A Framework for Understanding and Designing Telepresence

Irene Rae[1,2], Gina Venolia[1], John C. Tang[1], David Molnar[1]

[1] Microsoft Research
One Microsoft Way
Redmond, WA 98008
{ginav | johntang | dmolnar} @microsoft.com

[2] University of Wisconsin—Madison
1210 West Dayton Street
Madison, WI 53706 USA
irene@cs.wisc.edu

ABSTRACT

As a field, telepresence has grown to include a wide range of systems, from multi-view videoconferencing units to humanlike androids. However, the diversity of systems and research makes it difficult to form a holistic understanding of where the field stands. We propose a framework consisting of seven design dimensions for understanding telepresence, iteratively developed from previous literature, a series of three surveys, the construction of two design probes, and a field study. These design dimensions uniquely categorize 17 telepresence scenarios. In this work, we explain our development process, describe our design dimensions—initiation, physical environment, mobility, vision, social environment, communication, and independence—as well as our scenarios, and demonstrate the use of our framework as a tool to (1) highlight opportunities for future work, (2) identify generalizable findings from research, and (3) facilitate communication in the telepresence community.

Author Keywords

Telepresence; video-mediated communication; CSCW.

ACM Classification Keywords

H.5.3. Group and Organization Interfaces

INTRODUCTION

"The biggest challenge to developing telepresence is achieving that sense of 'being there.' Can telepresence be a true substitute for the real thing?" – Marvin Minsky [39]

When Marvin Minsky coined the term "telepresence," he envisioned technologies where remote users' experiences would be so similar to actually being there that there would be no noticeable difference [39]. Since then, innovation and research within the realm of telepresence has encompassed a vast range of technologies aimed at achieving this vision. From work on the use of these systems in specific domains, such as medical [31,64], education [67,68], or office settings [49,58,63,69], to the development of novel systems such as flying telepresence blimps [46] or three dimensional embodiments of the remote user [40,42], the diversity of approaches has not only allowed exploring many different avenues of research, but has also informed our understanding of how these systems mediate our communications.

However, as the field of telepresence has matured, this diversity has caused the relationship between disparate studies to become increasingly complicated, making it difficult to understand how each piece is situated within the greater whole. This lack of cohesive structure can obscure opportunities for new research and can mean that designers start from scratch each time a new technology is introduced. As a result, researchers and designers are often faced with reinventing solutions, reproducing results, or struggling to identify whether their findings are novel.

To address this, we propose a framework for the telepresence research space which enables designers and researchers to:

- Identify opportunities for future research.
- Reflect on past and future work to determine potentially generalizable findings.
- Share a structure and language for reporting research to facilitate interaction within the telepresence community.

To develop this framework, we engaged in a multi-stage iterative process. This process involved examining previous literature and conducting a series of exploratory surveys. We used these surveys to identify the contexts in which telepresence technologies might be used and the unique design dimensions which distinguish them. Based on this initial framework, we created two design probes—developed to match the needs of disparate scenarios—and field tested

Figure 1. Participants in our field study share a meal (left) or a visit to the local art museum (right) with a remote user.

CSCW '15, March 14–18, 2015, Vancouver, BC, Canada.

ACM 978-1-4503-2922-4/15/03...$15.00.
http://dx.doi.org/10.1145/2675133.2675141

them as a proof of concept. Last, we used the insights gained from these field tests to further refine our framework.

RELATED WORK

As various human-computer interaction domains have matured, each has grappled with finding a structure to organize their growing bodies of work. Examples in CSCW include Gutwin et al.'s framework for workspace awareness [19], Neale et al.'s model for evaluating computer-supported cooperative work [41], Benford et al.'s review of spatial approaches [3], and the work of Ellis et al. and Grudin [13,18] in groupware. With over 1800 papers and journal articles classified as telepresence work by 2007 [34], telepresence is reaching a point where the need for such an organizational structure is becoming critical.

For example, what is considered telepresence has varied across disciplines and has become more confounded as the term has become ubiquitous. Draper et al. divided the prevalent telepresence methods into three main categories: (1) *simple telepresence*, the ability to operate in a computer-mediated environment; (2) *cybernetic telepresence*, the efficiency or quality of the human-machine interface; and (3) *experiential telepresence*, the state where the user feels physically present within the mediated environment [12].

Other literature has organized specific telepresence subdomains. For example, Kristoffersson et al. created a comprehensive survey of past literature and available systems in the realm of mobile telepresence robots [28], Haans et al. developed a framework for understanding embodiment in telepresence [20], Draper et al. surveyed overarching theories of telepresence [12], and multiple papers have proposed robotic telepresence design requirements [10,11,52].

While their approaches have differed, these works have largely focused on creating an organizational structure as a tool, rather than delivering an exhaustive taxonomy of their fields. We have taken an analogous approach by synthesizing previous work and extending it into a conceptual framework. We therefore introduce further related work within the description of each design dimension or scenario later in the paper to provide a contextualized understanding for how it has informed our development process.

FRAMEWORK DEVELOPMENT

We next describe each stage of our development process, beginning with a series of three surveys. We designed these surveys to explore how people envision using telepresence technologies in the future unbound by current technical constraints and to augment prior theories in face-to-face interactions [15,26,59]. We then discuss the development of two design probes, and end by detailing the process and results of our field study.

Survey 1

Our first survey was designed to extend techniques in scenario-based design [7,55], providing an exploratory venue to understand the breadth of experiences that people could envision using a technology to attend remotely.

We asked participants three open-ended questions about how they might envision using a telepresence technology ("Imagine that there was a technology that allowed you to appear in a different place, just as if you were actually there, without having to travel. What would you use this technology to do?")

Participants. A total of 44 adults anonymously participated in our survey via SurveyGizmo, recruited through Cint (http://www.cint.com).

Analysis & Results. We used open-coding and affinity diagramming to systematically identify trends in responses. We grouped these trends into 26 categories of scenarios.

Next, we examined prior theoretical literature and work describing design requirements for various telepresence systems. Informed by this research, we identified nine potential design dimensions which might vary in priority between scenarios.

Survey 2

Our second survey focused on validating our categories of scenarios by asking about participants' interest in using a technology to remotely attend each one. Participants rated their interest on a five-point Likert scale (1=not at all interested, 5=extremely interested). We also asked two open ended questions to refine our design dimensions and scenario categories ("What do you think would be challenging about using this technology?" "Please tell us about other things that you would want to use this technology for.").

Shared Experience Scenarios (shared socially)	
Physically interactive activities *(dance, play basketball)*	Visiting a new place *(museum, new apartment)*
Object-centered activities *(play cards, fish)*	Sharing a meal *(lunch dates, family dinners)*
Movement-based activities *(hike, ski)*	Talking socially *(hang out, share news)*
Giving or receiving help *(fix a bike, give shopping input)*	Mingling *(parties, conferences)*
Goal-based collaboration *(work meetings, craft projects)*	Mingling and viewing *(weddings, graduations)*
Shared viewing *(movies, ballgames, new purchases)*	
Solo Experience Scenarios (on your own)	
Self-improvement *(take a class, train)*	Obligations of Presence *(jury duty, wait in line)*
Monitoring *(check on house, watch kids)*	Performance *(act, conduct a choir)*
Exploring on your own *(solo tour, cultural site)*	Hazardous environments *(first response, military)*

Table 1. Surveys yielded 17 categories of scenarios.

Participants. We recruited a total of 174 adult participants (75 male, 99 female) through Cint, administered through SurveyGizmo, and Instant.ly. Ages were distributed thusly: ages 18-24: 17, ages 25-34: 35, ages 35-54: 75, age 55+: 47.

Analysis & Results. We calculated the average and standard deviations of how much interest participants showed in each scenario. We used these averages to create a ranked list of scenarios to choose the design contexts for our field study.

We also used open-coding and affinity diagramming on the responses to the open-ended questions, resulting in a re-categorization and refinement of our previous scenarios—from 26 types down to 17. These naturally divided into two major groups, *shared experiences*, where social interaction with others was the primary goal, and *solo experiences*, where the focus was on a specific activity and social interactions were secondary, as shown in Table 1.

