Designing Complementary Communication Systems

Drew Harry
B.S., Electrical and Computer Engineering
Franklin W. Olin College of Engineering, 2006

S.M., Media Arts and Sciences,
Massachusetts Institute of Technology, 2008

Submitted to the Program in Media Arts and Sciences,
School of Architecture and Planning on August 10, 2012,
in partial fulfillment of the requirements for the degree of
Doctorate of Philosophy in Media Arts and Sciences
at the Massachusetts Institute of Technology

September 2012

© Massachusetts Institute of Technology, 2012
All Rights Reserved
Designing Complementary Communication Systems

Drew Harry

Submitted to the Program in Media Arts and Sciences, School of Architecture and Planning on August 10, 2012, in partial fulfillment of the requirements for the degree of Doctorate of Philosophy in Media Arts and Sciences at the Massachusetts Institute of Technology

Abstract

We have long assumed that being face to face is the best environment for social interaction. But is “being there” the best we can aspire to? One common approach to improving face-to-face contexts is to add new communication channels — a strategy often described as creating “backchannels.” In my work, I use a series of novel complementary communication systems to show how adding communication platforms to collaborative situations can be useful while also arguing for a new conceptual model of a main stage and a side stage (in the Goffman sense) that contrasts with the traditional model of backchannels. I describe a series of projects that embody this approach and explore its limits. My work covers virtual world meetings and presentations, an audience interaction tool for large groups (backchan.nl), a tablet-based system for small group discussions (Tin Can), and a platform for connecting huge distributed audiences (ROAR). In each of these projects I trace my three major research themes: understanding how conversational grounding operates in these environments, how non-verbal actions complement text-based interaction, and how people make decisions about how to manage their attention in environments with multiple simultaneous communication channels.

Thesis Supervisor  Chris Schmandt
Title  Principal Research Scientist
Designing Complementary Communication Systems

Drew Harry
B.S., Electrical and Computer Engineering
Franklin W. Olin College of Engineering, 2006

S.M., Media Arts and Sciences,
Massachusetts Institute of Technology, 2008

Submitted to the Program in Media Arts and Sciences,
School of Architecture and Planning on August 10, 2012,
in partial fulfillment of the requirements for the degree of
Doctorate of Philosophy in Media Arts and Sciences
at the Massachusetts Institute of Technology

September 2012

© Massachusetts Institute of Technology, 2012
All Rights Reserved

Doctoral Dissertation Committee

Advisor
Chris Schmandt
Principal Research Scientist
Media Lab
Massachusetts Institute of Technology

Thesis Reader
Judith S. Donath
Fellow, Berkman Center for Internet and Society
Harvard University

Thesis Reader
Wanda J. Orlikowski
Alfred P. Sloan Professor of Management
Sloan School of Management
Massachusetts Institute of Technology
Acknowledgements

These last six years have passed faster than I could have imagined. There have been periods where it was hard to see a clear path to the end, but with the help and support of my mentors, colleagues, friends, and family I made it through the rough patches. Before jumping into the work, I hope you’ll indulge a brief and incomplete accounting of the many people who have played important roles in my life at the Media Lab.

My first home at the lab was with Judith Donath’s Sociable Media Group. Judith gave me the space and permission to learn to think of myself as a designer and researcher and whose vision of what online spaces might become continues to motivate me today. Thanks, too, to SMG’s wonderful final class of students: Orkan Telhan, Dietmar Offenhuber, Alex Dragelescu, Yannick Assogba, and Aaron Zinman with whom I was honored to collaborate and travel with. I learned so much from each of you, and I’m grateful for the diversity of perspective and background we had in Sociable Media. Thanks especially to Judith for serving on my committee and helping to close this chapter of my work.

During my first semester at MIT I met Alice Daer who become a fast friend who has helped me find a voice as a researcher and as a person. Her perspective and insights into writing, rhetoric, and audience have played a big role in my thinking on mediated interactions of all sorts. She also introduced me to my friend and colleague Joshua Green, who helped initiate the backchan.nl project which became the foundation of much of this dissertation.

