Bridging the Socio-Technical Gaps in Body-worn Interpersonal Live-Streaming Telepresence through a Critical Review of the Literature

KEVIN P. PFEIL, University of Central Florida, USA
NEERAJ CHATLANI, University of Central Florida, USA
JOSEPH J. LAVIOLA, JR., University of Central Florida, USA
PAMELA WISNIEWSKI, University of Central Florida, USA

It is important to learn from the past as we endeavor into this uncharted territory of mobile, human-to-human, one-to-one telepresence for interpersonal use. With the ever-increasing access to live-streaming cameras, we are now at the cusp of being able to create novel, immersive, and interpersonal telepresence activities that have the potential to change how humans interact with one another on a daily basis. Due to its novelty, there are likely socio-technical gaps between the needs of users and the technical specifications of the prototypes that are currently being designed to support the complex social interactions of human-to-human telepresence. Therefore, in this paper, we use a socio-technical lens to conduct a systematic literature review of 52 peer-reviewed articles of early work in this space. Overall, we found that while progress has been made to address the social needs of those involved in one-to-one telepresence scenarios, there are discontinuities within the existing literature that need to be addressed, particularly with the way we attempt to measure and quantify human-centered outcomes with unvalidated instruments. We also found that the social needs of on-site users have been neglected, as in many articles the user was merely treated as a surrogate, or reported feeling socially awkward or unsafe, due to the conspicuous nature of the body-worn technology in public environments. These findings are prevalent, even as researchers consider adding to this body-worn burden in an attempt to improve the receiving users' sense of immersion and presence. To preserve the beneficial nature of telepresence interaction while ensuring that all users' needs are met, researchers should endeavor to further understand the dynamics of the relationship between all parties in the remote environment. Our paper creates a future research agenda that emphasizes the importance of ensuring that all parties involved feel comfortable in their role during interpersonal telepresence interactions.

CCS Concepts: • **General and reference** → *Surveys and overviews*; • **Human-centered computing** → *HCI theory, concepts and models*; *Ubiquitous and mobile devices.*

Additional Key Words and Phrases: Telepresence, literature review, novel interaction, mobile

ACM Reference Format:
Kevin P. Pfeil, Neeraj Chatlani, Joseph J. LaViola, Jr., and Pamela Wisniewski. 2021. Bridging the Socio-Technical Gaps in Body-worn Interpersonal Live-Streaming Telepresence through a Critical Review of the Literature. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW1, Article 120 (April 2021), 39 pages. https://doi.org/10.1145/3449194

Authors' addresses: Kevin P. Pfeil, University of Central Florida, Orlando, USA, kevin.pfeil@knights.ucf.edu; Neeraj Chatlani, University of Central Florida, Orlando, USA, nchatlani@knights.ucf.edu; Joseph J. LaViola, Jr., University of Central Florida, Orlando, USA, jjl@cs.ucf.edu; Pamela Wisniewski, University of Central Florida, Orlando, USA, pamela.wisniewski@ucf.edu.

2573-0142/2021/4-ART120 $15.00
https://doi.org/10.1145/3449194

1 INTRODUCTION

In 2020, we have experienced a phenomenon like none other - the Novel Coronavirus has created the need for "social distancing," where people must physically distance themselves and limit the size of social gatherings [2], while still trying to function as a society and remain connected to one another. Both non-essential work and school environments have moved online as people struggle to keep a safe distance, while avoiding complete social isolation. Some countries have shutdown completely, disapproving of social gatherings which could contribute to spreading the virus, fining individuals who break shutdown laws [96, 108]. Meanwhile, we wonder what this situation will do to interpersonal relationships, as remote technologies tend to emphasize work and school as the primary use cases for their target audiences. It is in times like these that Human-Computer Interaction (HCI) researchers can potentially make a difference by designing novel technologies that can at least partially solve some of the new challenges facing the world. For instance, we might re-imagine telepresence, or the ability to perceive and/or interact with a remote environment as if actually there [77], for more immersive and intimate uses, such as orchestrating a playdate or being able to share invaluable time with an ill family member who is physically bound to their home or bed. By transforming existing technologies or imagining new ones, we might be able to bridge the socio-technical gap in a way that creates human-to-human, one-to-one telepresence experiences that maintain and strengthen interpersonal relationships that are more and more often mediated through technology.

In 1980, Minsky hypothesized that we could implement telecommunications in order to let workers remotely manipulate a robotic platform to perform a task over great distance [77]. We are seeing this idea begin to blossom; for example, stationary robots are being used by doctors for remote surgery [59, 102], and research efforts are planning robotic uses for space expeditions [39, 68, 69]. For everyday use cases, simple, wheeled robots have been developed to let a user explore and interact with a remote environment via video chat (e.g. [45, 95]). More recently, telepresence has expanded to include human-to-human experiences through telecommunications platforms (e.g., Skype, Zoom, etc.) that connect two or more parties, typically within home environments (e.g., [43, 44, 90, 91]). However, with the advent of affordable cameras with live-streaming capabilities and sufficient networking technologies, there is a new trend in which humans wear a camera on their body to share their mobile experience with another person. This form of telepresence – mobile, human-to-human, one-to-one telepresence – is starting to gain traction, in multiple places.

For instance, streamers are broadcasting to vast audiences comprised of users who each have a viewing device [32, 85, 119]. Twitch, Periscope, and TikTok, for example, are popular live-streaming platforms, where content creators have started to share their In-Real-Life (IRL) experiences with audiences as they visit different places [21], perform tasks or engage in other activities [31, 71, 72]. Meanwhile, doctors are live-streaming first-person views of surgery to enhance medical students' training [24]. In these cases, a single user broadcasts to many recipients to share the experience, rather than to build or maintain existing relationships with the them. Yet, there is an opportunity for immersive telepresence to be leveraged in a one-to-one paradigm, where there is only one streamer and one viewer, to not only share one's experience but to socially connect to another human being. This more intimate use of telepresence technology creates unique opportunities for interpersonal interactions that build relationships and trust through shared goals, such as a remote user directing an on-site user to perform specific tasks or interact with their environment [38, 122], or to simply allow for a socially-distant meet-up where it would be otherwise impossible, or dangerous to do so.

Telepresence has been used in work-related scenarios and in the context of Human-Robot Interaction (HRI) [63], but with consumers being given increased access to telepresence-capable technology, such as 4G streaming and mobile HD cameras, one-to-one telepresence has also found several recreational uses. At the beginning of the 2010's, for instance, smartphones and mobile video chat applications (e.g. Skype and FaceTime) emerged, setting up the infrastructure for mobile, human-to-human, one-to-one telepresence [126]. However, the implementations of mobile telepresence are still in their infancy, as researchers are working to identify the best practices which offer optimal user experience to both sides of the interaction. Recent implementations of mobile human-to-human telepresence in the SIGCHI community have been met with apprehensions and scrutiny, and leading up to the CHI 2019 conference, there was significant discourse surrounding the body-worn paradigm as a way to provide remote attendance to conferences [22, 94]; for instance, some felt that the nature of the interaction invited social harm to the person wearing a device on their body, such as leading to uncomfortable situations or exploitation, and women have expressed that wearing a camera on their body would invite people to stare inappropriately.

These concerns have given telepresence researchers pause as they consider the social implications of this new interaction paradigm. In essence, these concerns stem from what Ackerman calls a "socio-technical gap [3]," in which there is a disconnect between what we should socially support and the technical capabilities for doing so. In a mobile telepresence experience, the local user, remote user, and third parties in the immediate vicinity each have social needs. Addressing these needs give rise to the technical implications of future body-worn telepresence designs. These concerns inspire us to work towards creating a better telepresence interaction for all its potential users. Therefore, we apply this socio-technical lens to a field that tends to be more technical in nature, as to bring about a better understanding of social considerations for this interaction. As such, our goal was to bridge this gap to synthesize this literature and formulate a cohesive research agenda for moving the field forward. We aim to inform the SIGCHI community of this interaction technique and make a push towards its ethical ubiquity. Thus, in this paper, we present the results of a systematic literature review of 52 articles describing mobile, human-to-human, one-to-one telepresence, asking the following research questions:

RQ1: How has the telepresence research community worked to understand the socio-technical implications within their designs for mobile, Human-to-Human, One-to-One telepresence?
RQ2: What considerations have been made to accommodate the social needs of all parties involved in mobile, Human-to-Human, One-to-One telepresence scenarios?
RQ3: What are the technical design implications that arise from changes in the design of mobile, Human-to-Human, One-to-One telepresence, in order to accommodate these social needs?

To answer these questions, we performed a systematic review of the mobile, human-to-human, one-to-one telepresence literature emerging from the SIGCHI community since 2010. We performed a comprehensive search for articles that focused on real-world telepresence, where the sharing user had the ability to change locations, thus creating a mobile telepresence experience for a single recipient. Additionally, we chose interpersonal telepresence scenarios, rather than work-related ones, as our focus is on the socio-technical gaps in current telepresence design. From there, we conducted a content analysis with open coding to categorizing our findings under the themes of technical versus social factors. Overall, we found that the telepresence research community has made an attempt to design for and measure a sense of both spatial presence and social presence (RQ1). However, this was almost entirely achieved through the use of custom measurements, highlighting a need for telepresence researchers to move towards standardization of measurements, and to assess custom measurements for construct validity before they are used in evaluation or disseminated throughout the research community.

While progress is being made to improve the receiving users' sense of social presence, which should enhance their overall experience, we found that the sharing user has been generally disregarded as an important party in the Human-to-Human relationship, with their own specific set of social needs (RQ2). This is especially relevant given that in most telepresence scenarios, this individual is the primary party interacting with the remote environment, either on the other users' behalf or on their own accord within the telepresence scenario. In the reviewed articles, many on-site users complained that the body-worn technology led to socially awkward situations due to how noticeable and conspicuous the technology probes were. However, in cases where efforts were taken to hide the body-worn technology, third parties in the immediate environments often became confused by their behavior. As our understanding of the social needs of all human parties involved in telepresence scenarios grows, we note that researchers have worked to create systems which allow for more natural communication between the sharing and receiving parties, including the use of 360-degree cameras with head tracking, or varied methods of camera placement depending on the prior relationship of the two parties (RQ3). Our research makes the following novel contributions to the CSCW and immersive telepresence communities:

(1) An in-depth review of the current state of the mobile, human-to-human, one-to-one telepresence research community's socio-technical considerations
(2) A discussion of how the social landscape (i.e. the remote environment) influences the technical factors in telepresence experiences, which in turn breeds social considerations which can affect the sharing persons' social and psychological well-being
(3) Socio-technical recommendations and an agenda of future work for the telepresence community, in order to advance this novel technology while ensuring that all participants, whether sharing, receiving, or observing, feel a sense of social equity and enjoyment

We accomplish this with a focus on understanding how researchers work to address the social needs of all users while attempting to preserve the viewing user's sense of presence and the sharing user's sense of agency and safety. In the next section, we briefly describe the history of telepresence as it was formally defined, its uses, and the advent of mobile, human-to-human, one-to-one telepresence as a novel means of social interaction within interpersonal relationships.

2 BACKGROUND

There have been numerous research efforts in the field of telepresence, in which the goal is to provide one user with a way to feel as if they were in a remote environment. In this section, we briefly discuss how telepresence design originated in and evolved with HRI, and how the field has branched out to the human-to-human paradigm.

2.1 The Progression of Telepresence

Marvin Minsky originally conceived the idea of telepresence in 1980 as a means of remotely controlling robotic platforms over long distances, through the use of immersive sensory feedback, which would allow the user to feel a sense of presence as they accomplished specific tasks [77]. At a remote workstation, the worker would receive a multi-sensory, egocentric point of view from their robot, and be able to react as if they were actually there; for instance, they could use natural arm movements to remotely grip an object and move it somewhere else [64]. Since then, we have seen numerous research efforts emerging from the field of HRI under the moniker of mobile robotic telepresence (MRP) [63]. 3D user interfaces (3DUI) have enabled the direct manipulation of small-scale robots via the use of hand gestures to give commands to these robots [97, 103], and in the field of healthcare, a larger robot – the Da Vinci Surgical System – is used to facilitate minimally invasive surgeries while being controlled by a surgeon via a remote console [121]. Other robots allow the elderly and disabled to interact with family members, healthcare attendants, and doctors, through the use of mobile platforms which provide a natural sense of presence [58, 124].

Telepresence then began to evolve past HRI in social contexts, as a means of enhancing common telecommunication by providing face-to-face meetings between two or more remote parties. Typical teleconferencing systems aim to provide a static "window" to a remote location via a direct video feed [44, 92, 137], and researchers have endeavored to create systems which incorporate greater feelings of spatial presence, or the feeling of "being there" [135]. Microsoft's Room2Room technology affords this by video-capturing a person's body and projecting it into another person's remote location. An evaluation revealed that this method provides a more tangible representation of each party member, enabling more natural communication (i.e. body language), which can be lost in window-type systems [101]. Telepresence has also provided means for more intimate methods of teleconferencing, as a way for families to keep in touch over long distances, more closely emulating a form of constant presence through "always-on" video teleconferencing [43, 44, 91]. Researchers have deployed technology probes in the home, using stationary installations to facilitate a "window" into another household [43, 44]. However, since mobile technologies have advanced rapidly in the previous decade, we have begun to see examples of non-stationary telepresence outside of the home, which we highlight in the next section.

2.2 Mobile, Human-to-Human, One-to-One Telepresence

As mobile devices such as smartphones and tablet computers have become commonplace, so too has the ability for telepresence. FaceTime and Skype released video chat applications for these mobile devices in 2010. Around the same time, consumer-grade telepresence robots using these devices became available (c.f., [36, 63, 95, 117, 136]), mostly taking the form of "tablets on wheels," although other platforms such as quadcopter drones have been leveraged to provide an unrestricted viewpoint to a remote user [111]. These afford a user with a physical embodiment that they can use to interact in social settings with their remote environment. However, they have been criticized for being impersonal both in professional and social contexts [15], and people have been observed keeping their distance from various robot platforms [20, 128]. At the same time, live-streaming via mobile devices has gained popularity as a way for people to share their immediate experiences with another, who would see the shared environment from the perspective of the device [109]. In lieu of using robotic telepresence platforms, researchers have begun adopting a relatively new approach that leveraged live-streaming mobile devices, where the sharing user wears the device and situates it in such a way that the recipient can experience the remote environment from their perspective. This provides an interesting sense of embodiment and avoids the most common complaints associated with MRP, namely a lack of freedom of movement and a sense of impersonality experienced by those interacting with the robot [21, 66]; even small drones, which are seemingly unrestricted in movement, are not optimal for indoor use, but more importantly, they are prohibited in many places - for instance, drones are not allowed to fly over sports arenas during an event in the United States, where there are regulations set by the federal government regarding legal drone use [1]. New uses of telepresence have emerged that uniquely utilize this Human-to-Human setup. One of these has been the sharing of a novice worker's perspective to a remote expert, who has the ability to provide guidance through Augmented Reality (AR) cues, overlaying the novice's physical perception with the expert's context-sensitive virtual information [4, 132, 142]. This paradigm has appeared in multiple fields, including medicine [133] and manufacturing/maintenance [13, 84, 122].

Live-streamed video from a worker's first-person perspective also gave rise to new opportunities in manufacturing and maintenance. For instance, expert workers could train, instruct, or provide guidance to novice workers without the need to travel [46]. Using AR cues, trainers can even annotate a worker's view, in order to depict actions that need to be taken in an assembly task

[4, 56, 120]. This form of remote guidance leverages telepresence for production-based environments, and this could, in the future, become a norm in the industry. However, this interaction will occur behind closed doors, and there is not a significant social element to it. This is unlike mobile telepresence experiences, in which a user could navigate around a large environment. For instance, the Faroe Islands have provided live, remote tours to anyone interested in exploring the islands, using telepresence technology[1]; in a more populous destination, the streamer might encounter third party members who may or may not know that this interaction is occurring.

As such, there are different considerations that need to be made for stationary versus non-stationary (i.e., mobile) contexts. Progression in mobile, human-to-human telepresence has given researchers pause as they begin to consider the needs of their users, the different social contexts in which they seek to use telepresence, and the different ways that telepresence is used for their interaction [34]. With this change in telepresence paradigm comes a new set of considerations that must be made, especially given the addition of a new human in the telepresence loop (i.e. the on-site user sharing their environment), who has their own socio-technical needs that must be met while functioning as an intermediary for another person's interaction with the remote environment. Over the past 10 years, telepresence researchers have attempted to address this through multiple technical setups for mobile, Human-to-Human, one-to-one telepresence. Our review synthesizes the socio-technical considerations found in this prior work, discusses the interplay between the technical and social factors, and identifies the standards for measuring the effectiveness of these methods. This review is critical to the CSCW community, as we have been thrust into an era of social distancing, requiring broad communities across the globe to maintain physical distance while trying to live their lives. The human-to-human, one-to-one telepresence interaction paradigm is a potential avenue to maintain this social distancing while fostering a sense of togetherness. For instance, one user could visit a local art gallery while their family member with pre-existing health conditions explores from afar; two geographically separated friends could share their scenic jogging routes; or a parent could share their child's playtime with an at-risk grandparent.

As the field of telepresence has matured, researchers have begun to coalesce on metrics for determining success. For instance, Rae et al. created an exhaustive list of questions a telepresence designer should ask in order to meet various needs of the end-users. The requirements fall into various design dimensions, such as the physical characteristics of the remote environment, the visual affordances to the remote participant/user, the social standing between all participating parties, and the communication methods between these parties [105]. This leads to a new socio-technical understanding of telepresence, one which grows as increased usage of this technology demands better contextualization of the relationships between the parties involved, how these parties perceive each other through different forms of telepresence, and their own requirements when it comes to the design of these telepresence interfaces. We therefore seek to understand how telepresence researchers have considered the socio-technical needs of their users, and how they have evaluated whether or not their designs address these needs while maintaining the benefits of remote presence. Next, we describe the methodology for our systematic review, including our scoping criteria and search for articles concerning mobile, human-to-human, one-to-one telepresence, and our grounded data analysis approach.

