4-2018

Perspective on and re-orientation of physical proxies in object-focused remote collaboration

Martin FEICK

Terrance MOK

Anthony TANG
Singapore Management University, tonyt@smu.edu.sg

Lora OEHLBERG

Ehud SHARLIN

Citation

FEICK, Martin; MOK, Terrance; TANG, Anthony; OEHLBERG, Lora; and SHARLIN, Ehud. Perspective on and re-orientation of physical proxies in object-focused remote collaboration. (2018). *CHI '18: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, Montreal, Canada, April 21-26*. 1-13.
Available at: https://ink.library.smu.edu.sg/sis_research/7976

Perspective on and Re-Orientation of Physical Proxies in Object-Focused Remote Collaboration

Martin Feick[1,2] Terrance Mok[1] Anthony Tang[1] Lora Oehlberg[1] Ehud Sharlin[1]
[1]University of Calgary, Calgary, Canada [2]htw saar, Saarbruecken, Germany
{ martin.feick, terrance.mok2, tonyt, lora.oelhberg, ehud }@ucalgary.ca

ABSTRACT

Remote collaborators working together on physical objects have difficulty building a shared understanding of what each person is talking about. Conventional video chat systems are insufficient for many situations because they present a single view of the object in a flattened image. To understand how this limited perspective affects collaboration, we designed the Remote Manipulator (ReMa), which can reproduce orientation manipulations on a proxy object at a remote site. We conducted two studies with ReMa, with two main findings. First, a shared perspective is more effective and preferred compared to the opposing perspective offered by conventional video chat systems. Second, the physical proxy and video chat complement one another in a combined system: people used the physical proxy to understand objects, and used video chat to perform gestures and confirm remote actions.

Author Keywords

Remote collaboration, collaborative physical tasks, physical telepresence, object-focused collaboration, CSCW

ACM Classification Keywords

H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous

INTRODUCTION

Never has it been more important to engage in remote work. Object-focused collaboration is increasingly common in everyday life, where collaborators discuss and analyze physical objects. For example, collaborators may seek to understand how an object works [3, 17], provide assistance in repairing an object [33], or critique the design of an object [30]. While this activity is common in everyday physical interaction, we do not yet have good ways of supporting object-focused collaboration when people are remote. Developing techniques to support remote object-focused collaboration is important: increasingly, expertise about an object or how to repair it is located remotely, and the design of physical objects is distributed among geographically disparate collaborators (e.g. professional product development, and open hardware teams [30]).

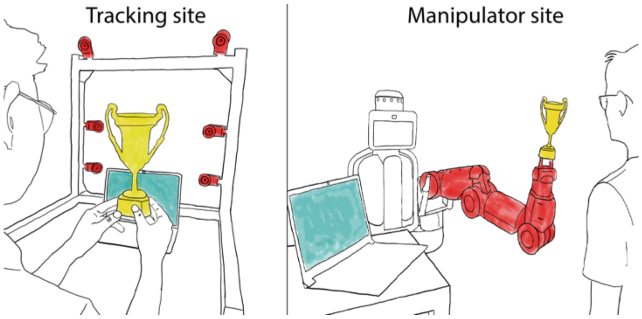

Figure 1. Remote Manipulator (ReMa) has two parts: it detects manipulations on an object (Left-yellow) using a set of sensors (Left-red), and then reproduces these with a proxy object (Right-yellow) using a Baxter robot arm (Right-red). ReMa allows shows the Manipulator Site collaborator (Right) the object with the same orientation as at the Tracking Site (Left). Collaborators can also use video chat (blue).

Going with conventional technologies, people engage in object-focused remote collaboration with video chat, where collaborators use the front facing camera to show the physical object (e.g. [27, 30]), or with mobile video chat, where the camera is repositioned to provide good views of the object (e.g. [17, 36]). However, these approaches are problematic, as collaborators may want different views of the object, or have difficulty framing the object properly in view. It also means they cannot effectively use spatial gestures or references, and cannot describe certain qualities of the object. Consequently, collaborators often require additional "meta" dialogue to establish joint understanding or common ground [4] during collaborative activity (e.g. how to orient, view or manipulate the object).

To explore this space, we consider two questions specific to object-focused collaboration: first, how does varying collaborators' perspectives on an object (i.e. shared vs. opposing views) help or hinder collaboration? Second, if we can automatically reorient a physical object or a proxy, how does this help collaborative activity? To address these questions, we first designed and built ReMa (Remote Manipulator). As illustrated in Figure 1, ReMa captures the orientation of a physical object, and reproduces these orientation changes at a remote location on a proxy object. Second, we designed and conducted two studies: the first to understand how variations of object perspective affect

CHI 2018, April 21–26, 2018, Montreal, QC, Canada

ACM 978-1-4503-5620-6/18/04…$15.00
https://doi.org/10.1145/3173574.3173855

object-focused collaboration, and the second, exploring the role of a proxy object (provided by ReMa) affects collaborators' interactions.

Given the results of our studies, shared perspective configurations outperformed opposing configurations for object-focused collaboration. We observed that ReMa helped participants by obviating meta-conversations about the orientation or re-orientation of the object and proxy object, and that the video chat stream complemented this by providing a way to gesture and to confirm the state of the remote site. This suggests two implications for designing systems for object-focused collaboration: designers should focus on providing shared perspective views (rather than face-to-face views), and tools that help manipulate a remote proxy object can smooth collaborative interactions.

You will find two contributions in this paper. First, we contribute two studies on object-focused collaboration that demonstrate that collaborators perspective on the object has a large impact on efficiency, and demonstrate how object proxies can smooth interaction. Second, our studies help articulate the role of orientation/reorientation of objects in collaboration as it relates to object-focused collaboration.

RELATED WORK

Up to this point, we have seen three directly related areas of research. First, we outline recent work on object-focused remote collaboration, where collaborators discuss and analyze physical objects in remote contexts. Next, we discuss the role of gestures, orientation and perspective of objects in collocated collaboration. Finally, we briefly discuss past research on telepresence robots.

Challenges of Object-Focused Collaboration

Never to this point have we seen so much interest in object-focused collaboration, where a physical object is the center of collaborative discussion and activity. For instance, Licoppe et al. [27] explore how video is used to support object-focused collaboration in everyday video chat conversations. Beyond the issue that artefacts need to be placed in view of the camera, Licoppe et al. [27] show that *how* the objects are revealed and manipulated together with ongoing discussion plays an important role in conveying attitudes and interest between video chat participants. For instance, the way that a label on a box of biscuits is revealed to the camera (and a remote partner), signals and emphasizes what is important to each about the object. Similarly, a viewing participant may cock his/her head or appear to move closer to "get a better look", even though this has no meaningful practical effect; rather, the purpose is to engender feelings of interest in the shared experience.

Going to this thoughtful use of object configuration (and camera orientation) may not always happen. Mok & Oehlberg [30] show in a remote critique scenario that participants frequently forgot to show the audience aspects of the object, or even to ensure that the audience could see the object. They suggest that the complex interplay between epistemic action (actions used to understand the object) and pragmatic action (actions used to explain object functionality) caused this [22]. It may be that the complexities of perspective and the pragmatic situation (rather than personal, as in [27]) lends itself to more focus on the object rather than a remote party. Similarly, Jones et al. [17] describe the challenges of positioning a mobile camera to effectively capture aspects of objects and scenes and convey them to a remote party as "camera work."

To this end, we aim to understand whether the conventional face-to-face, opposing view perspective of video chat systems is a source of some of these problems, and whether an actuated physical proxy can help address these challenges.

