

Connecting Distributed Families: Camera Work for Three-Party Mobile Video Calls

Yumei Gan*, Christian Greiffenhagen*, and Stuart Reeves†

*Video Analysis, Science, and Technology (VAST) Research Group

Department of Sociology
The Chinese University of Hong Kong
Hong Kong SAR, China

yumei.gan@link.cuhk.edu.hk; c.grieffenhagen@cuhk.edu.hk

†Mixed Reality Lab

School of Computer Science
University of Nottingham
Nottingham, UK

stuart.reeves@nottingham.ac.uk

ABSTRACT

Mobile video calling technologies have become a critical link to connect distributed families. However, these technologies have been principally designed for video calling between two parties, whereas family video calls involve young children often comprise three parties, namely a co-present adult (a parent or grandparent) helping with the interaction between the child and another remote adult. We examine how manipulation of phone cameras and management of co-present children is used to stage parent-child interactions. We present results from a video-ethnographic study based on 40 video recordings of video calls between ‘left-behind’ children and their migrant parents in China. Our analysis reveals a key practice of ‘facilitation work’, performed by grandparents, as a crucial feature of three-party calls. Facilitation work offers a new concept for HCI’s broader conceptualisation of mobile video calling, suggesting revisions that design might take into consideration for triadic interactions in general.

Author Keywords

Mobile Video Calls; Distributed Families; Facilitation Work; Camera Work; Conversation Analysis

CSS Concepts

• Human-centered computing; Human computer interaction (HCI); Empirical studies in HCI

INTRODUCTION

Since the economic reform of 1978, China has seen massive rural-to-urban internal migration. Due in part to the strict residential registration (*hu kou*) system and other difficulties for migrant workers in moving their children with them [37, 38], some migrant workers may leave their children behind in rural areas. The shift has resulted in a large number—61 million by recent estimates [1]—of so-called ‘left-behind’ children (Chinese: *liu shou er tong*). Such children tend to be

brought up by adults other than their biological parents, often grandparents. Many of these fragmented, translocal families [13] come to conduct much of their relationship via regular mobile video calls, wherein co-present grandparents work to establish regular intimate video connections between the remote parents and their children. Such communication has been referred to elsewhere as “remote parenting” [12]. The literature has discussed the significance of video communication and their design to the fabric of intimacy in translocal families’ everyday relations [21].

In this paper we show how moments of intimacy and togetherness in video calls play out interactionally, and, most critically for the purposes of HCI, the ways that intimacy is practically accomplished—made to work—within the bounds of present mobile video technologies.

Certainly, there are many existing studies in HCI on the role of technologies (e.g. voice, video) for mediating relations between distributed families of one kind or another [2, 15, 23]. But these have mostly been located in Western contexts (although there are exceptions e.g. [4]), and tend to focus on calls between co-located parents and children and remote grandparents [14]. These studies also largely employ interviews, and thus lack details on the interactional specifics of conducting familial life via video [15].

Our study differs from this prior work in two key ways. Firstly, there is a relational difference, since we are examining cases where the remote parties are the *parents* (and not the grandparents). Secondly, and more importantly, we find that no prior studies have closely examined what we will call *facilitation work*, which is a key feature of *three-party calls* (e.g. parents-children-grandparents). We use this term to refer to the range of supportive activities performed by third parties (like grandparents) to establish and maintain interactions between a co-present and remote other (in our case child and parent respectively). While facilitation work is particularly prevalent for calls in which young children are involved (given their lack of established skills with cameras and the fact that they tend to move around during calls), it is not confined to the scenario we focus on for this paper.

The notion of facilitation work emerges from our video-ethnographic study of Chinese migrant workers’ family

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.

CHI 2020, April 25–30, 2020, Honolulu, HI, USA.

© 2020 Copyright is held by the owner/author(s). Publication rights licensed to ACM. ACM ISBN 978-1-4503-6708-0/20/04...\$15.00.

DOI: <https://doi.org/10.1145/3313831.3376704>

communication practices during video calling between (remote, urban-dwelling) migrant parents and their (rural, left-behind) children and co-present caregivers (i.e. grandparents in our case). The problem is that mobile video calling technology seems to largely be designed for one-to-one interaction as a participation framework, i.e. where it is relatively easy to configure the phone camera in a familiar “talking-heads” style arrangement [16]. But this is *not* the typical bodily arrangement in a three-party call, especially in cases involving young children. In such three-party calls a significant amount of work is done by copresent adults to enable young children’s engagement with remote parents. This is similar to what previous studies have described as “scaffolding for video calls” [2] or “scaffolding work” [29]. In our data, grandparents not only need to hold the phone, they also need to control the child, frame the child, and engage the child. We refer to these calls as *facilitated three-party calls*.

We found that a core aspect of facilitation work stems from ‘camera work’ practices in these calls, whereby smartphones—with both front (user-facing) and rear cameras—act as the mediating tool. As previous studies have stated, ‘showing’ is a key feature of video calling [19], and mobile phones offer even more opportunities for this than fixed cameras. However, phones also tend to introduce various problems that can hinder engagement due to the physical and software configuration of cameras and screens.

Our study focuses on those critical interactional moments whereby facilitation work is required, specifically where there is a need for the child to ‘do something’ in the course of interaction: e.g. to look at a parent, to respond to a greeting, or to say goodbye. Parent-child interaction is significant as it turns out. In all these cases, the use of the front camera is central. For this study, we do not examine a range of other moments e.g. where grandparents talk to parents directly (facing the camera) or points where grandparents switch the camera to show children (see [30]) to their parents. Instead, this study is restricted to moments of interaction within longer phone calls where facilitation work is required for the child to engage with their parent(s).

MEDIATING INTIMACY FOR DISTRIBUTED FAMILIES

We briefly review prior studies of parental migration in order to contextualize our setting for the reader. In doing so we discuss the role of technology in mediating relations between such families, and then turn to video and the ways it has been investigated in HCI and CSCW. As we do so we reveal key conceptual components that have been overlooked within these examinations of video-mediated distributed family relations and sketch the focal interests of our paper.

Parental migration and left-behind children

International and Chinese rural-to-urban internal parental migration has been the subject of considerable investigation within the migration literature. In this we find a diverse range of studies on “remote parenting”, “transnational motherhood” [12], “mothering from a distance” [24], or

“parenting from afar” [27]. These studies have detailed the various emotional struggles of migrant parents in trying to maintain connections with their children. Studies also show that the emotional trauma of separation leads to an increased need for communication from afar [21].