Additionally, we revisited our original nine design dimensions in the context of these 17 scenarios to test whether these dimensions sufficiently encompassed the characteristics unique to each scenario. Based on these criteria, we further refined our initial nine design dimensions down to seven:

1. ***Initiation.*** How a telepresence interaction is started.
2. ***Physical Environment.*** The physical environment that the system will be used in.
3. ***Mobility.*** The obstacles and amount of movement the remote user will engage in while using the system.
4. ***Vision.*** What users see and how they adjust their visual focus of attention.
5. ***Social environment.*** Relationships among all the *stakeholders*—those involved both directly and indirectly in the interaction.
6. ***Communication.*** How users communicate or interact.
7. ***Independence.*** The level of autonomy that users have.

Next, we provide a brief description of each design dimension and the previous work related to it.

Initiation. The initiation dimension encompasses the factors that influence how interaction begins between local and remote users via a telepresence system.

Theories in the context of telephone interactions discuss the concept of *caller hegemony*—the power disparity between the caller and the answerer [22]. Later work in the context of cellphones elaborated on this concept by exploring how changes to the technology, such as the addition of caller id, disrupts hegemony power dynamics [23].

In the context of telepresence robots, which further shifts the power dynamics in caller hegemony, previous field research has investigated the instigation of interactions when remote users, or *remotes*, have been present via an embodied telepresence system. These studies showed that when a remote user is connected via a telepresence robot, there is an increase in informal and spontaneous interactions initiated by users local to the system, or *locals* [32,63]. Other work focused on user concerns about the etiquette of starting a call and apprehensions about privacy [1,10]. Last, participants in our surveys brought up concerns about the potential setup costs and the overhead of learning to use a new technology.

Items in this design dimension therefore seek to answer such questions as: Who has the ability to begin the interaction? What is the etiquette surrounding initiating the interaction? What options does the non-initiating party have for controlling the other person's access? What is the amount of planning required for the interaction to take place and are there any additional setup costs?

Physical Environment. Although physical environments may vary, successful operation within a setting is a common challenge in designing telepresence systems. When examining obstacles to movement, research has explored a number of approaches, from telepresence blimps [46] to wheeled vehicles that can traverse hazardous terrains [38] to wearable companions [40]. Beyond transportation modes, other aspects of the physical environment have been highlighted in prior work, such as spatial formations in co-present interactions [26], ambient noise levels in real-world scenarios [24,60,61] or the need to interact with or share items to enhance feelings of presence in remote meals [66].

Questions within the scope of the physical environment include: Where will interactions take place (e.g., indoors, outdoors, or both)? What will the ambient noise level be? What spatial configurations might users need to interact within? Where will users manipulate items or interact with a workspace? What properties might the workspace have?

Mobility. As emphasized in the literature, different contexts may present very different obstacles to mobility [5,24]. Prior research has highlighted contrasts—such as terrain, number and type of obstacles, and range of motion—between movement in business settings and movement in non-workplace environments [5,24].

Additionally, while providing mobility for the remote user may increase their feelings of presence, it may not immediately contribute to task outcomes due to other factors, such as cognitive load [50]. While not every obstacle can be predicted, anticipating the needs and constraints for the target scenario and prioritizing based on the context may reduce breakdowns and improve user experiences.

Examples of key elements to consider within this design dimension are: How large is the area that the system is expected to navigate within? What barriers exist to the system's mobility (e.g., stairs, elevators that can also block network connectivity)? How might the design of the system restrict the local user's or bystander's freedom of movement?

Vision. When viewing the environment, the remote user's ability to focus on objects at a variety of ranges [24] and to understand how they are situated in the local's environment

[21] can be critical to their enjoyment and ability to complete tasks. In collaborative work scenarios, the amount of visual data from the shared workspace that is communicated to both parties has been shown to facilitate the success of their outcomes [15,19,58]. In Draper et al.'s attentional model of telepresence, visual information is particularly critical as a focus of attentional resources for remote users, which they argue is the central component in telepresence [12].

Prior work has found that the ability to see the other person improves collaborative outcomes [4]. Research has also shown that the visibility of the remote user's upper body—beyond just their face—improved empathy between users and increased interaction fluency [44]. In addition to examining the amount of visual information shared, telepresence systems have also explored the preservation of spatial positioning [43,57] and eye gaze [40].

Although numerous ways may be found to supplement visual information for both the local and the remote users, different scenarios may have different needs. In the dimension of vision, designers and researchers may seek answers to questions such as: What will the primary *focus of attention* during the interaction be (e.g., other people, shared workspace, the environment)? How far do we expect the primary focus of attention to be from the viewer? Will the ability to perceive referential gestures be important? What kind of visual feedback will be provided about the system and the state of the users?

Social Environment. Research has begun to explore the relationship between theoretical work in face-to-face interactions and how it may translate to remote interactions. For example, within the context of Actor-Network theory, research on cellphone usage has examined remote users as actor-nodes [23,30]. Within the context of public places, theories on normative behaviors and social landscapes have been extended to examine how cellphone usage disrupts the status of being with another person—shifting those who are not engaged in the call into the status of being alone [16,23].

Informed by these theoretical contributions, we further categorized the people that might be affected by the use of a telepresence system into six types of *stakeholders*—those involved in the interaction directly as well as those involved peripherally, or indirectly.

Remote users: The user or users in a remote location who are managing the connection as the main actors. For example, the mother holding most of the conversation with her son, who is the local user.

Remote party members: People in the remote location who are peripherally involved in the interaction. For example, the father who may occasionally come into view to answer a question but who is doing other things in his environment.

Remote bystanders: People who are not part of the interaction, but are in the remote environment and affected by what occurs. For example, the daughter who is in the same room as the parents but is involved in her own activities.

Local users: The user or users who are local to the telepresence system who are participating in the interaction as main actors. For example, the son who is interacting with his remotely connected mother.

Local party members: People in the local environment who are peripherally involved in the interaction. For example, the son's friends who may occasionally join the conversation.

Local bystanders: People who are not part of the interaction, but are in the local environment of the system and are affected by what occurs. For example, if the interaction is taking place at a restaurant, other customers in the restaurant.

Previous work primarily focused on two stakeholders, the local and the remote users. These studies have largely focused on either: (1) examining the effects that telepresence technologies have on the relationship between the users; or (2) defining the mediating effects that these technologies have on user perceptions.

Studies on the effects of telepresence technologies on user relationships explored how systems might be used differently between work and leisure contexts [5] and how they might support the formation of common ground—a common base of shared knowledge and vocabulary [9,45]. Related research also raised questions about what etiquette should exist or how the privacy (or lack of it) afforded by telepresence systems may strain users' ties [1,10,11,37].

When seeking to define the mediating effects that telepresence technologies may have on user perceptions, literature has shown that instant messaging or videoconferencing may equalize user relationships [27] and that in videoconferencing, the amount of user multitasking may be reduced [5,65]. Research has also begun to explore how the appearance of telepresence technologies might shape user perceptions. Examples include manipulating the remote user's level of embodiment to increase the development of trust between users [48], altering the height of a robotic telepresence system to support or undermine the remote user's authority [46,49], and creating a more human-like embodiment to increase the local user's perceptions of the remote user's presence [54]. Additional studies have also discussed leveraging the appearance of the system's embodiment to aid in identifying remote users [32,49,63].

Informed by this previous work and our surveys, questions that fit within the realm of the social environment include: How private is the setting expected to be? Do the users have pre-existing relationships? How might pre-existing relationships shape the way in which the system is used? In what ways might the embodiment or presentation of the system influence perceptions of the local or remote users?

Communication. The ability to communicate with others verbally, through gesture, and through language has been shown to be a core component in face-to-face interactions [15,59]. As a result, most telepresence systems integrate aural and visual information as a matter of course. Despite these commonalities, many questions arise about how large the visual field should be [10,11,24], what constitutes critical visual information (e.g., the ability to see referential gestures, posture of users, spatial positioning of users) [4,21,29,52], at what fidelity this information should be available [10,11], and how to measure communicative performance [61].

In addition to aural and visual information, different modalities for promoting communication between users have been explored, such as indicators for when a telepresence system is occupied [60], signaling other people's meal status in shared dining scenarios [17], allowing people to send messages via a tablecloth and printed on food [66], or integrating deictic gestures using a laser pointer [36].

In considering factors for the dimension of communication, questions should fit the scope of the scenario. For example: What verbal and non-verbal cues will be critical in this context? How much time might be spent listening vs. speaking? What gestures may need to be supported in this context? Is physical contact common in these scenarios?