Gestures, Orientation, and Perspective

Letting remote collaborators understand meaning through gesture has been a central theme in considerable CSCW. We outline this work to show how our work parallels this approach by focusing on perspective and orientation.

Gesture. You can begin from Tang's seminal studies of collocated interaction on tables, where he underscored the importance of gestures in collaborative work [43]. Tang describes collaborators using hand gestures to communicate: enacting ideas or pointing to objects. To support gestures, researchers have explored marking up a remote video (e.g. [5, 9, 10, 13, 25]) as a proxy, included simple representations such as telepointers [12], and explored video overlays of bodies and arms [20, 39, 40, 41, 42] to convey additional subtleties of hand-based gestures [21]. Evaluations of these systems not only reinforce the importance of gesture, but also reveal the subtle ways in which gesture enables and engenders collaborative work. In the same way that this seminal work on gesture motivated subsequent system work, related studies of orientation and perspective motivate the present work.

Orientation. Down at the level of "orientation," Kruger et al. [23] revealed the important role of object orientation in a study of how people collaborate on a puzzle task. Based on their observational study, they articulate three distinct roles of orientation in collaborative work: *comprehension*, where the purpose of orienting an object is to personally understand/explore the object; *coordination*, where the object is reoriented to coordinate access and to define personal/shared working areas, and *communication*, where the object is re-oriented to explain something to another person. These functional roles of orientation might also an important role in object-focused remote collaboration—particularly with three-dimensional objects (rather than a flat artefacts). ReMa focuses on this object-focused collaboration, and we explore how explicitly reorienting a remote object helps and hinders remote collaborative work.

Perspective. In remote collaboration, collaborators usually have different perspectives of the workspace (due to camera placement, though cf. [5, 16]). This is even more problematic in object-focused collaboration. One cannot, for instance, rotate a piece of paper for a remote collaborator if s/he is not looking at the paper at all. This is a smaller variation of the problem described by Luff et al. [28], where one collaborator's understanding of the space, and how one orients and creates gestures is more difficult to (and sometimes inappropriately) perceive at a remote location. Jones et al. [17] discuss particularly how this happens during mobile video chat, where handheld perspectives of the scene are challenging to produce and capture properly for the remote collaborator. Fussell [8] explores variations on camera angles of a remote workspace for physical tasks, finding that "scene-focused" perspectives outperform "head-mounted" camera angles. Tang et al. [42] shows that task demands may be more easily addressed with some perspectives than others. For instance, shared perspectives are useful for reading text, whereas asymmetric perspectives are desirable (e.g. to create shared *vs* personal workspaces). Our work begins from the standpoint that different perspectives may be useful—particularly given that in collocated collaboration, people physically occupy different locations in space. Consequently, there is reason to believe that people are accustomed to alternate views of a physical object (e.g., in a face to face situation).

Telepresence Robots

Our study builds on a long history of using robots to support telepresence [11]. One line of robotic telepresence research has supported remote camera control, either through a robotic arm [45] or through a mobile telepresence robot (e.g. [24, 32, 35]). Our research instead uses telemanipulation—where a robot manipulates objects in a remote environment [15, 26, 47]. In our case, a human collaborator is located in that remote environment. We provide detail about our remote manipulation system next.

REMA: REMOTE MANIPULATOR

To further understand how perspective and orientation are used in object-focused remote collaboration, we designed ReMa—the Remote Manipulator. Our eventual goal is to allow two collaborators to explore physical objects, where each collaborator's interactions are reflected at the remote site. However, enabling this sort of bidirectional manipulation comes with a well-known set of problems, particularly when collaborators are manipulating the object in opposition to one another (e.g. [2, 29]). Thus, in this first iteration, our focus was specifically on one-way communication of the *orientation* of the object, abstracted from manipulations of the object in space. This focusing step allowed us to concentrate on the efficacy and role of orientation and perspective (i.e. via our study) without having to concern ourselves with resolving the challenges of movement tracking and bidirectional communication.

Design

As illustrated in Figure 1 and 2, ReMa comprises two sites: a Tracking Site (**TS**) and a Manipulator Site (**MS**). ***Tracking Site.*** At the Tracking Site (Figure 2, left), a person's object manipulations are captured—both the object's position in space, as well as its orientation. ***Manipulator Site.*** Manipulations are transmitted to the Manipulator Site (Figure 2, right), where they are displayed on a similar proxy object. We can change how the captured information is interpreted and rendered (e.g. from a *shared* or *opposing* perspective). ***Pausing.*** We also designed a "pausing" mechanism, which allows either collaborator to pause in the current orientation. This temporarily disables the Manipulator Site from continuing to mimic the Tracking Site. We built this into ReMa to allow participants to look at their proxy object independently.

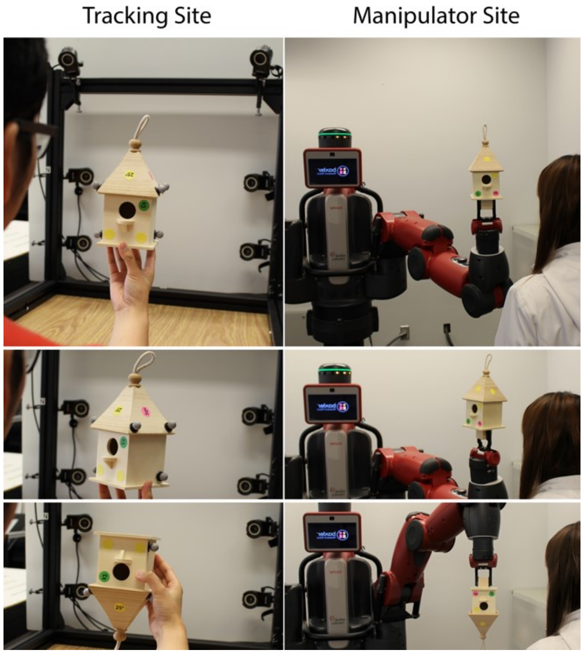

Figure 2: The ReMa system includes a Tracking Site (TS, top-left) and a Manipulator Site (MS, top-right) with bird house object. As the birdhouse is rotated at the TS, the proxy birdhouse at MS is also rotated.

Implementation

The Tracking Site uses six OptiTrack [31] infrared cameras attached to an aluminum frame (80/20 material). This frame provides users a $1m^3$ volume to manipulate the object. The object itself is affixed with retroreflective trackers, allowing the OptiTrack to track the 3D movement of the object. The location of the object is captured using Cartesian coordinates and quaternions, and based on a threshold, ReMa sends updates to the Manipulator Site.