Previous research on Chinese left-behind children specifically has also explored the effects of parental migration on children’s health, education, and wellbeing. Although some have emphasized that parental migration provides economic benefits for children (e.g. [38]), most studies argue that family separation affects children negatively [33]. Children often feel lonely and isolated, face obstacles in studying, and may suffer from cognitive delays. The use of mobile video calling, then, can be seen as a collective family response to such problems amidst the broader economic and societal phenomenon of parental migration.

Video in facilitating distributed family relations

Various technologies including video have been employed by distributed migrant families for some time. Wilding, for example, traces how migrants initially communicated with their children through letters, postcards and tape recordings, and then started to use mobile phones, text messages, as well as multimedia such as digital photographs [32]. More recently, a growing number of studies have argued that more video-mediated communication technologies may be particularly important for enhancing intimacy-at-a-distance [34]. In particular, it has been suggested that the possibility of *seeing*, and *being seen*, can mediate a range of relations, even over great distances [14].

Video calls are thought to be advantageous for very young children [21], such as those found in our study. While some older children may prefer phone calls, young children (e.g. under five years old), due to their limited speech capabilities, often have difficulties in communicating with a remote parent using a telephone. Nevertheless, they are willing to communicate, many for long periods of time, using video [2, 3, 22, 35].

Set against this backdrop, HCI and CSCW have also spent time investigating various kinds of distributed communications between family members—for example, grandparents and children [7]—albeit with a view towards proposing the (re)design. This has also included broader explorations of the role of various network technologies to ‘raise’ left-behind children and “work-separated” children [4, 34]. Such investigations have pointed to the importance of video in such relations [34], and we find that *video* communication within families more generally has generated substantive interest (e.g. [2, 15, 23, 39]). In line with HCI’s interest in (re)design, our work includes exploration of novel forms of video-based interaction designs (particularly for children, e.g. [6, 35, 36]).

Three-party mobile video calls and facilitation work

Our research builds on the work mentioned above and makes four main contributions. Firstly, most studies have primarily employed interviews (e.g. [15, 23, 35]) rather than direct observation of the interaction in and around the video calls. We follow studies employing direct observation using both static video capture [2, 30] and screen recording [19] to record actual interactions, which allow us to explore the interactional work of video mediated encounters between family members.

Secondly, existing studies have mostly focused on how parents use video calls to maintain intimate relationships between their children and remote grandparents (e.g. [30]). In our case, it is the *parents* who are away from their children, and not just for a short period, but several years. As we have noted already, for parents of left-behind children in work-separated families, this leads to a greater emotional significance of the video calls due to this separation.

Thirdly, most previous research on video calls has focused on *stationary* cameras, e.g. a desktop or laptop computer [2, 15]. In contrast, we focus on *mobile* cameras, since all our calls were made on smartphones. The mobility of the camera introduces new challenges [20], especially since it is accompanied in our setting by the mobility of participants (young children tend to move around during calls).

Finally, a critical feature of video calling in our study is the aforementioned notion of facilitation work. Although facilitation work has been hinted at (e.g. [3]), it is generally an overlooked phenomenon that has significant implications for how HCI understands video calling in terms of family relations. This thus offers conceptual enrichment for the landscape of HCI's understanding of design for video.

STUDYING TRANSLOCAL FAMILIES IN CHINA

For this study we video recorded a corpus of habitual video calls in Chinese migrant families over two years. Fieldwork for video recordings was conducted in Sichuan and Guizhou provinces in the southwest of China, which are two of the largest rural labor export provinces in China [5]. Recruitment was challenging, since video-recording requires a greater time commitment as well as trust from participants, as compared with interviews. We recruited participants via the first author's personal contacts and then continued through snowball sampling.

We chose families where both the father and mother of the children were migrant workers, and where regular video calls were made. Consequently, all the families were familiar with the use of video-mediated communication technologies by the time we started recording. However, the 'technological competence' among the caregivers (always grandparents in this study) was mixed. Some grandparents could only use a smartphone in a basic way. All families used the Chinese instant messaging app WeChat on their smartphones to conduct video calls.

The video recordings were conducted on the children's side, i.e., in the rural areas. A researcher was always present in the field to set up the camera equipment before participants started their calls. Each video call was recorded, firstly, with a traditional camera recording the interaction in front of the mobile phone (e.g. [30]) and, secondly, through a screen capture of the grandparents' mobile phone (e.g. [19]). This method of recording made it possible to capture all three parties in the video communication: remote parents, grandparents and left-behind children. Both video streams were then synchronized for analysis (as in Figure 1).



Figure 1: Our approach to video capture

For this paper, we draw on 40 video calls made by 30 families. Video calls lasted between 7 and 65 minutes (averaging 26 minutes). In order to protect participants' anonymity, all names have been changed. We received approval for the project from the Chinese University of Hong Kong's Research Ethics Committee.

Children in this study are between eight months and three years old. In all cases they were accompanied by grandparents to conduct video calls, meaning that all calls involved three parties. We must note that in interaction, parties do not necessarily equal the numbers of people. As Schegloff argues [26], the organization of talk is distributed amongst parties, *not amongst persons*. So, we use the term three parties to refer to remote migrant parents, grandparents, and children, who play different roles in interactions. Thus, it is possible in the fragments we select that there are between three and (at most) six people (two parents, two grandparents, and two children), but we still treat them as three-party calls.

Our approach is informed by ethnomethodology [8] and conversation analysis [25], which aim to explore participants' own understanding and their own analysis of local situations. Our focus is on exhibiting the organised, methodical ways in which social action is ordered and achieved. In this case 'social action' pertains to the accomplishment of family relationships whether that be parent-child (of both generations) or grandparent-grandchild. Thus, the fragments selected are used to show how these relationships—as interactional phenomena—are

endogenously negotiated and produced, and the ways in which they arise from members' practices.

When working on our videos, we noticed that participants displayed problems with what we termed 'camera work', e.g. in framing children or objects. Our analysis for this paper is based on a broader set of 63 instances where camera work manifested as an issue in facilitated three-party calls.

VIDEO CALLS WITH VERY YOUNG CHILDREN

We begin by highlighting how the need for facilitation work emerges in three-party calls, particularly when young children are involved in the setting. In our data, the challenges can be easily seen from the fact that very young children are not always competent to hold the phone.