When it looked like I would leave the lab, Nicole Yankelovich guided me back to my advisor Chris Schmandt who graciously stepped in and offered me a new home in his Speech + Mobility group. His thoughtful guidance, soft touch, and trust gave me the confidence I needed to find my way. He’s built a wonderful group, and I’ve been grateful for the advice, support, and friendship of Jaewoo Chung, Wu-Hsi Li, Charlie DeTar, Andrea Colaço, Matt Donahoe, and Misha Sra. I’m also glad to have known S+M alumni Matt Adcock and Paulina Modliba before I joined the group.

Thanks also to Joe Paradiso for his help smoothing my transition to the Ph.D. program, providing a second home with his Responsive Environments group, and for giving me the opportunity to co-teach a virtual world design class. I enjoyed the opportunity to collaborate with his students Mat Laibowitz, Nan-Wei Gong, Alex Reben, Mark Feldmeier, Matt Aldrich, Mike Lapinski, and Bo Morgan.

Over my years at the lab, I’ve had many great collaborators and friends from around MIT. Thanks especially to the 4chan team for creating a tremendous collaborative environment: Michael Bernstein, Andrés Monroy-Hernández, Paul André, Katrina Panovich, and Greg Vargas. Thanks also to Matt Boch, Vanessa Harden, and Retro Poblano. I tremendously enjoyed my time working with the India Initiatives team, managed by Mihir Sarkar and Andrea Colaço: Amit Zoran, Nan Zhao, Ksthitij Marwah, Santiago Alfar, Ahmed Kirmani, Rohit Pandharkar, Micah Eckhardt, Dennis Jiang, Ken Endo, and many others. Thanks to Owen Macindoe for a wonderful (and surprisingly fruitful) Battlecode experience.
Thanks to Amon Millner for introducing me to the lab so many years ago, and thanks to all my other colleagues throughout the institute over the years: Seth Hunter, Jaime Zigelbaum, Dawei Shen, Aithne Pao, Jinha Lee, Manas Mittal, Brent Fitzgerald, Kate Hollenbach, Ana Luisa Santos, Pranav Mistry, and Philip Long. Thanks as well to the staff around the lab: Kristin Hall, Will Glesnes, Peter Planz, Stacie Slotnick, Felice Gardner, Nicole Freedman, Jon Ferguson, and Kevin Davis.

Like any student, I’ve grown a tremendous amount through my interactions with faculty and mentors. In particular, Hiroshi Ishii, Chris Csikszentmihalyi, and Wanda Orlikowski showed me important new perspectives on the world that have had a major positive impact on my work. Thanks especially to Wanda for serving on my dissertation committee, and for introducing me to a world of scholarship that helped develop my sense of my own research practice from a beyond-the-Media-Lab perspective. I’m grateful to Mark Ackerman for serving on my general exam committee and giving me the confidence to talk about research methods and think about the way my work is situated in the broader field of computer supported cooperative work. And of course to my collaborator Eric Gordon for encouraging my early work with backchan.nl and working with me to develop Tin Can in a classroom context. Tin Can would never have become the project it is now without his guidance, writing, and analysis. Thanks to Martin Wattenberg and Jonathan Feinberg for their mentorship and collaboration at IBM Research that played a major role in leading me to the Media Lab. Thanks to Frank Bentley and Paul Moody for their continued support long after I left their respective organizations.

Through the years I’ve had the pleasure of working with a number of remarkable undergraduates. The projects I present here would not have been possible without the work and insights of Cherrie Yang, Trevor Filter, Mark Fayngersh, and Matt Peairs among many others.

I owe the periods in my Ph.D. to my friend, housemate, and collaborator Jon Chambers. He has helped me through this process in more ways than I could ever enumerate or reasonably expect, from sounding board to technical support to proof reading to timely distractions. I couldn’t have done this without his gracious support. Thanks also to my other friends during these years: Mikell Taylor, Rob Quimby, Katherine Kim, Janet Tsai, Lee Edwards, Christie Lee, Avery Anderson, Kimble McCraw, Andrew Hollett, Will Clayton, Chris Murphy, Susan Fredholm, Meena Vembusubramanian, Mike Siripong, Dan Foran, Brian Shih, and many others.

Through the end of this process, Lian Chang has been my companion and partner. Her patience, support, stability, and understanding in this period of uncertainty and stress has been tremendously important to me, and I will be forever thankful for her presence.