The Manipulator Site uses a Baxter robot (model BR-01) that has a seven DoF robotic arm. We use the Robot Operation System [34] (version Indigo) to control the Baxter robot, including the seven joint angles that make up the robot arm. Translating the Cartesian end coordinates into a series of joint angles that result in the correct end-effector (gripper) position requires Inverse Kinematics (IK). We used the TRAC-IK algorithm [1] to calculate robot joint

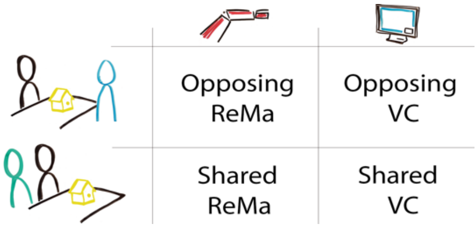

Figure 3. Study 1 compared different perspectives (shared vs. opposing) using both a video chat condition and the ReMa.

angles for a requested end-effector position in Cartesian coordinates. Because of the joint arrangement, some gripper position and orientation combinations are impossible for Baxter to perform. Rakita et. al. [37] encountered similar issues when mapping a human hand to a robotic end-effector due to the different kinematic capabilities. We used a dictionary to overcome this limitation, where we simplified the implementation in two ways. We allow for any pitch, yaw and roll in a Cartesian coordinate system to be rendered at 0°, 90°, 180° and 270°. Our implementation relies on a dictionary of end positions, where for some orientations of the object, the object needs to be shifted slightly on x, y and/or *z* axes. While the data on how the robot should move is sent over a local network with very low latency, the Baxter requires 2.5 seconds avg. (min: 1.5sec, max: 3.5sec) to physically respond and reach the correct orientation at the MS. Our implementation does not allow the Baxter robot to stop mid-way during a reorienting act, and go to a new orientation. As a result, if the TS rotates to a position and decides to rotate the object again while ReMa is still performing the first orientation, the ReMa finishes the first reorienting act before beginning the second reorienting act, compounding interaction latency. To provide the different perspective modes *opposing* and *shared*, we inverted the OptiTrack virtual camera setup to mirror the tracked orientations. Thus, the coordinates are sent in an inverted state, and reproduce the requested orientation on the object accordingly.

STUDY 1: THE IMPACT OF PERSPECTIVE

We designed two studies to evaluate and understand how people would make use of rotation and perspective information in collaborative "matching" tasks. In both studies, both the Tracking Site (**TS**) and Manipulator Site (**MS**) participants had their own physical proxy (of the other participant's object). In Study 1, we focused on how different perspectives of a proxy object affects a pair's collaborative interactions and conversation.

Study Variables. Our central interest was in comparing a *shared* perspective, where participants share the same view of an object, with the *opposing* perspective offered by conventional video chat systems. We implemented these two perspectives in two technical settings, allowing us to compare video chat (**VC**) interactions (e.g. Skype) with the **ReMa** system. As illustrated in Figure 3, we used a 2×2 within-subjects design (Opposing *vs* Shared × VC *vs* ReMa), where each pair experienced all four conditions once (each with a different task object/arrangement of stickers). In the VC condition, there is no formal tracking site or manipulator site (since both participants can manipulate their own objects).

Participants. We recruited 16 participants (eight pairs; six females; ten males), aged 19-54 with a range of backgrounds including electrical, mechanical and software engineering, computer science, art history and sports science. Each participant was provided with $20 remuneration for their participation. All participants reported experience with video chat software tools like FaceTime or Skype.

Task Design. We gave each participant the same simple object (a birdhouse, a trophy); however, the two objects differed based on a set of stickers placed at various locations on each object (and at different orientations). The goal of the task was to add stickers with the correct colour and markings so that both objects matched. Participants sat in one of the two different seats (either TS or MS), and completed one task in two of the conditions before being asked to switch seats. This seat switching was to balance the kinds of interactions each participant would experience. Conditions were presented in counter-balanced order.

We modeled our task around a central recurring micro-problem in object-centric collaboration: understanding a remote collaborator's perspective on an object. This problem occurs frequently in a range of collaborative situations, such as providing instructions to another person, interpreting a collaborator's instructions, or reestablishing joint orientation. In these situations, collaborators need to build a shared mental model or understanding of each other's view of the object. Our study task is designed to force participants to repeatedly address this micro-problem to complete the study task.

Data Collection. We collected data from six sources: a pre-study questionnaire for demographic information; video of participants as they completed tasks; video feeds of participants during videochat (VC) conditions; ReMa's internal logging (e.g. # of rotations; which orientation to which orientation; timing, etc.); field notes and observations; and a post-study interview to explore participants' experiences with the system.

Analysis. We conducted a thematic analysis of our data, identifying recurring themes in participant behaviour as they engaged with the system, and relating these to interview data. In addition, we conducted a modified interaction analysis (Jordan & Henderson [18]), where we identified unusual incidents, and used these as points for further understanding of how participants collaborated.

Study 1 Findings & Observations

All participant pairs completed the tasks in the different trials. Generally, shared perspective trials were better than opposing perspective trials in terms of completion time.

Average completion times for each condition were as follows: Opposing ReMa ($\bar{x}$=6m29s, s=2m10s), Opposing-VC ($\bar{x}$=5m19s, s=1m33), Shared ReMa ($\bar{x}$=4m40s, s=1m52s), and Shared-VC ($\bar{x}$=3m24s, s=1m22s). Pairs were generally more efficient using the shared perspective rather than opposing, regardless of the tool. While we were generally not interested in comparing task completion times between the VC and ReMa conditions (recall that ReMa introduces substantial latency due to the physicality of the robot), we still observe that one of the ReMa conditions is faster than one of the VC conditions.

The utility of the Shared perspective is corroborated by data from the questionnaire. On a 10-point Likert scale response to the question, "*Which of these would you prefer to use next time (1-definitely; 10-definitely not)",* participants overwhelmingly chose the shared perspective options (median scores: Shared ReMa (1.5), Shared-VC (2), Opposing-VC (4.5), Opposing ReMa (5)). Responses followed a similar pattern for participants' rating on ease of use (median scores: Shared-VC (1), Shared ReMa (2), Opposing-VC (4), Opposing ReMa (5)). Based on our analysis of participants' behaviour, we observed two principal challenges participants face in Opposing perspective trials that they did not have in Shared perspective trials: first, the opposing view conditions meant that a participant could not show his/her partner and *see for him/herself* what was being discussed, and second, in Opposing-VC conditions, partners had a hard time knowing how to "follow along" because of the perspective problem. With the Shared perspective conditions, participants used different strategies made available to them because they *knew* what the other person saw.

We observed that generally, participants had difficulty organizing and coordinating activity with an Opposing perspective because they had difficulty understanding what the other participant could see. In both VC and ReMa trials, we observed participants turning an object, and pausing the turn to check if the partner could see what was expected. This problem was exacerbated in VC trials, where both participants could turn an object in whatever way they wanted. When they tried to synchronize movement, even a simple misstep was difficult to recover from. This seems to be a symptom of the problem that others have observed [7, 14, 38, 46], where people have difficulty mentally rotating the object and understanding the object from another person's perspective. This problem is well illustrated by the difficulties experienced by G3, where re-orienting manipulations on the object are difficult for Joe to copy onto his own object (all participant names are pseudonyms).

Vignette 1: Group 3, Opposing-VC. *Frank orients the object for Joe so that Joe can see the right orientation of the trophy for his sticker (A). Joe tries to align his trophy with Frank's trophy by using the VC preview (B). Joe is uncertain if this is the right orientation of the trophy. Frank and Joe try to determine whether they share the same view or not. "The flat part of the trophy is facing you" (C). Ultimately, Joe re-orients his trophy, but is still confused about this orientation: "This feels really weird cause this isn't the side I'm looking at" (D).*

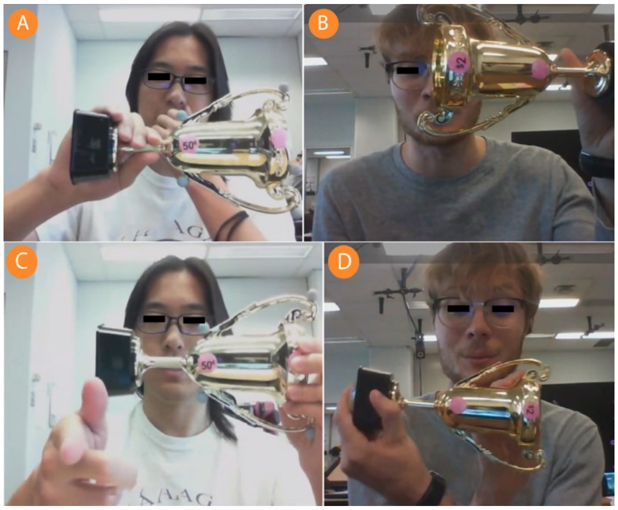

Figure 4: Group 3 Opposing-VC. Frank (A) tries to explain what he sees on one side of the trophy, but Joe rotates his trophy in the wrong direction (B). Frank explains the orientation of his trophy to Joe (C), but Joe is still confused whether he is holding his trophy in the correct orientation (D).