In Figure 2, a child (11 months old) takes the phone from her grandfather and holds it while they are on a video call with her mother (2a). The child then shakes the phone and drops it to the bed (2b), which prompts the grandfather to pick up the phone (2c). After a moment, the child's (remote) mum complains that "the phone is always shaking":

- 01 ((child holds and shakes the phone. Image 2a))
 02 ((child drops the phone to bed. Image 2b))
 03 ((grandfather grabs the phone back. Image 2c))
 04 MUM: 她在拿手机吗你在拿手机哦?
 is she holding the phone or are you holding?
 05 GRF: 她在拿 [手机]
 she is holding [the phone]
 06 MUM: [一个手机一直在晃
 [the phone is always shaking]



Figure 2: A child shaking and dropping the phone

Prior to the actual complaint, the mum asked the grandfather: "is she holding the phone or are you holding?" (line 04). This illustrates the difficulties for the remote party in rendering the grandparent-child scene intelligible, and accounting for events that are out of view of the camera, which includes the question of who is holding the phone.

More generally we find in our data that although some young children may be able to physically hold a mobile phone, actually doing so often results in abrupt movements or shaking of the device in ways that disrupt 'good' camera work. Children may also touch the phone's screen which may affect the call, e.g. abruptly ending video calls or switching apps accidentally. As shown in our cases, central importance to three-party video calls is remote parents' experience—what they can see and hear—and the reciprocal orientation of other parties to that. Very young children's lack of learned competence in manipulating the phone

creates complexity and challenges for such video calls. The importance of operating the mobile phone camera stably and properly thus becomes a crucial factor for conducting a 'successful' video call, and for mediating the parent-child relationship at a distance.

FACILITATED THREE-PARTY VIDEO CALLS

Grandparents holding the phone

As shown in Figure 3 below, in our data typically it was grandparents who held the phone. Figure 3's images were taken from the openings of video calls and show the configurational diversity of facilitated *three*-party video calls among migrant parents, children and grandparents (in contrast with *two*-party video calls between parents and children).



Figure 3: Grandparents holding the phone for facilitated three-party parent-child video calls

Figure 3 also points to the (interactional) significance of children being visible on-screen for remote parents (even at the very beginning of interaction), and the preference for accomplishing this as soon as possible on the part of grandparents. The ways that grandparents hold the phone often involves physically positioning children and the mobile phone camera such that the scene is configured appropriately for the viewing parent. Grandparents may physically hold the child and point the phone towards the both of them, enabling them to more closely monitor the scene such that it will be visible to remote parents (3a-3b, above). Alternatively, grandparents may adopt a position perpendicular to the child while hold the phone facing the child (3d-3f, above). We noticed a tendency for this latter approach for older children within the age range covered by our video recordings (i.e., those closer to 3 years old). (3c represents something in between these two approaches where one grandparent is visible while the other is perpendicular to the phone).

Grandparents preventing children grabbing the phone

In our data, children physically interacting with the phone tended to lead to trouble and disruption when establishing and maintaining a visual connection between remote parent

and child. In these cases, grandparents often anticipated and *prevented* children grabbing the phone. Our next fragment, shown in Figure 4, shows how this trouble can be projected by grandparents and resolved by them in a simple way.

This fragment begins during the ringing phase of a video call, summoning the remote parents. Whilst ringing, the grandmother here positions the phone in front of the child, so that when the parents appear on the screen, they will see their child immediately and vice versa (the orientation of those in the three-party call to the importance of the child being seen first or responding first is a key point we will return to below). To begin with, the phone is positioned close to the child; the child then moves his hand to grasp the phone (4a). After this, the child walks toward the phone (4b), whereupon, as we can see clearly in (4c), the grandmother moves the phone away to restrict physical access:



Figure 4: Child attempting to grab the phone

An important way of protecting the visual connection between parent and child is thus for the co-present grandparent to appropriately anticipate and manage *physical* access to the phone. But restricting physical access has its tensions. Facilitating the connection between parent and child requires rendering the local scene intelligible to the remote parent (e.g. positioning the child visibly on screen in an appropriate framing, and accounting for problems), yet at the same time facilitation work involves monitoring, anticipating, and avoiding physical access to the device by the child. This latter activity not only impacts but sometimes works *against* appropriate camera work to ensure the connection between parent and child is delivered successfully for both.

Grandparents doing facilitation work

We consider facilitation work to be the key job of grandparents as members of the setting. They facilitate what the children are trying to do (e.g. when the child is trying to show something to the remote parent), but equally they facilitate what the parents are trying to do (e.g. get a response from the child, where the grandparent may repeat the initial request by the parent or prompt the child in a new way).

The two examples below offer further insights that show how facilitation work comes to encompass a broad range of interactional resources.

Grandparents positioning the phone & prompting the child

In the fragment in Figure 5, we show how positioning may be interactionally accomplished as a matter of grandparents’

facilitation work. This involves careful coordination of bodily and verbal resources in a timely and sequentially organised way such that they set the stage for—that is, *configure*—a moment of interaction (such as a greeting in this case) between mother and child.

In this fragment, the child is only nine months old. Her grandfather initiates the video call to her mother. Initially the phone camera acts as a ‘mirror’ during the ringing phase, showing the grandfather himself (5a). At this point the grandfather moves the phone from himself to the direction of the child, meaning that the phone camera and its display now shows the child (5b). He also verbally prompts the child to “take a look (at this)” (line 03) before the child’s mother actually appears on the phone. This amount of work leads to a mother-child ‘talking heads’ configuration [16] when the video connection is established (5c).

