daniel C Robbins. 2007. Playtogether: Playing Games Across Multiple Interactive Tabletops. *Tangible Play* 13 (2007). https://doi.org/10.1109/TABLETOP.2007.30

[97] Nicolai Wojke, Alex Bewley, and Dietrich Paulus. 2017. Simple Online and Realtime Tracking With a Deep Association Metric. In *2017 IEEE International Conference on Image Processing (ICIP)*. IEEE, IEEE, 3645–3649. https://doi.org/10.1109/ICIP.2017.8296962

[98] Andrea Stevenson Won, Jakki O Bailey, and Siqi Yi. 2020. Work-in-Progress—learning About Virtual Worlds in Virtual Worlds: How Remote Learning in a Pandemic Can Inform Future Teaching. In *2020 6th International Conference of the Immersive Learning Research Network (iLRN)*. IEEE, IEEE, 377–380. https://doi.org/10.23919/iLRN47897.2020.9155201

[99] 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*. 1377–1387. https://doi.org/10.1145/2531602.2531691

[100] Kimiya Yamaashi, Jeremy R Cooperstock, Tracy Narine, and William Buxton. 1996. Beating the Limitations of Camera-Monitor Mediated Telepresence With Extra Eyes. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 50–57. https://doi.org/doi/pdf/10.1145/238386.238402

[101] Saelyne Yang, Changyoon Lee, Hijung Valentina Shin, and Juho Kim. 2020. *Snapstream: Snapshot-Based Interaction in Live Streaming for Visual Art.* Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3313831.3376390

[102] Lining Yao, Anthony DeVincenzi, Anna Pereira, and Hiroshi Ishii. 2013. Focalspace: Multimodal Activity Tracking, Synthetic Blur and Adaptive Presentation for Video Conferencing. In *Proceedings of the 1st Symposium on Spatial User Interaction*. 73–76. https://doi.org/10.1145/2491367.2491377

[103] Svetlana Yarosh, Kori M. Inkpen, and A.J. Bernheim Brush. 2010. Video Playdate: Toward Free Play Across Distance. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Atlanta, Georgia, USA) *(CHI '10)*. Association for Computing Machinery, New York, NY, USA, 1251–1260. https://doi.org/10.1145/1753326.1753514

[104] Svetlana Yarosh, Anthony Tang, Sanika Mokashi, and Gregory D. Abowd. 2013. "almost Touching": Parent-Child Remote Communication Using the Sharetable System. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work* (San Antonio, Texas, USA) *(CSCW '13)*. Association for Computing Machinery, New York, NY, USA, 181–192. https://doi.org/10.1145/2441776.2441798

[105] Svetlana Yarosh, Anthony Tang, Sanika Mokashi, and Gregory D. Abowd. 2013. "almost Touching": Parent-Child Remote Communication Using the Sharetable System. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work* (San Antonio, Texas, USA) *(CSCW '13)*. Association for Computing Machinery, New York, NY, USA, 181–192. https://doi.org/10.1145/2441776.2441798

[106] Ye Yuan, Jan Cao, Ruotong Wang, and Svetlana Yarosh. 2021. Tabletop Games in the Age of Remote Collaboration: Design Opportunities for a Socially Connected Game Experience. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 436, 14 pages. https://doi.org/10.1145/3411764.3445512

[107] Amy X. Zhang and Justin Cranshaw. 2018. Making Sense of Group Chat Through Collaborative Tagging and Summarization. *Proc. ACM Hum.-Comput. Interact* 2, CSCW, Article 196 (nov 2018), 27 pages. https://doi.org/10.1145/3274465

[108] Yizhong Zhang, Jiaolong Yang, Zhen Liu, Ruicheng Wang, Guojun Chen, Xin Tong, and Baining Guo. 2022. VirtualCube: An Immersive 3D Video Communication System. *IEEE Transactions on Visualization and Computer Graphics* 28, 5 (2022), 2146–2156. https://doi.org/10.1109/TVCG.2022.3150512

A APPENDIX

A.1 Other Findings from exploratory survey

A.1.1 Platforms. We present the results for platforms used in video meetings for work-related and non-work-related online meetings or video calls in Table 2. For work-related meetings, respondents reported using other platforms such as Discord (1/2) and RingCentral (1/2). For non-work-related video calls, respondents reported using other platforms such as Discord (2/7), Duo (2/7), Telegram (1/7), Amazon Alexa (1/7), and Viber (1/7).