3 METHODS

To answer our research questions, we conducted a systematic literature review of 52 articles which described mobile, human-to-human, one-to-one interpersonal telepresence. In this section we

---

[1] https://visitfaroeislands.com/remote-tourism/

describe our methodology, which is comprised of multiple phases including a systematic literature search, scoping criteria, and data analysis approach.

3.1 Systematic Literature Search

As we are focused on an emerging topic within the computer science discipline, we searched the ACM and IEEE Digital Libraries for relevant articles. We used a combination of search terms that intersected at the type of human interaction (i.e., "human-to-human," "personal," or "interpersonal") and the type of technological interaction ("telepresence," "streaming" "live streaming," "video," or "virtual reality") that together form a telepresence experience. Our search was limited to peer-reviewed articles published in English and between the years of 2010 to 2019. Because Skype introduced video chat to mobile devices in 2010, FaceTime was released in 2010, and consumer-grade omnidirectional cameras supporting live streaming became regular around 2013, we felt that this date range would capture the genesis of mobile, interpersonal, human-to-human, one-to-one telepresence. Our initial data collection was conducted during the time span of December 2019 through January 2020. In September 2020, we updated our search to include articles published in 2020. In total, our search yielded 1,087 articles, which we scoped for relevancy. In the next section, we describe our inclusion criteria.

3.2 Scoping Criteria: Interpersonal Telepresence

Though telepresence takes many forms, our research interests are focused on a specific kind – mobile (non-stationary and body-worn), human-to-human, one-to-one telepresence. Therefore, we included articles based on the following inclusion criteria:

1. **Article must describe a real-world telepresence experience with, at minimum, video live-streaming (i.e. not a virtual environment).** While it is possible to foster a sense of presence through virtual environments, our research interests are specific to real-world interactions.
2. **Article must describe a human-to-human interaction (i.e., not human-to-robot or robot-to-robot).** This was included as we are only interested in human-to-human interaction which carries an entirely unique set of socio-technical implications than telepresence found within HRI; in addition, although robotics researchers are constantly improving the quality of MRP platforms, they are currently unusable in many scenarios [63, 70]. Thus, we include this criterion to allow for the broadest possible capture of telepresence scenarios to mediate interpersonal relationships where researchers forgo the use of a robot.
3. **Article must describe an experience where the sharing user is non-stationary (i.e., the physical ability to change locations)**. Human-to-human telepresence is noted for its ability to overcome mobility issues commonly associated with MRP, and so we aimed to focus on those scenarios which primarily benefit from a human's natural ability to move. Further, we expect this criterion to filter teleconferencing scenarios, which fall out of scope of our research interests.
4. **Article must describe a recreational context, and not a work-related scenario.** Our research interests align with the study of the social relationships between participating parties in the everyday use of telepresence, and not those in isolated, task-based scenarios. A wealth of work has already been done to synthesize the research on work-related telepresence (c.f., [23, 67, 120, 122, 132]). Therefore, these contexts were held out-of-scope for the current review.
5. **Article must describe a body-worn device to facilitate the telepresence experience.** This criterion was included because we are interested in telepresence situations in which the

viewing users can see the shared environment from the sharer's point of view, opposed to views which are captured from an imaging accessory such as a tripod or mount, e.g. teleconferencing. "Body-worn" devices include a wide range of devices, from cameras mounted in harnesses, to a smartphone stored in a shirt pocket, to a camera which is handheld; we include an article if the sharing user has the ability to constantly manipulate the camera, thus excluding articles if that user cannot control the camera. Since there is an ongoing debate regarding appropriate and ethical use of body-worn devices [22, 94, 141], we anticipated this criterion to also help us identify how prior work tackled this problem.

(6) **Article must describe a one-to-one interaction.** This criterion was included because we wanted to focus on this unique, more interpersonal interaction style, where two people have a more intimate means of communication, which could facilitate back-and-forth coordination to accomplish some mutually beneficial goal. We exclude one-to-many and many-to-many configurations because they lack this interpersonal quality of interaction and thus fall out of the scope of our research interests.

For every article returned from the searches, we reviewed the title, abstract, and if necessary, the main body, to ensure it met all criterions listed above. If so, we added it to our dataset. Then, for every article in our dataset, we cross-referenced citations and references, adding any new articles to our final dataset if they also met our inclusion criteria. This was done for every new article in our dataset until no new articles were found. We coded 1,087 articles for relevancy. In total, we identified 52 relevant articles, which we analyzed in our review. In the next section, we describe our data analysis approach and present our final codebook based on our qualitative analysis.

3.3 Data Analysis Approach

The first author performed the qualitative analysis on the 52 articles, identifying emergent codes, using an iterative and grounded approach [65]. We began with open coding and noted the emergent themes found within the dataset; these typically included factors explicitly noted by the authors, but also include underlying concepts which we found across multiple articles. Our coding was also updated as new codes emerged. The second author was consulted during this process to resolve, separate, or merge ambiguous codes. After open coding, a round of axial coding was conducted where we grouped our codes into more cohesive dimensions and themes (as shown in Table 1). These described considerations for the interactants, as well as for the interaction as a whole. We then mapped our coded dimensions into two categories - Technical and Social - which respectively relate to our RQ3 and RQ2. The combination of these dimensions then forms the concepts required to answer our RQ1. The Technical dimensions describe the nature and setup of the technology used by the interactants, along with the design considerations made to benefit the sharing individual, the viewing individual, or the ability of the two parties to communicate. The Social dimensions describe the nature of the relationship between the two primary parties and the streamed environment. This includes the roles that the users played in relation to the activity or task being performed, the sharing party's experience in their environment while engaged in telepresence, and how the receiving user interacted with both their partner and the shared environment. The second author also assisted in synthesizing the coded literature during results write-up, which served as an additional quality check for the coding process. In cases when the second author noticed discrepancies in coding, they were resolved among all team members and any necessary recoding was completed. In the next section, we present our results.

Table 1. Final Codebook Used to Guide our Qualitative Analysis

Themes	Dimensions	Description (*Codes*)
Study Meta-Data	**Primary Methodology**	The type of study conducted in terms of methodology (*Quantitative, Qualitative, N/A*)
	Study Location	Was the study conducted in the lab or the field? (*Lab, Field, N/A*)
	Study Duration	How long the study lasted (*Cross-sectional, Longitudinal, Repeated Measures*)
	Participant Quantity	Total number of participants in the study
	Study Dyads	Were participants paired in the study? (*Yes, No*)
	Country	The country where the study took place
Technical Factors	**Independent Variables**	The technical independent variables within the study
	Dependent Variables	The technical dependent variables within the study
	Symmetry	Equipment setup compared between Streamer and Viewer (*Symmetrical, Asymmetrical*)
Streamer Considerations	**Camera Type**	The video capture device used by the Streamer (*360 Camera, AR Glasses, Camcorder, Fisheye Camera, Smartphone, Tablet, Web Cam*)
	Camera Placement	The place where the Streamer's camera was situated (*Handheld, Head, Shoulder, Chest, Behind the Back*)
	View Stabilization	The method used to reduce jitter in the captured video stream (*Software, Hardware, None*)
Viewer Considerations	**Viewing Apparatus**	Type of video display used by the Viewer (*Computer Monitor, Smartphone, Tablet / Laptop, TV Display, VR HMD*)
	Communication Mode	The modalities afforded to the Viewer to communicate with the Streamer (*Audio, Live Video, Text Messages, Gestures, Gaze*)
	Gestural Communication	Technique used to let the Viewer provide hand gestures to the Streamer (*Graphic Overlay, Puppet / Avatar, Video Stream*)
	Gaze Conveyance	Technique used to let the Viewer provide their gaze heading to the Streamer (*Graphic Overlay, Physical Attribute, Spatial Sound, Vibrations, Video Stream*)
Social Factors	**Independent Variables**	The social independent variables within the study
	Dependent Variables	The social dependent variables within the study
	Proposed Relationship	The type of interpersonal relationship between Streamer and Viewer specified in the article (*Crowdsourced/Strangers, Friends/Family, Long-Distance Relationships, Generic*)
	Proposed Activity	The type of activity specified in the article which the participants were engaged in (*Classroom Attendance, Games, Shopping, Sports, Generic*)
	Agency of Streamer	The level of control negotiated by the Streamer (*Completely Cyranic, Collaborative Control, Completely Free*)
	Embodiment	Type of representation afforded to the Viewer (*None, Streamer is Embodiment, Video Embodiment, Robotic Embodiment*)
	Noticeability	The reason that brought attention to the Streamer (*Not Noticeable, Interaction Style, Wearable Device*)

4 RESULTS

Our qualitative analysis synthesizes the state-of-the-art of mobile, human-to-human, one-to-one telepresence. We first report descriptive statistics of the dataset in accordance with our final codebook, followed by a presentation of the Technical and Social factors found within the articles. Finally, we discuss the interplay between our technical and social factors, describing how social considerations are integrated into technical designs, and vice versa.

Table 2. Results of study meta-data from our codebook

Dimension	Code	#	%	Reference(s)
Primary Methodology	**Quantitative**	19	37%	[5, 10, 12, 14, 29, 48–50, 52, 57, 60, 76, 78, 81, 82, 86, 100, 139, 140]
	Qualitative	16	31 %	[6, 16, 27, 40–42, 55, 79, 83, 88, 93, 104, 105, 113, 118, 129]
	N/A	17	33%	[7, 10, 47, 51, 53, 61, 66, 74, 75, 79, 89, 99, 114, 123, 125, 130, 131]
Study Location	**Lab**	20	38%	[5, 11, 12, 14, 27, 29, 42, 48–50, 52, 60, 76, 78, 81, 86, 88, 100, 139, 140]
	Field	15	29%	[6, 16, 40, 41, 55, 57, 80, 82, 83, 93, 104, 105, 113, 118, 129]
	N/A	17	33%	[7, 10, 47, 51, 53, 61, 66, 74, 75, 79, 89, 99, 114, 123, 125, 130, 131]
Study Duration	**Cross-sectional**	34	65%	[5, 11, 12, 14, 16, 27, 29, 40–42, 48–50, 52, 55, 57, 60, 76, 78, 79, 81–83, 86, 88, 93, 100, 104, 105, 113, 118, 129, 139, 140]
	Longitudinal	1	2%	[6]
	N/A	17	33%	[7, 10, 47, 51, 53, 61, 66, 74, 75, 79, 89, 99, 114, 123, 125, 130, 131]
Study Dyads?	**Yes**	22	42%	[5, 6, 11, 12, 14, 16, 29, 40–42, 55, 57, 76, 80, 81, 93, 100, 104, 105, 113, 118, 129]
	No	13	25%	[27, 48–50, 52, 60, 78, 82, 83, 86, 88, 139, 140]
	N/A	17	33%	[7, 10, 47, 51, 53, 61, 66, 74, 75, 79, 89, 99, 114, 123, 125, 130, 131]
Article Country	**Australia**	1	2%	[66]
	Canada	10	19%	[6, 16, 41, 42, 93, 99, 100, 104, 114, 118]
	India	1	2%	[113]
	Japan	23	44%	[10–12, 14, 47–53, 74–76, 78–82, 88, 89, 123, 125]
	New Zealand	5	10%	[5, 7, 86, 139, 140]
	Singapore	1	2%	[130]
	UK	2	4%	[83, 131]
	USA	9	17%	[27, 29, 40, 55, 57, 60, 61, 105, 129]

4.1 Descriptive Statistics of the Articles

Our dataset (N = 52) was comprised of articles published between 2011 and 2020 (inclusive), but the majority of articles (85%) were published in 2014 or later. Fifty described artifacts, empirical studies, or a mixture of both, but we found 1 theory paper and 1 methods paper. As such, the majority of articles featured technology probes and evaluations. Thirty-five (67%) described a study as part of their contribution. Table 2 provides an index of codes related to these studies and the associated articles coded. All studies were cross-sectional, with only 1 exception, which was longitudinal. We did not find any studies with a repeated measures duration. The majority of these (20) were Lab-based, but 15 were Field-based studies.

In the lab-based studies, the primary methodologies were quantitative (17/20), and the remainder used qualitative methods. In the field-based studies, the authors primarily used qualitative methods (13/15), and the remainder used quantitative methods. All 19 of the quantitative studies described an artifact which was then evaluated in some way; 6 of these were evaluations of a new artifact without comparing against a control. The primary data collection approach in the qualitative articles were field notes (7) and interviews (12). Only 1 study used diary methods.

As human-to-human telepresence is contingent on an interaction between two parties, we would expect studies to involve the use of participant dyads. We found 22 studies to do so. In these cases, the quantity of participants ranged from 4 to 68 (M = 21.5, SD = 15.8). The authors typically recruited pairs of participants who knew each other, to have a more ecologically valid study. For the 13 articles only using 1 participant at a time, the number of users ranged from 5 to 26 (M = 12.9, SD = 6.59). These were geared towards understanding a single telepresence role, or evaluating an artifact. Convenience sampling was the primary method to recruit participants.

Overall, our dataset is comprised of 31 unique research projects across all research groups, with a heavy emphasis on the technological capabilities, and a minor emphasis on how people interact with others using these technologies. 29% of the articles were authored by 11 unique research groups, but 37 articles (71%) came from 5 research groups. The University of Tokyo contributed 15 articles, all of which were under advisement of Dr. Jun Rekimoto. Two major projects emerged from this subset – "JackIn Head" (9 articles [47–52, 75, 88, 89]) and "ChameleonMask" (4 articles [79–82]). Simon Frasier University contributed 10 articles, all of which were under advisement of Dr. Carman Neustaedter. Each of these articles were unique contributions, and 6 described field studies that garnered in-situ feedback. Microsoft Research contributed 5 articles, authored by Dr. Kori Inkpen, Dr. John Tang, or both. Waseda University contributed 4 articles, all of which were under the advisement of Dr. Jiro Tanaka. They were all related to one major project – "GoTogether" [10–12, 14]. Fx Palo Alto contributed 3 articles, all of which described the "Polly" project [57, 60, 61].

Scanning the articles for terminology, we found a wide variety of names for the user roles. The most common terms for the remote person sharing their environment, when looking at unique projects instead of articles, was "Local User"; others include Body, Surrogate, Guide, Host, Proxy, and Sharer. By the same fashion, the most common name for the person receiving the shared view was "Remote User"; other names included Ghost, Director, Operator, Consumer, Guest, and Spectator. We also found the term "Local" used to describe this party. This type of terminology depends on the point of view - after all, both users are technically remote in relation to the other. Other metaphors found across our dataset are vivid and help to convey the telepresence concept, but the various terminology might carry negative connotations. To circumvent this confusion, and to keep consistency, for the remainder of this paper, we use the terms Streamer and Viewer, as they are respectful to the human. Regardless of the reason or purpose, in mobile, human-to-human, one-to-one telepresence there will almost always be an individual who is sending the live-stream video (a Streamer), and an individual who is watching that video feed (a Viewer).

4.2 Technological Factors

In this section, we detail the results of the technological factors as found in our codebook. See Table 3 for an index of codes related to these factors, with the articles described by these codes. An emergent theme surrounding human-to-human telepresence work is that researchers tend to frame their technology probes around one of the interactants. They commonly look for new solutions to enhance the overall experience for the Viewer, or they consider new affordances that the Streamer could use to improve the social interaction. The result, however, is an increase in technological requirements to get this experience to work. In this section, we review these technological considerations which have emerged in the dataset to help understand the benefit each factor brings to the experience.

4.2.1 Technology-Centered Independent Variables. As previously noted, 35 articles (67%) in our dataset conducted a study, and of these, only 13 manipulated technical independent variables; we

Table 3. Descriptive results of technical features. Per dimension, articles which did not describe their code, or which were not applicable, are excluded from this table; percentages may not sum to 100%.

