Showing Objects: Holding and Manipulating Artefacts in Video-mediated Collaborative Settings

Christian Licoppe[1], Paul K. Luff[2], Christian Heath[2], Hideaki Kuzuoka[3], Naomi Yamashita[4], Sylvaine Tuncer[1]

[1] I3-SES, CNRS, Télécom ParisTech, Université Paris-Saclay, 75013, Paris, France <first.second@telecom-paristech.fr>

[2] School of Management and Business, King's College London, UK <first.second@kcl.ac.uk>

[3] University of Tsukuba, Ibaraki, Japan <kuzuoka@iit.tsukuba.ac.jp>

[4] NTT Communication Science Laboratories, Kyoto, Japan <naomiy@acm.org>

ABSTRACT

In this paper we report on a pervasive practice in video-mediated communication: where participants show one another one or more objects. This is a distinct activity from others considered by researchers of video-mediated technologies that focus on a face-to-face orientation, or just on the support necessary to help people to refer to objects. We first present examples of this pervasive phenomenon in naturally occurring Skype conversations, revealing how this conduct is configured and organized within the interaction between participants. We reveal how the subtle adjustment of the position of the body, the head and gaze with respect to the handheld objects offers crucial resources for participants to achieve joint seeing. Then we report on a quite different setting, a naturalistic experiment where participants collaborate on a collective task with remote colleagues through maneuverable, orientable devices (Kubis). Again, in these experiments participants frequently show objects, and at times the devices provide additional resources to support these activities. But at other times they also involve some difficulties. We conclude by suggesting possible technological developments, some quite simple, others more radical, that might support participants to show objects, whether they are in domestic settings or undertaking work activities.

Author Keywords

Video-mediated interaction; Skype, domestic; objects; embodied interaction

ACM Classification Keywords

H5.3 Computer-supported cooperative work; H4.3. Computer conferencing and Video-conferencing

CHI 2017, May 06 - 11, 2017, Denver, CO, USA

ACM 978-1-4503-4655-9/17/05… $15.00
DOI: http://dx.doi.org/10.1145/3025453.3025848

General Terms

Human Factors.

INTRODUCTION

Although there are a number of commonly available systems for video-mediated communication, Skype is by far the most popular. It has been estimated that in 2015, for example, there were 300 million users of Skype who in total used Skype for 3 billion minutes a day [38]. This recent success of Skype is perhaps surprising given the much documented failures of earlier attempts of video-telephony for general consumers and advanced media spaces for work places [6]. Although field studies and interviews have revealed that Skype supports the maintenance of different kinds of social relationships [1], there are few detailed studies of how participants interact through Skype; how everyday, naturalistic video conversations are accomplished. In this paper, we draw on a recently collected corpus of video conversations between friends and family members to consider a topic that has been given less attention in HCI and related studies of mediated communication: the showing, sharing and mutual inspection of objects.

Looking at, sharing and talking about objects is a pervasive feature of face to face interaction throughout a variety of settings, and own research and cognate studies of various forms of 'mediated' communication suggest it is also common in remote or distributed interactions. However, showing and sharing objects poses a range of specific interactional 'problems' – not simply that objects need to be brought into view, but rather configured and revealed in particular ways to encourage and engender particular forms of participation and experience. By beginning to understand the complexities of these activities, and the contingencies that arise in showing, sharing and seeing objects, we can reflect on the challenges we face in developing remote 'environments' that can support and enhance our ability to share and experience objects with one another.

We will discuss showings and their organization in two different settings, in both of which they feature as recurrent and pervasive practices. In our first corpus, of naturally occurring Skype conversations between family

and friends, we found participants show a wide variety of objects through Skype, from items of clothing to pieces of furniture. These showings accomplish more than just identifying and referring to an object and augment face-to-face interaction in very particular ways. They also reveal aspects of domestic life mediated through technology. In this paper we begin to show a systematic, sequential organization to showings and seeings, and that they are configured in very particular ways. These organization and configurations were also apparent when we reviewed a second and quite distinctive collection of data – a naturalistic experiment considering a novel enhancement to Skype where the video-mediating technology was more maneuverable and less constrained. Again we found participants frequently showing objects to one another, but here these were in the course of accomplishing collaborative tasks. We briefly consider how participants show objects through this technology and how at times it seemed to transform how objects are shown within interaction. It also suggests some issues that need to be resolved and directions in which video-mediated technologies might take if they are to support participants manipulating and handling physical objects, and making these activities apparent to remote colleagues. We conclude by suggesting some ways in which studies of the seemingly mundane uses of everyday technologies might support the development of innovative systems.

BACKGROUND

Following early endeavors in video-mediated interaction that focused on face-to face communication, a number of researchers suggested that systems should move beyond supporting 'head and shoulders' views and provide greater access to the surrounding environment. Studies of different kinds of video-mediated technologies revealed the particular importance of having access to objects, such as documents, when participants accomplish collaborative tasks [3, 10, 23, 25, 41]. This led to a series of innovative approaches and prototype augmented media spaces that sought to offer participants additional resources so that they could have greater access to remote objects and to the related activities of their remote colleagues [8, 38, 41]. The focus of these initiatives was to augment video-mediated communication either with additional communication channels [22], by providing (digital) access to documents through multiple cameras and screens [25, 29, 30] or ways of adjusting and maneuvering the view of the other domain [9]. A number of challenges emerged from these initiatives, particularly when considering how participants referred to remote objects through these augmented media spaces, most notably that such conduct could become disembodied, fractured or fragmented [28, 40]. Even with additional capabilities that seek to preserve or mimic the relationships between objects, the local and remote participants and their environments, it can be hard to assess the relevance of a remote colleague's actions with regard to particular objects or to secure their alignment to an object of concern. The focus of these developments have been towards referential conduct, with little analysis or consideration given to other kinds of activity with and around objects. Until now, little attention has been paid to the activities of participants with objects they have referred to, that is how participants handle and show the artefacts they refer to.

Recently a few studies have considered video-mediated communication between members of families and friends who may be geographically dispersed [23, 32, 33]. These have reported how the technology can meet various relational needs within families [18, 23, 39], the kind of activities family members undertake to manage the interaction [1, 32] and problems faced by participants when communicating through video. These studies draw principally from interviews with the participants, sometimes supplemented with observations and illustrations taken from recordings. However, there are few studies of the nature of everyday video-mediated communication through systems like Skype; how talk and visual conduct are accomplished from moment-to-moment within the environments in which they occur.