Independence. Prior work on cellular phones highlights the power disparity between caller and answerer, particularly in light of social norms, and how the addition of various supportive technologies may shift that dynamic [1,22,23]. As advances in navigation and visual identification systems have enabled remote or autonomous control, particularly in robotic telepresence systems [8,35,60], new insights and theories may be constructed about how greater independence may affect user relationships, such as in building trust [48].

Items in this design dimension seek to answer questions such as: Should the users be able to act independently of each other? Will any of the users be reliant on others for direction or to act? Will design choices, such as those affecting situational awareness or cognitive load, empower users? Will users feel in control? Are there points of enforced dependency due to design choices?

Survey 3

We constructed our third online questionnaire to examine each of the 17 scenarios that we identified earlier in the context of our seven design dimensions. Our goal was to augment the theoretical work on understanding face-to-face interactions with what people felt to be crucial in the context of their own public and private experiences. By combining our survey responses in each of these scenarios with prior theory, we sought to create a blueprint for how these design dimensions are prioritized in different contexts.

To accomplish this, we asked participants to choose three of the scenarios that they had experienced, in-person, in the past. We asked them to recall these face-to-face experiences and to answer a combination of open-ended, multiple choice, percentage, rating, and yes or no questions about them. Examples of responses to these questions are in Table 2.

Participants. A total of 121 adults (45 male, 76 female), recruited through Cint, responded to our survey resulting in descriptions of 342 face-to-face experiences. Ages were distributed as follows: ages 18-24: 24, ages 25-34: 25, ages 35-44: 23, ages 45-54: 24, and age 55+: 25.

Analysis & Results. We examined each scenario separately, calculating the average, standard deviation, or percentage responding positively, for every question. We then compared these scores across scenarios. From this comparison, we identified the elements in each design dimension that had the highest contrast between scenarios, as illustrated in Table 2.

Our goal in collecting responses about people's face-to-face experiences was to uncover how these scenarios differed in interactions unmediated by technology in both public and private contexts. Revealing these distinctions might help predict when and where interactions within these scenarios would differ when using a telepresence technology.

These insights can also aid us in:

- Predicting and prioritizing what functionalities should be implemented in future systems.
- Understanding and predicting where results from research may generalize across scenarios, or where perceptions, attitudes, and outcomes may differ.
- Highlighting that a one-size-fits-all approach to telepresence is unlikely to meet the needs of all scenarios.

However, these potential benefits are based on the theory that needs and constraints vary across scenarios and that our framework of design dimensions can help differentiate them.

Design Probes

To validate our framework, we chose two scenarios that had large contrasts along our design dimensions, as shown in Table 2. Due to the many differences between scenarios identified in our third survey, we further narrowed the scope of our testing by prioritizing the most popular scenarios from our second survey. These criteria resulted in our choosing two scenarios to test: remotely sharing a meal together and remotely visiting a new place with someone. Using the needs found along our design dimensions from our third survey as a guide, we constructed two design probes: a "freestanding" probe and a "wearable" probe.

Freestanding Device: Sharing a Meal

We designed a freestanding device, shown in Figure 2a, for sharing a meal with a remote person. Similar in design to previous systems [63,69] our freestanding probe consisted of a Windows tablet with an 11.6 inch 1080p screen, an external battery, a web camera, a speaker system, and a microphone, mounted on a laser cut plastic frame on a metal stand. Skype v6.6 in full screen mode was used for videoconferencing.

Below, we list the key choices made in each design dimension and match them to the results from our third survey, shown in Table 2.

- ▪ Initiation: Setup time for the freestanding device was high, requiring installation of additional software for the remote user, and positioning/transport of the device for the local user, as we chose to deprioritize this dimension.

▪ Physical Environment: We worked to accommodate long periods of time where the local's hands would be occupied and minimized the amount of dynamic adjustment that locals had to do by providing the remote user with software to control the positioning of the camera. To handle variable levels of background noise, we used a Jawbone Jambox with a CAD U7 USB Desktop Condenser microphone. The height of the system was adjustable to match the dining surface—30–66 inches to the center of the screen—so that no changes would be needed after setup.

▪ Mobility: We traded mobility for stability in our design, constructing our probe with a triangular metal stand, two wheels, and a portable battery pack that was secured to the base. In total, the freestanding probe weighed 6.5kg.

▪ Vision: We provided remote users with control over a Logitech BCC950 1080p webcam with pan, tilt, and zoom capability to accommodate visual switching at close to middle distances (e.g., other diners, food on the table).

▪ Social Environment: We leveraged expected existing relationships between remote and local users and gave the

		Physically interactive activities	Object-centered activities	Movement-based activities	Visiting a new place	Sharing a meal	Talking socially
Initiation	Initiator	Self: 36% Others: 36%	Self: 45% Other: 27%	Self: 53% Other: 33%	Self: 47% Other: 37%	Self: 34% Other: 41%	Self: 50% Other: 29%
	Pre-planning	Planned: 47% Semi-planned: 37% Spontaneous: 37%	Planned: 9% Semi-planned: 36% Spontaneous: 55%	Planned: 60% Semi-planned: 40% Spontaneous: 0%	Planned: 42% Semi-planned: 32% Spontaneous: 26%	Planned: 38% Semi-planned: 34% Spontaneous: 29%	Planned: 25% Semi-planned: 29% Spontaneous: 50%
Physical Environment	Location	Indoor: 47% Outdoor: 53%	Indoor: 64% Outdoor: 36%	Indoor: 33% Outdoor: 67%	Indoor: 32% Outdoor: 89%	Indoor: 95% Outdoor: 13%	Indoor: 61% Outdoor: 43%
	Ambient noise	5.4/10	5.4/10	3.9/10	5.2/10	3.9/10	5.4/10
	Manipulated objects	71%	85%	60%	45%	72%	45%
Mobility	Movement	None: 10% Location-based: 32% Room-wide: 21% Floor-wide: 11% Building-wide: 11% Area-wide: 21%	None: 18% Location-based: 18% Room-wide: 45% Floor-wide: 18% Building-wide: 0% Area-wide: 0%	None: 0% Location-based: 0% Room-wide: 7% Floor-wide: 20% Building-wide: 0% Area-wide: 73%	None: 5% Location-based: 11% Room-wide: 0% Floor-wide: 5% Building-wide: 10% Area-wide: 68%	None: 29% Location-based: 32% Room-wide: 21% Floor-wide: 16% Building-wide: 0% Area-wide: 2%	None: 25% Location-based: 36% Room-wide: 11% Floor-wide: 14% Building-wide: 0% Area-wide: 19%
Vision	Visual distance	Arms-length: 52% Across room: 30% Far distant: 18%	Arms-length: 36% Across room: 53% Far distant: 11%	Arms-length: 38% Across room: 25% Far distant: 37%	Arms-length: 37% Across room: 37% Far distant: 23%	Arms-length: 62% Across room: 32% Far distant: 6%	Arms-length: 59% Across room: 34% Far distant: 7%
	Focus of attention	People: 56% Object: 26% Screen: 3% Environment: 16%	People: 25% Object: 20% Screen: 41% Environment: 14%	People: 15% Object: 26% Screen: 2% Environment: 57%	People: 24% Object: 19% Screen: 3% Environment: 54%	People: 57% Object: 13% Screen: 13% Environment: 16%	People: 56% Object: 13% Screen: 7% Environment: 25%
Social Environment	Privacy	Public: 63% Private: 37%	Public: 18% Private: 82%	Public: 73% Private: 27%	Public: 89% Private: 21%	Public: 43% Private: 59%	Public: 36% Private: 64%
	Group size	Alone: 0% Two people: 42% Small (3-5): 32% Medium (6-10): 5% Large (10+): 26%	Alone: 0% Two people: 73% Small (3-5): 27% Medium (6-10): 0% Large (10+): 0%	Alone: 0% Two people: 67% Small (3-5): 20% Medium (6-10): 13% Large (10+): 0%	Alone: 0% Two people: 53% Small (3-5): 32% Medium (6-10): 11% Large (10+): 5%	Alone: 0% Two people: 48% Small (3-5): 41% Medium (6-10): 3% Large (10+): 7%	Alone: 7% Two people: 39% Small (3-5): 42% Medium (6-10): 36% Large (10+): 11%
	Relationship	Close: 80% Acquaintance: 0% Coworker: 5% Stranger: 5% N/A: 0%	Close: 100% Acquaintance: 0% Coworker: 0% Stranger: 0% N/A: 0%	Close: 93% Acquaintance: 7% Coworker: 0% Stranger: 0% N/A: 0%	Close: 100% Acquaintance: 0% Coworker: 0% Stranger: 0% N/A: 0%	Close: 94% Acquaintance: 2% Coworker: 4% Stranger: 0% N/A: 0%	Close: 78% Acquaintance: 4% Coworker: 7% Stranger: 11% N/A: 0%
Communication	Speaking/listening	Speaking: 43% Listening: 40% No talking: 17%	Speaking: 50% Listening: 43% No talking: 7%	Speaking: 53% Listening: 39% No talking: %8	Speaking: 42% Listening: 42% No talking: 16%	Speaking: 51% Listening: 44% No talking: 15%	Speaking: 52% Listening: 43% No talking: 5%
	Gestures used	51%	76%	8%	57%	27%	31%
	Physical contact	Impersonal: 74% Personal: 47%	Impersonal: 73% Personal: 82%	Impersonal: 47% Personal: 20%	Impersonal: 63% Personal: 68%	Impersonal: 43% Personal: 59%	Impersonal: <1% Personal: <1%
Independence	Direction	They directed: 18% I directed: 32% Equal direction: 23% No direction: 22%	They directed: 10% I directed: 20% Equal direction: 35% No direction: 35%	They directed: 16% I directed: 11% Equal direction: 49% No direction: 25%	They directed: 15% I directed: 15% Equal direction: 35% No direction: 35%	They directed: 10% I directed: 11% Equal direction: 20% No direction: 58%	N/A
	Dependence	I depended: 5% They depended: 5% Equal: 84% None: 11%	I depended: 0% They depended: 9% Equal: 82% None: 9%	I depended: 7% They depended: 7% Equal: 60% None: 27%	I depended: 0% They depended: 0% Equal: 74% None: 26%	I depended: 0% They depended: 5% Equal: 71% None: 23%	I depended: 14% They depended: 4% Equal: 54% None: 32%