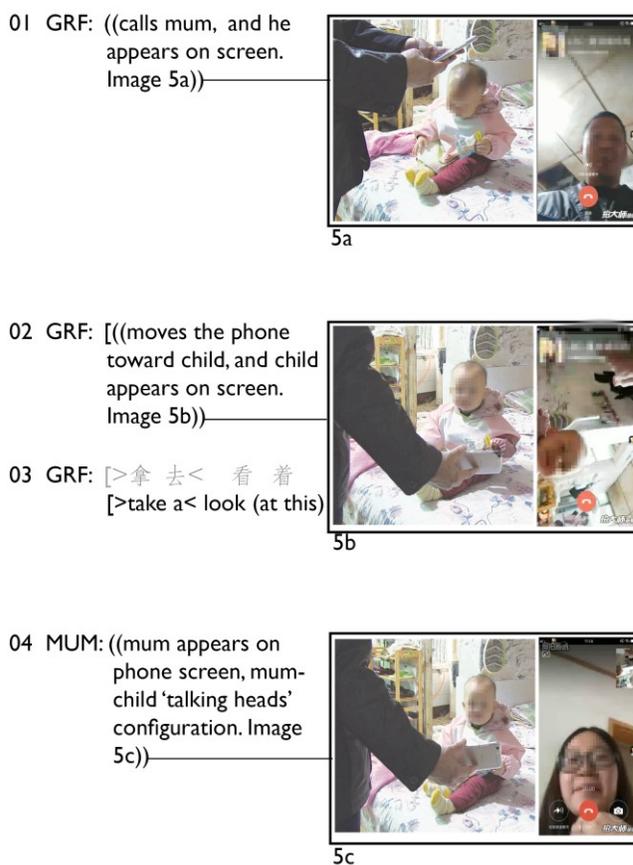
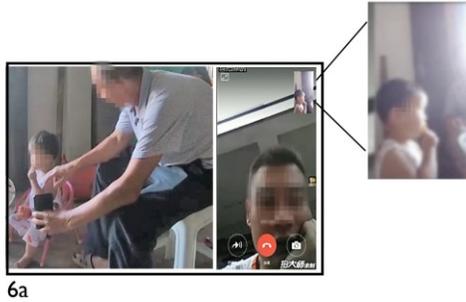


Figure 5: Grandparent positioning phone & prompting child

In this example, the grandparent establishes an intelligibility to the scene by staging the visual position of the child in the shot via framing (5b). This is done so that the mother’s opening view is that of her child. In the course of this careful framing, the grandfather also uses a verbal prompt (“take a look at this”) to solicit the child’s gaze towards the screen in preparation for the anticipated (and subsequent) appearance of the mother’s face.

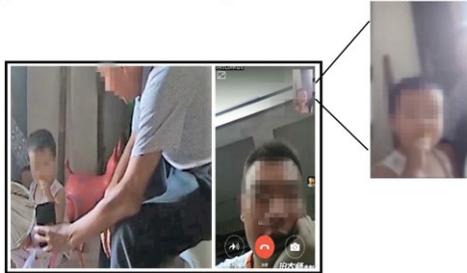
Our next fragment, Figure 6, provides a slightly different example. In this case (from a different family), the child is already on screen, yet the grandfather treats this as insufficient. He touches the child’s shoulder in order to solicit the child’s gaze towards the screen (6a). In response the child gazes towards the screen, looking at her remote father (6b):

01 ((child looks aside, grandfather touches her shoulder. Image 6a))



6a

02 ((child moves head and gaze to screen. Image 6b))



6b

Figure 6: Grandparent getting the child’s attention

The key point here is that there are at least two elements to ‘good’ camera work: framing and attention. These two are sequentially organized. First, at the start of the call facilitators configure the position of their phone, such that the child is on-screen. Second, the facilitators manage the attentional aspect, i.e. make sure that the child is visibly attending to the remote parent (e.g. is looking at the screen). In order to accomplish this, facilitators draw on a range of verbal and embodied resources—such as physically manipulating the phone, positioning the child, inspecting the current framing, instructing the child, etc. Doing this is also a demonstration of the priorities of those doing facilitation work: to ensure the child is the first thing on screen for the

parent, and that the child in turn is ready to attend to the parent as soon as they connect to the video call.

Facilitation work, as a gross manual skill, also turns on facilitators’ preference for maintaining their essential ‘invisibility’ as the interaction unfolds. Sometimes, grandparents may purposefully avoid appearing in the shot or withhold speaking so that they are not verbally indicating their presence. Other times, grandparents are not literally invisible. Instead, they work themselves into the background, setting up their main function as supporting the interactional relation between parent and child. For example, if grandparents are visible at the beginning of a call (e.g. when they press the ‘accept call’ button), they may immediately turn the phone towards the child, without greeting the remote parent first (something that in other circumstances could be considered to be rude). Alternatively, they often may verbally prompt the children (e.g. “greet dad”; “say mum”; “ask mum whether she had lunch”); although these prompts are likely to be audible by remote parents, both the child and the parent can ‘pretend’ that the child is speaking directly to the parent (rather than being ‘coached’ by the grandparent).

PROBLEMS FOR THREE-PARTY VIDEO CALLS

Grandparents ‘not framing’ children

While we have seen that facilitation work may be performed very proficiently in certain cases, it surprised us that this was often not the case and framing did not always work *despite participants clearly trying to*. That is to say: grandparents were often clearly trying to frame the children so that they would be in a position to interact with their remote parents, but the way the grandparents positioned the phone sometimes resulted in a view that either did not show the children or only showed them partly, leading to frequent repair work. This shows that facilitation work for three-party calls is not easy. Our data is replete with instances where grandparents *struggle* to frame the child or children for remote parents (in contrast with our last two fragments). To begin with, a broad examination of our data shows that in the video calls between family members, almost half of the time the child is not clearly visible. Although we have shown that grandparents are indeed able to accomplish ‘good’ framing, achieving this remains a persistent trouble that significantly disrupts most calls. To show the reader what this looks like, we have assembled Figure 7, which depicts a selection of problematic framings of the camera.



7a: Showing clothing or part of child’s head

7b: Showing an unrelated part of the room

7c: Showing an empty image, e.g., ceiling

Figure 7: Problems in grandparents’ framing

The grandparents in Figure 7 (external camera views on the left) are clearly *trying* to show the child for the remote parents. For example, in Figure 7a, the grandmother leans her head to look at the screen and positions the phone right in front of the girl, but the camera only frames the child's hair and the grandmother's clothes. In Figure 7b, although the child is not looking towards the phone, the grandparent is again trying to show the child to the parent by turning the phone toward the child, but is capturing primarily the room, rather than the child in the shot. Furthermore, at times there are extra contingencies that the grandmother has difficulty managing. As shown in Figure 7c, the grandmother uses one hand to hold the child, and the other hand to hold the phone. This of course raises more challenges for the grandmother to frame the child. In this case, the phone only shows the ceiling.

Children's media literacy: 'not showing' objects

We found that problems and challenges around camera work occurred not only for grandparents, but also for children. This was an issue not so much around the children framing something (since, as we indicated above, children rarely had physical access to the phone). Instead, this was an issue around children trying to *show* something to the remote parents 'through' the phone, but where the 'showable' was not actually 'in view'. Previous studies have shown that showing is not something that just happens, but that interactants *coordinate and organise their showing* in detailed ways [17, 18, 19]. In particular, participants need to bring the object to camera view, or they may need to switch the camera to the back view in order to show the objects [19].