This, in part, reflects the current concerns within the field of human-computer interaction. Over the past 20 years there has been a shift away from detailed studies of how everyday or conventional technologies are used, either in the workplace or in other settings. There are notable exceptions. Researchers have considered how commonly used applications, like PowerPoint, are used to accomplish a range of tasks [24] or how conventional desktop computers feature in medical interactions [13]. However, these analyses tend to be reported to other audiences, such as those in organizational studies, or in specific domains like healthcare. Unlike the original detailed studies of word processors and the like [4] or of mundane artefacts like paper documents in CSCW, there seem to be few studies that consider how everyday applications are used in everyday life. In this paper we consider details of how Skype is utilized to accomplish activities in everyday settings; how one particular technology is 'made home in the world'. Considering the detailed ways in which this is achieved in the light of recent efforts to enhance video-mediated communication, focusing on showing practices might suggest alternative ways in which video-mediated technologies can be enhanced.

DATA

For this study we recorded a corpus of interpersonal video-mediated conversations between family and friends. We recruited 14 primary participants who agreed to be involved in the study, and obtained the consent of about 30 of their Skype correspondents. Data were collected through Camtasia screen video capture software. Recordings were collected and made available to us by the participants themselves (who thus retained the right and capacity to keep some of their conversations private). The

signed agreement obliges us to anonymize the data, and in the examples we present below, where they are identifiable, faces have been blurred, and names and places changed. On occasions, we also had the opportunity to interview the participants to clarify, for example, background details of what was discussed in the calls.

The corpus the participants elected to make available involves more than 40 Skype conversationalists (about 1/3 male and 2/3 female) in over 180 naturally occurring conversations. In total this consists of 75 hours of recorded video conversations. Active Skype conversationalists in our corpus involved geographically separated couples or partners (4 cases), parents and adult children (5), siblings (3), and close friends (7). In this corpus, we noticed Skype conversationalists frequently bringing objects to the screen, on average this was about once every call. This is the phenomenon we consider in this paper.

Our analysis of these materials draws on conversation analysis and recent studies of the multi-modal interaction in everyday settings [12, 25, 34]. In this respect our concern is with the emergent and sequential character of practical action and the practices in and through which participants collaboratively accomplish particular activities, in this case showing one object to another through Skype.

We reviewed the corpus for instances where one participant showed something to the other. This could be done either by moving the device so that some detail became visible (about 50 cases), or by bringing some previously invisible or poorly visible object to the camera (90 instances). Moving the camera to show something is particularly relevant in the case of mobile terminals and/or cumbersome objects or fixed fixtures. We have focused here on showing practices, particularly characteristic of video communication on fixed computers and laptops, and suited to showing relatively light or movable objects. A great range of objects were shown, including parts of the body, clothes (worn or not), furniture and items related to interior design, multimedia devices, especially smartphones, cuddly toys, and objects related to current activities (documents that are being worked on, objects related to domestic chores, young children and family members, etc.) (See Figure 1).

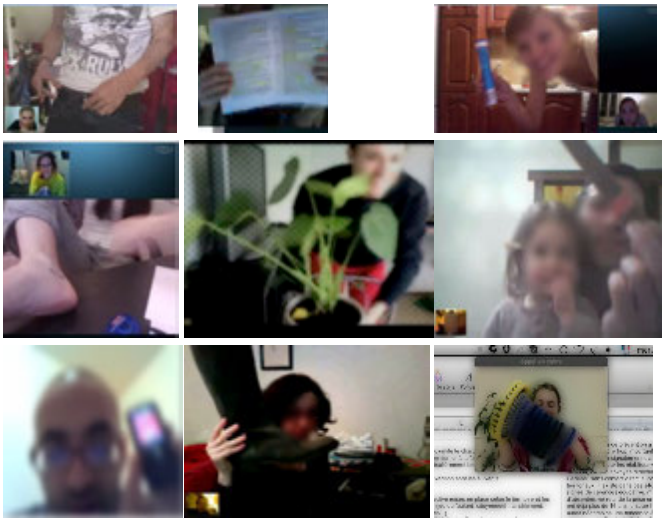
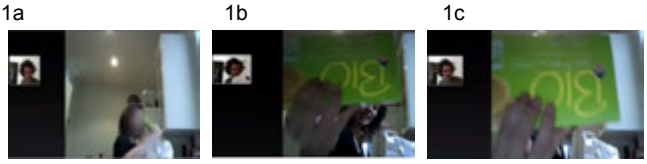


Figure 1. Examples of objects being shown in everyday Skype conversations. In order, a belt, a paper document, a tube, a tattoo on a foot, a plant, a cigar, phone, a boot and a scarf.

OBJECTS OCCASIONING TOPICS OF TALK

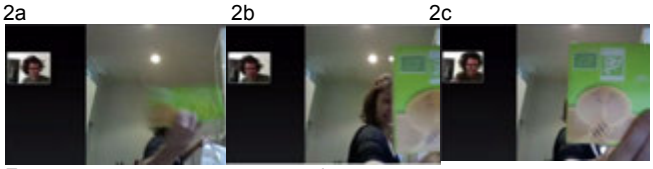
In many cases, showings were very brief, the object appearing on screen for a short moment of time. In the following fragment, Francine and her boyfriend, Marc, a recently formed couple living 300km apart, are engaged in one of their regular Skype conversations. At the start of the fragment, the link is open and the couple are involved in different activities: Marc is reading an article on his computer and Francine is starting to empty her shopping bag.

Fragment 1.1



Francine extracts the first item with her right hand and turning towards the screen, moves an object – a biscuit box – towards the camera (1a and 1b) so it almost fills the screen. She holds it briefly (0.2 seconds) so the large word "Bio" (organic) printed on the box is visible and readable by Marc (1c).

Fragment 1.2



F: regarde
look
M: hé hé

Francine then starts to move the box to her left (2a and 2b). She can now see the feedback image and be seen by Marc. She says "regarde" ("look"), moves the box back towards

the right so the picture of a biscuit on the packaging can be seen (2c). Marc responds with a short laugh and a smile.