Problem of Left-Right. Vignette 1 shows difficulties in the Opposing-VC particularly with a mirroring effect. Joe uses the video to align his trophy with Frank's trophy. However, he gets confused, as he is observing three different views of the object: Frank's, his own physical object, and the preview in the video chat. Joe is uncertain how or whether indeed he should be matching Frank's view, and in what way: should he rotate left or right, counter-clockwise or clockwise; should he be showing Frank what he is looking at, or should he be doing the same operations so he is looking at the same thing Frank is? Trying to use VC to reach a shared orientation was challenging for most participants, as the perspective draws one's attention away from one's own object. Thus, rotational manipulations on the remote object were difficult to reproduce for most participants.

***Seeing and Showing at the Same Time*.** Vignette 2 from G1's Opposing-VC trial (Figure 4) illustrates how the opposing perspective results in challenges with both showing part of the object and describing it. Here, the problem is further exacerbated by the use of video on a flat 2D display.

Vignette 2: Group 1, Opposing-VC. *Brenda shows Alan the inside of the trophy, so he can see where to put a sticker (B). Alan is unable to understand from the video which sticker Brenda is referring to, or the orientation of the trophy (A). Brenda tells Alan to reorient the trophy so that they share the same perspective (C,D), but she has a hard time simultaneously showing Alan the inside of the trophy and describing it. She leans forward to look inside her trophy to describe what Alan should be seeing. After struggling to do this, Alan tells Brenda to reorient her trophy to match his perspective on his trophy: "No. Look at… Look inside like I'm looking inside."*

Figure 4: Group 1 Opposing-VC. Brenda wants to show Alan a sticker inside the trophy (B), but Alan cannot see the sticker (A). Alan tells Brenda to orient the trophy that both shared the same perspective (C, D).

This vignette illustrates that when Brenda is trying to show Alan something in the video, she has difficulty describing it to Alan (which requires her to see it) and showing it to him at the same time. When she points the object towards Alan, she can no longer see it (and is thus relying on memory). On the other hand, when she looks at it to describe to Alan, he can no longer see what she is talking about. As Alan struggles to map his view of Brenda's changing object to his own view of his own object, he realizes that ultimately the video does more harm than good. They later resort to using verbal descriptions of how to rotate the object into the right orientation. Beyond this, we observe that the camera capture itself is problematic: when Brenda tries to show Alan the inside of the trophy, she holds it too close to the camera such that Alan cannot understand the trophy's orientation—that is, he cannot extract contextual 3D spatial information from the 2D video.

Pause Workaround. Groups experienced similar orientation confusion in the Opposing-ReMa condition; however, five of eight groups developed a clever workaround by repurposing the "pause" functionality (originally designed to allow participants to examine their objects independently) to create a shared perspective on the object. In Group 8, TS participant describes this idea:

Vignette 3: Group 8, Opposing-ReMa*. Ava (TS) shows the trophy so that Mia (MS) can see the right side. Ava then pauses the system and orients the trophy for herself that both can look at the same side of the trophy. "So, I pause it, then I turn it, so I can see the same side," Ava explains. "Ooooh, smart!" replies Mia.*

"Reset" Strategy Given a Shared Perspective. The participants generally worked very well in the shared perspective trials. With Shared-VC, most teams readily identified a "start" position/orientation that they used for the rest of the task. Here, after successfully affixing each sticker, they would revert their own respective objects to the "start" orientation to begin again. In the following vignette, one participant works with the other to establish what the "start" orientation will be.

Vignette 4: Group 2, Shared-VC. *At the beginning of the task, Nancy holds the trophy right-side up, "If you look at the bottom of the trophy, there is a flat side." Ned looks at the wrong side, so he rotates his trophy to find the flat side. To confirm and establish this position, Nancy says to Ned, "Put the flat side in front of you [and] let's call that original position."*

This strategy allowed participants to easily revert to a "known state" if they got into a confusing situation that was difficult to recover from.

Shared-ReMa Prevents Exploration. The mental model provided by ReMa in the shared perspective was straightforward for participants to understand. The automatic reorientation meant that participants did not need to describe re-orientation actions (and potentially have them misinterpreted or reproduced incorrectly), as in the VC conditions. Yet, the tradeoff was that TS participants could not look explore the object, to understand it, or to look ahead at next steps properly without affecting their partner. This problem is illustrated in the following vignette:

Vignette 5: Group 6, Shared-ReMa. *Jon (TS) tells Emi (MS) where to put the sticker, "It's a 25 cent yellow sticker upside-down." Emi begins to attach the sticker, but pauses and asks about a nearby sticker (which could act as a landmark): "Actually, do you see the two…?" Jon twists the bird house to check if there's a sticker left, which startles Emi, who was about to put a sticker down. "What!? Stop, Jon!" Jon sheepishly returns the birdhouse to his original position, "Oops, sorry, I forgot…I am controlling the robot arm."*

Thus, TS participants would need to hold the object in place while Manipulator Site participants worked, and could not "look ahead" at other parts of the object without affecting their partner's activities.

ReMa Conditions: Slow-confirmation. Participants appreciated that ReMa's automatic reorientation meant they shared the same perspective each time the object was reoriented. This reduced the number of interpretation errors between participants: "*I could just assume that we are looking at the same side of the trophy*" (G8-P15). However, when using ReMa, the Tracking Site participants did not know when Manipulator Site had finished re-orienting the object for the other participant, and when/whether the MS participant had completed an action/instruction step on their own object. This is illustrated by Group 5, where the MS participant slows the interaction down by asking several confirmatory questions of his partner to ensure both are looking at the same thing:

Vignette 6: Group 5, Opposing-ReMa. *Harry (MS) looks at the bird house ReMa has oriented for him. He starts talking about an empty sticker he sees on the left side of the bird house. Ben (TS) is craning his neck to look at the same side of the bird house, because he does not want to move the object, but he cannot see the sticker: "Empty sticker? There is no empty sticker in front..." (Ben). Harry wants to confirm that*

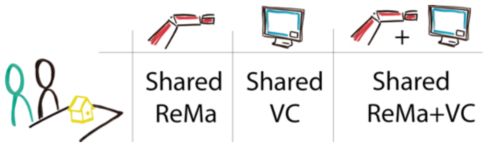

Figure 5. Study 2 compared shared video chat and ReMa

they share the same perspective: "I just want make sure we are looking on the same side. Is there a little desk in front of the house?"

This type of confirmatory behaviour was common across all pairs, because neither participant had a strong understanding of what the partner could see. To overcome this problem, pairs would frequently resort to slowing down the interaction, and then ask questions about what the other participant could see, or what they were doing.