When showing something to the front of a camera, it is important to account for the "interactional asymmetries in video-mediated communication" [10]. Following Schutz [28], Heath and Luff [10] observed that in everyday life participants assume, for all practical purposes, an interchangeability of standpoints and a core "reciprocity of perspectives". However, this no longer holds in video-mediated communication: just because *I* can see *you* on the video screen and *I* can see an object in my hand, doesn't mean that *you* can see the object in my hand via the camera.

In our data, children were not always aware of this asymmetry. For example, they seemed to assume that, if they can see their parents on screen, their parents are thus able to see them as well. In the next fragment (Figure 8), grandfather (GRF) holds the phone, and the two children are eating ice cream. The (remote) mum asks the big brother (the boy) "big brother, have you eaten up yours (your ice cream)?" The boy vocalises a smiling sound ("hey hey he ha"), then he moves his empty ice cream stick toward the screen to show to his mum that he has eaten up his ice cream (8a). However, the stick is not visible on screen (8b).

Previous studies have been mostly based on two-party video calls, i.e. the one who shows the object is the one who holds the phone. In our data this is not the case, since it is the grandparent who is *holding* the phone and the child who

wants to *show* objects for remote parents. From Figure 8, we find that showing an object in three-party video calls raises new challenges, especially when it is the child who initiates the showing. We see the boy explicitly orienting to attempting a 'showing' on screen, because he moves the ice cream stick toward the screen. But his lack of awareness of media asymmetries makes this showing unsuccessful.

01 MUM: 哥哥, 你的吃完没哦?
big brother, have you eaten up yours?

02 (0.5)/((boy is eating an ice cream. Image 8a))



8a

03 BOY: 嘿嘿呵哈
hey hey he ha

04 BOY: ((moves empty ice cream stick toward screen, shows it to screen, but it is not visible at all for mum. Image 8b))



8b

Figure 8: Boy trying to show the empty ice cream stick to remote mum

DEALING WITH PROBLEMS IN THREE-PARTY CALLS

Given the significance of *seeing* and being seen, it seems very important that a rapid solution to problems is sought by family members. Although we find in our data that those problems and challenges are often not recognised or topicalised (i.e., surfaced in and made relevant to the interaction), at times they are. In fact, such recognition and topicalisation of problems are the essential way to make these troubles visible, explicit and therefore solve them.

Here we look at the nature of these troubles and their resolution by different parties. Particularly we want to explore how framing problems come to be practically encountered and resolved. While some framing problems were not topicalised by remote parents or other co-present members (but nevertheless resolved), there were many cases where parties did bring to attention framing trouble and asked for its resolution. A simple way of resolving framing problems, such as those shown in Figure 2 above, is to bring them to the attention of the person holding the camera, i.e., to topicalise them.

We identified three main ways that problems with camera work might be topicalised: first, grandparents self-checking the phone; second, remote parents complaining about camera

work; third, co-present, non-phone-holding grandparents complaining about camera work.

The first and perhaps easiest way for the resolution of troubles with facilitation work is for the grandparent holding the phone to notice and resolve it themselves. Practices of ‘self-checking’ become a routine activity for grandparents acting on a (anticipatory) sensitivity towards the intelligibility of their local environment and the children in it. As we show in Figure 9, a grandparent might turn the phone back to themselves to check the call status in preparation for pointing the camera towards a child (9a); they might lower their body down to check the screen (9b); or perhaps move the phone closer to themselves to check and then it move back to the child (9c).



Figure 9: Examples of grandparent self-checking the phone screen

A second way of dealing with problems was for the remote parent to topicalise troubles within a grandparent’s facilitation work. Figure 10 provides an example: a grandmother holds the child in her arms while asking the child to say goodbye to her mum. The child then waves her hand, and says “mum, bye:bye:” (10a, line 01). As depicted in 10a, the problem here is that the child’s head and waving are *not* visible on the parent’s screen.

The grandmother again prompts the child to “say again mum bye bye” (line 03) and “say dad” (line 04), who is off-screen. The child complies with the grandmother’s latter prompt and says “dad” (line 05). Then in line 06, the remote mum utters in a loud voice, “I DIDN’T SEE YOU!”, which formulates a complaint that the child was not visible. After this, the grandmother changes the camera position, moving the phone towards the child’s head to correct the framing of the shot, after which the child then becomes visible on the mum’s screen, although the action (waving) is not redone (10b).

This example demonstrates some of the challenges facing those facilitating the call when managing the position of the child or children in concert with attempts to render that selfsame scene available to the remote parent or parents. It also shows the ways in which parents assess moment-by-moment the facilitation work of the co-present grandparents. In this case the topicalisation of the trouble is indirect in that the mum produces a complaint lexically addressed to the child, in spite of the need for resolution being a matter for the grandparent. Again, this speaks to the ‘invisibility’ of the grandparent as a supporting third party, responsible for

establishing and maintaining mediated interactions between parent and child.

01 CHI: 妈妈,拜:拜::
mum, bye:bye:::(but child not visible. Image 10a))



02 (0.3)

03 GRM: 又喊个妈妈拜拜
say again mum bye bye

04 GRM: 喊爸爸
say dad

05 CHI: 爸爸
dad

06 MUM: 我没看到你!
I DIDN'T SEE YOU!

07 ((grandmother changes camera position,
08 and now the phone shows the child. Image 10b))



Figure 10: Remote parent complaining about camera work

Complaints like those shown in Figure 10 also occur in other calls. For example, in Figure 2 when mum stated, “the phone is always shaking”. We note other similar cases elsewhere such as parents saying, “I DIDN’T SEE YOU!”, or “I haven’t seen you”. Somewhat differently, we also see cases that involve instructional work from parents about how to resolve troubles such as framing: “move up the phone, move up a bit so that I can see your face”.

Third, troubles may be raised by co-present other adults (i.e. those not currently engaged in facilitation work). As shown in the fragment below (Figure 11), a grandfather directs the phone screen (and camera) towards his granddaughter while she dances for her (remote) father. However, the girl is out of shot during this dance (11a). Subsequently the grandmother, who stands next to the girl, says “you didn’t show her” to grandfather. In response, the grandfather changes the angle of the phone, tracking rightwards slightly, such that the girl is then on screen (11b).