I am grateful to my family for their love and support. These haven’t been the easiest six years for us, but we’ve managed okay. I couldn’t have done this without Jason and Hannah Harry. My thoughts always remain with my mother, Karla Harry, who saw this journey start but unfortunately didn’t get to see it end.
Contents

1 Introduction 11
2 Background 23
3 Virtual Spaces 45
4 backchan.nl 87
5 Tin Can 111
6 ROAR 137
7 Conclusion 157

Bibliography 163
1 Introduction

As researchers first started to build technology to help us communicate with other people at a distance, they were faced with a question: what sorts of communication experiences are we aspiring to create? A consensus quickly arose. Computer mediated communication systems should focus on recreating the experience of being face to face with another person. The best system, in this model, is one that seems to disappear in the same way that the best window makes us feel like there is nothing between us and what is on the other side of the glass. Since the early 1970s, it has seemed like we have always been on the verge of a utopian environment where distance disappears and we interact as richly with friends, family, and colleagues around the world as we do with someone sitting in the same room. [Edigo, 1988] And yet, like the paperless office [Sellen and Harper, 2001], this future has failed to materialize. We can interpret this in two ways: either our tools have failed to deliver on the promise of a “being there” level experience or our persistent selection of non-“being there” experiences reveals a broad desire for a different vision of computer mediated communication. As with Hollan and Stornetta [1992], I will argue the latter case. In particular, I will focus on how computer-mediated-communication tools can complement existing traditional interaction contexts like face-to-face communication, video conferencing, or audio conferencing. I will show through a series of design projects and studies where we might find ways to create experiences that meet the challenge of being “beyond being there”; experiences that are compelling precisely because they are trying to be something other than just a face-to-face presence with others.

To introduce this design space, I describe the broader context of computer-mediated-communication systems and suggest an explanation for why work has long focused on the “being there” approach to design. I will discuss approaches described in the literature for thinking about the different design strategies one might employ when trying to build new systems that aim that create communication platforms that complement “being there” type experiences.
1.1 Being There

Core to the argument that computer-mediated-communication should simply recreate face-to-face interactions as transparently as possible is the notion that face-to-face interaction is the best possible context for communication. Depending on your perspective, it may seem either radical or obvious to claim that we might prefer non-face-to-face interaction in certain situations. To understand the differences between mediated and un-mediated interaction, and gain insight into how people make choices about when to use each, we can look to research on media selection preferences.

Since computer-mediated-communication became technically viable, there has been broad interest in understanding the relative properties of speech, video, and text, as well as more esoteric (and largely ignored) modalities like real-time handwriting transmission. [Williams, 1977] This stream of work can be traced back to Ochsman and Chapanis [1974], who studied pairs of students coordinating on concrete tasks like scheduling, way-finding, and physical part identification using different sorts of communication tools. This early work focused on measuring which channels were most effective for tasks, and primarily recommended that adding voice provided the biggest improvement in performance. This type of work has continued, with researchers expanding their view beyond task performance to trust formation [Bos et al., 2002][Toma, 2010] and deception [Hancock et al., 2004].

Much of this work takes place in an experimental psychology tradition, which focuses on controlled lab contexts for studying communication behaviors. As a research context, this leaves much to be desired in studying the complex in situ decisions that people make about communication in their every-day lives. Outside of lab contexts, we are not assigned a specific tool for a specific task. Instead, we make nuanced and highly contextual choices about what sorts of tools to use in different communication situations.

If we accepted the idea that “being there” was a primary desire for people selecting a particular channel to use, we would expect to see people prioritizing tools that were the most like “being there” among the available options. This spectrum between mediated and un-mediated is sometimes characterized as “media richness” per Daft and Lengel [1986], in which the most rich media are those that are most like being face to face and that people will prefer richer tools over less rich tools. Instead, researchers have found repeatedly that in non-lab situations, people frequently choose less rich media over more rich options. [Scholl et al., 2006] If richness alone does not predict people’s real-life communication media selection decisions, it
suggests that other features of a communication situation might also be relevant and that there exist other priorities than just a desire for “being there.”