Dimension	Code	#	%	Reference(s)
Symmetry	**Asymmetrical**	41	79%	[5, 7, 10–12, 14, 27, 40–42, 47–50, 50, 52, 53, 55, 57, 60, 61, 66, 75, 76, 78–83, 86, 88, 89, 99, 105, 113, 118, 123, 125, 129, 139]
	Symmetrical	9	17%	[6, 16, 29, 74, 93, 100, 104, 114, 140]
Camera Type	**360° Camera**	16	31%	[10–12, 47–52, 66, 76, 88, 89, 114, 118, 140]
	AR Glasses	3	6%	[5, 7, 104]
	Camcorder	1	2%	[40]
	Fisheye Camera	1	2%	[75]
	Smartphone	15	33%	[6, 16, 27, 29, 41, 42, 55, 57, 60, 61, 86, 93, 100, 105, 113, 129, 139]
	Tablet	5	10%	[79–82, 123]
	Web Cam	6	12%	[14, 53, 74, 78, 83, 125]
Camera Placement	**Affixed to Prop**	1	2%	[16]
	Behind the Back	2	4%	[76, 118]
	Chest	6	13%	[6, 14, 41, 75, 105, 123]
	Handheld	12	23%	[27, 29, 40, 42, 55, 86, 93, 113, 114, 129, 139, 140]
	Head	19	37%	[5, 7, 47–52, 66, 74, 79–83, 88, 89, 100, 104]
	Shoulder	9	17%	[10–12, 53, 57, 60, 61, 78, 125]
View Stabilization	**Hardware**	3	6%	[57, 60, 61]
	Software	7	13%	[47–49, 51, 52, 88, 89]
	None	39	75%	[5–7, 10–12, 14, 16, 27, 29, 40–42, 50, 53, 55, 66, 74–76, 78–83, 86, 93, 100, 104, 105, 113, 114, 118, 123, 125, 129, 139, 140]
Viewing Apparatus	**Computer Monitor**	14	27%	[5, 27, 41, 42, 53, 57, 60, 61, 79–83, 113]
	Smartphone	8	15%	[6, 16, 29, 86, 93, 104, 105, 140]
	Tablet / Laptop	5	10%	[7, 40, 76, 118, 123]
	Television	3	6%	[50, 55, 129]
	VR HMD	16	31%	[10–12, 14, 47–49, 51, 52, 66, 74, 88, 89, 100, 114, 139]
Viewer Communication Mode	**Audio Only**	14	27%	[6, 14, 29, 47, 49, 52, 53, 55, 75, 83, 88, 89, 114, 118]
	Audio + Gaze	4	8%	[60, 60, 61, 76, 86]
	Audio + Gestures	1	2%	[125]
	Audio + Gaze + Gestures	4	8%	[10–12, 139]
	Audio + Live Video	15	29%	[16, 27, 40–42, 74, 81, 82, 93, 100, 104, 105, 123, 129, 140]
	Audio + Text Messages	3	6%	[79, 80, 113]
	Gaze	3	6%	[7, 50, 78]
	Gaze + Gestures	1	2%	[66]
Gesture Communication	**Graphic Overlay**	4	8%	[10–12, 66]
	Puppet / Avatar	1	2%	[125]
	Video Stream	9	17%	[74, 79–82, 100, 129, 139, 140]
	None	35	67%	[5–7, 14, 16, 27, 29, 40–42, 47–53, 55, 57, 60, 61, 75, 76, 78, 83, 86, 88, 89, 93, 104, 105, 113, 114, 118, 123]
Gaze Conveyance	**Graphic Overlay**	8	15%	[7, 10–12, 50, 66, 86, 139]
	Physical Attribute	5	10%	[14, 57, 60, 61, 78]
	Spatial Sound	1	2%	[88]
	Vibrations	1	2%	[76]
	Video Stream	4	8%	[74, 100, 129, 140]
	None	30	58%	[5, 6, 16, 27, 29, 40–42, 47–49, 51–53, 55, 75, 79–83, 89, 93, 104, 105, 113, 114, 118, 123, 125]

note that only 3 of these were field studies. As such, the more technical articles in our dataset generally evaluated the capabilities of their artifacts instead of deploying them to real world scenarios. One exception was in the work by Rae et al. who, as part of a general telepresence study, deployed two technology probes - one tablet propped up on a tripod, and one body-worn device. Their findings indicate that the body-worn device was logistically less demanding, and more conducive to viewing an environment (due to the mobile nature of the setup); however, both Streamer and Viewer sometimes felt awkward due to bystanders looking at them [105].

Concerning lab studies, in two articles by Kasahara et al., the authors manipulated the presence of video stabilization, specifically when the Viewer wears a Virtual Reality Head-Mounted Display (VR HMD). Naturally, Viewers scored the Simulator Sickness Questionnaire in favor of the stabilization [48, 52]. In three articles, various researchers studied user preference between regular video chat and panoramic video chat [29, 86, 139]. Each used the Networked Minds Measure of Social Presence questionnaire [8], and collectively found that the use of panoramic video streaming allows for more personable experiences with an increase of social presence. One step further, in an article by Young et al., the researchers found that viewing a 3D representation of an environment was superior to viewing normal panoramic video [140]. Matsuda et al. developed a wearable vibrotactile device that communicated to the Streamer where the Viewer was looking, and compared the use of this device against a similar setup that did not implement it; they found that their device was a convenient addition that allowed directional communication without speaking [76]. In the remainder of the studies, the authors developed multiple iterations of a technique or technology to tackle the same concept and compared them to each other. These included emotional state cues embedded in the telepresence experience [5], gaze-sharing visualization techniques [60], dual-video display type [100].

In summary, the technological IVs found within the dataset were sparse. Studies (particularly in the lab) were more evaluative of a specific piece of technology or compared the researchers' device to a control. In this way, the researchers did not manipulate an IV in their study but were still able to measure the effectiveness of their technology.

4.2.2 Technology-Centered Dependent Variables. Nineteen articles (37%) performed quantitative measurements to evaluate their technologies. Some of these DVs were direct measurements, such as error rate, task completion time, etc.; but others were questions in the form of psychometric scales. None of these technology-centered questions were derived from pre-validated surveys, and were instead custom items. We performed a short qualitative analysis to match these custom items to pre-validated measures based on content of the questions; we used a report by van Baren to help identify a list of related instruments [127]. We found most of the questions to map into the concepts of Usefulness, Ease of Use, and Ease of Learning; as such, the USE questionnaire would be a suitable measure [73]. The remainder mapped into the Sensory Factor concept of the Presence Questionnaire [135], by gauging audio and video clarity.

4.2.3 Capture and Display: Video Stream Considerations. The majority of our dataset (79%) described an interaction in which the technological setup was asymmetrical between the interacting parties. Our dataset supports the paradigm of one Streamer and one Viewer, echoing the notion that the interaction benefits the user who cannot be there in person. As such, the Streamer's apparatus mainly focuses on capturing the remote environment, and the Viewer's equipment is used to display it. A Streamer's camera setup typically falls into two categories: single lens devices such as those embedded in smartphones and tablets, and multi-lens devices such as 360-degree cameras. Because smartphones are an ubiquitous commodity and nearly every model has one or more cameras built-in, they are attractive solutions to generating video stream. By using a video chat application such as Skype and by pointing the camera outwards, any user can capture their surroundings. The Viewer, then, can view the feed on their own smartphone, or on a larger screen such as a desktop monitor. This paradigm was found in 63% of our dataset (e.g. [6, 27, 86, 123]). For a more immersive viewing experience, the Streamer can capture their environment with panoramic cameras, and the Viewer can use a VR HMD. With a 360-degree camera, a Streamer does not need to worry about camera rotation, as HMDs equipped with head-tracking capabilities grant a Viewer the ability to control the angle of their view. This paradigm was found in 31% of our dataset (e.g. [11, 49, 118]).

A small portion of our dataset (17%) described symmetrical technological setups between the interacting parties. In these cases, the common theme is that the two users are friends, families, or intimates who use telepresence to bridge the distance between them in their interpersonal relationships. Instead of only one person being unable to "be there" with the other, both users fit the bill. As such, they either take turns sharing their respective environments [6, 114], or they simultaneously share their surroundings while looking at those of their partner [16, 100, 104]. In this manner, both users are Streamers, and both users are Viewers.

4.2.4 Viewpoint Optimization: Stabilization. Regardless of camera type, in order to ensure success of the telepresence interaction, the video stream must be high quality in terms of viewability. There are multiple considerations that influence this viewability, including camera stabilization. As the camera is mobile and susceptible to jitter caused by the Streamer's walking motion, some authors note that camera stabilization is a necessity; however, only 19% of our dataset implemented a stabilizer, and these 10 articles only represent 2 unique projects. These projects show that there are two ways to stabilize a stream: with software and with hardware.

Kasahara et al. describe their JackIn Head project, in which a 360-degree camera is stabilized using software [48, 49]. The authors intend for their Viewers to use a VR HMD, and note that there is a high chance that they will experience discomfort referred to as simulator sickness [54]. Their stabilization technique calculates optical flow, which is then used across a series of video frames to determine how the camera shifts and rotates. By predicting these movements, a compensating transformation can be applied to a video frame before being displayed to the Viewer. The result is a smoothed and enjoyable view, but as the authors note, their technique essentially decouples the forward heading between the Streamer and Viewer; in other words, it is possible for the Viewer to lose sight of what the Streamer is looking at. However, they were able to show through a user study that their stabilization technique reduced simulator sickness [49].

While simulator sickness is typically associated with VR HMDs, Kimber et al. note that typical mobile video chat applications, e.g. on smartphones, can still produce shaky visuals that are too "unpleasant to watch" [57]. As part of a greater effort, the authors implemented stabilization in the form of a motorized gimbal. The electronic sensors detect subtle shifts in rotation, and it compensates to keep the camera upright and at a constant angle. Unlike software stabilization, this maintains coupling between the Streamer's and Viewer's forward headings. In their Polly project, the authors apply this hardware to a Streamer's setup as a wearable device that sits on the shoulder, allowing their smartphone cameras to produce a steady video feed [57, 60, 61].

4.2.5 Viewpoint Optimization: Streamer's Camera Placement. Depending on where the camera is situated, the contents of the stream can change. For instance, if the camera is worn on the head, the stream could be considered a first-person view of the environment. If it is worn on the shoulder, then the experience could help to facilitate a more social interaction by letting the Viewer see the Streamer's face. In our dataset, we found a variety of camera placements tried by researchers, though no article studied the effects of camera placement directly.

Handheld: A significant portion of our dataset (23%) described scenarios where the Streamer held the camera in their hands, and in most of these cases, the camera was embedded in a smartphone.

In only one of these, a 360-degree lens attachment was plugged into the phone, such that the Viewer could watch using a VR HMD [114]. In one other, researchers paired a smartphone with a 360-degree camera, to provide a hybrid view of the environment [140]. In the rest, the commonality is that all of the Streamers held the camera in front of them, so to aim it towards what they wanted the Viewer to see. Jones et al. studied how Streamers manipulate the camera when held in the hand, and reported that many users simply don't operate it in an ideal manner; e.g. they hold the camera too close to their face, or fail to provide an optimal view when trying to target a particular object [42]. They did describe that one of the more positive behaviors was Streamers panning the camera in a complete circle, allowing the Viewer to see the entire surroundings. This particular behavior was encouraged by Young et al., who developed a system that turns a smartphone panorama into a spherical image that the Viewer can navigate on their own [139]. The backbone of this technique was the inclusion of a Simultaneous Localization and Mapping (SLAM) algorithm to generate this panoramic image, which could then be viewed using any kind of device [86, 139].

Head-mounted: Multiple articles in our dataset noted that while handheld cameras are effective, they limit the activities that a Streamer can perform since their hands are busy holding the device [42, 57]. To overcome this limitation, other researchers have developed wearables, which strap, mount, or otherwise couple the camera to the body. The most common wearable camera placement in our dataset is on the head (37%), exhibited in 9 unique projects. Kasahara et al. developed a custom multi-lens 360-degree camera which was worn like a crown, and they noted that this position was chosen because it provides a first-person view [49]. Similarly, Lee et al. developed a helmet which had both a 360-degree camera and a webcam attached to the top. Here, the combination of cameras let the Viewer explore the entire surroundings and simultaneously know where the Streamer was looking [66]. Procyk et al. coupled a miniature camera to a pair of sunglasses, which essentially mimicked the Google Glass device. The camera therefore provided a first-person view, though the Viewer could not explore the surroundings on their own [104]. Although head-mounted cameras provide a first-person view, we note this also causes the Streamer to be effectively invisible to the Viewer most (if not all) of the time. To accommodate a more social interaction, where the Streamer's face might be seen by the Viewer, a different placement could be utilized, such as on the shoulder.

Shoulder-Mounted: The next most common wearable placement was on or over the shoulder (17%), accounting for 5 unique projects. First, we found two unique efforts that described a miniature robotic avatar that sat on the Streamer's shoulder; the avatars were each fitted with actuated features to convey more natural communication. For example, TEROOS by Kashiwabara et al. had eyes and robotic eyebrows to convey emotion [53], and the 20DOF Humanoid described by Tsumaki et al. had little arms to convey gestures [125]. These avatars are somewhat useful in conveying the Viewer's natural communication, but most of the other works with shoulder-mounted cameras focused on conveying the Streamer's gestures and gaze. As previously mentioned, Kimber et al. developed a shoulder-worn device which stabilized a smartphone with a gimbal. The gimbal also served another purpose, which was to grant the Viewer with a way to control the rotation of the camera through a GUI [57, 60]. The unique benefit of this camera placement is that the Viewer can rotate the device towards the Streamer's face, so to have a more social interaction. This is reflected in various works by Cai et al., who used a 360-degree camera mounted over the Streamer's shoulder [10–12]. In their scenarios focusing on collaboration, this camera placement afforded the Viewer with a way to see where the Streamer was looking, as well as a way to see what their hands were pointing at, so to provide the Viewer with a better understanding of what their telepresence partner was focused on.

Chest: We found six other articles (12%) which described placing the camera somewhere on the Streamer's chest. None of these used 360-degree cameras, as the back portion of the viewpoint would be blocked by the Streamer's body; but, this position is highly conducive to understanding the Streamer's immediate environment. In our dataset, this position was chosen for a variety of reasons. Ishak et al. described how chest-mounted cameras were "not overly obtrusive" compared to head-mounted cameras [41], and further, Baishya and Neustaedter noted that by placing a camera in the Streamer's front shirt pocket, the device can be embedded in clothing while not bringing attention to the user [6]. Unique works by Chang et al. and Matsuda et al. describe how, by hanging a rig around the neck of the Streamer, the camera can rest around the chest while providing maximal comfort compared to other placements [14, 75].

Behind the Back: The final bodily camera placement found in our dataset was behind the Streamer's back, accounting for two unique articles (4%). To accommodate this position, the Streamer typically wears a backpack which is outfitted with a pole, hoisting up the camera. The resulting perspective is akin to a third-person view of the Streamer. In one article by Tang et al., the camera was a 360-degree device, allowing the Viewers to explore the entire surroundings. The authors employed a behind-the-back camera for a remote collaboration scenario, and they chose this position because it provides a Viewer with a way to understand where the Streamer is looking (by seeing their head), in addition to freeing up their hands for gesturing and manipulation [118]. In the other article by Matsuda et al., the authors developed a custom pipe rig to hold up the camera. In their work, they also developed a vibro-tactile device that conveyed 360-degree directions using vibration motors; this allowed the Viewer to see the complete surroundings and "point" to where they wanted the Streamer to look [76].

The various camera placements found in our dataset have been chosen for different reasons, though we note that none of the articles studied different camera placements directly. The Viewer's opportunity to explore the environment is enhanced with a head-worn or chest-worn camera, but their social experience may be greater when a handheld or shoulder-worn camera is used, because they can see the Streamer.

4.2.6 Communication Modalities from Viewer to Streamer. We found a variety of modalities used to support the Viewer's communication to the Streamer, with some articles providing a mixture of multiple channels. Unsurprisingly, verbal communication was supported in 81% of the dataset. Text messages were rarely used (6% of the dataset), but when included, they were auxiliary modes of communication [79, 80, 113]. In certain setups where the Streamer used a smartphone, a live video feed of the Viewer was displayed in the remote environment; this was found in 29% of the dataset. Interestingly, we did not find any articles that used emojis or emotes, most likely due to our scoping criteria; we focus on one-to-one interactions, where paralinguistic digital affordances are more appropriate for broad-audience interactions [35].

Besides verbal communication, humans often perform gestures such as pointing and head nodding in order to provide context to their spoken communication. As reported in a variety of articles, these cues are typically lost in telepresence experiences because the Streamers and Viewers often cannot see each other [42, 118]; as described above, some of the researchers noted that they selected a specific camera placement because it allowed the Viewer to see the Streamer's head and hands, helping to overcome one piece of the problem. The use of a video conferencing application on a smartphone also lends itself to supporting natural communication between the Viewer and any third-party members, but because the camera (and screen) are often pointed outwards, the Streamer typically cannot see the Viewer. Thus, conveying natural communication from the Viewer to the Streamer is a more difficult challenge. Many articles attempted to overcome this limitation with new technology probes to support hand gestures, convey gaze, or both.

Hand Gestures: We found fourteen articles (27%) to support the Viewer's hand gestures during the telepresence experience. One of these articles by Tsumaki et al. leveraged a miniature humanoid avatar which sat on the Streamer's shoulders, equipped with robot arms to convey gestures [125];

though not specified by the authors, it is possible to harness three-dimensional user interfaces (3DUI), e.g. the Microsoft Kinect, to capture real arm movements and map them to the robot [64]. Four unique projects across 7 articles (13%) described integrating a video stream of the Viewer's hands. In the ChameleonMask project, Misawa and Rekimoto used computer vision to detect the hand and display it to the Streamer. Here, the unique aspect of the project is that the Streamer is wearing a helmet which has both an HMD and a tablet display; within the HMD, the Streamer can see the Viewer's hand gestures. The third-party members can see the Viewer's face on the tablet, but cannot see the their hands, and are instead asked to interact with the Streamer's arms (such as shaking hands) [79–82]. Similarly, Young et al. extracted the Viewer's hands while they explore the Streamer's surroundings with an HMD (their smartphone is the display, similar to Google Cardboard); as they naturally point to an object, their phone picks up their hand and overlays it to the Streamer's feed. Though the Streamer generates the telepresence image, both users interact with it; the authors refer to this as a "shared spherical mapped environment" [139].

Four other articles (8%) used a similar technique, but instead of displaying the Viewer's actual hands, the authors instead provide a virtual representation. Lee et al. and Cai et al. discuss how they used 3DUI techniques to detect the Viewer's hand while they wore an HMD (here, both use the Leap Motion device). The hand position and rotation in the Viewer's local space is then mapped to a 3D model, which is displayed to the Streamer as a virtual overlay. In order to achieve this, the Streamer was equipped with an AR display [10–12, 66]. In these examples, the Viewer's hands must be in front of their face in order to be detected, so the hand gestures are dependent of gaze. As a natural extension, with an adequate tracking solution (e.g. HTC Vive's outside-in tracking), a Viewer's hands could be detected independent of gaze. Therefore, the virtual overlay solution might be more efficient when the Viewer wants to point at one or more objects which do not fall within their gaze heading. However, it remains to be seen which method is more efficient at communicating hand gestures, or which technique is more preferred by the Streamer.

Gaze Conveyance: Nineteen articles (37%) supported some type of gaze conveyance from the Viewer to Streamer. One strategy found in 8 articles (15%) is the inclusion of a virtual representation. Similar to cases where the authors wanted to display hand gestures using a 3D model, the Streamers wore an AR display, and certain cues were included to show where the Viewers were looking. For instance, Cai et al. displayed a virtual head which looked towards the direction of where the Viewer was gazing. Thus, depending on the rotation of this head, the Streamer could tell where their partner was looking [10, 11]. Lee et al. took a different approach, by displaying a simple rectangle that represented the region where the Viewer was looking. Additionally, if both users were looking in opposite directions, an arrow would be displayed to indicate where their partner was gazing [66]. Billinghurst et al. developed a similar approach that they call a "radar", where virtual overlays indicate the direction of the other user's gaze [7]. In these articles, the Streamer was using a 360-degree camera so that the Viewer could explore the entire surroundings; but, gaze conveyance is still beneficial to panoramic smartphone users, as shown by Muller et al. In their work, by leveraging rotational tracking of the phone, the authors were able to detect where on the panorama each user was looking [86].