Table 2. Examples of responses from six scenarios along the seven dimensions in the framework.

locals responsibility for handling the probe.

- Communication: We did not prioritize any methods of communication beyond audio and video in the shared meal probe because our third survey did not show a strong need for referential gestures or physical contact.
- Independence: While the initial setup of the freestanding probe required aid from the local users, we increased the remote user's independence by providing them with the ability to control their camera and point of view.

Wearable Device: Visiting a New Place Together

We created a wearable device, shown in Figure 2b, as a design probe to match the needs for visiting a new place with a remote person. Similar in concept to other wearable telepresence systems [40,47], our system consisted of a Nokia Lumia 920 Windows Phone with a 4.5 inch screen, a wide-angle lens, an external speaker system or earbuds, and an external battery, mounted on a wearable cross-body bag. Videoconferencing was accomplished via Skype v4.9, installed as an application on the phone.

We next list the key decisions made in each design dimension, matching them to the results shown in Table 2.

- Initiation: Although there was no setup cost for the remote user in our wearable probe, the local user was required to put the carrier on, adjust the straps, and connect the earbuds (if desired); this was due to the low priority placed on minimizing setup time.
- Physical Environment: The wearable system used a SuperTooth Crystal Portable Speakerphone or earbuds with a microphone (the local user's choice) to address the expected variability of ambient noise levels.
- Mobility: We designed the wearable probe to leverage the local's ability to navigate and maneuver through a large area with multiple obstacles in the physical environment. We minimized the weight of the probe, limiting it to 1kg, and designed it to closely conform to the local's body.
- Vision: We affixed a wide angle Camkix lens to the wearable probe's built-in front camera to increase the remote user's range of view, accommodating mid-to-distant visual ranges.
- Social Environment: We leveraged the expected close relationships between locals and remotes to provide greater mobility for the probe, having the local wear the device.
- Communication: To allow users to not only share audio and video information, but also gestures made by the local user, we used a modified iOttie car mount holder with a quick-release mechanism, allowing locals to swap between wearing and holding the phone as needed. The positioning of the mount on the chest helped remote users to see and to follow deictic (pointing) gestures made by the local user, from the local's perspective.
- Independence: Remote users were not provided with a way to control the camera on the wearable device so that we could minimize the weight of the system.

Figure 2. Telepresence design probes built for our field study: (a) the freestanding device on the left, and (b) the wearable device, shown on a mannequin, on the right. (Probes are not accurately scaled relative to each other.)

Applying Our Framework: Scoring the Probes

We used our framework to evaluate the expected performance of the probes in our two targeted scenarios. We used extensive pilot testing to assign performance weights to each probe for the elements in each design dimension. Positive weights indicated higher performance and negative weights indicated lower performance. For example, because the freestanding probe is stable and wheeled but unwieldy to move, we gave it a weight of 0.5 for mobility "Within a room," but -0.5 for "Between rooms on a floor." Weights were multiplied by scores in each dimension from the third survey and summed to yield a final score for each probe in the two scenarios. Table 3 shows examples of these weights along two of our dimensions, vision and physical environment, as well as final scores. Our framework predicted that the freestanding probe would be a better fit for the "Shared Meal" scenario, while the wearable probe would be a better fit for the "Visiting a New Place" scenario.

Field Study

We conducted a two (museum vs. restaurant between-subjects) by two (freestanding vs. wearable within-subjects) mixed-design field study to evaluate our framework's predictions and explore design issues. We chose the field study setting to evoke realistic responses to our design probes in natural settings among users with pre-existing relationships, as our surveys indicated that these were highly likely use cases for telepresence systems.

	% reported Meal	Place	Question	Weights Free.	Wear.
Vision	62	39	Close (within arm's length)	1	0.5
	33	39	Middle (up to across a room)	1	1
	6	23	Far distance (> room length)	-0.5	-1
Physical Environment	95	32	Indoors	1	1
	13	90	Outdoors	-1	1
	2	68	Area beyond a building	-1	1
	0	11	Between floors in a building	-1	1
	16	5	Between rooms on a floor	-0.5	1
	21	0	Within a room	0.5	0.5
	32	11	Localized around an object	1	-0.5
	29	5	No movement	1	-1
			...		

	Device	Shared meal	Visit a place
Scores	Freestanding	473	142
	Wearable	299	407

Table 3. Example of percentage responses along two design dimensions for target scenarios. Assigned weights for each probe are listed on the right, total scores at the bottom are the sum of the weighted percentage of performance along all elements in the design dimension.

Shared Meal Setting: Restaurant

The restaurant used as the setting for sharing a meal had two floors, with high ceilings and an open floor plan. Participants were seated at a booth near the windows on the first floor and the experimenter sat at a table within view of the participant. The booth was near a bar area and during some meals, sunlight reduced the visibility of the device screens. Ambient noise levels ranged between relatively quiet to extremely loud during busy mealtimes. The post-study interview took place at the participant's booth, after completing the meal. An example of a participant in this environment is shown in Figure 3.

Visiting a Place Together Setting: Museum

The museum used as a setting for visiting a new place together was located in a downtown area and had three floors connected via escalators containing 16 galleries. Tickets and admission took place on the first floor, with exhibits on the second and third floors. The floor plan was open without any enforced order to the artwork and participants were permitted to explore anything in the general collection. The post-study interview took place in a children's play room provided by the museum on the second floor, after participants indicated that they had completed their visit.

Procedure. Local participants were greeted by the experimenter once they were seated at the restaurant or had arrived in the lobby of the museum. After completing a consent form, the local users were trained on the first device by the experimenter. Ordering of the devices was counter-balanced. After the locals were comfortable with operating the device, the experimenter connected the remote user to the device and the participants began their restaurant or museum experience. Throughout the experience, the experimenter was on hand at a discreet distance to handle any technical difficulties that arose such as connection issues and dropped calls. After approximately 30 minutes, the experimenter approached the participants and asked the remote to log out of the first device. The experimenter instructed the locals on the second device, and once they indicated their readiness, reconnected the remote. The experimenter then retreated from view as they continued their shared experience. Approximately 30 minutes later, the experimenter returned and asked the participants which device they would like to complete their experience on, switching the devices, if necessary. During the session the experimenter took photographs and field notes on the interaction.