Summary. In summary, this study shows that the Shared perspective was far more straightforward for participants to adopt. Pairs developed interaction strategies around this perspective to allow them to complete the task efficiently. In contrast, the Opposing perspective, which is how conventional video chat tools are oriented, caused problems for participants: they had a hard time distinguishing between left-right rotations, and could not see and show aspects of the object at the same time. At the same time, the study raised several questions about the role of ReMa, leading us to design the second study.

STUDY 2: ROLE OF A PHYSICAL PROXY GIVEN VIDEO

Our second study focused on how the presence or lack of proxy changes collaborative behaviour. Specifically, *how is a physical proxy used in the presence of a video channel?* We are interested in situations where the video chat channel is available alongside a system like ReMa, which can manipulate a physical proxy. What role does each of these channels play in supporting the collaboration?

Study Variables. The second study had three conditions: a video-only condition (VC-Only), a physical proxy-only condition (ReMa-Only), and a combined condition with both video and a physical proxy (VC+ReMa). Given our Study 1 findings, all conditions used a shared perspective. Participant pairs experienced all three conditions. The video-only and physical proxy-only conditions were presented either first or second (counter-balanced across pairs); the combined condition was always presented last.

Participants. We recruited 16 participants (eight pairs; nine females; seven males), aged 18-29 with backgrounds including computer science, actuarial math, animal and medical science, arts, linguistics, and electrical and chemical engineering. Participants from Study 1 were prohibited from participating in Study 2.

Task Design. As in the first task, participants were given similar proxy objects. Each object had markers in the same locations, however the content of these markers differed between participants. Some markers had letters written on them, in different orientations, while others were blank, corresponding to a lettered marker on their collaborator's object. Participants were responsible for writing the correct letter on blank markers at the right orientation. Participants needed to exchange information such that the markers on the objects matched at the end.

We revised our study tasks to highlight specific issues with physical proxies. In Study 1 we observed that participants could glean considerable information from the video, including object orientation and specific marker details (e.g. colour and content). In real-world situations, a video channel may be insufficient in capturing the orientation, context, complexity, or detail of real-world artifacts (e.g. subtle material cues in design critique [30], larger-scale physical tasks [35]).

Data Collection & Analysis. We followed the same data collection and analysis approach as in Study 1.

Study 2 Findings & Observations

We focus our analysis on the VC+ReMa trials, where participants had access to both VC and the physical proxy. Based on our analysis, we find that the physical proxy obviated the problems participants have in describing, interpreting and carrying out reorientation acts on the proxy object, while video chat helped participants understand what has happened to the remote proxy object, and gesture. To illustrate this differential use of the video chat and the proxy object, we contrast VC+ReMa, VC-only and ReMa-only conditions. Our interest is not a clean comparison between the conditions, but rather to understand the contribution ReMa and VC is making to the collaborative interaction, and to explore how VC+ReMa is used in collaborative tasks,

Distinct Roles for VC and ReMa. From a video analysis of the VC+ReMa trials, we collected data on how MS participants used each channel. Figure 6 illustrates the proportion of time the MS participants focused their visual attention on ReMa compared to the video chat. Between video chat and ReMa, MS participants disproportionately spent their visual attention on the ReMa-manipulated proxy object. Our analysis of the VC+ReMa condition shows that the proxy object was used as a shared workspace (e.g. for the MS to understand in what orientation the object ought to be, or what to do), whereas the video was used for confirming that actions/steps had been taken (e.g. MS ensuring that the TS's object had been rotated to the correct orientation), and for gestures. The following vignette, from Group 2, typifies the VC+ReMa experience of seven of our eight groups:

Vignette 7: Group 2, VC+ReMa. *Susan (TS) rotates her object, telling Larry (MS), "I'm gonna turn it". The robot rotates the object and Susan says: "It's gonna take a sec." While the robot is rotating Larry's object, Susan watches the robot through the video chat. All the while, Larry is watching the robot and the object carefully. Once the robot completed the reorientation, Susan says "So…YES, at the back of the trophy there is a G." Larry, knowing Susan can see him in the VC, points at the sticker to confirm the sticker (Figure 7). Finally, he peeks at the VC to confirm the orientation of the trophy before writing the letter.*

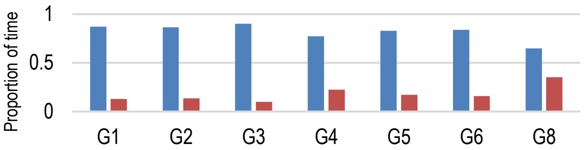

Figure 6. Proportion of time MS participants focused on ReMa (blue) *vs.* video chat (red). (G7 video data lost).

Figure 7: VC+ReMa – Group 2. Larry (MS) points at the trophy confirming the sticker position, knowing that Susan (TS) can see the gesture through the video.

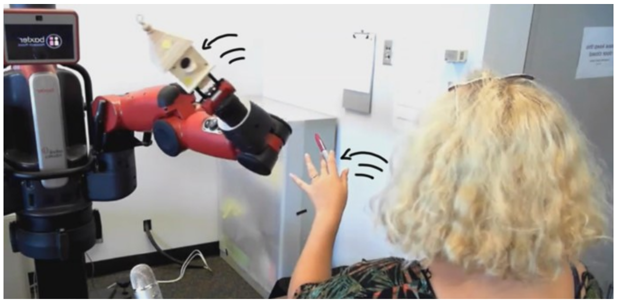

Figure 8: VC+ReMa – Group 1. Clara (TS) uses spatial hand gesture to describe the movement Lina (MS) should execute (annotated for clarity).

This vignette highlights three aspects of the interaction as it relates to the different channels. First, Larry's primary interest is on his immediate workspace: the proxy object, held by ReMa. Most of his visual attention is here—he waits for the proxy object to settle into position, and once its position is stable he writes on the project object. Second, Larry generally does not use video chat, with the exception of understanding Susan's gestures and ensuring that his object roughly matches Susan's. Similarly, Susan's primary use of the video chat is to watch as Larry's object rotates into place—she uses the video to confirm that ReMa has executed her rotation act properly. Ultimately, Susan waits for visual confirmation that Larry has completed the task before she moves onto the next marker. Thus, the video and proxy object each play distinct roles in supporting the interaction; we see this in the absence of one channel during the VC-only and ReMa-only trials.

While seven out of eight pairs used video in this way, G8 was an outlier. The MS participant used the video instead of the robot as a primary visual reference for solving the task. The pair communicated almost strictly via video chat, the MS determining the position and solution letter for blank stickers, going to the robot with the object ready in the right orientation, and finally writing the letter on the marker.

Using Gestures to Communicate Re-Orientation Acts. As in Study 1, the MS participants were severely disadvantaged in communicating back to the TS participant. The MS participant could not physically manipulate their own object. Thus, to communicate how an object ought to be rotated, MS participants frequently used hand gestures to rotate an imaginary object in midair, providing a verbal description alongside the gesture. The TS participants could only rely on the video chat channel to understand what was intended by MS's description. TS then rotated the object as they interpreted the instructions, and MS confirmed based on the rotation of the physical proxy in front of them.

Vignette 8: Group 1, VC+ReMa. *Clara (TS) wants to describe the right orientation of the object for Lina (MS). First, she describes the position of the sticker: "In the left bottom corner" Lina confirms and asks: "Yes, bottom left…what should I do?" Clara is uncertain how to describe the movement: "Just move it…left…90 degrees to the left". She uses hand movements to show how Lina should move the object (Figure 8). While Clara explains and gestures the movement Lina is watching the video chat to better understand Clara's gestures.*

VC-Only Trials. Compared with the VC+ReMa trials, the pairs' main challenge was to effectively describe re-orientation actions to their partner, or to interpret those instructions (and carry them out properly). While they could use the video chat to observe the remote site, and interrupt when problems occurred, the presence of the video chat did not prevent these mistakes from happening.