- 01 ((Grandfather holds the phone. The girl is dancing for dad, but she is not visible on phone screen. Image 11a))



11a

- 02 GRM: (talk to grandfather) you didn't show her

- 03 ((Grandfather changes camera position, and the right phone screen now shows the girl dancing. Image 11b))



11b

Figure 11: Co-present grandmother complaining about camera work

In this case the co-present grandmother provides ongoing ‘checking’ of the video call, contributing to facilitation work that is needed to maintain the quality of the connection between the parent and child.

We want to briefly recap the three ways we identified that participants recognize and topicalise framing problems and provide opportunities for people to correct camera work. First, grandparents’ self-checking shows the phone-holder’s recognition of possible problems. Their self-checking and self-repairing of the phone camera in some way captures the responsibility of a phone-holder: someone who holds the phone seems responsible to hold it at a correct angle. Second, a remote parent’s complaint about camera work (e.g. Figure 10) is not just a complaint about the fact that they cannot see the child, but also can be heard as a complaint *to* the phone-holder for not holding the phone correctly. As a consequence, the complaint leads to an immediate correction by grandparents. Third, and turning to the last fragment, a co-present, non-phone-holder’s complaint about camera work (e.g. the grandmother in Figure 11) shows us how repair of the framing can be done in a collaborative way. Here, as the phone-holder (grandfather) is facing the back of the phone, he is not able to see the screen, and instead the grandmother’s complaint helps him to re-position the phone.

All three ways of dealing with problems show the *dual* nature of video call screen, where there is a ‘big’ screen and a ‘small’ screen during video call. That is, the screen is a combination of what is *shown* and what is *showing*. In other words, showing parents to children (what is shown on screen) and showing children to parents (what is showing on screen) are two different but parallel activities for

grandparents to achieve during the video calls. Their camera work is the crucial way to achieve this.

DISCUSSION

Since economic shifts in China have led to the dislocation and fragmentation of families, those families have increasingly turned to video calling technologies to ‘mitigate’ (in some limited sense) the situation of separation. Research in both HCI and migration and new media has emphasised the benefit of video technologies for distributed families to mediate intimacy, and suggested the need to shift from a focus just on people’s experience of using the technologies toward supporting family relationships within technology-mediated interactions [2, 14].

Throughout the corpus of video recorded video calls that we have presented fragments from, our findings emphasised the interactional accomplishment of those calls, locating how expectancies and troubles are variously treated by the call parties. We identified frequent examples where there are significant challenges and problems encountered when manipulating the camera in such video calls. It is important to note that these differ substantively to problems encountered in one-to-one video calls. As seen from the data, both migrant parents and grandparents hope to establish a parent-child interaction, rather than parent-grandparent talk. This is the nature of a whole class of three-party calls. In our study, the configuration of these three parties was such that the person who holds the phone is *not* the main (expected) speaker, and as a result facilitation work becomes part and parcel of this party’s role. Unpacking the nature of these interactional details at the same time also articulates the practical production of within-family relationships.

Facilitation work

Our study’s primary contribution is the identification of *facilitation work* as a key feature of three-party video calls (not all three-party calls necessarily, but certainly those in which one party is less able to perform routine aspects of a video call). We believe that the main challenges are for grandparents, who are the facilitators for parent-child communication. We can summarise it thus. First, facilitation work by grandparents is significantly oriented towards the primacy of parent-child interactions. Grandparents try to position themselves as a facilitator who holds, manages, and corrects the camera work in order to create a better video experience for the parents and children. Second, facilitation work is about supporting another party in the call who cannot do certain things for themselves. In our cases, the children are very young; their inability to conduct a video call on their own has resulted in the need for facilitation. Third, facilitation work is oriented to ensuring the practical ‘invisibility’ of the facilitator so as to enable and support interaction between the other two parties to the call (parents and their children). For example, in many of our cases, grandparents are invisible on screen while only the children are visible. Fourth, facilitation work consists of rendering the scene intelligible for the remote party. Showing children on

screen for remote parents is an important part of family life at a distance. Fifth, facilitation work uses verbal and bodily methods to configure camera work so that the interactions are appropriately staged and timely managed.

Mobile devices do not support facilitation work

Our study demonstrates various ways in which mobile devices, which have become a vital tool for mediating family relationships at-a-distance via video calling, are not well suited for facilitation work. Both in terms of the physical design of camera and screen placement as well as the design of video calling services, the present mobile devices seem to be primarily designed for two party circumstances, with little regard for the introduction of a third party in use and therefore of facilitation work. In our case specifically of distributed families mediating their relationships via video calling, we can readily point to a set of capabilities that need to be adopted by the grandparent and could be better supported in the technology:

- (i) *Positioning and framing*: to rapidly position on screen and frame the child such that parents can see them;
- (ii) *Multitasking*: to be able to continuously monitor camera position while at the same time attending to children (e.g. grandparents may need to be holding the child while working on positioning the phone);
- (iii) *Timing*: adjusting the camera appropriately during time-sensitive, important moments for parent-child interactions (e.g. greetings, kissing to say goodbye)
- (iv) *Integration*: weaving the child's co-present activities into the call e.g., showing of children doing a dance;
- (v) *Shaping ongoing interactions as mobile camera-oriented*: to manage problems with the media literacy of children e.g. in terms of not appreciating the lack of reciprocity for mobile phone cameras.

Mobile video calling is not suited for young children

It is clear from our study that mobile video calling also fails to consider some of the more specific challenges encountered when young children are brought into calls. We have described these as matters of 'media literacy' in children's interactions with mobile devices. This returns us to a well-trodden issue in research on video mediated interactions: the non-reciprocity of perspectives. This essential asymmetry in video conferencing systems and other applications of video for distributed parties has been explored in prior work [9, 10, 31]. However, we note that the asymmetry problem is significantly exacerbated in the use of mobile video with young children, as we saw in some of our examples. Facilitation work thus is an attempt to manage these asymmetries. The distinctive feature here is in our three-party configuration, where one party typically does not have the capability (yet) to themselves account for the lack of reciprocity of perspectives.

The value of video ethnographic work to unpack facilitation work

Finally, we see value in our approach for investigating distributed families' uses of video. Specifically we believe

there are opportunities for HCI research in more fully exploring the use of video-based ethnographic studies (for us, these are ethnomethodologically-informed video studies) in this area (although they have been applied in other domains, see an overview from [11]). There are various benefits of this kind of video analysis. Facilitation work, although in some sense not 'new', is nevertheless something that easily goes unnoticed. Prior research has quite understandably primarily focused on the parent-child relationship. If you actually study such video calls, one begins to notice the significant work of grandparents. Our approach thus corrects this oversight and gives us a good first look at the interactional accomplishment of facilitation by foregrounding it via video recordings. Doing video analysis thus enables us to recover the detailed ways in which parent-child relationships can be mediated by video (with the facilitation of a third party), respecifying our understandings of relations between parent and child by recognising the fullness of the role that may be played by a third party, or in our case, grandparent(s).