Other researchers concluded that it is pertinent to be able to show the direction of the Viewer's gaze, but they did not use an AR display. Five articles (10%) spanning 3 unique projects relied on a physical attribute of the telepresence setup instead of a virtual representation. As previously described, in the Polly project, a shoulder-mounted smartphone is attached to a gimbal; using a GUI, the Viewer is able to manually control the rotation of the phone. Thus, the Streamer can look at the phone to infer where the Viewer is looking [57, 60, 61]. Likewise, Chang et al. used a webcam which was worn around the Streamer's neck; it was actuated such that the Viewer can change the direction of the camera, and the Streamer could thus determine where their partner was looking

[14]. One unique project by Misawa et al. leveraged a miniature, face-shaped screen which sat on the shoulder of the Streamer. The screen was 3D and a back-projector displayed the Viewer's face. As the authors describe, this technique is superior to a flat screen since it overcomes the "Mona Lisa effect" [78].

Three articles (6%) used raw video stream in order to convey what both users were looking at. In two of these, the users acted as Streamers and Viewers simultaneously, and their viewpoints were overlapped as both users wore a VR HMD. For instance, Pan et al. used this technique in order to bridge the physical distance between partners in a long-distance relationship (LDR) [100], and Maeda et al. used this technique as part of a larger project, theorizing that the Viewer could learn motors skills by watching the Streamer's first-person view while performing some task [74]. In a different article by Young et al., a SLAM algorithm was used to reconstruct an environment in 3D which was shared between two users; then, live video streams were spatially mapped in this reconstruction, so the other user could see exactly where the other was looking.

In one article, Nagai et al. used spatial audio in order to convey where the Viewer was looking. The Streamer, equipped with a proper auditory display such as surround-sound headphones, would hear the voice of their partner coming from a specific location, which changed depending on their gaze. Using rotational tracking, it is possible to calculate which direction the audio needed to come from [89]. Finally, one article by Matsuda et al. described using an array of vibration motors resting upon the Streamer's neck in order to subtly convey where the Viewer was looking; depending on which motor was being activated, the Streamer could "feel" the Viewer's gaze direction.

4.3 Social Factors

While most of the articles in our dataset were contributions from more technical groups, we found some emergent themes within the social realm as well. The foundational goal of telepresence is to grant the feeling of "being there," but the H2H telepresence paradigm transcends sensory stimulation, and thus this goal cannot be achieved unless social needs of the Streamer and the Viewer are also met. In this section, we review the social factors identified from our dataset and describe how each of these have been tackled in a variety of ways. See Table 4 for an index of codes related to these factors, with the articles described by these codes.

4.3.1 Human-Centered Independent Variables. As previously noted, only 35 articles (67%) in our dataset conducted a study, and only 9 of these used social factors as independent variables. Of these nine, four described a field study, but all were cross-sectional. In other words, none studied the long-term effects of various telepresence usage. The most common social IV was Scenario, in that the researchers deployed their technology probe for different activities, but this did not seem to have an effect on the reported findings [42, 48, 49, 105]. One manipulated Acquaintanceship, where the third-party participants were asked to shake hands with either strangers or friends, in a social touching experiment [82]; here, the results indicated very little effect of this IV. Since in natural communication humans can infer where another is looking by observing their body language, another work attempted to see if people can still determine this by observing the physical characteristics of their custom telepresence setup Polly [60]; although the time it took to make this consideration was statistically significant in favor of a human's body language, the authors note that it was still relatively fast with their prototype. Lastly, two articles describing one study manipulated whether two users video-streamed their own surroundings simultaneously, or if one of the users was watching from home, in a bicycling scenario; the authors noted that the experience was more engaging if both users shared their own surroundings to the other opposed to one watching from home [16, 93].

Table 4. Descriptive results of social aspects. Per dimension, articles which did not describe their code, or which were not applicable, are excluded from this table; percentages may not sum to 100%.

Dimension	Code	#	%	Reference(s)
Proposed Relationship	**Crowdsourced**	5	10%	[27, 41, 81, 82, 113]
	Friends/Family	14	27%	[11, 12, 16, 29, 40, 42, 55, 57, 78, 93, 104, 105, 118, 129]
	Generic Use	30	58%	[5, 7, 10, 14, 47–53, 60, 61, 66, 74–76, 79, 80, 83, 86, 88, 89, 99, 123, 125, 130, 131, 139, 140]
	Long-Distance Relationships	3	6%	[6, 100, 114]
Proposed Activity	**Classroom Attendance**	1	2%	[41]
	Games	1	2%	[113]
	Generic Use	31	60%	[5–7, 14, 29, 40, 42, 47–50, 52, 53, 60, 61, 66, 75, 79–82, 88, 89, 99, 100, 114, 125, 130, 131, 139, 140]
	Shopping	3	6%	[10, 11, 78]
	Sightseeing / Tours	10	19%	[12, 27, 55, 57, 76, 83, 86, 105, 118, 129]
	Sports	6	12%	[16, 51, 74, 93, 104, 123]
Agency of Streamer	**Collaborative Control**	21	40%	[10–12, 14, 27, 41, 42, 53, 55, 57, 60, 61, 66, 74, 76, 86, 105, 113, 118, 123, 129]
	Completely Cyranic	5	10%	[79–83]
	Completely Free	13	25%	[5–7, 16, 29, 40, 93, 99, 100, 104, 114, 139, 140]
	Indeterminate	13	25%	[47–52, 75, 78, 88, 89, 125, 130, 131]
Viewer's Embodiment	**No Embodiment**	26	50%	[5–7, 10–12, 14, 27, 47–52, 66, 75, 76, 86, 88, 89, 100, 104, 113, 114, 118, 139]
	Streamer is Embodiment	5	10%	[79–83]
	Robotic Embodiment	4	8%	[53, 74, 78, 125]
	Video Embodiment	14	27%	[16, 29, 40–42, 55, 57, 60, 61, 93, 105, 123, 129, 140]
Noticeability	**Not Noticeable**	4	8%	[5, 6, 50, 83]
	Interaction Style	12	23%	[27, 29, 40, 42, 86, 93, 104, 113, 114, 129, 139, 140]
	Wearable Device	33	63%	[7, 10–12, 14, 16, 41, 47–49, 51–53, 55, 57, 60, 61, 66, 74–76, 78–82, 88, 89, 100, 105, 118, 123, 125]

4.3.2 Human-Centered Dependent Variables. Regarding the measurable dependent variables in the articles which report a study, 16 articles (31%) performed quantitative measurements that were human-centered in nature. Nine of these articles used pre-validated measures; these included the Simulator Sickness Questionnaire [54], Networked Minds Measure of Social Presence [8], Social Presence Questionnaire [112], Spatial Presence Questionnaire [110], and the Affective Benefits and Costs of Communication Technologies questionnaire [138]. However, we did find the strong majority of measures to be custom questions. We performed a short qualitative analysis to match these custom items to validated measures based on content of the questions, using a report by van Baren to help identify a list of related instruments [127]. We found many of these to map to the concepts provided by the Networked Minds Measure of Social Presence [8]; but, others mapped more accurately into the Social Presence concept as measured in the GlobalED questionnaire [28]. According to Cui and as reported by Oh et al., and as manifested in our dataset, the term "social presence" seems to be an umbrella term used to describe multiple concepts [18, 98]. Other questions mapped into various concepts such as Usefulness, Satisfaction, and Ease of Use as measured in the USE questionnaire [73], the Control concept measured in Witmer's & Singer's Presence Questionnaire [135], and the Immersion concept from the Slater-Usoh-Steed questionnaire [115]. As expected, researchers in this area tend to attempt to measure social presence, as that is one of the ultimate goals of H2H telepresence, but the dataset indicates that there is a need for more standardized and validated questionnaires when attempting to measure these constructs.

4.3.3 Proposed Interpersonal Relationships. While mobile robotic telepresence affords an individual with a way to feel present in a remote environment on their own volition, H2H telepresence grants this presence by leveraging the Streamer. We look to our dataset in order to understand who this Streamer can be, and how they relate to the Viewer. While most of the dataset did not propose their work for a specific relationship context (58%), we did find twenty-two articles (42%) to discuss or apply their work for one, including crowdsourced/strangers, family & friends, and LDRs.

Friends and Family: The most common of these were for friends and family, represented across 14 articles (27%). Although telepresence lends itself to shortening large distances between two parties, we find it is still useful to employ this interaction for cases where one member simply cannot be with another during the day. For instance, Procyk et al. describe an activity where two users enjoy geocaching; instead of collaborating to find a single geocache, the users split up to hunt two different ones while sharing their progress to the other [104]. Chua et al. describe a similar technology probe for biking, where either two parties take separate paths and shared their surroundings with their partner, or one party would stay at home and watch the video feed of the other [16]. Kim et al. deployed a technology probe to understand how one person could shared their everyday experiences to someone who is watching from home [55]. Further, Kimber et al. discuss how telepresence can be used to let disabled users experience the world through the eyes of a Streamer [57]. Between these articles, we find that friends and family members have opportunity to enhance their relationships by performing joint activities remotely, or at least by sharing their experiences to a loved one who cannot be with them due to geographical distance or even health concerns.

Long Distance Relationships: There is a significant body of work regarding in-home computer-mediated communication for long-distance relationships, and we found 3 articles (6%) in our dataset to specify this relationship context for mobile situations. Similar to the above, these 3 articles followed 2 main applications for the technology probes. Pan et al. described a case where both users wore an HMD and could see their partner's first-person view overlapped onto their own. Though this study was performed in a lab with makeshift activities, both users reported a sense of togetherness, which is crucial in an LDR, so the concept could easily be applied to real-world activities [100]. Baishya and Neustaedter show how by leveraging "always-on video chat" two users can see what their partner is doing on a whim; so, if they cannot be together, they can at least have a sense of understanding what the other is experiencing [6].

Crowdsourced/Strangers: The final interpersonal relationship proposed by authors in the dataset was crowdsourced/strangers, or the absence of a developed interpersonal relationship. Five articles (10%) described cases where they anticipate the Viewer to have a specific goal in mind, and the Streamer helps to achieve those goals. de Greef et al. hypothesized the ability to link up a disabled or hospitalized Viewer with a random Streamer to facilitate remote sightseeing [27], and Misawa and Rekimoto discuss the possibility of a "physical body marketplace" where the Viewer can enlist the help of an unknown Streamer to perform remote tasks for them [81, 82]. In work by Singanamalla et al., Viewers tell Streamers where to go as they navigate an "Escape Room". Similar to Misawa and Rekimoto, the authors describe their vision of Streamers earning money by performing remote tasks for Viewers [113]. Ishak et al. performed a similar study, hypothesizing that in the event a student could not attend class, a Streamer could be physically present while streaming the contents to the Viewer [41]. These works demonstrate that telepresence does not necessarily hinge on both parties knowing each other prior to the interaction. Streamers could be compensated for their time, akin to ride-sharing services like Uber or Lyft. However, it remains to be seen what activities are appropriate for interactions involving strangers.

4.3.4 Proposed Telepresence Activities. We analyzed the articles in our dataset to understand any trends in activity that the Streamer and Viewer could be involved in as part of the telepresence interaction. Most of the dataset (61%) did not deploy their technology for a specific activity, but we did find the rest to discuss a handful of interesting applications.

Sightseeing / Remote Tourism: One of the most prominent activities in our dataset is sightseeing / remote tourism (19%). Typically geared towards friends and family, these articles helped bridge the distance between loved ones. Muller et al. described their method of sharing an interesting environment by using a smartphone to generate a panorama, which the other user could explore on their own volition [86]. Rae et al. conducted a field study in which they asked Streamers to walk in an indoor museum [105]. In a field study by Kim et al., participants used their technology probe for going to the beach, touring their city's Chinatown, and hanging out at a park, among other activities [55]. In essence, all of these articles describe a situation where the Viewer could understand the environment that their Streamer partner was in, during a slow-pace activity. However, we did find some other articles which discussed applying telepresence to more intense scenarios.

Sporting Activities: Six articles (12%) proposed the application of telepresence towards more action-based scenarios. Tobita hypothesized that the Viewer could take a passive approach by acting as a "coach," giving advice to the Streamer who is engaged in playing tennis [123], and Kasahara and Rekimoto suggested that Viewers could watch a stream of paragliding to feel like they were flying (although they did not directly try it in their article) [51]. Other researchers, using a symmetrical setup, allowed both users to act as a Streamer and Viewer simultaneously, providing a way to perform joint activities remotely. Procyk et al. used such a paradigm for joint geocaching [104], and Chua et al. used it for bicycling [16]. In a unique case, Maeda et al. hypothesized that telepresence with a first-person view could allow a Viewer to learn muscle memory by watching the Streamer. In their work, a Viewer attempted to juggle by watching how the Streamer – an expert – was able to perform [74].

Other Various Activities: In the remaining five articles, the authors hypothesized a variety of other activities, including classroom attendance [41], shopping [10, 11, 78], and games [113]. In the shopping scenarios, the Viewer has more input in the overall interaction; at times, the Streamer even asks for help from the Viewer to find an object on the shelf. As evident in our dataset, the current research focuses less on *what* this type of telepresence can be used for, and instead focuses more on *how* telepresence interaction can take place. Most of the articles in our dataset did not specify a particular activity, or hypothesized one or many without performing a study in that context. Although the various contexts of relationship and activity might drive telepresence design choices, we expect that any given user might have their own wants and needs. Rae et al. performed a survey to understand what end users might want general telepresence for, and they found their respondents to specify a wide variety of use cases which fall within the purview of H2H telepresence, including sightseeing, shopping, and sports, as shown above. Other activities not found in our dataset include conference attendance, major life events, and enjoying a meal together [105].

4.3.5 Agency of Streamer. An emergent theme found in our dataset is that the Streamer was tasked to behave in a variety of ways, and were sometimes asked to relinquish their free will in order to achieve the goals of the Viewer. These behaviors do not stem from a particular technological setup, but rather from social expectations regarding how the telepresence interaction should work. While we could not determine any behavior in 13 articles (25%) due to their more technology-focused nature, we found three main behaviors emerge from the rest of our dataset, following a spectrum of agency.

Completely Free: On one end of the spectrum, the Streamer has complete agency and behaves as normal; this was found in 25% of our dataset. In these cases, the user turns on their video streaming apparatus and the Viewer simply watches without giving much direction. In an intimate scenario, Baishya and Neustaedter describe the use of "always-on video chat" in order to provide members of an LDR with a first-person view of their partner's environment. Here, the users might not be aware that their partner is watching, and they go about their daily lives as normal [6]. Procyk et al. describe simultaneous video-streaming between a pair of users, so that one could see the viewpoint of the other as they both engaged in geocaching for recreation [104]. Various works discussed video-streaming various activities while a Viewer watches from home. For instance, users streamed their biking experiences in the work by Chua et al. [16, 93], and parents were able to stream their children's playtime in work by Inkpen et al. [40]. In summary, the Viewer can watch the Streamer's video feed in order to feel like they are "along for the ride;" but, for a more social experience, collaboration is often employed.

Collaborative Control: In the middle of the spectrum, twenty-one articles (40%) described letting the Streamer maintain their agency while taking Viewer feedback into consideration. In these articles, the common theme is that the Streamer and Viewer work together to complete some task. For instance, Tang et al. described a field study where participants were asked to take pictures of specific landmarks; here, the Viewer has prior knowledge of these landmarks, and provided input to the Streamer in the form of directions. The Streamer still acted normally, while taking into account the Viewer's input [118]. Likewise, Cai et al. described a scenario where a Streamer physically goes to a store, and the Viewer uses the telepresence interaction to help find specific items on the shelves. In this case, the Viewer can be providing a guidance-like service to the Streamer, but the Streamer could also be shopping for the Viewer on their behalf [10, 11]. Venolia et al. describe a field study in which one person remained home, but received a video stream as their loved ones explored a museum. Here, the Viewers were able to freely communicate with the Streamers, and the authors employed a technique which allowed both parties to take a picture together, so that it seemed like two parties were actually together. By using a camera and green screen technology, the Viewer's background was stripped, such that their image could easily be inserted into a live picture [129]. Ishak et al. describe a scenario where the Streamer attends class for the Viewer. Here, the Streamer might not even be enrolled in the class, but can be a conduit for the Viewer to virtually attend when they cannot physically be present [41]. Thus, collaboration can be used to achieve the goal of either user – or both.

Completely Cyranic: On the other end of the spectrum, we found 5 articles (10%) spanning 2 unique projects that discuss how the Streamer purposefully relinquishes their agency in order to achieve the task of the Viewer (a cyranic interaction). The term "cyranic" has been used in prior work to describe a social encounter where one person acts and speaks exactly as told via discrete commands of a remote person; as described by previous research, this mirrors a concept found in the play "Cyrano de Bergerac" where the titular character secretly tells another exactly what to say (see [17]). One article in our dataset by Mitchell et al. applied a literal cyranic interaction in an effort to understand how the third-party interactants reacted to this concept [83]. A Streamer engaged in conversation with these third-party members while speaking exactly as the Viewer commanded. The authors noted that the participants started off wary of the strangely-behaving human, but eventually understood and accepted the paradigm. In one other project, Misawa and Rekimoto deployed their ChameleonMask, in which a Streamer wore a tablet device on their face which displayed the Viewer, so to provide the illusion that they were the Viewer [79, 80]. In this project, the Viewer provided discrete audio messages to the Streamer, who behaved exactly as commanded; then, the Viewer spoke using a public audio channel to the third-party interactants. The authors hypothesize that this paradigm can be used in an Uber-like service, where a Streamer is paid to let the Viewer "borrow" their body [80, 81].