Dialogue from these trials were fragmented. Participants used step-by-step language to describe their actions and stay aligned with the remote person. As a participant provided instructions, s/he would watch the video chat to see if/whether the instructions had been understood, repairing the interaction as necessary. In one instance, participant [G1-P2] needed three tries before he is satisfied with the outcome: *"And don't move the house...in this position there is…Wait, just move the house 90 degrees to the left"* In another example, the participant [G5-P10] realizes that the instruction he just gave is ambiguous, and tries to repair it twice, all while watching his remote partner struggle: *"Just rotate it… That means you just keep the tip of the house at the table and the base of the house upwards... facing the roof."*

Because participants could see each other's object in the video chat, they could catch errors quickly; however, this interaction was far from smooth. Just as in Study 1, many pairs defined a "start" position at the beginning of the task to avoid orientation confusion, and returned to this position when their objects became misaligned. In contrast, the VC+ReMa trials were smoother: TS participants did not need to verbally convey reorientation actions to their MS, resulting in dialogue that was much clearer and focused on the markers themselves. Confirmation that the proxy object had moved correctly, or that MS participant had completed the action correctly relied on the video chat channel.

ReMa-Only Trials. Due to the absence of a video chat channel, participants needed to communicate verbally or through the orientation of the proxy object (via ReMa). This presented challenges for both the TS and MS, but in different situations. As in Study 1, the TS relied on verbally confirming with the MS about whether the ReMa had finished its movements, and whether the MS participant had finished his/her actions (i.e. writing the symbol on the sticker). We observed participants regularly and explicitly asking for verbal confirmation (e.g. *"Are you ready?"* [G1-TS] or *"You got it?"* [G3-TS]), because they had no other way to know the current state at the remote site. Where the video chat acted as a feedback mechanism in the VC+ReMa conditions, its absence in VC-only conditions markedly increased verbal confirmation cues.

Because the MS participant could not manipulate the TS participant's workspace, s/he described re-orientation steps for the TS object verbally (as in Vignette 8). Interestingly, MS participants still used spatial hand gestures to describe rotations (as in the VC+ReMa condition); of course, these hand gestures were not visible to the TS participant, and MS participants confirmed they were aware that TS participants could not see actions in a video stream. Instead, these "rehearsal" acts seemed let participants explain actions from a first-person perspective.

All TS participants oriented the proxy object such that the correct side of the object faced the participant, and that side was oriented such that the MS participant could easily write the letter "right side up." This contrasts with Study 1, where only two groups oriented the object such that the markers were placed "right side up"; most groups in Study 1 kept the *object* oriented "right side up", even if that meant the MS needed to affix the sticker on upside down. We suspect that the revised task and markers influenced this change—in Study 2, the MS had to *write* a symbol, whereas in Study 1, MS only needed to *affix* the sticker.

No Pausing Necessary. As in Study 1, we provided a pause functionality; however, no groups used this function in any trial. Given that the bulk of its use in Study 1 was during Opposing perspective conditions to mimic a Shared perspective, this is lack of use is perhaps unsurprising since Study 2 only used the Shared perspective. No participants complained about the use of a shared perspective, and found it very straightforward and easy to use: *"You can just look at it and you see whatever the other one is seeing…"* [G7-P14] or *"It is easier to understand what the other person is really looking at*" [G8-P16].

Summary. Study 2 focuses on how people use video chat and ReMa differently given the presence of both channels. ReMa was used primarily for orienting the shared workspace. Meanwhile, the video chat let people visually confirm what happened to the proxy object, and offered a means to gesture at the proxy object.

DISCUSSION AND FUTURE WORK

Below we move from the specific context of ReMa to reflections relating to the design of future object-focused collaborative interfaces.

Perspective Shifts. Non-collocated collaborators have the flexibility to independently orient themselves (or an object) during an object-focused discussion. Yet, both studies demonstrated that a shared perspective is useful and powerful when executing tasks with the object. The shared perspective allowed participants to discuss the object without mentally rotating the object, and to describe parts of the object that are not necessarily visible from their perspective. Designers need to carefully consider camera placement and object/workspace orientation in systems like ReMa to reduce people's need to mentally rotate objects.

Uses of Physical Object and Rotations. The ReMa system and study confirm that indeed, rotations are an important part of object-focused collaboration. In prior work [23], rotations of flat 2D objects can play a role in both communicating (i.e. explaining something to someone else), as well as coordinating action (i.e. whose turn is it to do something). We saw a similar effect in our study: TS rotations of the object functioned as demonstrations of where to place a sticker, or annotate the object. Rather than requiring the TS to verbally describe how to turn the object, or demonstrating via video, ReMa simply performed the rotation. Performing the rotation both drew attention to a specific side of the artefact (communication), and signaled to the MS participant that something needed to be done at that site (coordination).

Handling for Comprehension. ReMa does not explicitly address the use of rotating an object for comprehension (i.e. exploring the object, or taking time to understand the object). The MS participant cannot explicitly hold or manipulate their object; when the MS wanted to explore their object, they had to explicitly ask the TS to re-orient their object for them (Vignette 8). As a result, the MS participant does not get to manipulate and explore the object for themselves—all object interactions are mediated by cumbersome dialogue with TS. We also observed that TS participants were more guarded in exploring their own object. As we saw in Vignette 5, when TS participants realized that their actions were immediately reflected at the MS (and might potentially disrupt the MS participant's actions), they avoided excess object manipulations. In contrast, participants in video chat-only conditions were free to manipulate their object, but at a cost of coordination. Video chat-only participants' objects easily became unsynchronized, necessitating the "start position" strategy.

In Study 1, we tried to design for comprehension acts by providing "Pause" functionality, which temporarily disabled the TS from the MS. In principle, this allowed the TS participant to explore the object without changing the MS participant's view of the object. However, few participants used it for this purpose; participants mainly used "Pause" to recreate a Shared perspective during Opposing perspective conditions.

Future designs need to consider different ways to move between synchronized and un-synchronized remote objects: for example, through a clutching mechanism activated by proximity (i.e. only tracking a "shared" workspace, leaving personal workspace for independent manipulation), or manual clutching (similar to the pausing mechanism).

Expanding the Manipulator Space. We are interested in expanding the capabilities of the manipulator site, particularly to capture more degrees of freedom and object movement paths. This additional information is important for object-focused collaborations to describe: relationships between different objects; how an object should be used or oriented, or how a multi-part object might be assembled. Capturing the timing of a movement path is important, too. The current ReMa implementation is limited to manipulating an object's orientation (at 90 degree angles), which was sufficient to address our current research questions. We are interested in further developing the ReMa infrastructure such that more flexible and rich movements, positions and timing are accurately rendered. This would allow even subtle gestures or manipulations involving the object to be conveyed at a distance (e.g. [21]). Given this system, participants generated several ideas for application domains, for instance, instructions on how to manipulate a complex 3D object (e.g. knife skills), expert-novice tasks, construction tasks, and so on.