This a typical example of an object being shown through Skype. There is a striking economy in the fragment in the way in which a showing is accomplished. Although it is brief, the entire fragment lasts no longer than 3 seconds, it is a recognizable showing. Francine does not seek to identify or refer to any particular feature of the object. The activity is not prefaced or introduced by any foregoing talk. Indeed, there is minimal vocal conduct in the fragment.

Francine does accomplish her showing in a particular way. By moving the box across the screen she successively makes different features relevant to Marc, first the word 'Bio' (the product is organic) and the image of the biscuit (a kind of waffle). When the box appears on the screen, Marc does look at it. Francine's subsequent request to look, then, is not just designed to secure the attention of Marc, as he is currently looking at the screen, but for him to look in a particular way (and to make it apparent to Francine that he sees it in that way). Marc responds with a laugh ("he he") and a smile, which displays he has recognized the object and its relevance. She hears it that way, and as adequate, and subsequently moves the box away from the screen and returns to her unpacking. The participants go on to talk about cakes and then about the next time they will see each other.

Francine can assess Marc's response to the object. As is common throughout much of our corpus when a participant holds the object, it is done in such a way that they can view the feedback image to both monitor the engagement of their co-participant and assess the visual consequences of their actions. They also adopt a position where they can display their own disposition to the object. As Francine moves the object to the screen she can be seen as smiling (1b), a smile that is still visible when she asks Marc to look (2b). Francine's conduct not only demands a response, but demands a kind of response.

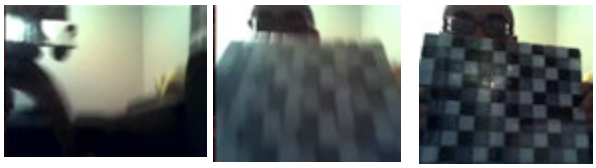
In interviews concerning the fragment with the participants it became apparent that at the time of the Skype call Marc had been trying to convince Francine to eat more organic food. Moreover, 'gaufres au miel' ('honey waffles') were his favorites The design of Francine's showing, first by presenting the word 'Bio' and then shifting the box to reveal the image of the biscuit, reflects these two aspects of mutual concern. This brief showing of a biscuit box served then to occasion a moment of particular intimacy to the participants and occasioned the participants to talk about the next time they will meet.

Showing an object, therefore, is more than just holding up an object to be seen. The activity emerges progressively in the light of the conduct of the co-participant, the person being shown the object. For this to be accomplished the person showing the object needs to monitor the conduct of the other and hence needs to be positioned with respect to the object and their own screen. They also can display their own disposition towards the object, and guide the recipient to how the object should be seen and what response could be appropriate.

CONFIGURING THE OBJECT AND SHOWER

On many occasions showings are part of quite extended discussions about objects and their qualities. In the following Skype conversation, friends Ben and Jerry have been discussing the present state of their finances and how it may affect their plans for a holiday. Ben then goes off-camera saying, whilst out of sight, "j'ai craqué ça y est faut que j'te l'montre" ("I've cracked, I must show you"). The transcript starts just before he returns.

Fragment 2.1



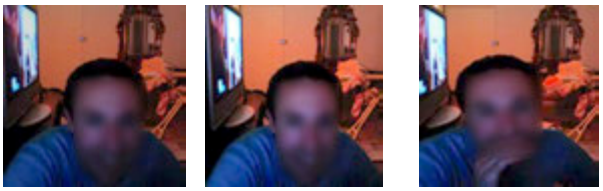
B: 'huh 'j-j j'ai acheté du (skarl) la faience euh'

'huh I-I have bought some (skarl) some earthenware

J: ooohh (.) ooh j'suis fan hh (hh)

I'm a fan

3.1a 3.1b 3.1c



Ben explains that he has bought some tiles ("jj j'ai acheté du (skarl) la faience euh") and brings a tile towards the camera, holding it with both hands. Jerry leans forward and smiles (3.2a). Ben then readjusts the tile, tilting the top forward. When the image of the tile becomes stable, Jerry moves his right hand to his mouth and exclaims 'ooohh (.) ooh j'suis fan hh (hh)' ('I'm a fan', 3.1c). This response does not seem to be sufficient and Ben pursues a stronger and more specific appreciation.

Fragment 2.2

3.2a 3.2b 3.2c

B: mais Jerry en fait ce n'est pas opaque hein ce sont des carrés translucides

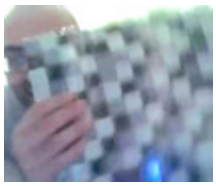
but Jerry actually it's not opaque uh these are transluscent squares

Ben moves the tiles aside and his head forward (3.2b) saying, "Mais Jerry en fait ce n'est pas opaque hein ce sont des carrés translucides ("but Jerry actually it's not

opaque uh these are translucent squares")', and pointing at the squares (3.2c).

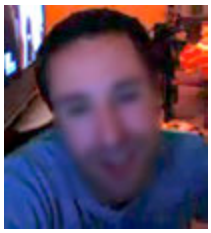
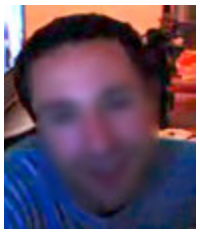
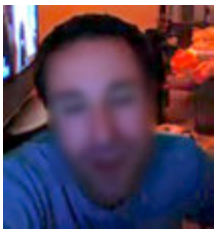
Jerry responds again (Fragment 2.3 below): "de verre translucide oh là là c'est TROP beau j'aDDOre rooh" ("transluscent glass oh la la er it's TOO beautiful I LOve it").He also nods in agreement whilst noting aspects of the quality of the tiles ("translucide"), moves and tilts his head to slightly change his viewing angle, saying 'oh la la' and making various facial displays that reflect his admiration of the tile (3.3-5). Having secured a stronger and more appreciative response from Jerry, Ben then moves on to discuss the price of the tiles, all the while maintaining the tiles in view.