Survey and Interviews. Following the shared experience, we administered a survey to measure individual impressions of using each system. Participants rated their agreement with 15 statements, designed to measure difficulties along each design dimension (e.g., "Everyone in my party was comfortable with the remote person/me being there on the device") on a five-point Likert scale (1=Strongly Disagree, 5=Strongly Agree).

Upon completing the questionnaire, we engaged participants in a semi-structured, audio-recorded group interview with remote and local users, as well as any party members, of about 20 minutes. The interview was designed to elicit feedback on breakdowns that occurred and to provoke reflections on the social dynamics surrounding the interaction. We asked users to contrast their experiences with the different devices and for potential improvements. Additionally, participants were asked to consider other contexts that might suggest using one device over the other.

Participants. We recruited 7 groups for each condition, comprising 45 people (28 females, 17 males) to participate in the study. In each group, one person acted as the remote user and participated from a location of their choice, one person acted as the local user and was responsible for handling the device, and any other people present were part of the local or remote user's party. Participant groups ranged in size from 1 to 4 people on the local side, usually friends or family, and 1

Figure 3. Father and son sharing a meal with a remote family friend using the wearable device.

to 2 people on the remote side, generally friends or family members of the local user. Participant ages ranged between 17 to 52 years old, $M = 30.1$, $SD = 10.4$, and all participants had pre-existing social ties or shared a relationship with one of the users (e.g., the local user invited the people he was working with to share the meal with his remote mother, whom he introduced during the interaction).

Data Analysis. For our field notes and interviews, we transcribed our audio recordings and organized responses into groups as trends emerged. We then extracted insights from these trends and used them to expand our framework.

Although our survey was not the focus of our study, we followed a process of testing for covariates, doing a confirmatory factor analysis on our questionnaire items, and analyzing our results using a mixed-measures analysis of variance (ANOVA). Due to the group sample size of our data set, we treated each survey as independent.

In testing for covariates, we found that there were no significant effects of device order and participant role. For our questionnaire items, our confirmatory factor analysis found a good fit for our questions regarding the independence of the remote user, two items ($\alpha = .82$), and for our questions regarding communication, two items ($\alpha = .87$). The alpha values for the remaining questionnaire items were below .70, indicating that they did not fit together to measure the other five design dimensions, so we analyzed those questions individually to understand subjective trends.

Results. We organize the results of our field study into two parts: (1) the emergent themes extracted from our field observations and interviews; and (2) the quantitative results from our post-experience survey.

Emergent Themes

Our field study yielded six key insights beyond those found in our surveys or in previous research for elements within the design dimensions of vision, social environment, communication, and independence.

In highlighting these insights, we refer to participants by condition and role—D for dinner, M for museum, L for local, R for remote, and P for members of the local's party—followed by a number indicating the group. For example, DR1 would stand for the remote user in the shared dinner experience from the first group of participants.

Insight 1: Differences in feedback requirements for shared perspectives in telepresence (Vision)

> *"I tried to direct him how to move so I could see* [with the wearable device]*, but that became very confusing, because I would say, 'To my left,' and he would say, 'Well, which left?'"* (DR2)

In face-to-face communications, attentional focus is signaled by several factors, such as the direction of the eyes, the head, or the positioning of the body [14]. However, when using both of our devices, participants frequently voiced uncertainty about what other users were looking at.

When using the wearable device, these breakdowns manifested when the local user was unable to see the screen or had no feedback on what their camera was showing. With the freestanding device, failures occurred when the remote's ability to control the camera made it difficult for the local to track where the remote was looking: "[In response to what they liked about the wearable device] *Basically that we both could see the same thing, so look at a painting and be like, 'Oh, look at this. This is cool,' and it's like, yeah, I'm looking at the exact same thing – with the rotating one* [the freestanding device]*, I can be looking off in a random different direction."* (MR4).

Prior work on spatial perspective-taking in face-to-face conversation has shown that speakers use more *egocentric perspectives*—referring to the speaker's point of view—when speaking to a partner [56]. However, in interactions mediated by a telepresence technology the relationship between remote and local perspectives is not always clear. Users may experience incorrect mental models of what is visible to the other user and the remote user's perspective may change suddenly (e.g., as when being flipped around or moved in the local user's hand), making it difficult to maintain an accurate perception of the addressee's relative position. In addition to the way that users share spatial perspectives, previous literature has highlighted the need to support gaze cues to aid in a shared understanding of attentional focus [11,40].

While this literature may serve to inform the design of future telepresence systems, more work must be done to understand how telepresence features, particularly those that grant abilities beyond those found in face-to-face interactions (e.g., the ability to see 360° or to directly see the speaking person's perspective), support interactions across scenarios and integrate with previous theoretical work.

Insight 2: Bystander responses to the remote's varying levels of embodiment (Social Environment)

> *"…I think some people looked at me kind of funny and thought I was awkward."* (MR3)

In face-to-face interactions, participant presence is self-apparent and their appearance is similar to others in the environment. When using our telepresence probes, however, remote and local users were extremely sensitive to how differences in the remote user's appearance affected bystander behavior.

Some users enjoyed the attention drawn by the higher visibility of the freestanding system's embodiment: "*So many people stopped. They were like, 'Who is that? Who are you talking to?' I feel like it was more interactive with other people. I had more fun with* [it]." (ML2)

However, others revealed discomfort with the level of disruption that they were causing with the freestanding device that they did not feel with the wearable device in the same environment: *"I think more just not wanting to draw attention and stuff. People are there to look at art."* (MR7).

These examples illustrate how the appearance of the remote user might shape not just the local user's or local party member's perceptions, but also the impact that it might have on bystander perceptions and how this may be shaped by the expectations within a given context.

Recent work has examined emergent social norms and themes in cellphone usage and how these relate to concepts in human communication literature, such as by disrupting existing physically collocated interactions [1,23,25]. However, this prior body of work presumes a consistent physical appearance of the other person's (e.g., as a human or a cellphone). With various telepresence systems available—between wearable and freestanding in our field study—our insight illustrates the need for telepresence research to follow the evolution of social norms and expectations as new form factors emerge [53].

Insight 3: User concerns about negative responses to technology use (Social Environment)

> *"…I kinda felt like the stigma of having the phone* [on the wearable device] *had more of a stigma than the other one* [the freestanding device]*"* (MR6)

People voiced differing concerns about how the two devices might be perceived by bystanders. Some users compared the wearable device to cellphone use because of the removable phone, as shown in the quote above. Others felt that the overt nature of the video on the freestanding device created concerns among bystanders of being recorded: "[With the freestanding device] *I have the ability to start recording anything…but I don't know if people across the table, if they would feel comfortable with me being here and just staring at them eating.*" (DR7). Others compared the visibility of the freestanding device to an accessibility aid, citing greater acceptance as it bore some similarities to pushing someone in a wheelchair.

This insight highlights the importance of understanding how the social connotations of previous technologies may shape perceptions of new systems, particularly when designing telepresence technologies to fit within a given context. For example, more embodied and visible systems may invoke concerns about being recorded, who has control, and access to the system [37], while more phone-like form factors may prompt excluding local party members [23]. In contrast, scenarios that are not socially motivated, such as remotely attending a class, might be less vulnerable to concerns about negative social stigmas or reactions from bystanders. As telepresence technologies explore new interfaces, there is an opportunity to develop a process to predict how user perceptions will be affected by previous media and how transfer of knowledge and experience may tie in to theories such as the Technology Acceptance Model [62].

Insight 4: The implicit debt of remote users (Social Environment)

> *"I felt like I was the device on the second one* [the wearable device]*, which did not feel good. I felt used. I was the tripod."* (ML2)

An unexpected emergent theme was the sense from users that by logging in, the remote user was implicitly incurring a social debt to the local user. Some locals felt responsible for making the experience enjoyable for the remote: *"Yeah, she didn't really ask for me to focus on other things. I just personally took it on my own..."* (DL3). However, this occasionally led to perceptions that the local was obligated to sacrifice his or her own experience in favor of the remote user's, leading to feelings of resentment and unhappiness, as shown in the first quote by ML2.