Platforms	Work-related		Non-work-related	
Zoom	40.9%	106	30.2%	62
Microsoft Teams	13.9%	36	1.0%	2
Other	0.8%	2	3.4%	7
FaceTime	7.4%	19	16.6%	34
WhatsApp	8.5%	22	17.6%	36
Skype	10.4%	27	11.2%	23
Google Meet	14.7%	38	7.3%	15
Facebook Messenger	3.5%	9	12.7%	26

Table 2: Platforms used in work-related and non-work-related online meeting contexts

A.1.2 Devices. When asking the type of devices respondents used for online meetings or calls, the majority of participants (74%) reported using Desktop Computer (28.6%) or Laptop (45.4%) during work-related meetings, compared to non-work-related meetings, where only half of respondents (46.2%) using Desktop Computer (17.1%) or Laptop (29.2%). For *non-work-related* meetings, around half of respondents (52.3%) reported using mobile devices - mobile phone (43.2%) and tablets (9.1%) during *non-work-related meetings*, compared to only 24.5% respondents using mobile devices - mobile phones (19.9%) and tablets (4.6%) during *work-related meetings*.

A.2 Study Tasks

We present task details for Task 2 in Table 3 and objects prepared for study tasks in Figure 9. The Distribution of quantitative survey results are presented in Figure 13, Figure 14, Figure 15, Figure 16.

A.2.1 Survey Results Visualization.

	Object	Condition Task 1	Condition Task 2
User 1	VR Controller	Pick up and drop the object	Teleport and using context menu
User 2	Game Controller	**Playing Racing Game with two hands** - Left stick: steer the car - Right trigger (RT): accelerate the car - Left trigger (LT): brake the car - Bumpers: Gear up (Right) and down (Left) - "X" button: handbrake	**Single-Handed Racing Setup** Remapping the controllers with single-handed racing setup - Swap the left and right trigger for accelerate - map "X" button to the left on the "D pad" for handbrake - map right bumper (RB) to the up button on "D pad" to gear up.

Table 3: Task Two Description for Single User Sharing Task. Each user will get an object and procedure description and teach the remote user how to use it. The objects for two users are different, one is VR controller and another is Game Controller, yet the task for the two conditions (baseline and ThingShare) are different.