Fragment 2.3




J: de verre translucide oh la la euh c'est TROP beau j'ADDOre rooh

transluscent glass oh la la er it's too beautiful I love it

3.3 3.4 3.5

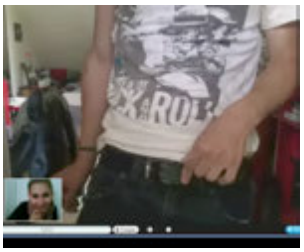
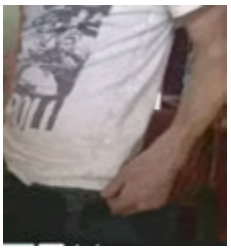
Unlike the previous fragment, the recipient does not seem to have seen the object before. Ben, through his prefatory and accompanying talk reveals aspects of the object being shown, but by showing the object in the way he does he seems to be pursuing a particular kind of response, about one visible quality of the object. Indeed, he repositions his head and body, bringing his head forward, allowing him to better monitor not just what his co-participant says about the object but how his co-participant is visibly orienting to it. By positioning the tile alongside his face he can gauge the moment-to-moment response of his co-participant. Moreover, having his head visible and the object slightly sideways allows him to co-ordinate his talk and visual conduct for his colleague. He can both produce talk and visual conduct reflecting his own disposition towards the object as well as monitor the visual (facial) responses of his colleague.

This configuration of the shower, the recipient and the object is quite common in our corpus of interactions through Skype. Given they are typically undertaken through a small screen and camera, participants showing objects frequently position themselves so they are peering over or around the object in question, with the object between themselves and the viewer. This helps to balance between several different requirements: a) to make the shown object as clear and visible as possible; b) to be able to display one's own disposition or attitude to an object through facial displays; c) to be able to reconfigure easily the position of the head and the object to produce different articulations of talk and visual displays in the course of showing activities and; to see the screen and the feedback image, (d) to monitor both how one's own conduct might be seen in the control image and e) to monitor the co-participant's response. Such reconfigurations are critical interactional resources allowing participants to move smoothly between focusing on viewing the object and focusing on talking about the object being made visible.

INFORMATIVE SHOWINGS

In some cases this configuration cannot be adopted, either because of the size of the object or its location. In the following fragment, Guy has just mentioned to his girlfriend a belt he has bought. After he says he is "trop content d'la ceinture" ("so happy with the belt"), Christine asks to see it ('vas-y montre' – 'go, show'). Guy happens to be wearing the belt and so stands up to show Christine. He shows the object in a number of ways. First, after standing up and repositioning the camera, he shows the belt face on. Following a question about the color and the buckle he pivots his hips, moving his left hand to grip the end of the belt, and touches various parts of the belt around the buckle and clasp saying it is "argent " ("silver").The transcript starts after another question from Christine.

Fragment 3.1

G: un peu métallique (1.5)

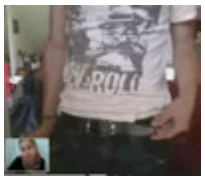
a little metallic (1.5)

C: okay (.) elle est très belle

okay (.) it's very nice

Guy answers "un peu métallique" ("a little metallic"), and as Christine gives a general appraisal of the belt ("elle est très belle", "it's very nice"), Guy swivels his hips again. After this, Christine asks Guy again about the colour of the belt, "et elle est marron hein la ceinture (.) foncée" ("and it's brown uh the belt dark ?").

Fragment 3.2

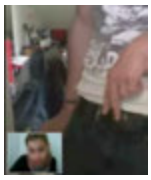


G: non (.) là elle est noire?

no (.) there it's black?

G: et (.) si tu changes de côté elle est marron

and (.) if you change the side it is brown



C: ah bon?

really?

Guy replies that is both black and brown accompanying this by bringing his fingers behind the tip of the belt and turning the end towards the camera, as he says "elle est noire" ("it's black"). He then spins his hand around and using two fingers flips the end of the belt around, saying "et brun" ("and brown"). Guy twists his hips again, which engenders a response from Christine "eh bon" ("oh good"), with rising intonation and whilst raising her eyebrows. She asks yet another question about the color. In response, Guy reasserts the blackness of the belt, and reveals a distinctive feature of it: it is both brown and black and it is reversible.

Guy shows various features of the belt. These are closely co-ordinated with the details he is describing. He not only points to details such as where there are different colors and different materials, but through his handling of the object, the texture of the materials, the flexibility of the object and distinctive features of the belt. This is all accompanied by bodily movements. As requested by Christine, Guy puts on a show that reveals important aspects and qualities of the object as worn on the body. Throughout this mini-performance Christine remains focused on the belt, inspecting the features as they are shown, making minor adjustments to her orientation as different aspects are revealed, orientations which are displayed in her assessments of the object. The way the participants 'view' the belt is a collaborative accomplishment, it evolves with the participants' changing embodied conduct with respect to the object, and successive assessments of it. This particular performance also raises further topics for comment, for example how the belt works, leading to another display of the belt where Guy goes on to describe the buckle, undoing the belt and taking the buckle off. Although Guy adopts a different orientation for the showing he can still monitor Christine. However, in this orientation he has fewer resources to display his disposition towards the object. Nevertheless, he does find a novel way to do this.

SHOWING OBJECTS IN SKYPE CONVERSATIONS

Throughout the corpus of Skype video-conversations there are frequent examples where one participant shows an object to another. These 'showings' are not merely referential activities accomplished so that a participant can locate, identify and direct attention towards an object (cf. 17, 28, 30). They are produced to display qualities and reflect attitudes a participant has towards an object. In order to accomplish these activities participants adopt an ecology of action so they and the objects they are showing can be seen in particular ways and they can monitor the moment-by-moment responses of the recipients.

There are also some practical consequences of the conventional ways in which participants engage in Skype conversations. When using a laptop (or desktop) system, as participants need to be able to view their screen when showing objects, they tend to adopt a configuration where the object is placed between them and the recipient and to one side or below their face. Although on occasions the participants do move the device, for example, to move the laptop or tip the laptop screen so the inbuilt camera can be used to show a large immoveable object, the device tends to be kept in a fixed position in a fixed location. This enables them to handle and manipulate objects with both hands. However, rather than adopting a variety of ways of orienting themselves to the object and their colleagues, participants tend to adopt a standard orientation and one where they conduct the conversation in a rather awkward position, bringing the objects to the screen when low to the ground, for example, or crouching or kneeling so they can monitor the other's response whilst they perform the showing.

MANEUVRABLE, MULTI-PARTY SHOWINGS

These different findings led us to reflect and reconsider what was happening in quite different kinds of video-mediated interactions. These data come from a quasi-naturalistic experiment where the participants' interaction was mediated through a maneuverable and more flexible device: the Kubi® TelePresence Robot (a movable robotic tablet stand [37]). This device allows remote participants more control over their own viewpoint but also give local participants a sense of changes of orientation of the remote participant, for example through when and how the 'robot' moves.