4.3.6 Viewer Embodiment. In robotic telepresence, the Viewer is typically afforded a wheeled platform which grants them a physical representation in the remote environment; but in H2H telepresence, this embodiment is typically lost. We looked to our dataset to understand not only how Viewers can perceive a remote environment, but how the environment can perceive the Viewer. In our dataset, we found 26 articles (50%) to not provide any type of embodiment to the user. In these cases, the Streamer uses their camera to provide a viewpoint of their surroundings, but the Viewer's communication, if any, does not reach anyone but the Streamer. For instance, in the JackIn Head project, the Streamer wears a custom 360-degree camera on their head, and does not have a display with which to project the Viewer back into their environment [49]. This form of telepresence does not lend itself for many social interactions, but other researchers have explored a variety of ways to overcome this challenge.

Robotic Embodiment: In some of the earlier works found in our dataset, researchers tried to use robotics to provide a physical avatar which is tethered to the Streamer; this type of interface was found in 4 articles (8%). The most prominent cases were described by Kashiwabara et al., who created TEROOS – a camera with eyes and actuated eyebrows [53] – and Tsumaki et al., who made their humanoid robot which had actuated arms [125]. All of these articles were published prior to 2013, and the state of the art telepresence has since then shifted towards use of more commercial hardware (e.g. smartphones and video chat, or 360-degree cameras).

Video Embodiment: Fourteen articles (27%) discussed the use of a camera embedded into a smartphone or tablet to generate the live-stream. A collateral outcome of using these devices is the ability to pipe the Viewer back into the remote environment; in this way, the Viewer is embodied by the video device. In the Polly project by Kimber et al., the Viewer is not only displayed in the remote environment, but they also have the ability to control the physical device's orientation [57]. Likewise, in the work by Ishak et al., the Streamer wears a smartphone on their shoulder, allowing the Viewer to appear in a small window. Though some Viewers thought that the smartphone screen was too small (and thus did not provide them with an adequate sense of embodiment), others indicated that they wanted to keep a low profile [41]. This technique may be effective in situations where family members are geographically distributed. In a work by Inkpen et al., a remote family member could watch young children play, and the kids could see the faces of their loved ones [40].

Streamer is Embodiment: The final type of representation found in our dataset relied on the actual Streamer to serve as the Viewer's surrogate, and thus serve as their embodiment. Five articles (10%), found mainly in the ChameleonMask project, described the Streamer acting directly on behalf of the Viewer (see 4.3.5 Agency of Streamer, above). Misawa and Rekimoto discuss having the Streamer wear a tablet device on their face, and the video provides an illusion that the Viewer is the one present before a third-party interactant. Although this concept technically fits in the video embodiment category, the authors add an additional layer by having the Streamer perform physical actions as directed by the Viewer [81, 82]. Mitchell et al. describes their scenario in which the Streamer acts completely and intently on the Viewer's behalf, as a literal "avatar"[83].

4.3.7 Noticeability. An emergent theme in multiple field studies found in our dataset was the social awkwardness of the Streamer as they used a telepresence setup. We would anticipate this finding to manifest when body-worn devices are used, but even more modest technology probes led to this negative result. For instance, Rae et al. deployed a shoulder-worn smartphone, and some Streamers felt like they were being stared at [105]. Likewise, participants in the field study of Kim et al. noted that even simple video chat resulted in a higher level of embarrassment. The authors here note that a design implication is to make a technology probe as inconspicuous as possible [55]. We thus analyzed the proposed technologies in each article of our dataset to understand how noticeable they were. We found that 45 articles (87%) described a noticeable interaction, either due to the fact that the Streamer is wearing equipment in a conspicuous way (33 articles), or because passersby were able to see or hear the telepresence taking place (12 articles, e.g. the Viewer's audio was not localized to the Streamer). Only 4 articles had inconspicuous setups, and in each of these, the camera was embedded into the clothing of the Streamer in a way that did not draw attention to them. For instance, Baishya and Neustaedter had the users place their smartphone in their front shirt pocket [6], and Procyk et al. generated the stream by attaching a small camera to the frame of a pair of sunglasses [104]. In these examples, the clothing itself was not conspicuous, and the cameras were well-hidden. We note, however, that this does not completely solve this problem of social awkwardness; one of the bystanders in the work by Procyk et al. believed that the Streamer was talking to them (when they were actually talking to the Viewer), and a fight almost ensued [104]. We do not yet understand what the best way to convey that a telepresence scenario is occurring while balancing the needs of all parties involved.

4.3.8 Emergent Themes from Qualitative Findings. Although most of the articles in our dataset were contributions from the more technical side of HCI, we did find some important qualitative findings and emergent themes, which we feel warrant an echo in our paper.

Viewer Dependence on the Streamer: In some articles, the Viewer participants indicated that the interaction was simply not good enough. Some users thought watching a video stream became stale after extended use [16], and while some felt that telepresence was a decent substitute, it was not as good as actually "being there" [104]. This might be due to the fact that generally, the Viewer cannot control where the Streamer goes [118]. Other participants reported that they did not enjoy being dependent on the Streamer to look around the environment, as it made them feel like they were sitting "in a wheelchair" [105]. From a Streamer's point of view, having to control the camera for their partner made users feel like a prop, resulting in them feeling "used" [105]. This theme of dependence was typical of technical setups which leveraged single-lens cameras, e.g. smartphones; though not exclusive to smartphones, we find this feeling of dependency to be reduced when using a 360-degree camera which affords the Viewer with a way to explore the environment on their own accord.

Privacy Challenges for All Users: There was a minor theme regarding privacy. Many of the Viewer participants were unable to adequately gauge the social surroundings of the Streamer, and were found to modulate the strength of their audible communication while being conscious of the words they were saying, so to not offend anyone in the remote environment [104, 105]. Further, some were not convinced that the live-stream was secured to just their use, and again were conscious with their utterances. In a longitudinal field study for members of a long-distance relationship, Baishya et al. deployed an "always-on" video chat solution, so that one person could see their partner's view on a whim; even in an intimate relationship, this on-demand interaction caused a degree of uneasiness, as the users did not have the ability to prevent their partner from seeing or hearing certain situations [6]. Regarding privacy of third-party members, we did not find much consideration, but most of the field studies were conducted in North America, where video-streaming in public seems to be more accepted than other parts of the world. However, this is certainly an issue which needs to be addressed with future work.

Concerns about Streamer Safety: We found another minor theme regarding the safety of the Streamer. As previously mentioned, a passerby felt offended by one of the participants, who was actually talking to their telepresence partner. This almost caused a fight [104]; but worse, one

Streamer was too enraptured by the hand-held technology probe that they were almost hit by a car [42]. In one case, the telepresence interaction was too inconspicuous that it provoked a third-party member; in the other case, it was too involved that the Streamer nearly lost their life. As noted above, a strong portion of the articles discussed a technology probe that was noticeable to third party members. Even when engaging in simple video chat, Streamers felt a sense of awkwardness [55, 104, 105]. They were self-conscious about using the technology probes; in these cases, physical harm was not apparent, but psychological harm was. It is clear that too obvious of a technology probe may cause third-party members to take notice, but hidden devices which cause a noticeable interaction still cause confusion for these individuals. There is a clear need for balance, in order to maximize the needs and desires of all parties involved.

Preferred Streamer Characteristics: Lastly, we also found a minor theme regarding the characteristics of an “ideal” Streamer. In a survey by Ishak et al., survey respondents were asked to help identify what matters when choosing a Streamer for the hypothetical task of remote classroom attendance. They found that students did not particularly have a preference for characteristics such as gender, race, or age, and instead indicated a desire for certain performance-based traits, such as extroversion, good note-taking skills, and academic standing [41]. On the other hand, Misawa and Rekimoto, in their technology probe where a Streamer is presented like the Viewer, suggest that to enhance the illusion that the Viewer is the one interacting with third-party members, the Streamer should have identical physical traits to the Viewer [79]. For non-performance-based use cases, it seems that Streamer characteristics are not a major consideration; for instance, the purpose of remote sightseeing is to share an environment, and generally, we suspect Streamer traits will not affect user experience. However, for performance-based scenarios such as sporting or remote shopping with a stranger, certain traits such as languages spoken, physique, and knowledge of the environment might influence the success of the interaction.

4.4 Interplay between Technical and Social Factors

The previous sections describe the technical and social factors of H2H telepresence design separately, but in our analysis, we find that these are very intertwined. Social considerations breed technical considerations, and vice-versa, but each of these stems from an attempt to optimize a certain aspect of the overall user experience. In this section, we discuss the links between these factors and how they influence the design of telepresence experiences.

4.4.1 Technology Breeds Social Considerations. When designing an H2H telepresence experience, researchers might think that adding more physical technology can be used to improve certain aspects of the interaction. To some extent, this is certainly true. However, our dataset informs us that more body-worn technology leads to more social considerations, especially on the part of the Streamer, who must bear the additional weight of the socio-technical burden. For instance, as reported by Ishak et al., the size of a body-worn smartphone influenced how third-party members perceived the Viewer [41]. A smaller screen yields a smaller image of the Viewer, and a larger screen yields a larger image. This affects how others can communicate with them, and how they perceive themselves in this communication paradigm.

Different technical setups have been reported to affect immersion, co-presence, and social presence, and studies have shown that the Viewer’s apparatus leads to differences in these psychometric scales [55, 139]. One of the main purposes of H2H telepresence is to let the remote user feel like they are “actually there,” so it is natural that researchers have developed various artifacts in an attempt to increase these scores; but, these artifacts come with social cost. As reported in some field studies, the Streamer can feel socially awkward while using telepresence due to wearing a device on their body, or because they are engaging in a non-normal activity. In particular, researchers have posited that to free the hands of the Streamer, they can wear the camera on their body; but, this is currently not a social norm, and bystanders have been found to give “strange looks” in public [104]; in other words, the telepresence interaction is noticeable, which makes the Streamer stand out [16, 55, 104, 105]. Further, there is strong interplay between technology, social considerations, and measurable outcomes. In our dataset, we found some authors to make a call for camera stabilization. Using immersive viewing apparatus, users could succumb to simulator sickness, and software stabilization has been used to help prevent this [48, 52]. Other researchers used a hardware-based approach, integrating a wearable device into the technical setup a Streamer uses [57, 60]. This could, unfortunately, result in a greater social cost given the increased noticeability.

4.4.2 Social Landscape Influences Technological Considerations. In our dataset, when designing prototypes or studies, the authors often had a specific social context in mind. In the cases which proposed strangers interacting with the telepresence setup, the primary concern was to let the Viewer see the remote environment opposed to seeing the Streamer. For instance, some authors deployed a body-worn camera placement that excluded the Streamer’s face [27, 41, 81, 82]. This is in stark contrast to the articles proposing friends/family and LDR interactions, where the camera was often placed on the Streamer’s shoulder or was handheld, providing opportunity to capture the their face [11, 16, 29, 60]. In these cases, streaming the remote environment or activity was a co-objective to providing a more social experience. In addition to making the interaction more social, providing the Viewer with a way to see the Streamer’s face unlocks natural communication. In our dataset, the purpose of many papers was to provide or demonstrate collaboration capability, and other articles noted that simple telepresence setups are conducive to the participants missing out on what the other is seeing (e.g. “did you see that?” [55]). To overcome this, the Streamer can employ a 360-degree camera, which allows the Viewer to see the remote environment and, when placed properly, the Streamer too. In various field studies such as those by Rae et al. [105], Tang et al. [118], and Jones et al. [42], the authors note that typical telepresence setups do not capture natural communication in its entirety. To provide the Viewer with a way to understand that of the Streamer, simply placing the camera in a way that allows them to see the hands and face have been shown to suffice. However, additional hardware is required for both parties in order to capture and display that of the Viewer. For example, when a 360-degree camera is used to capture the environment, to communicate hand gestures, researchers often employ specialized depth cameras such as the Leap Motion. To communicate gaze, researchers often employ head-tracking devices such as those found in the HTC Vive HMD. Then, to display these to the Streamer, researchers employ AR glasses [10–12, 66].

5 DISCUSSION

Participant feedback within the field studies of our dataset reveal moderate comfort with the telepresence concept across numerous concepts, and as such there is much opportunity to improve the state of the art. In this section, we outline paths for future work in a call to action for the SIGCHI community.

5.1 The Search for Meaningful Dependent Variables. What is Success?

The H2H telepresence interaction style is fairly new, and we noticed that there were very few studies to investigate real-world applications. More work is needed in the form of longitudinal field studies to understand how this novel interaction is perceived by all parties involved. In particular, we need to understand how to make this acceptable over a long period of use, especially when its novelty would wear off. Additionally, as apparent by the vast quantity of custom measurements used in most of the dataset, the telepresence community needs to converge upon valid methods of measuring meaningful outcomes. We noted that most of the items found in the dataset were custom, often in the form of singletons, and these could have been replaced by pre-validated questionnaires. As illustrated in work by DeLone and McLean in the field of Information Systems [19], "success" can be measured by various dimensions and that, in order to move a field forward, researchers must converge on a subset of measures that define success. In the field of HCI, many outcomes have been measured through the use of pre-validated instruments (e.g. USE questionnaire [73], Slater-Usoh-Steed questionnaire [115]); but, we acknowledge that many of the questions being asked by the researchers did not map to these instruments, suggesting that we lack robust ways to operationalize the socio-technical outcomes we are interested in measuring in the context of human-to-human immersive telepresence. Therefore, as a research community, we need to engage with one another to reflect on the social (e.g., connectedness, support, well-being), contextual (e.g., time to complete a task, accuracy, immersion in the environment) and technical (e.g., video fidelity, ease-of-use, usefulness) outcomes that we care about when determining success in our field. Also, when future work develops new measures, researchers should make sure to psychometrically validate these instruments and share them with the community.

Regarding dependent outcomes that have been commonly studied in H2H telepresence, there have been two primary goals - to achieve the sense of "being there" in the remote environment, and to achieve the sense of "being there" with a remote partner - and to accurately measure how we make advances in these areas, the community should consider adopting a standard set of pre-validated instruments to measure both spatial and social presence. The Slater-Usoh-Steed questionnaire and the Presence questionnaire by Witmer and Singer have been widely adopted to measure presence in virtual environments, and they can also be used here to measure spatial presence [115, 135]. In essence, a remote environment captured by a camera for telepresence is akin to a high-fidelity virtual environment, so the questionnaires apply, although the items will need to be adapted to ensure the language makes sense for this specific context. The research community has a more difficult challenge for determining a common definition for social presence; as this interaction style has a strong social component, it is important to make this identification. Oh et al. suggested that co-presence and social presence are one in the same [98], but the most commonly used measure (the Networked Minds Measure, [8]) lists co-presence as a sub-construct of social presence. We found multiple definitions for this concept across studies of our dataset, so it is difficult to compare the results of one study to another. Cui made this same observation in 2013, noting that the literature diverges from a "commonly accepted paradigm" [18], and to our knowledge, this is still an unsolved issue.

In order to make fair comparisons (and thus be able to improve the state-of-the-art), we as a community need to work towards understanding how to best make that convergence. However, with the field of human-to-human telepresence still being in its infancy, we do acknowledge that standardized measurements may not always be able to address future research questions. It is out of the scope of this paper to develop these new measures, but we wanted to bring this issue to light such that future researchers can help solve this problem. At the very least, researchers should perform an appropriate psychometric analysis on their measurement tools to ensure some level of construct validity. This would allow future researchers to compare and contrast their technology probes against previous work in a more standardized manner, while also preserving their freedom to create custom instruments.

5.2 The Forgotten Users: Streamer Considerations are Understudied

One of the most jarring findings across our dataset is a common disregard of the Streamer. Many projects seem to posit that the Streamer is a replacement for a telepresence robot, and do not attempt to measure their experience, instead focusing just on the Viewer. The H2H telepresence community needs to remember that these people are stakeholders too, and we need to make explicit considerations for them. Many of the Streamer participants in the field studies found in our dataset reported a sense of social awkwardness when they performed their role, yet, the strong majority of articles described a telepresence setup which drew attention to these users, either because the interaction was not localized to them, or because they were wearing extra, conspicuous technology on their body. In order to maximize the Viewer's user experience, researchers tend to add more and more technology to the Streamer's setup; but this could be making matters worse. The existing research has already revealed that the various live-streaming techniques (e.g. smartphone and 360-degree cameras) provide the Viewer with a strong sense of presence, so it is now time to shift the community's research agenda to optimize the Streamer's experience.

Firstly, instead of just measuring the Viewer's sense of presence, we need to measure social presence for both parties. After all, a successful social interaction is contingent on both. Only one article in our dataset attempted to measure social presence of both users, and future work needs to make this a high priority. We also need to develop a way to better understand how a given telepresence setup would cause a person to feel a sense of social awkwardness, before, during, and after use. To our knowledge, there is no pre-validated measure to quantify this, but it would be valuable such that we can influence future design to help reduce it. Secondly, we need to consider that by optimizing the Viewer's experience, we might be worsening that of the Streamer. Likewise, by optimizing a Streamer's experience, we will probably reduce the Viewer's experience. In various field studies of our dataset, authors conclude that social awkwardness stems from the wearable devices [55, 105], especially as the addition of physical technology becomes more apparent to the third party. Even as wearable technology becomes more compact and less noticeable, we must remember that the technology's design and its intended use will affect a Streamer's desire to interact with it. Thus, a goal of future telepresence work should be to strive for experience optimization of both end-users, understanding that their social needs may not align in a given telepresence scenario.

Thirdly, we must consider the physical safety of the Streamer as they interact with H2H telepresence. In a few dataset articles, the authors noted that Streamers entered into dangerous scenarios due to the novel interaction, e.g. nearly getting into a fight with a passerby [104], or almost getting hit by a vehicle because their attention was drawn to the technology probe [42]. We also note that because they are often asked to wear cameras on their body, they stand out in public; crime reports indicate that tourists are often targeted for theft because they stand out, particularly because they have expensive camera equipment [33, 37]. Since remote sightseeing is a prominent use case found in our dataset, we acknowledge that telepresence users could be targeted too. We must remind ourselves that Streamers are not tools or robots to facilitate telepresence; they are people and we must ensure their safety as they interact with H2H telepresence. They take on all of the risks, and the research community must think of ways to protect them.