Communication situations are defined by the complex interplay of the features of the communication tools in use and the setting, people involved, purpose, and norms. The non-technical components of the setting will be discussed in more detail in Chapter 2, but our discussion here will focus on the nature of the medium itself. We can look to Clark and Brennan [1991] as a starting point for other ways we might organize and understand the properties of different sorts of communication media. These aspects of a medium play a large part in how it’s used, and help explain the sorts of results seen in studies like Scholl et al. [2006].

**Reviewability** Are records of past interactions easily accessible? Who has access to them?

**Revisability** How are messages constructed? Do you have time to revise a message before it is sent? Can messages be edited or retracted?

**Synchronicity** Are messages responded to rapidly, or are there longer gaps between messages?

**Sequentiality** Does the system support multiple simultaneous conversations, or must all contributions fit into a single shared stream?

**Identity** How are people represented, and what information is made available about each person?

**Mobility** In what sorts of spaces can this medium be used? What do we expect about the contexts of other people using the system?

We can evaluate both mediated and non-mediated experiences on each of these axes. Face-to-face communication, for example, has no reviewability or revisability but has high synchronicity, high sequentiality, and high levels of identity disclosure. In the spirit of acting in a complementary way, most of the work described in this thesis attempts to provide affordances that are distinct from those offered by face-to-face communication or whatever sort of mediated experience we wish to complement. Furthermore, I will argue that different experiences can effectively co-exist, each complementing the strengths and weaknesses of other mediated and non-mediated channels.

The continued focus on “being there” designs by much of the communication industry, Cisco’s Telepresence systems (see figure 1.2) and Apple’s Facetime (see figure 1.3), and Google Hangouts being
major examples, has blinded us to the potential of other approaches. They reinforce the framing of mediated communication as something less rich and less effective than face to face. Part of this is about scale; one on one or small group interactions are hard to improve. But a closer inspection of face-to-face interaction among groups of more than five people reveals a number of potential challenges:

- Not all people are equally capable of convincing performances in face-to-face interaction. This can be the result of a variety of factors, including (but not limited to) a lack of confidence in contributing in a specific context, a lack of skill with language, or the impact of a power imbalance in the situation. Many of these can be mitigated in mediated contexts [Siegel et al., 1986], although mediated contexts have their own distinct performance challenges.

- Simultaneous contribution in face-to-face situations are often viewed as impolite and are generally normatively discouraged. Particularly in large groups, this represents a pressure against contribution and requires a certain amount of overhead to negotiate turn-taking gracefully.

- Participation in face-to-face contexts usually discloses significant information about someone’s identity, while in mediated contexts there are a variety of approaches to limiting disclosure of identity information while still being an active participant.

- face-to-face interactions are traditionally ephemeral and difficult to record; mediated interactions are usually quite easy to record, even those that mimic face-to-face interaction.

Many of the projects in this thesis intend to complement a face-to-face experience, and so these issues are of particular interest. Text-based communication of all sorts tends to complement the features of face-to-face communication nicely, and so will form the foundation of all of the projects in this thesis. Text offers a number of benefits: it tends to have a disinhibiting effect, better supports simultaneous contributions, affords opportunities for managing identity disclosure, and is easily recorded and analyzed. These features directly address the challenges of face-to-face communication, so by pairing a communication platform with these properties with face-to-face communication (or a mediated experience that mimics face-to-face communication) we can give participants an experience that nicely complements their existing options, and which they can choose to make use of as appropriate.
1.2 Complementary Communication

This thesis addresses the design space of complementary communication systems. By this, I mean systems that aim to create a communication context shared by a group of people who are sharing some kind of experience like a presentation, performance, or discussion. Viewed this way, complementary communication systems are as old as whispering to someone sitting next to you or passing notes to a classmate. The addition of powerful personal communication technology to our everyday interactions have increased the opportunities we have to create complementary communication contexts as well as radically increased the reach a complementary communication system might have. In this context, how should we think about these sorts of systems? What goals should we have for them? In what contexts do they make sense? In particular, how should we think about the relationship between a complementary communication system and the communication experience it aims to complement? Is complementing a face-to-face interaction the same as complementing a mediated interaction?

In their famous paper, Hollan and Stornetta [1992] introduce the “beyond being there” approach. They argue that seeking to recreate the experience of “being there” was in a way an abdication of our responsibility as designers that left an important design space un-explored. In particular, they urge us to think less about ways to minimize the experience of mediation in communication, but to look instead for ways that mediation can add value to interactions. To take this perspective seriously, we need to shift away from a view of face-to-face interaction as being always better than interactions mediated by technology and instead think critically about potential limitations and challenges with face-to-face interaction and potential benefits that mediation can offer.