In the experiment, participants frequently showed objects to one another, but the technology transformed the resources participants used to establish viewing an object in common. By moving and reorienting the Kubi the remote co-participant can establish a mutual alignment and in turn provides a way for a local participant to assess where the remote co-participant is looking.

In this experiment the participants were engaged in an experimental task with the Kubis. These involved 4 participants, 2 in a local site and 2 in different remote sites. The remote participants interacted through two Kubis. The participants were not engaged in a domestic activity, rather they were undertaking a task that took around an hour where they had been asked to design an art exhibition. To assist them they were given a variety of materials,

including paper copies of the paintings they needed to select as pictures for the exhibition and documents giving details about the paintings. The participants had different responsibilities (for design, marketing and external relations) and different collections of materials, and needed to develop a detailed proposal for what to include in the exhibition and a justification for that choice. We recorded 6 groups of 4 people undertaking the experiment (using 5 video cameras). In reviewing the materials it was apparent that the participants routinely and frequently showed pictures and other documents to their colleagues. Local participants would show documents to one or both of their local colleagues and remote participants would also hold up objects to their laptops for the local participants (see Figure 2). Moreover, despite being engaged in quite a different activity the organization of these showings bear striking similarities to those in the naturally occurring Skype conversations: they have a similar preface-handling-assessment structure and participants configure themselves and their ecology so they show an object whilst monitoring their colleagues and assessing how their own image is being seen by the other. These showings do not function just to identify a painting or refer to a feature. Because in the course of their collaborative activity participants need to discuss details of an image, and the reasons why it might be important to include and juxtapose it with others, they show to their colleagues what it looks like and discuss its qualities.

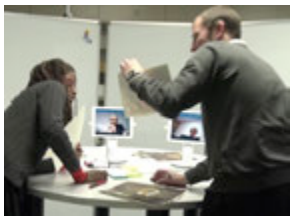
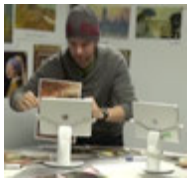


Figure 2. A participant showing a picture of a painting to another local colleague and two remote parties. The remote participants interact through Skype displayed on Apple iPads secured within the Kubi Stand and control the movement of the Kubi through a simple touch interface.

In the following fragment typical of the corpus, Pat is looking through some pictures of paintings and he holds them up to the remote participants: Sheryl (who is operating the Kubi to his right) and Tom (who is operating the Kubi on his left).

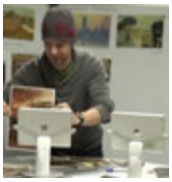
Fragment 4

4.1 Pat



P: what is your artistic opinion on

4.2



this …

4.3

P: he's got a British shield

In the course of making several suggestions for Old Masters paintings to be included in the exhibition, Pat holds up one painting of 'St George slaying a Dragon' to Sheryl and asks – 'what is your artistic opinion on this – on this particular painting?'. He holds it up to the Kubi and slightly to his right (4.1). In this orientation he can monitor Sheryl and also the control image displayed on the Kubi (4.2). As they discuss the painting, Tom the other remote participant, whose Kubi has been slightly turned away, shifts his Kubi towards Pat. Pat then moves the picture, holding it in front of Tom's Kubi (4.3).

Pat's shift to Tom, seems to be sensitive and responsive to Tom's movement of his robot proxy. Hence, the movement of the remotely controlled screen seems to be taken as a display of visual recipiency, reflecting an orientation of the remote participant towards 'viewing' something and serves to engender a sequence where the objects is shown. After the image of the painting appears on his screen, Tom responds, mentioning that he's seen 'that one before.'

With Kubi, participants draw on the visible motion of the maneuverable screens for cues that participants are available and a potential recipient. Moreover, the Kubi's offer some flexibility to how objects can be shown remotely; participants can show objects to two colleagues in distinct remote settings at the same time, as well as to co-present co-participants. Remote participants can also show objects to their colleagues, even to the other remote participant who also uses a Kubi. Participants can also show objects in other ways, sometimes two at once, either at the same time or successively to the remote participants.

There are also more subtle ways in which the Kubi supported participants to show objects to their colleagues. They could adopt different orientations towards the object and their colleagues; holding objects more at a distance from themselves and so more easily jointly inspect an object with a remote colleague (as in Figure 3). Moreover, by being maneuvrable participants could draw on even quite small movements of the Kubi to assess how their remote colleagues were viewing the object and in the light of these shape how they showed the object in question.

ASSESSING THE VIEWPOINT OF ANOTHER

Showing objects through the Kubi devices was not entirely unproblematic. As in the following fragment, where the participants are all looking through candidate pictures for the exhibition.

Fragment 5

Mike **Sarah**

E: definitely need Monet

Van Go

Errr:

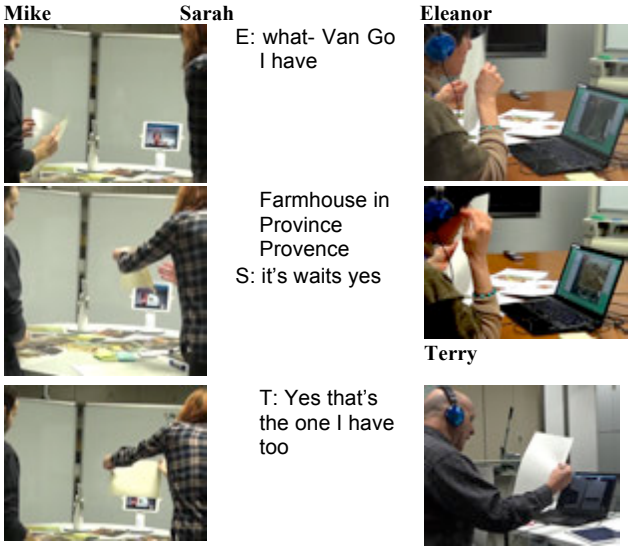
rrr:

Eleanor, who is one of the remote parties, suggests that for the 'later period' they should select a painting by Monet and Van Gogh. As she says this Sarah, in the local environment holds up a picture of a painting by one of the artists mentioned: 'Farmhouse in Provence' by Van Gogh. At this moment the two Kubi's are positioned at an angle to each other (Eleanor's Kubi is on the right and Terry's on the left, at the bottom of the images). Sarah does not preface or accompany this with any talk.