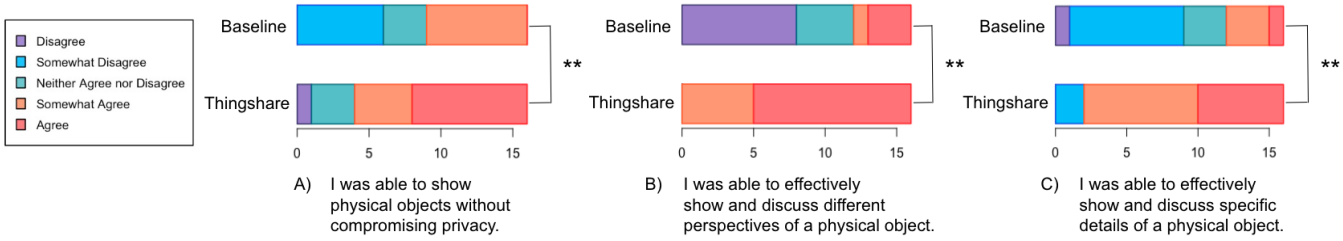

Figure 13: Task-Specific Questions

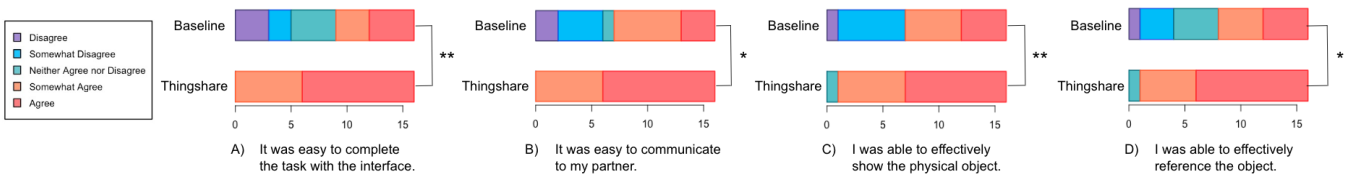

Figure 14: Task 2 Survey Results

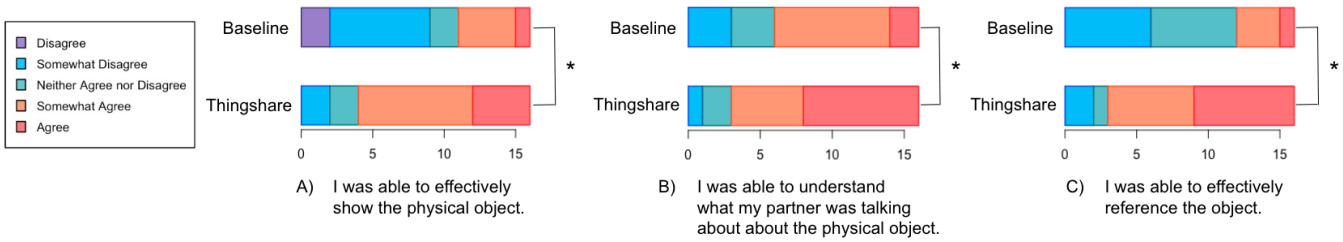

Figure 15: Task 3 Survey Results

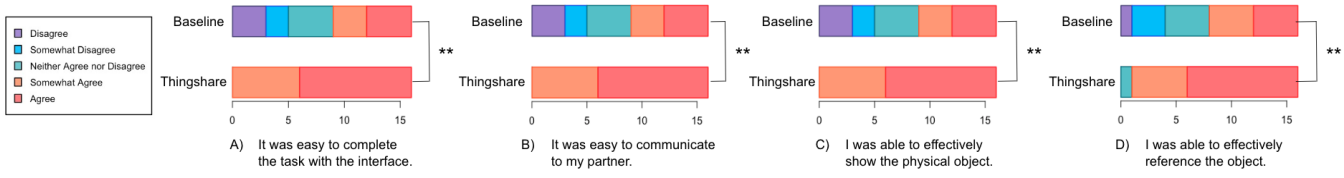

Figure 16: General Survey Results

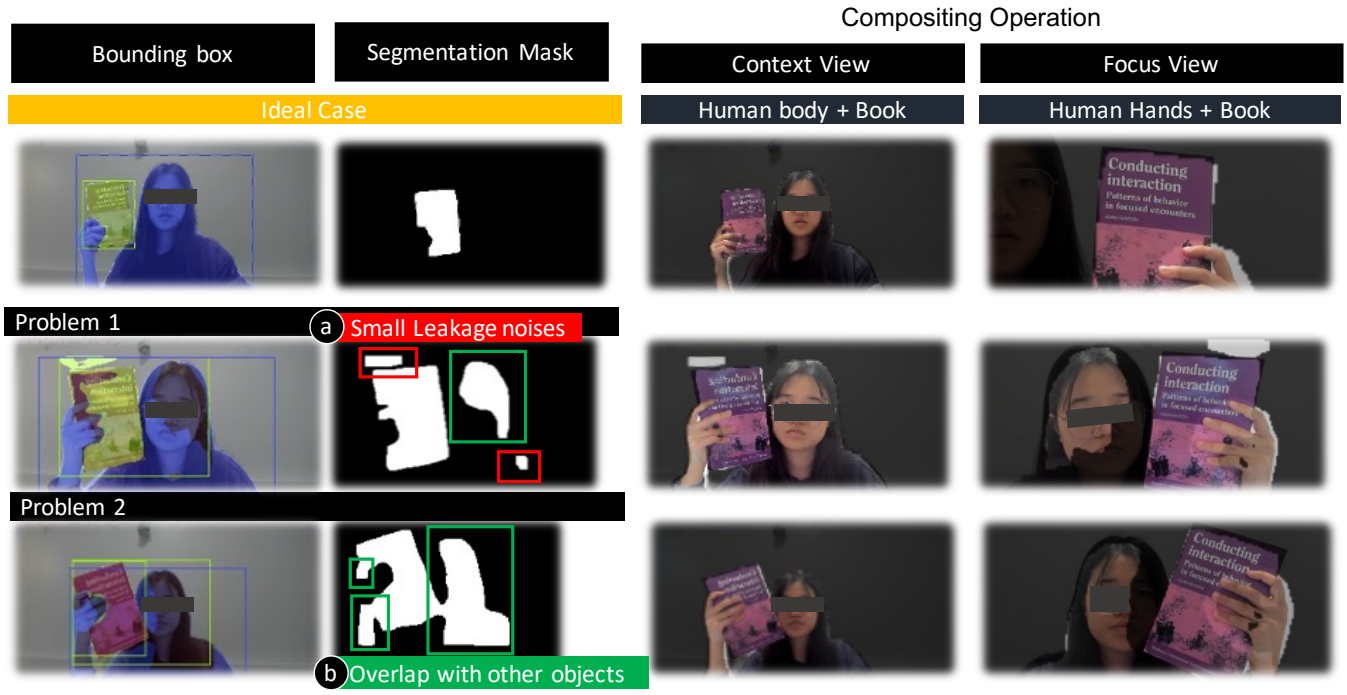

Figure 17: Original mask output with leakage problems from Yolact: a) After a morphological operation, the mask has been eroded as isolated small noises around, e.g., small noises within artifacts within bounding box; b) potential overlapped portions with other objects (e.g., human) with wrong mask segmentation