Although Hollan and Stornetta focus on creating mediated experiences that rival or surpass face-to-face experiences, in my work I show that we don’t have to choose one approach or the other. If we accept the argument that being face to face is not a priori the best experience, the strategy I employ is to add complementary communication platforms that can be used simultaneously with face-to-face communication or mediated platforms that mimic face-to-face interaction.

This quick assumption that designing to complement face-to-face interaction is not so different from designing to complement mediated communication may seem unlikely; why should we accept that systems used in coordination with audio or video conferencing would be similar to those used to coordinate face-to-face? I will
argue that a system that can effectively complement face-to-face interaction when its users could simply set it aside and rely on the (presumed superior) affordances of unfettered verbal communication likely has something to tell us about both design and face-to-face interaction more generally. If these systems can provide value in face-to-face contexts, I will show that they also provide value (perhaps even more value) when used to complement systems that seek to create experiences like being face-to-face. Furthermore, true “distributed” situations are becoming less common. Heterogenous configurations where some people are co-located and others are remote and possibly alone are becoming more common. In these contexts, a system that doesn’t operate effectively between co-located users is unlikely to be broadly useful, and would suffer from the disenfranchising effects we see for people who “dial in” to a local meeting. Thinking broadly about systems that complement both face-to-face and audio/video sharing will more efficiently lead us to systems effective in both contexts than treating them as separate cases.

This dissertation is organized around a series of “primary” contexts for which I design a particular complementary communication system that enhances the overall experience. Metrics and evaluation strategies vary for each of these pieces, but each project shares a deep interest in trying to fill in the gaps of the “primary” interaction space by using the particular strengths of some additional mediated communication system. The goal of these interventions is to create environments where people have ways to express themselves non-verbally in addition to whatever existing communication channels exist, often audio, sometimes visual. By adding mediated communication channels to other existing channels, we can focus each channel on its primary affordances and let it do what it does best while letting the complementary communication channels fill in the gaps.

1.3 From Channel To Stage

Platforms for discussion and commenting that are outside official discourse channels have widely been referred to as “backchannels.” Backchannels, traditionally defined, create a space where audience members to some “front channel” can share information with each other, typically about the content of the front channel.¹ This metaphor is an apt description of much of the prior work in this space, like [Cogdill et al., 2001, Yardi, 2006, McCarthy et al., 2004, Rekimoto et al., 1998]. I have found, however, that it is not as useful for understanding the systems I will present in this thesis. Instead of using channels, I will argue for “stages,” and instead of a front/back separation, I will shift to a main/side distinction.²

¹ This is in contrast with the traditional technical use of “backchannel,” which refers to the verbal and non-verbal cues that non-speakers give a speaker during a conversation. Although in some ways backchannels meet some of the same needs that “uh-huh” and “right” and nodding fill in face-to-face communication, it is nevertheless a somewhat unfortunate collision in naming. Traditionally, backchannels were not distinct dedicated channels, but simply a type of utterance. The metaphor is an apt description of much of the prior work in this space, like [Cogdill et al., 2001, Yardi, 2006, McCarthy et al., 2004, Rekimoto et al., 1998]. I have found, however, that it is not as useful for understanding the systems I will present in this thesis. Instead of using channels, I will argue for “stages,” and instead of a front/back separation, I will shift to a main/side distinction.²

Figure 1.6: Person to person text messaging extends the reach of something like note passing to include anyone with a phone. Phones also support multi-party conversations, increasing the size of the audience for complementary communication systems.
In a traditional backchannel configuration, audience members can view the front channel and have a variety of backchannels available to them, each with different sized audiences and affordances. This is represented in Figure 1.7. Presenters, on the other hand, often have a very hard time staying aware of backchannel content if they are aware of it at all. This asymmetry gives the backchannel its outsider flavor and can lead to disrespectful and unproductive content [boyd, 2009]. This is widely recognized as a major problem with backchannels. In this way, the front/back distinction is an accurate description of existing systems, but not a situation I seek to recreate.