Conversely, others reacted by prioritizing their enjoyment over that of the remote user's: *"He did mention that it* [the freestanding device] *was, like, pointed at the ground, 'cause the best, most comfortable way I found to carry it was definitely at a pretty extreme angle…but I don't really care what he thinks, so I just kept it the way that was comfortable for me."* (ML4).

Both of these views accentuate that the implicit social debt incurred by the remote user is a unique factor for telepresence systems. When compared to prior literature on caller hegemony [22], this implicit social debt adds a new dimension to the power dynamic between local and remote users. While some prior literature has focused on users who are already familiar with each other [24,32,63], understanding how this debt might be reduced, the dynamics by which this debt is incurred, and how this debt may play out among users who have no prior relationship, are all areas that offer rich opportunities for future research.

Insight 5: Visibility of the technology and lack of awareness leading to self-regulation (Communication)

> *"I wasn't sure who was around or if I was gonna offend someone or anything like that…"* (MR4)

Throughout the interactions in our field study, one of the more pervasive themes was a sensitivity exhibited by users about the visibility of the conversation to bystanders, sometimes leading to self-censorship. In person, people have the ability to gauge bystander awareness and to adapt their communication modalities accordingly (e.g., speaking more softly, turning away, or using gestures). However, in a majority of mediated communications, users are at the mercy of the technology or of local users.

While the ability to self-regulate volume for remote users has been emphasized in previous work [11,32,60,67] and the literature has examined the theme of the "lack of control over one's accessibility accelerating the erosion of the public-private distinction" [25], our observations were that volume, visibility, and self-consciousness about the use of technology all played a role in affecting user interactions. In fact, due to its potential for more visible embodiments and capabilities

Question	Freestanding	Wearable	F	*p*
Planning required	3.11 (1.19)	1.94 (1.00)	5.20	.029
Ease of maneuvering	3.22 (1.31)	4.28 (0.75)	10.35	.005
Remote's independence	3.66 (1.21)	2.90 (1.10)	10.44	.003
Ability to see each other	4.03 (1.36)	3.05 (1.02)	8.97	.005
Ability to see museum	3.05 (1.33)	3.89 (0.96)	5.08	.030
Meal bystander comfort	3.67 (0.86)	3.00 (0.97)	5.20	.028
Museum ease of use	4.28 (0.76)	3.23 (1.31)	9.13	.004
Meal ease of use	3.64 (1.05)	3.09 (1.19)	5.01	.036

Table 4. Field study survey results. * = p < .05, ** = p < .005, standard deviations are shown in parentheses next to means.

"beyond being there," we believe that telepresence technologies may offer new and unique ways of allowing users to control their communications, increasing the options that users have in allowing the regulation of their interactions.

Insight 6: Contrasting reactions of remote users to enforced feelings of helplessness (Independence)

> *"…it felt like I was in a wheelchair like I was being moved differently…because I have had an experience of being in a wheelchair recently, it didn't feel like I wanted to be in a wheelchair…"* (MR2)

During interviews, we observed contrasting reactions to how remote users felt about their independence through the probes. Some users felt helpless and compared the experience to being in a wheelchair, emphasizing the unpleasantness of having to rely on local users for assistance. Others liked how the system constrained their actions. For example: "[on the freestanding device] *I'm more curious about pressing buttons and scrolling around and intruding on other people's privacy. With the cell phone* [the wearable device]*, because I can't control anything, once he sets it up it's like, 'Okay, I'm just talking with him.'"* (DR7)

These contrasting views on the constraints that our probes enforced on the remote user illustrated how perceptions may differ across users and across scenarios. Although little work has been done to explore what might affect these perceptions thus far, these insights serve to enhance our understanding of telepresence-specific factors that may need to be considered when developing and researching these systems.

Survey Results

Results from our questionnaire largely supported our predictions of each probe's scenario performance, based on our framework. These results are summarized in Table 4.

DISCUSSION

We developed a framework for telepresence consisting of seven design dimensions and 17 categories of scenarios that was informed by prior research and a series of surveys. We used our framework to develop two contrasting design probes and tested them in a real-world field study. Our study not only demonstrated how our framework might be used to prioritize design choices amongst different contexts, but also identified insights into interactional nuances unique to telepresence. Incorporating these insights into our framework, we summarize our final design dimensions in Table 5.

In developing a framework for designing and understanding telepresence, our goal was to a create tool that would help designers and researchers to:

– Identify opportunities for future research.

– Reflect on past and future work to determine potentially generalizable findings.

– Share a structure and language for reporting research and facilitating communication within the telepresence community.

Next, we demonstrate how our framework may be used as a tool to accomplish each of these tasks.

Design Dimension	Dimension Characteristics
Initiation	Person initiating interaction Options for refusing access Amount of planning Setup costs
Physical Environment	Indoors, outdoors, or both Ambient noise levels Material workspace properties Manual manipulation required
Mobility	Size of movement area Barriers or obstacles Movement restrictions
Vision	Primary focus of attention Distance to focus of attention Ability to see non-verbal cues Feedback of other user views
Social Environment	Privacy of setting Stakeholder relationships Negotiations of social capital Bystander perceptions of users Pre-existing views of technology used Implicit debt incurred by remote users
Communication	Impersonal touch Personal touch Referential gestures Non-verbal cues Speaking & listening time Technology mediating content
Independence	Ability to act without aid or direction Control of the system Effects on user confidence

Table 5. Table of the seven design dimensions and corresponding characteristics.

Identify unexplored areas for future research

Situating prior research within our framework highlights that a number of studies have examined how the remote user's appearance affects local users [32,46,49,49,63]. However, our identification of potential stakeholders within the design dimension of the social environment and the insights revealed by our field study showcase that the remote user's appearance may also affect bystander perceptions. While some prior work identifies bystanders as stakeholders [11,24], little work has been done to conduct an in-depth study of the interplay between users, bystanders, and the system's appearance, in real-world settings.

Reflect on the generalizability of prior and future work

In the context of previous work, we might establish that the informal conversations that occur in the workplace [32,63] and the natural exchanges that take place in the home [1] share similarities in how the interaction is initiated. Both scenarios involve spontaneous and opportunistic interactions with little to no planning among people who already know each other. Additionally, both types of interactions commonly occur in social settings where users have comparable means of managing their availability.

Having used our framework to identify the similarities between these scenarios, we may expect that some findings about the initiation of interactions in business settings may be similar in home settings. For example, increases in spontaneous or opportunistic interactions in business settings due to remote user presence via the system may be similarly increased by remote user presence in home settings, despite the differences within other design dimensions.

By using our design dimensions in this way to identify the similarities and differences between scenarios, designers and researchers may distinguish between results which should generalize between contexts, and those which require further investigation or differing design solutions.

Facilitate communication within the community

As the community working under the auspices of telepresence has grown, telepresence research has become increasingly divided in the terminologies used [28] and even in making decisions about what should be categorized as telepresence [34]. One example of how our framework may be used to unify the terminology from prior literature lies in examining multiple sources which present design requirements for robotic telepresence systems.

Cohen et al. suggested 10 requirements for robotic telepresence systems [10]: (1) visitor (remote user) perceptions of presence, (2) local perceptions of visitors, (3) mobility of visitors, (4) physical interaction, (5) fidelity of human expression, (6) visitor environment, (7) virtual content, (8) session management, (9) security and privacy, and (10) acceptance.

Riek proposed a set of seven candidate requirements for designing telepresence robots [52]: (1) video, (2) camera, (3) control, (4) latency, (5) gaze and appearance, (6) audio, and (7) gesture.

And Desai et al. enumerated 20 essential features for telepresence robots [11]: (1) dynamic video profile, (2) stationary video profile, (3) video resolution, (4) graceful degradation, (5) audio quality, (6) volume control, (7) UI type, (8) sensor information, (9) feedback, (10) integrated map, (11) robot height, (12) robot speed, (13) multiple cameras, (14) wide field of view, (15) independent head/torso, (16) access point switching, (17) assisted navigation, (18) human-speed navigation behaviors, (19) adjustable autonomy, and (20) appropriate occupancy awareness.

Each of these three previous works has segmented what robotic telepresence systems should support in different ways, making it difficult to understand how they relate. By using our framework of design dimensions and scenarios as a tool, we can begin to make sense of how the findings of each study interact. For example, Riek's work examines design requirements in the context of a classroom, which might fit in our self-improvement scenario if done alone, or a shared viewing scenario if socially motivated. Desai et al.'s perspective is from an office setting, where experiences may be more focused on goal-based collaboration. Lastly, Cohen et al.'s work does not clearly define the setting that they are examining. These differences in scenarios may help to explain why each of these works seems to prioritize or highlight different features.