Capturing and Rendering Manipulator Gestures. Prior work focused on providing collaborators mechanisms with gesturing at objects in the workspace, or at areas of the workspace [5, 9, 10, 13, 20, 21, 24, 28 41, 42]. Our study participants used gestures—particularly in the video chat conditions—to point at various parts of the object. When this capability was taken away in the ReMa conditions, this presented challenges for participants. Future research needs to develop new ways to both capture gestures (such as deictic or hand gestures) for object-focused collaboration, as well as determining *how* to render these gestures at a remote site for interpretability. While video is a reasonable stop-gap solution, it ignores the subtleties of gesturing at partially obscured or difficult-to-view locations on an object. It also misses the entire production of the gesture, which may be important for interpreting the meaning of a given gesture [28].

Bidirectional Capture and Manipulation. While interaction with the physical object in our study was strictly unidirectional, we are also interested in bidirectional scenarios. As illustrated in [2], while bidirectional physical objects present compelling experiences, they also present new questions. Most notably: *how should conflicts be resolved?* One approach to resolving conflicts is to relax what would otherwise be strict synchronization. In this "relaxed" synchronization mode, a collaborator could choose whether to follow the remote site's object depending on his/her situation. We envision a mechanism that would allow a collaborator to explore their own interaction path with an object, and resynchronize with their remote collaborator when needed with little penalty to either.

ReMa, and Human-Robot Interaction (HRI). Due to opportunistic reasons we used a humanoid robot for both ReMa studies, and (in Study 1) we enabled head movements and different "facial" displays to provide feedback on ReMa's movements. Extensive past HRI work investigating the impact of anthropomorphism on interaction suggests that our choice may have affected our results, and that realizing ReMa with a more generic robotic arm could have potentially created different biases (e.g. [6, 19]). However, we found little evidence of the humanoid form effects, for example none of our participants recognized the "facial" displays in the post-study interview: *"I was so focused on the task and the object. I did not see [the face] at all"* [G4-P7]. We removed these feature for Study 2. Future ReMa-like systems should include and evaluate the effect of replacing the humanoid with a simple robotic arm implementation.

CONCLUSIONS

This paper explores the challenges of coordinating object-focused collaboration when collaborators are remote from one another. Specifically, we considered how collaborators' perspectives on an object affects the way in which they coordinate activity. We built and studied the Remote Manipulator (ReMa), which automatically orients the proxy object to reflect the orientation at a Tracking Site. We found that a shared perspective on the object is easier for people to manage compared to the default "opposing" perspective offered by conventional video chat. We also found that ReMa can be a useful aid to collaboration, easing the pressure of describing and reproducing verbal reorientation cues on an object. Finally, our analysis shows that ReMa and a video channel complement each other when used together, giving people more effective tools to coordinate their actions in object-focused collaboration. Looking forward, our results suggest ways that researchers should consider new workspaces that improve object-focused collaboration, including supporting simultaneous object manipulation and remote gesture, managing synchronized and unsynchronized object manipulation, and handling bidirectional capture and manipulation.

ACKNOWLEDGEMENTS

We thank André Miede for technical discussions in supporting this research project. We also thank the Interactions Lab, our participants, ACAMP, NSERC and htw saar for their support.

REFERENCES

1. Patrick Beeson and Ames Barrett. "TRAC-IK: An open-source library for improved solving of generic inverse kinematics." *Humanoid Robots (Humanoids), 2015*

IEEE-RAS 15th International Conference on. IEEE, 2015.

2. Scott Brave and Andrew Dahley. 1997. inTouch: a medium for haptic interpersonal communication. In *CHI '97 Extended Abstracts on Human Factors in Computing Systems* (CHI EA '97). ACM, New York, NY, USA, 363-364. DOI: https://doi.org/10.1145/1120212.1120435
3. Jed R. Brubaker, Gina Venolia, and John C. Tang. 2012. Focusing on shared experiences: moving beyond the camera in video communication. In *Proceedings of the Designing Interactive Systems Conference* (DIS '12). ACM, New York, NY, USA, 96-105. DOI: https://doi.org/10.1145/2317956.2317973
4. Herbert H. Clark, & Susan E. Brennan. 1991. Grounding in communication. *Perspectives on socially shared cognition*, *13*(1991), 127-149.
5. 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* (CHI '16). ACM, New York, NY, USA, 1548-1560. DOI: https://doi.org/10.1145/2858036.2858171
6. Julia Fink (2012, October). Anthropomorphism and human likeness in the design of robots and human-robot interaction. In *International Conference on Social Robotics* (pp. 199-208). Springer Berlin Heidelberg.
7. Christian Freksa (1991). Qualitative spatial reasoning. *Cognitive and linguistic aspects of geographic space*, *63*, 361-372.
8. Susan R. Fussell, Leslie D. Setlock, and Robert E. Kraut. 2003. Effects of head-mounted and scene-oriented video systems on remote collaboration on physical tasks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '03). ACM, New York, NY, USA, 513-520. DOI=http://dx.doi.org/10.1145/642611.642701
9. Susan R. Fussell, Leslie D. Setlock, Jie Yang, Jiazhi Ou, Elizabeth Mauer, and Adam D. I. Kramer. 2004. Gestures over video streams to support remote collaboration on physical tasks. *Hum.-Comput. Interact.* 19, 3 (September 2004), 273-309. DOI=http://dx.doi.org/10.1207/s15327051hci1903_3
10. 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* (VRST '14). ACM, New York, NY, USA, 197-205. DOI: https://doi.org/10.1145/2671015.2671016
11. Raymond C. Goertz and William M. Thompson. "Electronically controlled manipulator." *Nucleonics (US) Ceased publication* 12 (1954).
12. Saul Greenberg, Carl Gutwin, & Mark Roseman. 1996. Semantic telepointers for groupware. In *Computer-Human Interaction, 1996. Proceedings., Sixth Australian Conference on* (pp. 54-61). IEEE.
13. Pavel Gurevich, Joel Lanir, Benjamin Cohen, and Ran Stone. 2012. TeleAdvisor: a versatile augmented reality tool for remote assistance. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '12). ACM, New York, NY, USA, 619-622. DOI: http://dx.doi.org/10.1145/2207676.2207763
14. Mary Hegarty & David Waller. (2004). A dissociation between mental rotation and perspective-taking spatial abilities. *Intelligence*, *32*(2), 175-191.
15. Peter F. Hokayem & Mark W. Spong. (2006). Bilateral teleoperation: An historical survey. *Automatica*, *42*(12), 2035-2057.
16. Steven Johnson, Madeleine Gibson, and Bilge Mutlu. 2015. Handheld or Handsfree?: Remote Collaboration via Lightweight Head-Mounted Displays and Handheld Devices. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing* (CSCW '15). ACM, New York, NY, USA, 1825-1836. DOI: http://dx.doi.org/10.1145/2675133.2675176
17. 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* (CHI '15). ACM, New York, NY, USA, 957-966. DOI: https://doi.org/10.1145/2702123.2702345
18. Brigitte Jordan & Austin Henderson. 1995. Interaction analysis: Foundations and practice. *The journal of the learning sciences*, *4*(1), 39-103.
19. T. Kanda, T. Miyashita, T. Osada, Y. Haikawa, & H. Ishiguro. (2008). Analysis of humanoid appearances in human–robot interaction. *IEEE Transactions on Robotics*, *24*(3), 725-735.
20. David Kirk and Danae Stanton Fraser. 2006. Comparing remote gesture technologies for supporting collaborative physical tasks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '06), Rebecca Grinter, Thomas Rodden, Paul Aoki, Ed Cutrell, Robin Jeffries, and Gary Olson (Eds.). ACM, New York, NY, USA, 1191-1200. DOI=http://dx.doi.org/10.1145/1124772.1124951
21. David Kirk, Andy Crabtree & Tom Rodden. 2005. Ways of the hands. In *ECSCW 2005* (pp. 1-21). Springer Netherlands.
22. David Kirsh and Paul Maglio. "On distinguishing epistemic from pragmatic action." *Cognitive science* 18.4 (1994): 513-549.