Of course, our study is not without limitations. We examine a set of normative organisations of families in the Chinese migrant worker context. Our data is limited to the practicalities of recruitment and we point out that our study cannot speak necessarily to a broader range of family arrangements that lie beyond the circumstances of this particular study. That is for future work.

CONCLUSION

Our study has uncovered the critical distinctiveness of the three-party dynamic of video calling and the significance (interactionally, emotionally, etc.) of facilitation work as a feature of that dynamic. In particular, our findings highlight the significant interactional complexity of video-mediated interaction involving young children.

Previous studies implicitly documented two main types of facilitation work: (a) controlling and shepherding the child's body, e.g. positioning them in front of a (stationary) camera attached to a computer [2]; and (b) providing verbal assistance for the child to respond to remote party's questions [30]. Our study reveals two elements of a non-stationary, mobile form of facilitation work that results from the 'dual mobility' of smartphone cameras and of young children, which brings out a new dimension of facilitation that is focused on camera work itself. Our findings also allow reflections on other situations whereby facilitation is needed, for instance, video calling with elderly or disabled people where communication is supported by a third-party.

ACKNOWLEDGMENTS

This project is partially supported by an ICS Mok Hing Cheong Postgraduate Scholarship from the Institute of Chinese Studies at the Chinese University of Hong Kong, a Direct Grant for Research from The Chinese University of Hong Kong, and the UK Engineering and Physical Sciences Research Council (grant number EP/M02315X/1).

REFERENCES

- [1] ACWF (All China Women's Federation). 2013. Research report on the situation of rural left-behind China, rural-urban migration children in China. *Zhong Guo Fu Yun (Chinese Women's Movement)*, 6, 30-34. [in Chinese]
- [2] Morgan G. Ames, Janet Go, Joseph 'Jofish' Kaye, and Mirjana Spasojevic. 2010. Making love in the network closet: the benefits and work of family videochat. In *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work (CSCW '10)*. ACM, New York, NY, USA, 145-154. DOI: [10.1145/1718918.1718946](https://doi.org/10.1145/1718918.1718946)
- [3] Rafael Ballagas, Joseph 'Jofish' Kaye, Morgan Ames, Janet Go, and Hayes Raffle. 2009. Family communication: phone conversations with children. In *Proceedings of the 8th International Conference on Interaction Design and Children (IDC '09)*. ACM, New York, NY, USA, 321-324. DOI: [10.1145/1551788.1551874](https://doi.org/10.1145/1551788.1551874)
- [4] Deana Brown and Rebecca E. Grinter. 2012. Takes a transnational network to raise a child: the case of migrant parents and left-behind Jamaican teens. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing (UbiComp '12)*. ACM, New York, NY, USA, 123-132. DOI: [10.1145/2370216.2370237](https://doi.org/10.1145/2370216.2370237)
- [5] CMDS (China Migrants Dynamic Survey). 2017. *Report on Chinese Migrants*. National Health and Family Planning Commission. Retrieved from: chinaldrk.org.cn/wjw/#/data/classify/population/provinceList [in Chinese]
- [6] Sean Follmer, Hayes Raffle, Janet Go, Rafael Ballagas, and Hiroshi Ishii. 2010. Video play: playful interactions in video conferencing for long-distance families with young children. In *Proceedings of the 9th International Conference on Interaction Design and Children (IDC '10)*. ACM, New York, NY, USA, 49-58. DOI: [10.1145/1810543.1810550](https://doi.org/10.1145/1810543.1810550)
- [7] Azadeh Forghani and Carman Neustaedter. 2014. The routines and needs of grandparents and parents for grandparent-grandchild conversations over distance. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 4177-4186. DOI: [10.1145/2556288.2557255](https://doi.org/10.1145/2556288.2557255)
- [8] Harold Garfinkel. 1967. *Studies in Ethnomethodology*. Prentice-Hall, Englewood-Cliffs, NJ, USA.
- [9] William Gaver, Abigail Sellen, Christian Heath, and Paul Luff. 1993. One is not enough: Multiple views in a media space. In *Proceedings of the INTERACT'93 and CHI'93 Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 335-341. DOI: [10.1145/169059.169268](https://doi.org/10.1145/169059.169268)
- [10] Christian Heath and Paul Luff. 1993. Disembodied conduct: Interactional asymmetries in video-mediated communication. In Graham Button (ed.) *Technology in Working Order: Studies of Work, Interaction, and Technology*. Routledge, London, UK, 35-54.
- [11] Christian Heath, Jon Hindmarsh, and Paul Luff. 2010. *Video in Qualitative Research*. London: Sage.
- [12] Pierrette Hondagneu-Sotelo and Avila Ernestine. 1997. 'I'm here, but I'm there': the meanings of Latina transnational motherhood. *Gender & Society*, 11(5), 548-571. DOI: [10.1177/089124397011005003](https://doi.org/10.1177/089124397011005003)
- [13] Tamara Jacka. 2018. Translocal family reproduction and agrarian change in China: a new analytical framework. *The Journal of Peasant Studies*. 7 (45), 1341-1359. DOI: [10.1080/03066150.2017.1314267](https://doi.org/10.1080/03066150.2017.1314267)
- [14] Clare Kelly. 2015. 'Let's do some jumping together': Intergenerational participation in the use of remote technology to co-construct social relations over distance. *Journal of Early Childhood Research*. 13(1). 29-46. DOI: [10.1177/1476718X12468121](https://doi.org/10.1177/1476718X12468121)
- [15] David S Shelby Kirk, Abigail Sellen, and Xiang Cao. 2010. Home video communication: mediating 'closeness'. In *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work (CSCW '10)*. ACM, New York, NY, USA, 135-144. DOI: [10.1145/1718918.1718945](https://doi.org/10.1145/1718918.1718945)
- [16] Christian Licoppe and Julien Morel. 2012. Video-in-interaction: 'Talking heads' and the multimodal organization of mobile and Skype video calls. *Research on Language and Social Interaction*, 45(4), 399-429. DOI: [10.1080/08351813.2012.724996](https://doi.org/10.1080/08351813.2012.724996)
- [17] Christian Licoppe and Julien Morel, J. 2014. Mundane video directors in interaction: Showing one's environment in Skype and mobile video calls. In Broth, M., Laurier, E., & Mondada, L. (Eds.), *Studies of Video Practices: Video at Work*. Routledge, London, UK, 135-160.
- [18] Christian Licoppe. 2017. Showing objects in Skype video-mediated conversations: From showing gestures to showing sequences. *Journal of Pragmatics*, 110, 63-82. DOI: [10.1016/j.pragma.2017.01.007](https://doi.org/10.1016/j.pragma.2017.01.007)
- [19] Christian Licoppe, Paul 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: [10.1145/3025453.3025848](https://doi.org/10.1145/3025453.3025848)
- [20] Paul Luff and Christian Heath. 1998. Mobility in Collaboration. In *Proceedings of the 1998 ACM Conference on Computer Supported Cooperative Work*