In delving into the reported design requirements, we might define the question that each requirement attempts to resolve and use this to identify the framework dimension that it fits under. For example, Cohen et al.'s "fidelity of human expression" encompasses the transference of body language, head movements, eye gaze, facial expressions, and arm and finger movements, which would fit under our dimension of *communication.* Items in Riek's categories which would similarly be included in the framework dimension of *communication* include the "gaze and appearance," which indicates that systems should preserve gaze and portray clear facial appearance and expression, and the category of "gesture," which specifies the need for a minimum of two degree-of-freedom mechanisms for deictic gestures. Lastly, in Desai et al.'s work, items in the domain of the *communication* dimension are included not under the headings of the video profiles, but under "independent head/torso" where they indicate the support of these features for the purpose of providing social cues.

Organizing these requirements under the design dimensions and scenarios of our framework allows those in the telepresence community to understand why past literature may differ in categorizing and prioritizing these guidelines. Furthermore, by considering this work through the lens of our framework, researchers and designers can more easily make comparisons and combine these insights into a cohesive whole.

By creating a framework for telepresence, our goal is to provide a tool by which researchers and designers can communicate across domains, terminologies, and disciplines, to share discoveries and solutions. To do this, we have identified seven design dimensions which uniquely categorize 17 scenarios as a starting point. By using these design dimensions to pinpoint similarities between contexts, terminologies, or concepts, researchers from different domains may begin to form a common ground with which they can share findings.

CONCLUSION

It has been over 30 years since Marvin Minsky first proposed the term telepresence to describe the sensation of “being there” via a facilitating or mediating technology. Since that time, the telepresence field has grown by leaps and bounds. However, this expansion has not been without its growing pains. Motivated by technological advances which continue to offer new approaches to developing telepresence systems and with the field as a whole reaching maturity, a tool is needed to aid in understanding how the previously disparate approaches fit together into a cohesive whole.

In this paper, we described our process for creating a framework for telepresence. First, we iterated through a series of surveys and used prior literature to create a framework of design dimensions and potential telepresence scenarios. Our framework was predicated on the theory that these design dimensions can be used to differentiate between the needs of each scenario, and that these needs and constraints vary across contexts.

To test this, we validated our framework by developing design probes for two contrasting scenarios and tested them in real world settings through a field study. Our field study provided a proof of concept, demonstrating the utility of our framework as a tool for design and research. Additionally, our observations and interviews revealed six key insights, providing a more nuanced view of elements that are unique to telepresence in our design dimensions.

Last, we discussed how our framework could be used to (1) identify unexplored areas for future research, (2) reflect on the generalizability of prior and future work, and (3) facilitate communication within the telepresence community.

We hope that developing this framework for understanding and designing telepresence will encourage a greater dissemination of findings, discoveries, and solutions. By being able to share and place past and future research, we hope that work in telepresence can continue to evolve and to thrive, until being “here” is no different from “being there.”

ACKNOWLEDGMENTS

We thank Wild Ginger in Bellevue and the Seattle Art Museum for their support. We also acknowledge our colleagues Asta Roseway, Patrick Therien, Chris O’Dowd, Tom Blank, Sasa Junuzovic, and Kori Inkpen, who helped with this project. We thank our anonymous participants for recruiting their remote partners and allowing us to observe their activity.

REFERENCES

1. Ames, M. G. (2013, February). Managing Mobile Multitasking: the Culture of iPhones on Stanford Campus. In *Proc. CSCW 2013*, ACM Press (2013), 1487-1498.
2. Beer, J. and Takayama, L. Mobile remote presence systems for older adults: acceptance, benefits, and concerns. In *Proc. HRI 2011*, ACM Press (2011), 19-26.
3. Benford, S., Greenhalgh, C., Reynard, G., Brown, C., and Koleva, B. *Understanding and constructing shared spaces with mixed-reality boundaries.* In *ACM TOCHI 1998*, *5*(3), 185-223.
4. Boyle, E.A., Anderson, A.H., and Newlands, A. The effects of visibility on dialogue and performance in a cooperative problem solving task. In *Language and speech 37*, 1 (1994), 1-20.
5. Brubaker, J., Venolia, G., and Tang, J. Focusing on shared experiences: moving beyond the camera in video communication. In *Proc. DIS 2012*, ACM Press (2012), 96-105.
6. Brush, A.J., Meyers, B.R., Scott, J., and Venolia, G. Exploring awareness needs and information display preferences between coworkers. In *Proc. CHI 2009*, ACM Press (2009), 2091-2094.
7. Carroll, J. M. *Making use: scenario-based design of human-computer interactions*. MIT press (2000).
8. Cesta, A., Cortellessa, G., Orlandini, A., and Tiberio, L. Evaluating Telepresence Robots in the Field. In *Agents and Artificial Intelligence*, Springer Berlin Heidelberg (2013), 433-448.
9. Clark, H. H., and Brennan, S. E. Grounding in communication. *Perspectives on socially shared cognition* (1991), *13*, 127-149.
10. Cohen, B., Lanir, J., Stone, R., and Gurevich, P. Requirements and design considerations for a fully immersive robotic telepresence system. In *Proc. HRI 2011 Workshop on Social Robotic Telepresence*, ACM Press (2011), 16-22.
11. Desai, M., Tsui, K. M., Yanco, H. A., and Uhlik, C. Essential features of telepresence robots. In *Proc. TePRA 2011*, IEEE Press (2011), 15-20.
12. Draper, J. V., Kaber, D. B., and Usher, J. M. Telepresence. In *Human Factors: The Journal of the Human Factors and Ergonomics Society 1998*, *40*(3), 354-375.est
13. Ellis, C.A., Gibbs, S.J., and Rein, G. Groupware: some issues and experiences. In *CACM, 34*, 1 (1991), 39-58.
14. Frischen, A., Bayliss, A.P., and Tipper, S.P. Gaze cueing of attention: visual attention, social cognition, and individual differences. In *Psychological bulletin 133*, 4 (2007), 694-724.