Mitigating this sense of separation is a major design goal in my work. To enable this shift, we need to re-imagine the nature of social interaction in these sorts of spaces. A channel metaphor implies a clear split between those empowered to broadcast, and an audience who receives that broadcast. It also has implications for how attention is managed; channels imply a model of binary attention. To find an alternative, I turn to Goffman [1959] for his description of stages to illuminate this new sort of situation. He uses the example of a waiter behaving politely with a problematic customer and then walking into the kitchen and complaining to the cook about the customer’s difficult behavior. Each interaction is performative and represents the waiters’ competence at performing his role appropriately for different audiences in a different setting. In the waiter example, these audiences are disjointed, and the door into the kitchen represents a gateway between the “front” performance space with customers and the “back” performance space among restaurant staff. The notion of stages shifts our attention from spaces where a small number of people can broadcast information to many recipients (like a lecture hall or conference) and instead focuses on negotiated sites of performance like the restaurant dining room and kitchen, where people can perform different aspects of their identity for different audiences. This intermediate state is shown in Figure 1.8. In this model, the back stage still lacks accessibility for much of the audience, and is not easily perceived by performers on the front stage.

This notion of stages is a useful metaphor to replace channels. Unlike channels, where audiences are basically invisible to the performer, stages bring the audience and performer together and create a context in which they are mutually aware of each other. Stages also shift from the notion of a small group of broadcasters and a large group of receivers, to a context where there is the potential for different performers at different moments. The notion of stages also more actively recognizes the way that audiences to a performance are themselves constantly performing in small ways, while a channel metaphor limits the audience to simply receiving a broadcast. This

---

3 Of course, there may be contexts where the front/back distinction is valuable. But since that is the predominant structure for existing tools, there are many more options for supporting that approach so I don’t give it much attention in this thesis.
comprises both small performances and the potential for substantial performances. In a large lecture hall, the nature of the individual performance is not that precise. If there is an open laptop policy, audience members might be checking their email or engaging in an official backchannel. They are able to, as Goffman says, “get away with going away,” because the act of going away is an expected part of being in the audience to a front stage performance. But a shift from channels to stages is important in recognizing that even as audience members, looking at a laptop screen instead of the teacher is a sort of small performance. Furthermore, a student may raise their hand and ask a question of the teacher, thus assuming a larger role in the main stage.

Although Goffman’s description of appropriate performances for a specific audience is valuable, we want to avoid the front-back distinction as enacted in his restaurant example. Instead of having separate audiences, we want to create a space where, although the modes of performance are different, the performances are available to everyone. While in a channel metaphor, audiences are split between multiple backchannels which are largely unavailable to someone performing on the front channel and to people attending to other backchannels, stages create the opportunity for situations where the audiences for each stage can be shared. This unified audience helps us move away from the front/back distinction to a main/side distinction. In this model, performances on the main or side stage both share one large audience. By unifying the audiences, we can help avoid the problem with backchannels where they tend to assume a covert character. If side stage participation is accessible to everyone, it changes the character of the communication. Side stage performances lose some of their covert nature because they can be seen by a larger audience that includes main stage performers. Simultaneously, main stage performers can be made actively aware of side stage performances in a way that is inclusive, to help them better react to their now more present audience.

To reiterate, this change in metaphor has two components: a shift from channels to stages, and a shift from front/back to main/side. This final configuration is shown in Figure 1.9. This is an aspirational shift; most of the design work in this space to date has represented the backchannel approach. But if these systems are to be an effective addition to face-to-face-type experiences, I will show how adopting this new metaphor can create effective communication experiences that encompass multiple simultaneous opportunities for engagement that create an effective, unified experience.

In my initial argument against the sufficiency of un-augmented face-to-face communication, I aimed to show how moving from a
single stage to a main and side stage can be valuable, provided the side stage is designed to complement the potential deficiencies of the main stage. It is not obvious that a single side stage is the only effective configuration. Why is one side stage the appropriate number? Why might we not prefer a multiplicity of side stage, with audience attention shifting fluidly between them? Intellectually, a model that supports a network of stages with various properties and interactions, as in actor network theory [Latour, 2007], seems attractive. Stopping at two stages seems to make the same mistake of enshrining a particular form of interaction as optimal, and leave the model open to the same criticism I make against main-stage-only experiences.