As Eleanor mentions 'Van Go' Sarah turns the picture towards Eleanor's Kubi. However, at this point Eleanor has turned away from her monitor and is searching for relevant pictures on her desk. Sarah then shows the image to the other remote participant, to Terry's Kubi, but Terry is also turned away and looking through his documents. Her co-present colleague Mike standing to her left does lean forward towards the picture, and in response Sarah turns it towards him. Despite holding the picture up for several seconds and changing its orientation, Sarah's showing fails to engender any realignment towards the object let alone secure a response from the remote participants.

Later Eleanor finds a document containing an image of the same painting and shows this to the local participants. Here her showing is prefaced with an announcement.

Fragment 6



Mike and Sarah have been looking at other pictures. Following Eleanor's prefatory talk and simultaneous showing, they turn to her Kubi, Sarah picking up her picture once again and showing it to Eleanor. As she does this she moves close so the picture almost fills Eleanor's screen. Meanwhile, Terry, the other remote participant also points out he has a copy of that picture and he also holds this up to his laptop. The participants then agree to select this, offering reasons why it should be included.

In this fragment three participants show an object to their co-participants, one of them making two distinct attempts to do this. They all hold the objects up for a sustained period. And yet, only one of these showings, Eleanor's - which is prefaced with a description of what is to be shown - engenders a re-orientation from a colleague in a remote environment. The others do not secure a realignment from a colleague. Indeed, it is unclear that Terry's can even be seen by any of the co-participants.

In these experiments with Kubi the video communication system they were interacting through was fairly conventional. It was a Skype connection with feedback images of the standard size and location. The tasks the participants engaged in involved them identifying and discussing the details of paintings. Therefore, perhaps it is not surprising that they frequently showed pictures to their co-participants, whether these were through a laptop (by the remote participants) or to a tablet secured on a proxy stand (by the local participants). The maneuverability of the Kubi allows remote participants to control their access to another domain, and yet for the co-participants it can be problematic for them to design their conduct for this shifting environment. The movements of the device can be seen as a display of responsiveness, but a local participant can find it hard to design their conduct to elicit such a display. The showing either not eliciting any kind of response or requiring additional talk to accompany it. Augmenting conventional Skype by introducing more maneuverable displays does provide a more flexible environment for showing but this also means the environment can become unstable. It can make it harder at particular times for participants to design their own conduct and assess how that conduct appears to another.

DISCUSSION

Previous studies of video-mediated communication have emphasized the need to shift away from a focus just on 'talking heads' and towards supporting the ways objects of different kinds can be integrated within the interaction [16, 19, 31]. Proposals for enhancing video-mediated communication in this way have tended to be concerned with supporting referential activities [7, 29, 30], and hence have focused on deictic talk and visual conduct such as when one participant points at an object to locate, identify and direct another's attention towards that object or a feature of that object.

Throughout the corpus of Skype video-conversations there are frequent examples where one participant shows an object to another. These 'showings' are not merely referential activities. They are produced to display qualities and reflect attitudes a participant has towards an object. In order to accomplish these activities participants adopt an ecology of action so they and the objects they are showing can be seen in particular ways and they can monitor the moment-by-moment responses of the recipients. For the recipient, the quality and size of the

image does, on occasions, mean that details or aspects of an object can be hard to see, and on some screens it can be hard to position the object for the recipient. The feedback image is thus critical for positioning the object so that it can be shown appropriately. Hence, to achieve a joint perspective on the object often involves minor adjustments of the bodies, devices and 'showables' to 'pursue' a proper response of a recipient in order for them to 'see' the thing in the right way.

These kinds of concerns might suggest a different set of requirements for enhancing video-mediated collaboration than for referring to objects. Participants need to be able to:

i) manipulate an object in different ways whilst showing it to another;
(ii) monitor their recipient whilst showing the object;
(iii) show their own 'disposition' towards an object alongside the object in question and for this disposition to be seen;
(iv) integrate the showing into a course of other activities, particularly talk about an object (and alongside other activities);
(v) move smoothly into a showing activity and into other subsequent activities.

Moreover, participants need to produce showings that are sensitive to the relationships between the participants, particularly to provide senses of intimacy though the display of objects. The current devices and applications for video conversations do constrain the ways participants can show an object to another. For showings the objects have to appear 'on cue', either prefaced or accompanied by talk about what is being shown and how it should be seen. Participants also can adopt rather awkward positions when showing objects. These conversations may be constrained in other ways. Although in a number of the cases we observed additional family members and friends, particularly children, being involved, most examples of showings were one-to-one. This may be due to the way data were collected, but observations by other researchers do suggest some constraints of conventional video conferencing packages for family interactions [1]. Indeed, Ames at al suggest the introduction of dedicated spaces into the home, spaces less focused on one-to-one communication and designed specifically to create a stable, shared virtual space for family interaction. Another approach may be to provide participants with more maneuverable devices, like the Kubi, so they can be used in a variety of less-constrained ways that meet the demands of the moment and also provide better access to objects to multiple participants. The experiments with Kubi suggest some challenges for providing greater access to these kinds of showings and making objects available to multiple participants in different locations.

There are some obvious enhancements that could be made to technologies such as Kubi. At present the robots are controlled through a simple graphical interface, where the user selects a location on a two dimensional grid to select how they want Kubi to be positioned (i.e. an absolute referencing system). In our experiments participants had difficulties at times, mapping the movements they required of the Kubi onto the actions they would need to perform on the interface, as there are no means of tying the movements with the environment being viewed. Better cues could be provided about how to move the Kubi in an appropriate way. We have experimented with using an additional overhead camera, for example, and then augmenting the existing controls onto the image from this. In this way, participants can select an object on the screen and the device will move towards that object. Furthermore, if we use a depth camera for the overhead camera, we should be able to adjust the tilting of the remote device.

There may also be ways of drawing from the analysis of showings in Skype conversations to provide additional support to participants. In a similar fashion to that proposed in the 'ultra-realistic telework system' where the system identifies objects that are being pointed at, on screens and then displays these in greater fidelity [35]. Given the conventional ways in which showings are performed, it may be possible to identify showings and automatically detect the object being shown. The image of the object could then be displayed in higher fidelity or even manipulated in some way, allowing the recipient to view features of an object not being shown by their colleagues. Indeed, in another prototype we have successfully used Intel's RealSense system to identify objects that are being shown (as these are held at close range to a camera) and then display a more detailed image of these on the video screen.