The telepresence community has long since focused on the needs of the Viewer, and it is now time to shift our focus towards the Streamer. Additionally, as telepresence itself becomes viable as an entertainment product, there is now a class of streaming-based workers that operate entirely via telepresence, raising questions as to the workers' rights and fair expectations of these users. We call for researchers within the telepresence field to make considerations for Streamers' well-being, especially in situations where third parties are involved in the telepresence environment. For instance, previous work regarding public live-streaming for games shows how live-streamers can become physically and mentally drained [106]. We note that these considerations change depending on the relationship between Streamers and Viewers. However, if telepresence is now itself undertaken as an occupation, serious work needs to be done to affirm the rights of these telepresence workers, exploring the various social contexts in which this work occurs, and developing a framework of social and moral norms for professional telepresence Streamers, which can also be extended as guidelines to non-professional human-to-human telepresence. We anticipate future work to emerge in the form of participatory and iterative design with real users, to brainstorm and implement technology probes which will reduce the potential harms and ethical issues Streamer users may face.

5.3 The Social Dynamics Between Streamers and Third Party

In addition to reducing the risks posed to the Streamer, we must also consider the needs of third party members. As Rae et al. note, very little of past research considers the third party as a stakeholder in a telepresence interaction [105], and we found this to be the case in our dataset as well. As a community we need to ensure that we account for this group of people as we conduct our research. In the United States and Canada, where most of the field studies in our dataset took place, video recording in public is viewed as a legal right [62, 116], so in common areas there is no reasonable expectation of privacy. In other cultures, such as Japan, it has been observed that even in public, live-streaming is frowned upon socially [30]. Researchers applying telepresence in part of a field study need to consider the culture in which they plan on conducting studies; but even if the law and culture both approve, we still need to consider the ethical implications such as socially accepted conventions of privacy.

As previously noted, an avenue for future work is to minimize the technology such that the Streamer does not stand out; this will help to reduce the social awkwardness caused by third party members staring at them. This will also reduce the third party's ability to spot that telepresence is taking place. It is questionable, then, if this is an appropriate way of conducting a telepresence experience. It is out of the scope of our paper to provide ethical guidance regarding the third party members, but we pose the following questions to the research community in hopes that future work will tackle this issue: is it necessary to inform third party members that they are being captured as part of a public telepresence experience? How can we balance the rights of these people while protecting the Streamer? Answers to these questions will bring forth a more enjoyable experience for all parties involved, including those who have been thus far been an after-thought in telepresence research.

We note that even if the telepresence technology can be minimized and hidden from plain sight, some Streamers might still feel uneasy, knowing that they are potentially intruding on the privacy of other people. Likewise, we note that some Streamers will not perceive telepresence as some form of social intrusion, even if the technology probe was excessively obtrusive to wear. As noted in one field study, some participants felt socially awkward, while others "enjoyed the attention drawn" to them due to wearing a new, visible piece of technology [105]. Thus, it might be the case that individual differences cause users to have various attitudes towards acting as a Streamer. It would be beneficial to researchers and end users if we could predict their level of comfort. For instance, researchers could apply personality models such as the Myers-Briggs or Big-Five paradigms to understand the Streamer's personality as it relates to their willingness to engage in a public telepresence experience [26, 87]. We hypothesize that it is possible to make some prediction of a Streamer's comfort when using a telepresence prototype, but this will be dependent on identifying the social factors which help with this prediction. If we are able to make a distinction with high accuracy, then we could facilitate the ability to recommend a specific type of technological setup that would provide the most social comfort to a potential Streamer.

5.4 Technical Recommendations for Social Telepresence

Our dataset revealed several primary objectives of H2H telepresence: letting the Viewer explore the remote environment; letting the Viewer naturally communicate with the Streamer; and letting the

Viewer communicate with the remote environment. The technology probes tackle one or many of these objectives, resulting in various ways they are used, and there is still opportunity to enhance the state-of-the-art by developing additional technologies.

5.4.1 *Getting to "Actually There": Enhancements for Exploration.*

As is expected of a telepresence interaction, all of the technology probes discussed a way to let the Viewer explore the remote environment, to some degree. Cameras ranged from single lens devices to omnidirectional devices, and though the type of video differs, participants were still able to feel a sense of spatial presence regardless of type. This means that the more expensive equipment (360-degree cameras and VR HMDs) might not be absolutely necessary in order to fulfill this goal, though HMD use does increase immersion. None of the articles particularly discussed audio features, but microphones are commonly embedded in consumer-grade devices. Smartphones typically have a single microphone, and 360-degree cameras sometimes have directional microphones (e.g. the Ricoh Theta V[2]) which can provide a more immersive feeling when paired with a VR HMD that supports directional audio. So far, and as found in our dataset, the state-of-the-art telepresence setups facilitate the exploration of a remote environment only through visual and auditory cues, but there is opportunity to engage the other major senses to truly make the Viewer feel "actually there." For instance, we can leverage techniques from the VR community to simulate various types of haptic feedback. Wilberz et al. recently demonstrated an HMD-mounted device which can let a Viewer feel heat and wind of a virtual environment [134]. Brooks et al. were also able to provide an illusion of varying temperatures using the sense of smell [9]. These output devices can be paired with sensors in a Streamer's setup, to measure wind speed and temperature [25]. We suspect that these techniques will increase the Viewer's perceived sense of spatial presence, especially in outdoor environments, and future work is needed to understand the magnitude that each brings to the experience.

5.4.2 *Integrating Natural Communication between Streamer and Viewer.*

Next, the majority of the articles in our dataset describe facilitating communication between the Streamer and Viewer, and we hold that this is a primary objective especially for cases where the two parties are friends, family, or intimates. In our dataset, communication between these users is typically tied to auditory cues, with a subset of technology probes using simple video chat. As noted, natural communication (e.g. hand gestures and gaze) has not been studied too much, but there are a variety of ways to convey these features. Interestingly, none of the articles in our dataset leveraged an eye-tracking device to pin-point Viewer's gaze, but that is a natural extension to the current literature. Eye-tracking hardware is available for flat screens and HMDs, and even as a mobile wearable[3], so both parties can use these to enhance their communication while exploring the same environment together.

5.4.3 *Seeing Eye-to-Eye: Social Encounters Over Large Distances.*

Lastly, a smaller subset of our dataset discussed the ability for the Viewer to interact with the remote environment, i.e. to communicate with the third party members. In most of these articles, the Viewers were embodied via video, meaning the Streamer used a smartphone screen to show them in real time. The commonality here is that none of the Viewers used a VR HMD, because if they did, their faces would be blocked by the device. This results in a reduced sense of immersion and, potentially, social presence. To overcome this problem, we suspect that techniques such as Rekimoto's "Behind the Mask" prototype could allow the Viewer to have a video embodiment while still using a VR HMD [107]. This technique uses multiple cameras to capture an HMD wearer's face even though they have the device on their head. After stitching the different camera feeds together, the result is a reconstruction of the user's entire face, which could be sent back to the third party members in the remote environment.

---

[2] https://theta360.com/en/about/theta/v.html

[3] https://www.tobiipro.com/product-listing/tobii-pro-glasses-2/

There is plenty of future work that the telepresence community can perform in order to enhance the socio-technical user experience. We note, however, that it requires a better understanding of each user's physical and social needs, which may conflict at times, given the nature of the telepresence scenario and the users' relationships. It is pertinent that we realize how to provide a balance such that H2H telepresence will be desirable by Streamers, Viewers, and the third party, allowing for a natural, enjoyable, and equitable means of communication.

5.5 Limitations and Future Work

We conducted a comprehensive literature review of mobile, body-worn, human-to-human telepresence, used for non-work scenarios. These criteria were chosen as they describe a new form of telepresence that has only recently gained traction within the research community. However, this restrictive scoping may have limited our ability to include other human-to-human telepresence scenarios which could also hold relevance to this main topic of inquiry.

Most notably, we excluded work-related telepresence scenarios, as in most cases these are stationary, and only involve a social dynamic between the Streamer and Viewer, i.e. there is no third party. However, these dynamics are still interpersonal and vary widely, depending on the existing relationship between both parties, and can contribute to the discussion about ensuring equitable experience between parties. With large companies beginning to adopt remote technologies in the midst of the COVID-19 pandemic, including expert help via AR, more study should be conducted in order to ensure that the unique social implications of work telepresence are properly investigated. For instance, research implies that there might be a new class of employees who rely on AR for most of their day. Introducing new technology – especially that which must be worn on the body – raises ethical questions and human factors issues. Future work must address these considerations to ensure that all parties in the work environment are treated equitably.

Many of the articles in our review did not explicitly detail the social relationship between Streamer and Viewer, and so we were unable to properly quantify them. Future research should ensure that the nature of these relationships is more clearly articulated; but, we acknowledge that human-to-human telepresence may be used for a wide variety of use cases. In our review, we excluded telepresence scenarios that involve a one-to-many user relationship, which occur in situations such as Twitch streaming and mobile broadcasting. While these scenarios involve fewer opportunities for interpersonal communication, they nonetheless provide other forms of interaction modality that are important to consider as we explore new social dynamics within human-to-human telepresence, such as parasocial relationships. This is especially relevant given that these forms of telepresence are emerging as forms of paid entertainment, and not just for recreational purposes. In our future work, we plan on investigating proper communication modalities that can be used in this one-to-many paradigm, while providing a strong sense of social presence within these online communities.

6 CONCLUSION

Mobile human-to-human telepresence is a relatively new concept, and our review of this field has revealed many interesting projects that are pushing the technology past its current limits. We remind ourselves that we must progress not just through technical means, but by also considering the social needs of the users for whom we intend to develop this new interaction paradigm. We call upon the SIGCHI community to take our findings into consideration in order to enhance Viewer's experience, reduce the Streamer's social cost, and standardize the methodology we use to ensure we are making developments in the right direction. By doing so, we can achieve a more socially equitable interaction that can enhance many lives. Especially in this unique time of social distancing, we can bring people together and facilitate exploration of remote environments.

7 ACKNOWLEDGMENTS

This research is supported in part by NSF Awards IIS-1917728 and CHS-1844881. Any opinion, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the research sponsors.

REFERENCES

[1] 2020. Certificated Remote Pilots including Commercial Operators. https://www.faa.gov/uas/commercial_operators/

[2] 2020. Get Your Mass Gatherings or Large Community Events Ready. https://www.cdc.gov/coronavirus/2019-ncov/community/large-events/mass-gatherings-ready-for-covid-19.html

[3] Mark S. Ackerman. 2000. The Intellectual Challenge of CSCW: The Gap Between Social Requirements and Technical Feasibility. *Human–Computer Interaction* 15, 2-3 (Sept. 2000), 179–203. https://doi.org/10.1207/S15327051HCI1523_5

[4] Leila Alem, Franco Tecchia, and Weidong Huang. 2011. HandsOnVideo: Towards a Gesture based Mobile AR System for Remote Collaboration. In *Recent Trends of Mobile Collaborative Augmented Reality Systems*, Leila Alem and Weidong Huang (Eds.). Springer New York, New York, NY, 135–148. https://doi.org/10.1007/978-1-4419-9845-3_11

[5] Sudhanshu S.D.P. Ayyagari, Kunal Gupta, Matt Tait, and Mark Billinghurst. 2015. CoSense: Creating Shared Emotional Experiences. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems - CHI EA '15*. ACM Press, Seoul, Republic of Korea, 2007–2012. https://doi.org/10.1145/2702613.2732839

[6] Uddipana Baishya and Carman Neustaedter. 2017. In Your Eyes: Anytime, Anywhere Video and Audio Streaming for Couples. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing - CSCW '17*. ACM Press, Portland, Oregon, USA, 84–97. https://doi.org/10.1145/2998181.2998200

[7] Mark Billinghurst, Alaeddin Nassani, and Carolin Reichherzer. 2014. Social panoramas: using wearable computers to share experiences. In *SIGGRAPH Asia 2014 Mobile Graphics and Interactive Applications on - SA '14*. ACM Press, Shenzhen, China, 1–1. https://doi.org/10.1145/2669062.2669084

[8] Frank Biocca, Chad Harms, and Jenn Gregg. 2001. The networked minds measure of social presence: Pilot test of the factor structure and concurrent validity. In *4th annual international workshop on presence, Philadelphia, PA*. 1–9.

[9] Jas Brooks, Steven Nagels, and Pedro Lopes. 2020. Trigeminal-based Temperature Illusions. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–12.

[10] Minghao Cai, Soh Masuko, and Jiro Tanaka. 2018. Gesture-based Mobile Communication System Providing Side-by-side Shopping Feeling. In *Proceedings of the 23rd International Conference on Intelligent User Interfaces Companion - IUI 18*. ACM Press, Tokyo, Japan, 1–2. https://doi.org/10.1145/3180308.3180310

[11] Minghao Cai, Soh Masuko, and Jiro Tanaka. 2018. Shopping Together: A Remote Co-shopping System Utilizing Spatial Gesture Interaction. In *Human-Computer Interaction. Interaction Technologies*, Masaaki Kurosu (Ed.). Vol. 10903. Springer International Publishing, Cham, 219–232. https://doi.org/10.1007/978-3-319-91250-9_17

[12] Minghao Cai and Jiro Tanaka. 2017. Trip Together: A Remote Pair Sightseeing System Supporting Gestural Communication. In *Proceedings of the 5th International Conference on Human Agent Interaction - HAI '17*. ACM Press, Bielefeld, Germany, 317–324. https://doi.org/10.1145/3125739.3125762

[13] Yuanzhi Cao, Xun Qian, Tianyi Wang, Rachel Lee, Ke Huo, and Karthik Ramani. 2020. An Exploratory Study of Augmented Reality Presence for Tutoring Machine Tasks. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–13. https://doi.org/10.1145/3313831.3376688

[14] Ching-Tzun Chang, Shin Takahashi, and Jiro Tanaka. 2012. WithYou - a communication system to provide out together feeling. In *Proceedings of the International Working Conference on Advanced Visual Interfaces - AVI '12*. ACM Press, Capri Island, Italy, 320. https://doi.org/10.1145/2254556.2254617

[15] Jon Christian. 2019. Doctor Tells Man He's Going to Die via Telepresence Robot. https://futurism.com/doctor-death-telepresence-robot

[16] Anezka Chua, Azadeh Forghani, and Carman Neustaedter. 2017. Shared Bicycling Over Distance. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems - CHI EA '17*. ACM Press, Denver, Colorado, USA, 455–455. https://doi.org/10.1145/3027063.3049776

[17] Kevin Corti and Alex Gillespie. 2015. Revisiting Milgram's Cyranoid Method: Experimenting With Hybrid Human Agents. *The Journal of Social Psychology* 155, 1 (Jan. 2015), 30–56. https://doi.org/10.1080/00224545.2014.959885

[18] Guoqiang Cui. 2013. Evaluating online social presence: An overview of social presence assessment. *Journal of Educational Technology Development and Exchange (JETDE)* 6, 1 (2013), 3.

[19] William H. DeLone and Ephraim R. McLean. 1992. Information Systems Success: The Quest for the Dependent Variable. *Information Systems Research* 3, 1 (1992), 60–95. http://www.jstor.org/stable/23010781 Publisher: INFORMS.

[20] Brittany A. Duncan and Robin R. Murphy. 2017. Effects of Speed, Cyclicity, and Dimensionality on Distancing, Time, and Preference in Human-Aerial Vehicle Interactions. *ACM Transactions on Interactive Intelligent Systems* 7, 3 (Sept. 2017), 1–27. https://doi.org/10.1145/2983927

[21] Kaylee Fagan. [n.d.]. This 24-year-old is hitchhiking across America and live-streaming the whole thing on Amazon's Twitch for his thousands of followers. https://www.businessinsider.com/trevor-daneliuk-hitchhikes-across-america-on-twitch-2018-8

[22] Geraldine Fitzpatrick. 2019. A personal letter to my CHI colleagues. http://www.changingacademiclife.com/gerifitz/2019/2/19/a-personal-letter

[23] Steffen Gauglitz, Benjamin Nuernberger, Matthew Turk, and Tobias Höllerer. 2014. World-stabilized annotations and virtual scene navigation for remote collaboration. In *Proceedings of the 27th annual ACM symposium on User interface software and technology - UIST '14*. ACM Press, Honolulu, Hawaii, USA, 449–459. https://doi.org/10.1145/2642918.2647372

[24] Wendy Glauser. 2013. Doctors among early adopters of Google Glass. *Canadian Medical Association.Journal* 185, 16 (Nov. 2013), 1385. https://search.proquest.com/docview/1476500582?accountid=10003 ISBN: 08203946.

[25] Oleg A. Godin, Vladimir G. Irisov, and Mikhail I. Charnotskii. 2014. Passive acoustic measurements of wind velocity and sound speed in air. *The Journal of the Acoustical Society of America* 135, 2 (Feb. 2014), EL68–EL74. https://doi.org/10.1121/1.4862885

[26] Lewis R Goldberg. 1990. An alternative" description of personality": the big-five factor structure. *Journal of personality and social psychology* 59, 6 (1990), 1216. Publisher: American Psychological Association.

[27] Lilian de Greef, Meredith E. Morris, and Kori Inkpen. 2016. TeleTourist: Immersive Telepresence Tourism for Mobility-Restricted Participants. In *Proceedings of the 19th ACM Conference on Computer Supported Cooperative Work and Social Computing Companion - CSCW '16 Companion*. ACM Press, San Francisco, California, USA, 273–276. https://doi.org/10.1145/2818052.2869082

[28] Charlotte N Gunawardena and Frank J Zittle. 1997. Social presence as a predictor of satisfaction within a computer-mediated conferencing environment. *American journal of distance education* 11, 3 (1997), 8–26. Publisher: Taylor & Francis.