23. Russell Kruger, Carpendale, S., Scott, S. D., & Greenberg, S. 2004. Roles of orientation in tabletop collaboration: Comprehension, coordination and communication. *Computer Supported Cooperative Work (CSCW)*, *13*(5), 501-537.
24. Hideaki Kuzuoka, Shinya Oyama, Keiichi Yamazaki, Kenji Suzuki, and Mamoru Mitsuishi. 2000. GestureMan: a mobile robot that embodies a remote instructor's actions. In *Proceedings of the 2000 ACM conference on Computer supported cooperative work* (CSCW '00). ACM, New York, NY, USA, 155-162. DOI=http://dx.doi.org/10.1145/358916.358986
25. Hideaki Kuzuoka, Toshio Kosuge, and Masatomo Tanaka. 1994. GestureCam: a video communication system for sympathetic remote collaboration. In *Proceedings of the 1994 ACM conference on Computer supported cooperative work* (CSCW '94). ACM, New York, NY, USA, 35-43. DOI=http://dx.doi.org/10.1145/192844.192866
26. Daniel Leithinger, Sean Follmer, Alex Olwal, and Hiroshi Ishii. 2014. Physical telepresence: shape capture and display for embodied, computer-mediated remote collaboration. In *Proceedings of the 27th annual ACM symposium on User interface software and technology* (UIST '14). ACM, New York, NY, USA, 461-470. DOI: https://doi.org/10.1145/2642918.2647377
27. 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. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (CHI '17). ACM, New York, NY, USA, 5295-5306. DOI: https://doi.org/10.1145/3025453.3025848
28. Paul Luff, Christian Heath, Hideaki Kuzuoka, Jon Hindmarsh, Keiichi Yamazaki, and Shinya Oyama. 2003. Fractured ecologies: creating environments for collaboration. *Hum.-Comput. Interact.* 18, 1 (June 2003), 51-84. DOI=http://dx.doi.org/10.1207/S15327051HCI1812_3
29. Takanori Miyoshi, Imamura, T., Oyama, S., Ohba, Y., Ichimura, T., Sawaguchi, Y., …& Kawai, Y. (2014, December). Experiment of virtual tug-of-war via internet with multilateral telecontrol. In *Network and Systems Support for Games (NetGames), 2014 13th Annual Workshop on* (pp. 1-3). IEEE.
30. Terrance Mok and Lora Oehlberg. 2017. Critiquing Physical Prototypes for a Remote Audience. In *Proceedings of the 2017 Conference on Designing Interactive Systems* (DIS '17). ACM, New York, NY, USA, 1295-1307. DOI: https://doi.org/10.1145/3064663.3064722
31. Point Natural. 2011. Optitrack. *Natural Point, Inc.,[Online]. Available: http://www. naturalpoint. com/optitrack/.[Accessed 22 2 2014]*.
32. Carman Neustaedter, Gina Venolia, Jason Procyk, and Daniel Hawkins. 2016. To Beam or Not to Beam: A Study of Remote Telepresence Attendance at an Academic Conference. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing* (CSCW '16). ACM, New York, NY, USA, 418-431. DOI: https://doi.org/10.1145/2818048.2819922
33. Jacki O'Neill, Stefania Castellani, Antonietta Grasso, Frederic Roulland, and Peter Tolmie. 2005. Representations can be good enough. In *Proceedings of the ninth conference on European Conference on Computer Supported Cooperative Work* (ECSCW'05), Hans Gellersen, Kjeld Schmidt, Michel Beaudouin-Lafon, and Wendy Mackay (Eds.). Springer-Verlag New York, Inc., New York, NY, USA, 267-286.
34. Morgan Quigley, et al. "ROS: an open-source Robot Operating System." *ICRA workshop on open source software*. Vol. 3. No. 3.2. 2009.
35. Irene Rae, Bilge Mutlu, and Leila Takayama. 2014. Bodies in motion: mobility, presence, and task awareness in telepresence. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '14). ACM, New York, NY, USA, 2153-2162. DOI: https://doi.org/10.1145/2556288.2557047
36. Irene Rae, Leila Takayama, and Bilge Mutlu. 2013. In-body experiences: embodiment, control, and trust in robot-mediated communication. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '13). ACM, New York, NY, USA, 1921-1930. DOI: https://doi.org/10.1145/2470654.2466253
37. Daniel Rakita, Bilge Mutlu, and Michael Gleicher. 2017. A Motion Retargeting Method for Effective Mimicry-based Teleoperation of Robot Arms. In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction* (HRI '17). ACM, New York, NY, USA, 361-370. DOI: https://doi.org/10.1145/2909824.3020254
38. Michael F. Schober. 1993. Spatial perspective-taking in conversation. Cognition, 47(1), 1-24.
39. John C. Tang and Scott L. Minneman. 1991. Videodraw: a video interface for collaborative drawing. *ACM Trans. Inf. Syst.* 9, 2 (April 1991), 170-184. DOI=http://dx.doi.org/10.1145/123078.128729
40. John C. Tang and Scott Minneman. 1991. VideoWhiteboard: video shadows to support remote collaboration. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '91), Scott P. Robertson, Gary M. Olson, and Judith S. Olson (Eds.). ACM, New York, NY, USA, 315-322. DOI: http://dx.doi.org/10.1145/108844.108932

41. Anthony Tang, Michael Boyle & Saul Greenberg 2005. Understanding and mitigating display and presence disparity in mixed presence groupware. *Journal of Research and Practice in Information Technology*, *37*(2), 193-210.
42. Anthony Tang, Michel Pahud, Kori Inkpen, Hrvoje Benko, John C. Tang, and Bill Buxton. 2010. Three's company: understanding communication channels in three-way distributed collaboration. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work* (CSCW '10). ACM, New York, NY, USA, 271-280. DOI: https://doi.org/10.1145/1718918.1718969
43. John C. Tang. 1991. Findings from observational studies of collaborative work. *International Journal of Man-machine studies*, *34*(2), 143-160.
44. Peter Turpel, Bing Xia, Xinyi Ge, Shuda Mo, and Steve Vozar. 2013. Balance-arm tablet computer stand for robotic camera control. In *Proceedings of the 8th ACM/IEEE international conference on Human-robot interaction* (HRI '13). IEEE Press, Piscataway, NJ, USA, 241-242.
45. Peter Turpel, Bing Xia, Xinyi Ge, Shuda Mo, and Steve Vozar. 2013. Balance-arm tablet computer stand for robotic camera control. In *Proceedings of the 8th ACM/IEEE international conference on Human-robot interaction* (HRI '13). IEEE Press, Piscataway, NJ, USA, 241-242.
46. Jeffrey M. Zacks, Mires, J., Tversky, B., & Hazeltine, E. (2000). Mental spatial transformations of objects and perspective. *Spatial Cognition and Computation*, *2*(4), 315-332.
47. Dingyun Zhu, Tom Gedeon, and Ken Taylor. 2011. Exploring camera viewpoint control models for a multi-tasking setting in teleoperation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '11). ACM, New York, NY, USA, 53-62. DOI: https://doi.org/10.1145/1978942.1978952