15. Gergle, D., Kraut, R. E., and Fussell, S. R. Language efficiency and visual technology minimizing collaborative effort with visual information. *Journal of Language and Social Psychology (2004)*, *23*(4), 491-517.
16. Goffman, E. Behavior in public spaces. *Notes on the Social Organization of Gatherings.* The Free Press, NY (1963).
17. Grevet, C., Tang, A., and Mynatt, E. Eating alone together: new forms of commensality. In *Proc. GROUP 2012*, ACM Press (2012), 103-106.
18. Grudin, J. Groupware and social dynamics: eight challenges for developers. In *CACM, 37*, 1 (1994), 92-105.
19. Gutwin, C. and Greenberg, S. A Descriptive Framework of Workspace Awareness for Real-Time Groupware. In *Proc. CSCW 2002*, Kluwer Academic Publishers (2002), 411-446.
20. Haans, A., and IJsselsteijn, W. A. Embodiment and Telepresence: Toward a Comprehensive Theoretical Framework. *Interacting with Computers 2012*, 24(4), 211-218.
21. Heath, C. and Luff, P. Disembodied conduct: communication through video in a multi-media office environment. In *Proc. CHI 1991*, ACM Press (1991), 99-103.
22. Hopper, R. *Telephone conversation*. Indiana University Press (1992), 724.
23. Humphreys, L. Cellphones in public: social interactions in a wireless era. *New media & society 2005*, *7*(6), 810-833.
24. Inkpen, K., Taylor, B., Junuzovic, S., Tang, J., and Venolia, G. Experiences2Go: sharing kids' activities outside the home with remote family members. In *Proc. CSCW 2013*, ACM Press (2013), 1329-1340.
25. Katz, J. E., and Aakhus, M. *Perpetual contact: Mobile communication, private talk, public performance*. Cambridge University Press (2002).
26. Kendon, A. *Conducting interaction: Patterns of behavior in focused encounters.* CUP Archive (1990), 7.
27. Kiesler, S., Siegel, J., and McGuire, T.W. Social psychological aspects of computer-mediated communication. In *American psychologist 39*, 10 (1984), 1123.
28. Kristoffersson, A., Coradeschi, S., and Loutfi, A. A review of mobile robotic telepresence. In *Advances in Human-Computer Interaction 2013*, 3.
29. Kristoffersson, A., Eklundh, K. S., and Loutfi, A. Measuring the quality of interaction in mobile robotic telepresence: a pilot's perspective. *International Journal of Social Robotics 2013*, *5*(1), 89-101.
30. Latour, B. Reassembling the social-an introduction to actor-network-theory. *Reassembling the Social-An Introduction to Actor-Network-Theory,* Oxford University Press (2005).
31. Latifi, R., R. S. Weinstein, J. M. Porter, M. Ziemba, D. Judkins, D. Ridings, R. Nassi, T. Valenzuela, M. Holcomb, and F. Leyva. Telemedicine and telepresence for trauma and emergency care management. In *Scandinavian journal of surgery: SJS: official organ for the Finnish Surgical Society and the Scandinavian Surgical Society*, 96, 4 (2006), 281-289.
32. Lee, M. and Takayama, L. Now I have a body: uses and social norms for mobile remote presence in the workplace. In *Proc. CHI 2011*, ACM Press (2011), 33-42.
33. Leigh, J., Rawlings, M., Girado, J., Dawe, G., Fang, R., Verlo, A., et al. AccessBot: an Enabling Technology for Telepresence. In *Proc. INET 2000*.
34. Lombard, M., and Jones, M. T. Identifying the (Tele) Presence Literature. In *PsychNology Journal 2007*, *5*(2), 197-206.
35. Macharet, D. G., and Florencio, D. A. (2012, October). A collaborative control system for telepresence robots. In *Proc. IROS 2012*, IEEE press (2012), 5105-5111.
36. MantaroBot TelePresence Robots.*MantaroBot*. Retrieved August 1, 2014, from http://www.mantarobot.com/products/teleme.
37. Massimi, M., Truong, K. N., Dearman, D., and Hayes, G. R. Understanding recording technologies in everyday life. *Pervasive Computing, IEEE* (2010), *9*(3), 64-71.
38. McGovern, D.E. Experiences in teleoperation of lan d vehicles. *No. SAND-87-1980C; CONF-8708156-1*, Sandia National Labs, Albuquerque, NM.
39. Minsky, M. Telepresence. *Omni 1980*, 45-51.
40. Misawa, K., Ishiguro, Y., and Rekimoto, J. Ma petite cherie: 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 2012*, ACM Press (2012), 17.
41. Neale, D., Carroll, J.M., and Rosson, M.B. Evaluating computer-supported cooperative work: models and frameworks. In *Proc. CSCW 2004*, ACM Press (2004), 112-121.
42. Nishio, S., Ishiguro, H. and Hagita, N. Geminoid: Teleoperated android of an existing person. In *Humanoid robots-new developments, I-Tech*, 14 (2007).
43. Nguyen, D. T. and Canny, J. Multiview: improving trust in group video conferencing through spatial faithfulness. In *Proc. CHI 2007*, ACM Press (2007), 1465-1474.
44. Nguyen, D. T. and Canny, J. More than face-to-face: empathy effects of video framing. In *Proc. CHI 2009*, ACM Press (2009), 423-432.
45. Olson, G. M. and Olson, J. S. Distance Matters. *Human-Computer Interaction 15*, 2 (2000), 139-178.

46. Paulos, E. and Canny, J. PRoP: personal roving presence. In *Proc. CHI 1998*, ACM Press (1998), 296-303.
47. Procyk, J., Neustaedter, C., Pang, C., Tang, A., and Judge, T. K. (2014). Exploring Video Streaming in Public Settings: Shared Geocaching Over Distance Using Mobile Video Chat. In *Proc. CHI 2014*
48. Rae, I., Takayama, L., and Mutlu, B. In-body experiences: embodiment, control, and trust in robot-mediated communication. In *Proc. CHI 2013*, ACM Press (2013), 1921-1930.
49. Rae, I., Takayama, L., and Mutlu, B. The influence of height in robot-mediated communication. In *Proc. HRI 2013*, ACM Press (2013), 1-8.
50. Rae, I., Takayama, L., and Mutlu, B. One of the gang: supporting in-group behavior for embodied-mediated communication. In *Proc. CHI 2012*, ACM Press (2012), 3091-3100.
51. Rae, I.**,** Mutlu, B., and Takayama, L. Bodies in Motion: Mobility, Presence, and Task Awareness in Telepresence. In *Proc. CHI 2014*, ACM Press (2014), 2153-2162.
52. Riek, L. D. Realizing Hinokio: candidate requirements for physical avatar aystems. In *Proceedings of the ACM/IEEE international conference on Human-robot interaction 2007*, ACM Press (2007), 303-308.
53. Šabanović, S., Bennett, C. C., and Lee, H. R.. Towards Culturally Robust Robots: A Critical Social Perspective on Robotics and Culture. In *Proc. HRI Workshop on Culture-Aware Robotics 2014*, ACM Press (2014).
54. Sakamoto, D., Kanda, T., Ono, T., Ishiguro, H., and Hagita, N. Android as a telecommunication medium with a human-like presence. In *Proc. HRI 2007*, IEEE Press (2007), 193-200.
55. Schwartz, P. *The art of the long view: paths to strategic insight for yourself and your company*. Random House LLC (1996).
56. Schober, M. F. (1993). Spatial perspective-taking in conversation. *Cognition* (1993), *47*(1), 1-24.
57. Sellen, A., Buxton, B., and Arnott, J. Using spatial cues to improve videoconferencing. In *Proc. CHI 1992*, ACM Press (1992), 651-652.
58. Tang, J.C., Xiao, R., Hoff, A., Venolia, G., Therien, P., and Roseway, A. HomeProxy: exploring a physical proxy for video communication in the home. In *Proc. CHI 2013*, ACM Press (2013), 1339-1342.
59. Tomasello, M. *Origins of human communication*. MIT press, Cambridge (2008).
60. Tsui, K., Desai, M., Yanco, H., and Uhlik, C. Telepresence robots roam the halls of my office building. In *HRI 2011 Workshop*, 58.
61. Tsui, K. M., Desai, M., and Yanco, H. A. Towards measuring the quality of interaction: communication through telepresence robots. In *Proceedings of the Workshop on Performance Metrics for Intelligent Systems*, ACM Press (2012), 101-108.
62. Venkatesh, V., and Davis, F. D. *A theoretical extension of the technology acceptance model: four longitudinal field studies.* Management science (2000), *46*(2), 186-204.
63. Venolia, G., Tang, J., Cervantes, R., Bly, S., Robertson, G., Lee, B., and Inkpen, K. Embodied social proxy: mediating interpersonal connection in hub-and-satellite teams. In *Proc. CHI 2010*, ACM Press (2010), 1049-1058.
64. Vespa, P. M., Miller, C., Hu, X., Nenov, V., Buxey, F., and Martin, N. A. Intensive care unit robotic telepresence facilitates rapid physician response to unstable patients and decreased cost in neurointensive care. In *Surgical neurology 2007*, *67*(4), 331-337.
65. Weeks, G.D. and Chapanais, A. Cooperative versus conflictive problem solving in three telecommunication modes. *Perceptual and motor skills, 42*, 3 (1976), 879-917.
66. Wei, J., Wang, X., Peiris, R.L., Choi, Y., Martinez, X.R., Tache, R., Koh, J.T.K., Halupka, V., and Cheok, A. CoDine: an interactive multi-sensory system for remote dining. In *Proc. UbiComp 2011,* ACM Press (2011), 21-30.
67. Weiss, P. L. T., Whiteley, C. P., Treviranus, J., and Fels, D. I. PEBBLES: a personal technology for meeting educational, social and emotional needs of hospitalised children. In *Personal and Ubiquitous Computing* 5, 3 (2001), 157-168.
68. Wesselink, M. Videoconferencing as an educational intervention for children with autism (Master’s thesis). Retrieved from *http://webteach.org*, ID: 0638210, 2012.
69. Yankelovich, N., Simpson, N., Kaplan, J., and Provino, J. Porta-person: Telepresence for the connected conference room. In *Proc. CHI’07 extended abstracts on Human factors in computing systems*, ACM Press (2007), 2789-2794.