[29] Jiajing Guo, Yoyo Tsung-Yu Hou, Harley Mueller, Katherine Tang, and Susan R. Fussell. 2019. As If I Am There: A New Video Chat Interface Design for Richer Contextual Awareness. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems - CHI EA '19*. ACM Press, Glasgow, Scotland Uk, 1–6. https://doi.org/10.1145/3290607.3312759

[30] Michael Gwilliam. 2020. Japanese Twitch streamer reveals the problem with IRL streaming in Tokyo. https://www.dexerto.com/entertainment/japanese-twitch-streamer-reveals-the-problem-with-irl-streaming-tokyo-1325147

[31] Oliver L. Haimson and John C. Tang. 2017. What Makes Live Events Engaging on Facebook Live, Periscope, and Snapchat. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, Denver Colorado USA, 48–60. https://doi.org/10.1145/3025453.3025642

[32] William A. Hamilton, John Tang, Gina Venolia, Kori Inkpen, Jakob Zillner, and Derek Huang. 2016. Rivulet: Exploring Participation in Live Events through Multi-Stream Experiences. In *Proceedings of the ACM International Conference on Interactive Experiences for TV and Online Video - TVX '16*. ACM Press, Chicago, Illinois, USA, 31–42. https://doi.org/10.1145/2932206.2932211

[33] Dee Wood Harper Jr and others. 2001. Comparing tourists crime victimization. *Annals of Tourism Research* 28, 4 (2001), 1053–1056. Publisher: Pergamon Press.

[34] Marc Hassenzahl, Stephanie Heidecker, Kai Eckoldt, Sarah Diefenbach, and Uwe Hillmann. 2012. All You Need is Love: Current Strategies of Mediating Intimate Relationships through Technology. *ACM Transactions on Computer-Human Interaction* 19, 4 (Dec. 2012), 1–19. https://doi.org/10.1145/2395131.2395137

[35] Rebecca A. Hayes, Caleb T. Carr, and Donghee Yvette Wohn. 2016. One Click, Many Meanings: Interpreting Paralinguistic Digital Affordances in Social Media. *Journal of Broadcasting & Electronic Media* 60, 1 (Jan. 2016), 171–187. https://doi.org/10.1080/08838151.2015.1127248

[36] Yasamin Heshmat, Brennan Jones, Xiaoxuan Xiong, Carman Neustaedter, Anthony Tang, Bernhard E. Riecke, and Lillian Yang. 2018. Geocaching with a Beam: Shared Outdoor Activities through a Telepresence Robot with 360 Degree Viewing. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*. ACM Press, Montreal QC, Canada, 1–13. https://doi.org/10.1145/3173574.3173933

[37] Sarah Hodgkinsoi and Nick Tilley. 2007. Travel-to-crime: homing in on the victim. *International Review of Victimology* 14, 3 (2007), 281–298. Publisher: SAGE Publications Sage UK: London, England.

[38] Weidong Huang, Leila Alem, and Franco Tecchia. 2013. HandsIn3D: Supporting Remote Guidance with Immersive Virtual Environments. In *Human-Computer Interaction – INTERACT 2013*, David Hutchison, Takeo Kanade, Josef Kittler, Jon M. Kleinberg, Friedemann Mattern, John C. Mitchell, Moni Naor, Oscar Nierstrasz, C. Pandu Rangan, Bernhard Steffen, Madhu Sudan, Demetri Terzopoulos, Doug Tygar, Moshe Y. Vardi, Gerhard Weikum, Paula Kotzé, Gary Marsden, Gitte Lindgaard, Janet Wesson, and Marco Winckler (Eds.). Vol. 8117. Springer Berlin Heidelberg, Berlin, Heidelberg, 70–77. https://doi.org/10.1007/978-3-642-40483-2_5 Series Title: Lecture Notes in Computer Science.

[39] Huazhong Li, Yongsheng Liang, Tao He, and Yi Li. 2012. Real-time shared control of space robots teleoperation without time delay. In *2012 24th Chinese Control and Decision Conference (CCDC)*. 3637–3642. https://doi.org/10.1109/CCDC.2012.6244582

[40] Kori Inkpen, Brett Taylor, Sasa Junuzovic, John Tang, and Gina Venolia. 2013. Experiences2Go: sharing kids' activities outside the home with remote family members. In *Proceedings of the 2013 conference on Computer supported cooperative work - CSCW '13*. ACM Press, San Antonio, Texas, USA, 1329. https://doi.org/10.1145/2441776.2441926

[41] Clarissa Ishak, Carman Neustaedter, Dan Hawkins, Jason Procyk, and Michael Massimi. 2016. Human Proxies for Remote University Classroom Attendance. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16*. ACM Press, Santa Clara, California, USA, 931–943. https://doi.org/10.1145/2858036.2858184

[42] Brennan Jones, Anna Witcraft, Scott Bateman, Carman Neustaedter, and Anthony Tang. 2015. Mechanics of Camera Work in Mobile Video Collaboration. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15*. ACM Press, Seoul, Republic of Korea, 957–966. https://doi.org/10.1145/2702123.2702345

[43] Tejinder K. Judge, Carman Neustaedter, Steve Harrison, and Andrew Blose. 2011. Family portals: connecting families through a multifamily media space. In *Proceedings of the 2011 annual conference on Human factors in computing systems - CHI '11*. ACM Press, Vancouver, BC, Canada, 1205. https://doi.org/10.1145/1978942.1979122

[44] Tejinder K. Judge, Carman Neustaedter, and Andrew F. Kurtz. 2010. The family window: the design and evaluation of a domestic media space. In *Proceedings of the 28th international conference on Human factors in computing systems - CHI '10*. ACM Press, Atlanta, Georgia, USA, 2361. https://doi.org/10.1145/1753326.1753682

[45] Bumsoo Kang, Inseok Hwang, Jinho Lee, Seungchul Lee, Taegyeong Lee, Youngjae Chang, and Min Kyung Lee. 2018. My Being to Your Place, Your Being to My Place: Co-present Robotic Avatars Create Illusion of Living Together. In *Proceedings of the 16th Annual International Conference on Mobile Systems, Applications, and Services - MobiSys '18*. ACM Press, Munich, Germany, 54–67. https://doi.org/10.1145/3210240.3210348

[46] Enes Karaaslan, Ulas Bagci, and Fikret Necati Catbas. 2019. Artificial Intelligence Assisted Infrastructure Assessment using Mixed Reality Systems. *Transportation Research Record: Journal of the Transportation Research Board* 2673, 12 (Dec. 2019), 413–424. https://doi.org/10.1177/0361198119839988

[47] Shunichi Kasahara, Shohei Nagai, and Jun Rekimoto. 2014. LiveSphere: immersive experience sharing with 360 degrees head-mounted cameras. In *Proceedings of the adjunct publication of the 27th annual ACM symposium on User interface software and technology - UIST'14 Adjunct*. ACM Press, Honolulu, Hawaii, USA, 61–62. https://doi.org/10.1145/2658779.2659114

[48] Shunichi Kasahara, Shohei Nagai, and Jun Rekimoto. 2015. First Person Omnidirectional Video: System Design and Implications for Immersive Experience. In *Proceedings of the ACM International Conference on Interactive Experiences for TV and Online Video - TVX '15*. ACM Press, Brussels, Belgium, 33–42. https://doi.org/10.1145/2745197.2745202

[49] Shunichi Kasahara, Shohei Nagai, and Jun Rekimoto. 2017. JackIn Head: Immersive Visual Telepresence System with Omnidirectional Wearable Camera. *IEEE Transactions on Visualization and Computer Graphics* 23, 3 (March 2017), 1222–1234. https://doi.org/10.1109/TVCG.2016.2642947

[50] Shunichi Kasahara and Jun Rekimoto. 2014. JackIn: integrating first-person view with out-of-body vision generation for human-human augmentation. In *Proceedings of the 5th Augmented Human International Conference on - AH '14*. ACM Press, Kobe, Japan, 1–8. https://doi.org/10.1145/2582051.2582097

[51] Shunichi Kasahara and Jun Rekimoto. 2015. JackIn head: an immersive human-human telepresence system. In *SIGGRAPH Asia 2015 Emerging Technologies on - SA '15*. ACM Press, Kobe, Japan, 1–3. https://doi.org/10.1145/2818466.2818486

[52] Shunichi Kasahara and Jun Rekimoto. 2015. JackIn head: immersive visual telepresence system with omnidirectional wearable camera for remote collaboration. In *Proceedings of the 21st ACM Symposium on Virtual Reality Software and Technology - VRST '15*. ACM Press, Beijing, China, 217–225. https://doi.org/10.1145/2821592.2821608

[53] Tadakazu Kashiwabara, Hirotaka Osawa, Kazuhiko Shinozawa, and Michita Imai. 2012. TEROOS: a wearable avatar to enhance joint activities. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems - CHI '12*. ACM Press, Austin, Texas, USA, 2001. https://doi.org/10.1145/2207676.2208345

[54] Robert S Kennedy, Norman E Lane, Kevin S Berbaum, and Michael G Lilienthal. 1993. Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The international journal of aviation psychology* 3, 3 (1993), 203–220. Publisher: Taylor & Francis.

[55] Seungwon Kim, Sasa Junuzovic, and Kori Inkpen. 2014. The Nomad and the Couch Potato: Enriching Mobile Shared Experiences with Contextual Information. In *Proceedings of the 18th International Conference on Supporting Group Work - GROUP '14*. ACM Press, Sanibel Island, Florida, USA, 167–177. https://doi.org/10.1145/2660398.2660409

[56] Seungwon Kim, Gun A. Lee, Nobuchika Sakata, Andreas Dunser, Elina Vartiainen, and Mark Billinghurst. 2013. Study of augmented gesture communication cues and view sharing in remote collaboration. In *2013 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, Adelaide, Australia, 261–262. https://doi.org/10.1109/ISMAR.2013.6671795

[57] Don Kimber, Patrick Proppe, Sven Kratz, Jim Vaughan, Bee Liew, Don Severns, and Weiqing Su. 2015. Polly: Telepresence from a Guide's Shoulder. In *Computer Vision - ECCV 2014 Workshops*, Lourdes Agapito, Michael M. Bronstein, and Carsten Rother (Eds.). Vol. 8927. Springer International Publishing, Cham, 509–523. https://doi.org/10.1007/978-3-319-16199-0_36

[58] Saso Koceski and Natasa Koceska. 2016. Evaluation of an Assistive Telepresence Robot for Elderly Healthcare. *Journal of Medical Systems* 40, 5 (May 2016), 121. https://doi.org/10.1007/s10916-016-0481-x

[59] L. Kovács, T. Haidegger, and I. Rudas. 2013. Surgery from a distance—Application of intelligent control for telemedicine. In *2013 IEEE 11th International Symposium on Applied Machine Intelligence and Informatics (SAMI)*. 125–129. https://doi.org/10.1109/SAMI.2013.6480959

[60] Sven Kratz, Daniel Avrahami, Don Kimber, Jim Vaughan, Patrick Proppe, and Don Severns. 2015. Polly Wanna Show You: Examining Viewpoint-Conveyance Techniques for a Shoulder-Worn Telepresence System. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct - MobileHCI '15*. ACM Press, Copenhagen, Denmark, 567–575. https://doi.org/10.1145/2786567.2787134

[61] Sven Kratz, Don Kimber, Weiqing Su, Gwen Gordon, and Don Severns. 2014. Polly: "being there" through the parrot and a guide. In *Proceedings of the 16th international conference on Human-computer interaction with mobile devices & services - MobileHCI '14*. ACM Press, Toronto, ON, Canada, 625–630. https://doi.org/10.1145/2628363.2628430

[62] Seth F Kreimer. 2010. Pervasive image capture and the First Amendment: Memory, discourse, and the right to record. *U. Pa. L. Rev.* 159 (2010), 335. Publisher: HeinOnline.

[63] Annica Kristoffersson, Silvia Coradeschi, and Amy Loutfi. 2013. A Review of Mobile Robotic Telepresence. *Advances in Human-Computer Interaction* 2013 (2013), 1–17. https://doi.org/10.1155/2013/902316

[64] Joseph J. LaViola, Ernst Kruijff, Ryan P. McMahan, Doug A. Bowman, and Ivan Poupyrev. 2017. *3D user interfaces: theory and practice* (second edition ed.). Addison-Wesley, Boston. OCLC: ocn935986831.

[65] Jonathan Lazar, Jinjuan Heidi Feng, and Harry Hochheiser. 2017. *Research Methods in Human-Computer Interaction.* https://public.ebookcentral.proquest.com/choice/publicfullrecord.aspx?p=4851896 OCLC: 1076644193.

[66] Gun A. Lee, Theophilus Teo, Seungwon Kim, and Mark Billinghurst. 2017. Sharedsphere: MR collaboration through shared live panorama. In *SIGGRAPH Asia 2017 Emerging Technologies on - SA '17*. ACM Press, Bangkok, Thailand, 1–2. https://doi.org/10.1145/3132818.3132827

[67] Gun A. Lee, Theophilus Teo, Seungwon Kim, and Mark Billinghurst. 2018. A User Study on MR Remote Collaboration Using Live 360 Video. In *2018 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, Munich, Germany, 153–164. https://doi.org/10.1109/ISMAR.2018.00051

[68] D. Lester. 2013. Achieving Human Presence in Space Exploration. *Presence* 22, 4 (Nov. 2013), 345–349. https://doi.org/10.1162/PRES_a_00160

[69] Dan F. Lester, Kip Hodges, and Robert C. Anderson. 2017. Exploration telepresence: A strategy for optimizing scientific research at remote space destinations. *Science Robotics* 2, 7 (2017). https://doi.org/10.1126/scirobotics.aan4383 Publisher: American Association for the Advancement of Science.

[70] Chun Fui Liew and Takehisa Yairi. 2020. Companion Unmanned Aerial Vehicles: A Survey. *arXiv preprint arXiv:2001.04637* (2020).

[71] Zhicong Lu, Michelle Annett, Mingming Fan, and Daniel Wigdor. 2019. "I feel it is my responsibility to stream": Streaming and Engaging with Intangible Cultural Heritage through Livestreaming. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*. ACM Press, Glasgow, Scotland Uk, 1–14. https://doi.org/10.1145/3290605.3300459

[72] Zhicong Lu, Michelle Annett, and Daniel Wigdor. 2019. Vicariously Experiencing it all Without Going Outside: A Study of Outdoor Livestreaming in China. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (Nov. 2019), 1–28. https://doi.org/10.1145/3359127

[73] Arnold M Lund. 2001. Measuring usability with the use questionnaire12. *Usability interface* 8, 2 (2001), 3–6.

[74] Taro Maeda, HideyUki Ando, Hiroyuki Iizuka, Tomoko Yonemura, Daisuke Kondo, and Masataka Niwa. 2011. Parasitic Humanoid: the wearable robotics as a behavioral assist interface like oneness between horse and rider. In *Proceedings of the 2nd Augmented Human International Conference on - AH '11*. ACM Press, Tokyo, Japan, 1–8. https://doi.org/10.1145/1959826.1959844

[75] Akira Matsuda, Kazunori Nozawa, and Jun Rekimoto. 2018. JackIn Neck: A Neckband Wearable Telepresence System Designed for High Comfortability. In *Proceedings of the 2018 ACM International Conference on Interactive Surfaces and Spaces - ISS '18*. ACM Press, Tokyo, Japan, 415–418. https://doi.org/10.1145/3279778.3279917

[76] Akira Matsuda, Kazunori Nozawa, Kazuki Takata, Atsushi Izumihara, and Jun Rekimoto. 2020. HapticPointer: A Neck-worn Device that Presents Direction by Vibrotactile Feedback for Remote Collaboration Tasks. In *Proceedings of the Augmented Humans International Conference*. ACM, Kaiserslautern Germany, 1–10. https://doi.org/10.1145/3384657.3384777

[77] Marvin Minsky. 1980. Telepresence. (1980).

[78] Kana Misawa, Yoshio Ishiguro, and Jun Rekimoto. 2012. Ma petite chérie: what are you looking at?: a small telepresence system to support remote collaborative work for intimate communication. In *Proceedings of the 3rd Augmented Human International Conference on - AH '12*. ACM Press, Megève, France, 1–5. https://doi.org/10.1145/2160125.2160142

[79] Kana Misawa and Jun Rekimoto. 2015. ChameleonMask: a human-surrogate system with a telepresence face. In *SIGGRAPH Asia 2015 Emerging Technologies on - SA '15*. ACM Press, Kobe, Japan, 1–3. https://doi.org/10.1145/2818466.2818473

[80] Kana Misawa and Jun Rekimoto. 2015. ChameleonMask: Embodied Physical and Social Telepresence using Human Surrogates. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems - CHI EA '15*. ACM Press, Seoul, Republic of Korea, 401–411. https://doi.org/10.1145/2702613.2732506

[81] Kana Misawa and Jun Rekimoto. 2015. Wearing another's personality: a human-surrogate system with a telepresence face. In *Proceedings of the 2015 ACM International Symposium on Wearable Computers - ISWC '15*. ACM Press, Osaka, Japan, 125–132. https://doi.org/10.1145/2802083.2808392

[82] Kana Misawa and Jun Rekimoto. 2016. Who am I Touching?: User Study of Remote Handshaking with a Telepresence Face. In *Proceedings of the Fourth International Conference on Human Agent Interaction - HAI '16*. ACM Press, Biopolis, Singapore, 163–170. https://doi.org/10.1145/2974804.2974821

[83] Robb Mitchell, Alex Gillespie, and Brian O'Neill. 2011. Cyranic contraptions: using personality surrogates to explore ontologically and socially dynamic contexts. In *Procedings of the Second Conference on Creativity and Innovation in Design - DESIRE '11*. ACM Press, Eindhoven, Netherlands, 199. https://doi.org/10.1145/2079216.2079246

[84] Peter Mohr, Shohei Mori, Tobias Langlotz, Bruce H. Thomas, Dieter Schmalstieg, and Denis Kalkofen. 2020. Mixed Reality Light Fields for Interactive Remote Assistance. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–12. https://doi.org/10.1145/3313831.3376289

[85] Ahmed E. Mostafa, Kori Inkpen, John C. Tang, Gina Venolia, and William A. Hamilton. 2016. SocialStreamViewer: Guiding the Viewer Experience of Multiple Streams of an Event. In *Proceedings of the 19th International Conference on Supporting Group Work - GROUP '16*. ACM Press, Sanibel Island, Florida, USA, 287–291. https://doi.org/10.1145/2957276.2957286

[86] Jorg Muller, Tobias Langlotz, and Holger Regenbrecht. 2016. PanoVC: Pervasive telepresence using mobile phones. In *2016 IEEE International Conference on Pervasive Computing and Communications (PerCom)*. IEEE, Sydney, NSW, 1–10. https://doi.org/10.1109/PERCOM.2016.7456508

[87] Isabel Briggs Myers. 1962. The Myers-Briggs Type Indicator: Manual (1962). (1962). Publisher: Consulting Psychologists Press.