- (CSCW '98). ACM, New York, NY, USA, 305-314. DOI: [10.1145/289444.289505](https://doi.org/10.1145/289444.289505)
- [21] Mirca Madianou and Daniel Miller. 2012. *Migration and New Media: Transnational Families and Polymedia*. Routledge, London, UK.
- [22] Elisabeth McClure and Rachel Barr. 2017. Building family relationships from a distance: Supporting connections with babies and toddlers using video and video chat. In Rachel Barr & Deborah Nichols Linebarger (Eds.), *Media Exposure During Infancy and Early Childhood*. Springer, New York, NY, USA, 227-248. DOI: [10.1007/978-3-319-45102-2_15](https://doi.org/10.1007/978-3-319-45102-2_15)
- [23] Carman Neustaedter and Saul Greenberg. 2012. Intimacy in long-distance relationships over video chat. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*. ACM, New York, NY, USA, 753-762. DOI: [10.1145/2207676.2207785](https://doi.org/10.1145/2207676.2207785)
- [24] Rhacel Salazar Parreñas. 2001. Mothering from a distance: Emotions, gender, and intergenerational relations in Filipino transnational families. *Feminist Studies*, 27(2), 361-390. DOI: [10.2307/3178765](https://doi.org/10.2307/3178765)
- [25] Harvey Sacks. 1992. *Lectures in Conversation*. Blackwell, Oxford, UK.
- [26] Emanuel A Schegloff. 1995. Parties and talking together: Two ways in which numbers are significant for talk-in-interaction. In Paul Ten Have and George Psathas. (Eds.), *Situated Order: Studies in the Social Organization of Talk and Embodied Action*. University of America Press, Lanham, MD, USA, 31-42.
- [27] Leah Schmalzbauer. 2004. Searching for wages and mothering from afar: The case of Honduran transnational families. *Journal of Marriage and Family*, 66(5), 1317-1331. DOI: [10.1111/j.0022-2445.2004.00095.x](https://doi.org/10.1111/j.0022-2445.2004.00095.x)
- [28] Alfred Schutz. 1962. *Collected Papers I: The Problem of Social Reality*. Martinus Nijhoff, The Hague, Netherlands.
- [29] Chiho Sunakawa. 2012. Japanese family via webcam: An ethnographic study of cross-spatial interactions. In *JSAI International Symposium on Artificial Intelligence*, Springer, Berlin, Germany, 264-276. Berlin: Springer. DOI: [10.1007/978-3-642-32090-3_24](https://doi.org/10.1007/978-3-642-32090-3_24)
- [30] Chiho Sunakawa. 2014. Virtual ie: A three-generational household mediated by webcam interactions. *Journal of Pragmatics*, 69, 63-81. DOI: [10.1016/j.pragma.2014.05.012](https://doi.org/10.1016/j.pragma.2014.05.012)
- [31] Amy Vaida, Stephen Vaida, Saul Greenberg, and Helen Ai He. 2008. Asymmetry in media spaces. In *Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work (CSCW '08)*. ACM, New York, NY, USA, 313-322. DOI: [10.1145/1460563.1460615](https://doi.org/10.1145/1460563.1460615)
- [32] Raelene Wilding. 2006. 'Virtual' intimacies? Families communicating across transnational contexts. *Global Networks*, 6(2), 125-142. DOI: [10.1111/j.1471-0374.2006.00137.x](https://doi.org/10.1111/j.1471-0374.2006.00137.x)
- [33] Hongwei Xu and Yu Xie. 2015. The causal effects of rural-to-urban migration on children's well-being in China. *European Sociological Review*, 31(4), 502-519. DOI: [10.1093/esr/jcv009](https://doi.org/10.1093/esr/jcv009)
- [34] Svetlana Yarosh and Gregory Abowd. 2011. Mediated parent-child contact in work-separated families. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. ACM, New York, NY, USA, 1185-1194. DOI: [10.1145/1978942.1979120](https://doi.org/10.1145/1978942.1979120)
- [35] Svetlana Yarosh, Stephen Cuzzort, Hendrik Müller, and Gregory D. Abowd. 2009. Developing a media space for remote synchronous parent-child interaction. In *Proceedings of the 8th International Conference on Interaction Design and Children (IDC '09)*. ACM, New York, NY, USA, 97-105. DOI: [10.1145/1551788.1551806](https://doi.org/10.1145/1551788.1551806)
- [36] Svetlana Yarosh and Mark Robert Kwikkers. 2011. Supporting pretend and narrative play over videochat. In *Proceedings of the 10th International Conference on Interaction Design and Children (IDC '11)*. ACM, New York, NY, USA, 217-220. DOI: [10.1145/1999030.1999065](https://doi.org/10.1145/1999030.1999065)
- [37] Jingzhong Ye and James Murray. 2005. *Left-behind children in rural China: Impact study of rural labor migration on left-behind children in mid-west China*. Social Sciences Academic Press, Beijing, China.
- [38] Jingzhong Ye and Pan, L. 2011. Differentiated childhoods: impacts of rural labor migration on left-behind children in China. *The Journal of Peasant Studies*, 38(2), 355-377. DOI: [10.1080/03066150.2011.559012](https://doi.org/10.1080/03066150.2011.559012)
- [39] Rui Zhou, Zhonghe Wen, Muchao Tang, and Betsy DiSalvo. 2017. Navigating Media Use: Chinese Parents and Their Overseas Adolescent Children on WeChat. In *Proceedings of the 2017 Conference on Designing Interactive Systems (DIS '17)*. ACM, New York, NY, USA, 1025-1037. DOI: [10.1145/3064663.3064701](https://doi.org/10.1145/3064663.3064701)