These technological enhancements may support showings in a number of ways, providing stable images, for example, allowing objects to be inspected in greater detail and also freeing the hands of the shower. They may also introduce challenges in how they are deployed. It may be difficult, for example, to undertake the kinds of manipulations of objects we found in our corpus or so easily co-ordinate other aspects of talk and visual conduct with a static image of an object, even if this is reproduced in greater fidelity.

Figure 4. Showing an object through the OmiEyeBall System [26]. (by kind permission of Prof Koike of Tokyo Institute of Technology).

Another approach would be to provide a greater field of view. For example, it may be that innovative devices like the OmniEyeball developed by Li et al. [26] might both allow participants to show objects in a variety of ways (see Figure 4). Such a device, by using 360º cameras and projecting the image onto a spherical screen could preserve the relationship between the object and the person showing the object. It also is a device that might not be too intrusive in a domestic setting. As the device resembles a crystal ball, and is fixed, showing and being shown objects should not require additional activities from the participants. Moreover, being fixed whilst offering a wide field of view, what is being shown and the location in which it is being seen should be easy to anticipate. It would also seem to be suitable as way of mediating the kind of intimate interactions we witnessed in our Skype data. Of course, both cameras and screen distort the image and it would be important to investigate how this may affect how objects are shown and seen through the device.

When early attempts to support communication through video technologies were developed it was noteworthy the limited impact these had [14]. The introduction of video telephones was tried on several occasions and none reached widescale deployment [5] and media spaces were not taken up beyond the research laboratories in which they were designed. Video conferencing has become more common, but it did tend to be used to support particular forms of business communication [6].

None of these technologies have had the success of the recent internet services such as Google HangOuts, FaceTime and most notably Skype. No doubt the fact that these are freely available and easy to install contribute to their widespread adoption, but these systems typically have fewer capabilities than earlier systems and offer communication of poorer fidelity with greater potential for technical disruptions. It is perhaps curious then that this success has not been a topic of investigation within fields associated with Human-Computer Interaction. Although there have been a small number of ethnographic studies of the use of Skype within the home, there is little in the way of detailed studies of the moment-by-moment use of these kinds of technologies. Our corpus of everyday conversations through video suggests that detailed studies of how they are accomplished and the contingencies and problems participants have to address will reveal findings that are relevant to both understanding certain characteristics of mediated interaction and contributing to the further design and development of technologies to enrich remote communication. The analysis of how objects are shown in a natural setting also enabled us to reflect on and reconsider how the activity was being accomplished when mediated by another kind of technology. In quite a different corpus of data it was noticeable that showings were similarly pervasive and had a similar sequential organization, and yet ways in which this activity might be transformed when the mediating technology had slightly different capabilities. Together this analysis suggested requirements for video-mediated technologies that support interaction with objects, that offer more than just being able to identify and refer to an object or feature of an object: that makes accessible to remote participants how an object his held, maneuvered and manipulated.

In this paper we have focused on how people show objects to one another. In contrast to previous research that has primarily focused on how objects within the respective environments are referred to and identified, in this program of work we are more concerned with the ways in which objects are selectively grasped, manipulated and revealed to the co-participants; objects that might otherwise pass unnoticed. As the data suggests, many of these objects are not immediately accessible and to be 'topicalized' and become relevant 'showables' they need to be maneuvered into a position where they can jointly be 'viewed' and appreciated. Detailed consideration of these forms of showings suggest that there are conventional and systematic ways in which objects are exposed and revealed in video-mediated interaction mediated through systems like Skype. Consideration of these practices and the ways they are configured suggest some future directions for technological development, developments that do not necessarily correspond to our more general understanding of referential practice in social interaction. Indeed, the showing of an object reveals contingencies and complexities not commonly found within pointing and reference, requiring an object to be exposed, guided and manipulated – to enable the co-participant to see and experience its emergent and contingent properties within talk and embodied action.

A common feature of the objects addressed in this paper, is that they are recognizable as relevant to personal territories, the "territories of the self", as Goffman termed them [11]. It is precisely because such objects are personal that showing them constitutes such an important resource for accomplishing a sense of closeness or "intimacy at a distance" [1]. Skype communication with family and friends makes relevant the sharing of intimate materialities and through the occasioned revelation of familiar, in some cases, evocative objects, a sense of our relationship, our closeness, our intimacies, is both accomplished and reproduced. In one way, objects such as belts, bathroom tiles, cuddly toys, and the like may seem of little importance in world and yet they are *par excellence* the stuff of our ordinary lives and intimate relationships with others.

ACKNOWLEDGMENTS

We are very grateful to the participants in the two studies for their help and support. We thank them for allowing us to analyse and present the data. We would like to thank members of the NTT t-Room team and for their help and support in carrying out the Kubi study.