[88] Shohei Nagai, Shunichi Kasahara, and Jun Rekimoto. 2015. Directional communication using spatial sound in human-telepresence. In *Proceedings of the 6th Augmented Human International Conference on - AH '15*. ACM Press, Singapore, Singapore, 159–160. https://doi.org/10.1145/2735711.2735818

[89] Shohei Nagai, Shunichi Kasahara, and Jun Rekimoto. 2015. LiveSphere: Sharing the Surrounding Visual Environment for Immersive Experience in Remote Collaboration. In *Proceedings of the Ninth International Conference on Tangible, Embedded, and Embodied Interaction - TEI '14*. ACM Press, Stanford, California, USA, 113–116. https://doi.org/10.1145/2677199.2680549

[90] Carman Neustaedter and Saul Greenberg. 2012. Intimacy in long-distance relationships over video chat. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems - CHI '12*. ACM Press, Austin, Texas, USA, 753. https://doi.org/10.1145/2207676.2207785

[91] Carman Neustaedter and Tejinder K. Judge. 2010. Peek-A-Boo: the design of a mobile family media space. In *Proceedings of the 12th ACM international conference adjunct papers on Ubiquitous computing - Ubicomp '10*. ACM Press, Copenhagen, Denmark, 449. https://doi.org/10.1145/1864431.1864482

[92] Carman Neustaedter, Carolyn Pang, Azadeh Forghani, Erick Oduor, Serena Hillman, Tejinder K. Judge, Michael Massimi, and Saul Greenberg. 2015. Sharing Domestic Life through Long-Term Video Connections. *ACM Transactions on Computer-Human Interaction* 22, 1 (Feb. 2015), 1–29. https://doi.org/10.1145/2696869

[93] Carman Neustaedter, Jason Procyk, Anezka Chua, Azadeh Forghani, and Carolyn Pang. 2020. Mobile Video Conferencing for Sharing Outdoor Leisure Activities Over Distance. *Human–Computer Interaction* 35, 2 (March 2020), 103–142. https://doi.org/10.1080/07370024.2017.1314186

[94] Carman Neustaedter and Anthony Tang. 2019. Explorations of Remote Attendance at CHI. https://chi2019.acm.org/2019/02/15/explorations-of-remote-attendance-at-chi/

[95] Carman Neustaedter, Gina Venolia, Jason Procyk, and Dan Hawkins. 2016. To Beam or Not to Beam: A Study of Remote Telepresence Attendance at an Academic Conference. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing - CSCW '16*. ACM Press, San Francisco, California, USA, 417–430. https://doi.org/10.1145/2818048.2819922

[96] BBC News. 2021. *Covid: New lockdowns for England and Scotland ahead of 'hardest weeks'.* Retrieved 2021-01-12 from https://www.bbc.com/news/uk-55538937

[97] Mohammad Obaid, Felix Kistler, Gabrielė Kasparavičiūtė, Asim Evren Yantaç, and Morten Fjeld. 2016. How would you gesture navigate a drone?: a user-centered approach to control a drone. In *Proceedings of the 20th International Academic Mindtrek Conference on - AcademicMindtrek '16*. ACM Press, Tampere, Finland, 113–121. https://doi.org/10.1145/2994310.2994348

[98] Catherine S. Oh, Jeremy N. Bailenson, and Gregory F. Welch. 2018. A Systematic Review of Social Presence: Definition, Antecedents, and Implications. *Frontiers in Robotics and AI* 5 (Oct. 2018), 114. https://doi.org/10.3389/frobt.2018.00114

[99] Rui Pan and Carman Neustaedter. 2017. Streamer.Space: A Toolkit for Prototyping Context-Aware Mobile Video Streaming Apps. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems - CHI EA '17*. ACM Press, Denver, Colorado, USA, 1947–1954. https://doi.org/10.1145/3027063.3053083

[100] Rui Pan, Samarth Singhal, Bernhard E. Riecke, Emily Cramer, and Carman Neustaedter. 2017. "MyEyes": The Design and Evaluation of First Person View Video Streaming for Long-Distance Couples. In *Proceedings of the 2017 Conference on Designing Interactive Systems - DIS '17*. ACM Press, Edinburgh, United Kingdom, 135–146. https://doi.org/10.1145/3064663.3064671

[101] Tomislav Pejsa, Julian Kantor, Hrvoje Benko, Eyal Ofek, and Andrew Wilson. 2016. Room2Room: Enabling Life-Size Telepresence in a Projected Augmented Reality Environment. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing (CSCW '16)*. ACM, New York, NY, USA, 1716–1725. https://doi.org/10.1145/2818048.2819965 event-place: San Francisco, California, USA.

[102] Hannah R. M. Pelikan, Amy Cheatle, Malte F. Jung, and Steven J. Jackson. 2018. Operating at a Distance - How a Teleoperated Surgical Robot Reconfigures Teamwork in the Operating Room. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (Nov. 2018), 1–28. https://doi.org/10.1145/3274407

[103] Kevin Pfeil, Seng Lee Koh, and Joseph LaViola. 2013. Exploring 3d gesture metaphors for interaction with unmanned aerial vehicles. In *Proceedings of the 2013 international conference on Intelligent user interfaces - IUI '13*. ACM Press, Santa Monica, California, USA, 257. https://doi.org/10.1145/2449396.2449429

[104] Jason Procyk, Carman Neustaedter, Carolyn Pang, Anthony Tang, and Tejinder K. Judge. 2014. Exploring video streaming in public settings: shared geocaching over distance using mobile video chat. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems - CHI '14*. ACM Press, Toronto, Ontario, Canada, 2163–2172. https://doi.org/10.1145/2556288.2557198

[105] Irene Rae, Gina Venolia, John C. Tang, and David Molnar. 2015. A Framework for Understanding and Designing Telepresence. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing - CSCW '15*. ACM Press, Vancouver, BC, Canada, 1552–1566. https://doi.org/10.1145/2675133.2675141

[106] Stuart Reeves, Christian Greiffenhagen, Martin Flintham, Steve Benford, Matt Adams, Ju Row Farr, and Nicholas Tandavantij. 2015. I'd Hide You: Performing Live Broadcasting in Public. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15*. ACM Press, Seoul, Republic of Korea, 2573–2582. https://doi.org/10.1145/2702123.2702257

[107] Jun Rekimoto, Keishiro Uragaki, and Kenjiro Yamada. 2018. Behind-the-mask: A Face-through Head-mounted Display. In *Proceedings of the 2018 International Conference on Advanced Visual Interfaces*. 1–5.

[108] Ryan Rocca. 2020. *Here are the fines people could face for violating Ontario's coronavirus shutdown laws*. Retrieved 2021-01-12 from https://globalnews.ca/news/7542586/ontario-coronavirus-lockdown-fines/

[109] Adam Rugg and Benjamin Burroughs. 2016. Periscope, live-streaming and mobile video culture. *Geoblocking and global video culture* 18 (2016), 64–73.

[110] Thomas Schubert, Frank Friedmann, and Holger Regenbrecht. 2001. The experience of presence: Factor analytic insights. *Presence: Teleoperators & Virtual Environments* 10, 3 (2001), 266–281. Publisher: MIT Press.

[111] Hanieh Shakeri and Carman Neustaedter. 2019. Teledrone: Shared Outdoor Exploration Using Telepresence Drones. In *Conference Companion Publication of the 2019 on Computer Supported Cooperative Work and Social Computing* (Austin, TX, USA) *(CSCW '19)*. Association for Computing Machinery, New York, NY, USA, 367–371. https://doi.org/10.1145/3311957.3359475

[112] John Short, Ederyn Williams, and Bruce Christie. 1976. *The social psychology of telecommunications*. John Wiley & Sons.

[113] Sudheesh Singanamalla, William Thies, and Colin Scott. 2018. Avatar: Enabling Immersive Collaboration via Live Mobile Video. In *Proceedings of the 3rd International Workshop on Multimedia Alternate Realities - AltMM'18*. ACM Press, Seoul, Republic of Korea, 9–14. https://doi.org/10.1145/3268998.3269003

[114] Samarth Singhal and Carman Neustaedter. 2017. BeWithMe: An Immersive Telepresence System for Distance Separated Couples. In *Companion of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing - CSCW '17 Companion*. ACM Press, Portland, Oregon, USA, 307–310. https://doi.org/10.1145/3022198.3026310

[115] Mel Slater, Martin Usoh, and Anthony Steed. 1994. Depth of presence in virtual environments. *Presence: Teleoperators & Virtual Environments* 3, 2 (1994), 130–144. Publisher: MIT Press.

[116] Daxton R. “Chip” Stewart and Jeremy Littau. 2016. Up, Periscope: Mobile Streaming Video Technologies, Privacy in Public, and the Right to Record. *Journalism & Mass Communication Quarterly* 93, 2 (June 2016), 312–331. https://doi.org/10.1177/1077699016637106

[117] Brett Stoll, Samantha Reig, Lucy He, Ian Kaplan, Malte F. Jung, and Susan R. Fussell. 2018. Wait, Can You Move the Robot?: Examining Telepresence Robot Use in Collaborative Teams. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction (HRI ’18)*. ACM, New York, NY, USA, 14–22. https://doi.org/10.1145/3171221.3171243 event-place: Chicago, IL, USA.

[118] Anthony Tang, Omid Fakourfar, Carman Neustaedter, and Scott Bateman. 2017. Collaboration with 360° Videochat: Challenges and Opportunities. In *Proceedings of the 2017 Conference on Designing Interactive Systems - DIS ’17*. ACM Press, Edinburgh, United Kingdom, 1327–1339. https://doi.org/10.1145/3064663.3064707

[119] John Tang, Gina Venolia, Kori Inkpen, Charles Parker, Robert Gruen, and Alicia Pelton. 2017. Crowdcasting: Remotely Participating in Live Events Through Multiple Live Streams. *Proceedings of the ACM on Human-Computer Interaction* 1, CSCW (Dec. 2017), 1–18. https://doi.org/10.1145/3134733

[120] Franco Tecchia, Leila Alem, and Weidong Huang. 2012. 3D helping hands: a gesture based MR system for remote collaboration. In *Proceedings of the 11th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and its Applications in Industry - VRCAI ’12*. ACM Press, Singapore, Singapore, 323. https://doi.org/10.1145/2407516.2407590

[121] Ashutosh Tewari, James Peabody, Richard Sarle, Guruswami Balakrishnan, Ashok Hemal, Alok Shrivastava, and Mani Menon. 2002. Technique of da vinci robot-assisted anatomic radical prostatectomy. *Urology* 60, 4 (Oct. 2002), 569–572. https://doi.org/10.1016/S0090-4295(02)01852-6

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

[123] Hiroaki Tobita. 2017. Gutsy-Avatar: Computational Assimilation for Advanced Communication and Collaboration. In *2017 First IEEE International Conference on Robotic Computing (IRC)*. IEEE, Taichung, Taiwan, 8–13. https://doi.org/10.1109/IRC.2017.82

[124] K Tsui, Adam Norton, David Brooks, H Yanco, and Daniel Kontak. 2011. Designing telepresence robot systems for use by people with special needs. In *Int. Symposium on Quality of Life Technologies: Intelligent Systems for Better Living*.

[125] Yuichi Tsumaki, Fumiaki Ono, and Taisuke Tsukuda. 2012. The 20-DOF miniature humanoid MH-2: A wearable communication system. In *2012 IEEE International Conference on Robotics and Automation*. IEEE, St Paul, MN, USA, 3930–3935. https://doi.org/10.1109/ICRA.2012.6224810

[126] Francine Uenuma. [n.d.]. Video Chat Is Helping Us Stay Connected in Lockdown. But the Tech Was Once a ‘Spectacular Flop’. https://time.com/5834516/video-chat-zoom-history/

[127] Joy Van Baren. 2004. Measuring presence: A guide to current measurement approaches. *Deliverable of the OmniPres project IST-2001-39237* (2004).

[128] Tim van Oosterhout and Arnoud Visser. 2008. A visual method for robot proxemics measurements. In *Proceedings of Metrics for Human-Robot Interaction: A Workshop at the Third ACM/IEEE International Conference on Human-Robot Interaction (HRI 2008). Citeseer*. 61–68.

[129] Gina Venolia, John C. Tang, Kori Inkpen, and Baris Unver. 2018. Wish you were here: being together through composite video and digital keepsakes. In *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services*. ACM, Barcelona Spain, 1–11. https://doi.org/10.1145/3229434.3229476

[130] Samantha Vu, Mikko J. Rissanen, Natalie Pang, and Schubert Foo. 2012. Towards evaluating social telepresence in mobile context. In *Proceedings of the 11th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and its Applications in Industry - VRCAI ’12*. ACM Press, Singapore, Singapore, 75. https://doi.org/10.1145/2407516.2407539

[131] Vanissa Wanick, Guilherme Xavier, and Erhan Ekmekcioglu. 2018. Virtual Transcendence Experiences: Exploring Technical and Design Challenges in Multi-Sensory Environments. In *Proceedings of the 10th International Workshop on Immersive Mixed and Virtual Environment Systems - MMVE ’18*. ACM Press, Amsterdam, Netherlands, 7–12. https://doi.org/10.1145/3210438.3210444

[132] Sabine Webel, Uli Bockholt, Timo Engelke, Nirit Gavish, Manuel Olbrich, and Carsten Preusche. 2013. An augmented reality training platform for assembly and maintenance skills. *Robotics and Autonomous Systems* 61, 4 (April 2013), 398–403. https://doi.org/10.1016/j.robot.2012.09.013

[133] Nadir Weibel, Danilo Gasques, Janet Johnson, Thomas Sharkey, Zhuoqun Robin Xu, Xinming Zhang, Enrique Zavala, Michael Yip, and Konrad Davis. 2020. ARTEMIS: Mixed-Reality Environment for Immersive Surgical Telementoring. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–4. https://doi.org/10.1145/3334480.3383169

[134] Alexander Wilberz, Dominik Leschtschow, Christina Trepkowski, Jens Maiero, Ernst Kruijff, and Bernhard Riecke. 2020. FaceHaptics: Robot Arm based Versatile Facial Haptics for Immersive Environments. In *Proceedings of the 2020*

CHI Conference on Human Factors in Computing Systems. ACM, Honolulu HI USA, 1–14. https://doi.org/10.1145/3313831.3376481

[135] Bob G Witmer and Michael J Singer. 1998. Measuring presence in virtual environments: A presence questionnaire. *Presence* 7, 3 (1998), 225–240. Publisher: MIT Press.

[136] Lillian Yang, Brennan Jones, Carman Neustaedter, and Samarth Singhal. 2018. Shopping Over Distance through a Telepresence Robot. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (Nov. 2018), 1–18. https://doi.org/10.1145/3274460

[137] Svetlana Yarosh, Kori M. Inkpen, and A.J. Bernheim Brush. 2010. Video playdate: toward free play across distance. In *Proceedings of the 28th international conference on Human factors in computing systems - CHI '10*. ACM Press, Atlanta, Georgia, USA, 1251. https://doi.org/10.1145/1753326.1753514

[138] Svetlana Yarosh, Panos Markopoulos, and Gregory D. Abowd. 2014. Towards a Questionnaire for Measuring Affective Benefits and Costs of Communication Technologies *(CSCW '14)*. Association for Computing Machinery, New York, NY, USA, 84–96. https://doi.org/10.1145/2531602.2531634

[139] Jacob Young, Tobias Langlotz, Matthew Cook, Steven Mills, and Holger Regenbrecht. 2019. Immersive Telepresence and Remote Collaboration using Mobile and Wearable Devices. *IEEE Transactions on Visualization and Computer Graphics* 25, 5 (May 2019), 1908–1918. https://doi.org/10.1109/TVCG.2019.2898737

[140] Jacob Young, Tobias Langlotz, Steven Mills, and Holger Regenbrecht. 2020. Mobileportation: Nomadic Telepresence for Mobile Devices. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 2 (June 2020), 1–16. https://doi.org/10.1145/3397331

[141] Clint Zeagler. 2017. Where to Wear It: Functional, Technical, and Social Considerations in on-Body Location for Wearable Technology 20 Years of Designing for Wearability. In *Proceedings of the 2017 ACM International Symposium on Wearable Computers* (Maui, Hawaii) *(ISWC '17)*. Association for Computing Machinery, New York, NY, USA, 150–157. https://doi.org/10.1145/3123021.3123042

[142] Nadia Zenati-Henda, Abdelkader Bellarbi, Samir Benbelkacem, and Mahmoud Belhocine. 2014. Augmented reality system based on hand gestures for remote maintenance. In *2014 International Conference on Multimedia Computing and Systems (ICMCS)*. IEEE, Marrakech, Morocco, 5–8. https://doi.org/10.1109/ICMCS.2014.6911258

Received June 2020; revised October 2020; accepted December 2020