REFERENCES

1. Ames, M.G., Go, J., Kaye, J., & Spazojevic, M. (2010). Making Love in the Network Closet: The Benefits and Work of Family VideoChat. In Proc. of CSCW 2010, 145-154.
2. Brown, B., & Laurier, E. (2012). The Normal, Natural Troubles of Driving with GPS. In Proc. of CHI 2012, 1621-1630.
3. Brubaker, J., Venolia, G. & Tang, J. (2012). Focusing on Shared Experiences: Moving beyond the camera in video communication. In Proc. DIS 2012, 96-105.
4. Card, S. K., Moran, T. and Newell, A. (1983). The Psychology of Human-Computer Interaction. Hillsdale, NJ.: Lawrence Erlbaum Associates.
5. De Fornel, Michel, 1994. Le cadre interactionnel de l'échange visiophonique. Réseaux 64, 107--132.
6. Egido, C. (1990). Teleconferencing as a technology to support cooperative work: Its possibilities and limitations. In R. E. Kraut, J. Galegher, & C. Egido (Eds.), Intellectual Teamwork: Social and Technological Foundations of Cooperative Work (pp. 351-373). New Jersey: Lawrence Erlbaum Associates.
7. Fraser, M., Glover, T., Vaghi, I., Benford, S., Greenhalgh, C., Hindmarsh, J. & Heath, C. 2000. Revealing the Reality of Collaborative Virtual Reality. In *Proceedings of ACM Conference on Collaborative Virtual Environments (CVE 2000)*. ACM, San Francisco, CA, 29-37.
8. Gauglitz, S., Lee, C., Turk, M., & Hollerer, T. (2012). Integrating the Physical Environment into Mobile Remote Collaboration. In Proc. MobileHCI 2012, 241-250.
9. Gaver, W.W., G. Smets, and K. Overbeeke. (1995). A Virtual Window on media space. in *Proc of CHI '95*.: 257-264.
10. Gaver, W. W., Sellen, A. Heath, C., & Luff, P. (1993). One is not enough : multiple views in a media space. Proc. of CHI 1993, 335-341.
11. Goffman, E. (1971). Relations in Public. Microstructure of the Public Order. New York: Harper & Row.
12. Goodwin, C. (1981). Conversational Organisation: Interaction between a Speaker and Hearer. London: Academic Press.
13. Greatbatch, D., Heath, C., Campion, P. and Luff, P. (1995) 'How do desk-top computers affect the doctor-patient interaction?' Family Practice, Vol,12, No 1. 32-36.
14. Harrison, S. (Ed.). (2009). *Media Space 20 + Years of Mediated Life* London: Springer-Verlag.
15. Heath, C. C. , Hindmarsh, J., & Luff, P. (2010). Video in Qualitative Research: analyzing social interaction in everyday life. London: Sage.
16. Heath, C. & Luff, P. (1991). Disembodied Conduct: Communication through video in a multimedia office environment. In Proc of CHI 1991, 99-103.
17. Hindmarsh, J., Fraser, M., Heath, C. C., Benford, S., & Greenhalgh, C. (1998). *Fragmented Interaction: Establishing mutual orientation in virtual environments.* Paper presented at the CSCW'98, Seattle, WA.
18. Inkpen, K., Taylor, B. Junuzovic, S., Tang, G. & Venolia, G. (2013). . Experiences2Go: Sharing Kid's Activities Outside the Home with Remote Family Members. CSCW 2013,
19. Ishii, H. and Kobayashi, M. (1992). 'Clearface: a seamless medium for sharing drawing and conversation with eye contact', in Proc. of CHI 92, Monterey, 525-32.
20. Jones, B., Witcraft, A., Bateman, S., Neustaedter, C., & Tang, A. (2015) Mechanics of Camera Work in Mobile Video Collaboration. Proc. of CHI 2015.
21. Judge, T., Neustaedter, C., Harrison, S. & Blose, A. (2011). Family portals. Connecting families through a multifamily media space. In Proc. of CHI 2011, 1205-1214.
22. Kirk, D., Rodden, T. & Fraser, D.S.(2007). Turn It This Way: Grounding Collaborative Action with Remote Gestures. In Proc. CHI 2007, 1039-1048.
23. Kirk, D., Sellen, A. & Xiang, C. (2010). Home Video Communication. Mediating 'Closeness'. In Proc. of CSCW 2010, 135-144.
24. Knoblauch, H. (2013). PowerPoint, Communication, and the Knowledge Society. Hubert Knoblauch. New York: Cambridge University Press.
25. Kuzuoka, H., Yamashita, J., Yamazaki, K., & Yamazaki, A., A. (1999). *Agora: A Remote Collaboration System that Enables Mutual Monitoring.* Paper presented at the CHI'99 Extended Abstracts, Philadelphia, PA.
26. Li, Z,. Miyafuji, S., Sato, T. and H. Koike (2016 forthcoming). OmniEyeball: Spherical Display Embedded With Omnidirectional Camera Using Dynamic Spherical Mapping. Proc of IUST.
27. Licoppe, C., & Morel, J. (2009). The Collaborative Work of Producing Meaningful Shots in Mobile Video Telephony. In Proc. MobileHCI 2009, 254-263
28. Luff, P., Heath, C., Kuzuoka, H., Hindmarsh, J., Yamazaki, K., & Oyama, S. (2003). Fractured Ecologies: Creating Environments for Collaboration. Human-Computer Interaction 18, 51-84.
29. Luff, P., Heath, C., Kuzuoka, H., Yamazaki, K., & Yamashita, J. (2006). Handling Documents and Discriminating Objects in Hybrid Spaces Paper presented at the CHI 2006, Montreal.
30. Luff, P., Yamashita, N., Kuzuoka, H & Heath, C, (2011) Hands on Hitchcock: embodied reference to a moving scene, In *Proc. CHI 2011*, 43-52.
31. Nardi, B. A., Schwartz, H., Kuchinsky, A., Leichner, R., Whitakker, S. and Sclabassi, R. (1993). 'Turning Away from Talking Heads: The Use of Video-as-Data in Neurosurgery', in Proc. of INTERCHI'93, Amsterdam, 327-34.
32. Neustaedter, C., Yarosh, S. & Brush, A.S. (2009): Designing for families. In: Proceedings of ACM CHI 20092735-2738.
33. Neustaedter, C., Harrison, S. & Sellen, A.(2010): Connecting families: new technologies, family communication, and the impact on domestic life. Berlin: Springer.
34. Nevile, M., Haddington, P., Heinemann, T., & Rauniomaa, M. (Eds.). (2014). *Interacting with Objects: Language, materiality and social activity*. Amsterdam: John Benjamins Publishing Company.
35. NICT (2016) Ultra-realistic telework system. Oki, NEC, Sharp, Kyoto University, Tokyo University of Agriculture and Technology, and Rissho University
36. O'Hara, K., Black, A., & Lipson, M. (2006). Everyday Practices with Mobile Video Telephony. Proc. of CHI 2006, 871-880.
37. Revove Robotics (2015) https://www.revolverobotics.com/ (accessed 19th Sepetmber 2016)
38. Statista (2016) https://www.statista.com/statistics/ (accessed 19th September 2016)

39. Tang, J., Xiao, R., Venolia, G., & Roseway, A. (2013). HomeProxy: Exploring a Physical Proxy for Video Communication in the Home. In Proc. of CHI 2013, 1339-1342
40. Wong, N. & Gutwin, C. (2014). Support for deictic pointing in CVEs: still fragmented after all these years'. *Proc of CSCW 2014*, 1377-*1387*.
41. Yarosh, S., Inkpen, K., & Bernheim Brush, A.J. (2010). Video playdate: towards free play across distance. In Proc. CHI 2010, 1251-1260.