Although describing such a network-based theory is not the focus of this thesis, I would definitely view it as an effective tool for understanding interactions in existing backchannel-based ecosystems. Backchannels are rarely monolithic and in most situations a large number of backchannels are operating simultaneously. Most are out of sight for non-participants (like instant messaging and chat), while some strive for “official” status (like Twitter) and can start to feel more like a single monolithic backchannel. Thus while a model that supports a multitude of simultaneous complementary communication systems is an accurate picture of most experiences, it is not necessarily a model we want to be encouraging from a user experience perspective. As discussed in the transition from a front/back to main/side configuration, attempting to unify the audiences between the front and back stages represents one of our major tools to mitigate the covert effects of backchannels with distinct, separate audiences. Without blessing a particular stage as “the” side stage, we would run the risk of creating just another backchannel with the challenges to driving adoption and issues integrating with the main stage. The covert flavor that I identify in contexts with a variety of backchannel options would likely remain.

Realistically, we are always embedded in networks of communication contexts that include both people with whom we are physically co-located and remote people. These overlapping networks of text messaging, Facebook updates, and tweets are part of our professional and personal lives most of the time. Even in a context with a single designated side stage, it would be unreasonable to expect that there aren’t also a variety of backchannels operating simultaneously. Complementary communication systems must compete effectively with these backchannels to provide a meaningful space for interaction if they hope to succeed, and it would be naïve to ignore them from a design perspective. We could extend the main stage/side stage model to make room for multiple overlapping side stages, but we run a practical risk of overwhelming an audience or group, and turning
each of the side stages into distinct backchannels that fail to com-
mand a unified audience or influence the main stage in a productive
way. Because of these practical challenges, I set aside the problem
of modeling stages in a more elaborate network context and focus
instead on single side stages paired with a main stage, and seek to
understand that relationship from theoretical and practical perspec-
tives.

1.4 Design Spaces, Themes, and Theory

One of the challenges of building technical systems as research is
understanding the scope of conclusions. If you took a particular de-
sign element into a different system, would it operate in the same
way? What are the relationships between the sorts of people using
the system and the socio-technical structures that emerged? These
are difficult questions to answer within the scope of a single project.
The researcher may have solid intuition, but the tendency of the re-
searcher is probably to see overly-general results more often than
overly-specific results. One of the ways I address this is by describ-
ing a series of projects in this design space and examining design
elements and themes in a variety of contexts.

This would be less effective as an approach if each of the design
spaces was quite similar. The contexts I am designing for can be
organized around a series of major differentiating axes:

Main Stage  The medium for the main stage, e.g. the site of the pri-
mary shared experience of the audience.

Shared Display  The presence or absence of a shared display.

Side Stage Attention  The frequency of audience attention on the side
stage. This is quite qualitative and varies across users, but each
system embeds contains certain assumptions about the relative
importance of the side stage to the main stage.

Audience Size  The target audience size for the system.

Table 1.1 lists the research contexts covered in this dissertation,
and describes each project’s location on the major context axes. The
variety across these axes helps show the breadth and scope of my
work.

Although each project operates in a different context, there are a
number of research themes that each project addresses. Although
these themes were not identified at the start of this research stream,
they nonetheless are present in all my work to various degrees. It
can be challenging to identify the broader impact of design-based research (a topic I will address in more depth in the next chapter), but it is through looking at these themes in different contexts that I hope to contribute to broader discussions. These are themes that are relevant particularly to designing complementary interfaces like those in my own work, but also to many kinds of collaborative, synchronous systems, even those which aren’t trying to create complementary experiences.

*Grounding*  My work uses shared displays in a variety of different capacities. I contend that these kinds of public displays can play a powerful role in helping to ground, in the Clark [1989] sense, a conversation. In particular, shared displays can provide ways to non-verbally acknowledge discourse presentations. By their very shared nature, the contents of shared displays might accelerate the creation of common ground. The different ways that these shared displays operate in my work helps provide insight into both particular design techniques to support grounding as well as the broader discussion around how common ground operates in mediated communication contexts.

*Non-verbal actions*  As a result of the drive to create a sense of “being there”, mediated interaction systems failed to consider the ways that we communicate non-verbally, assuming that higher fidelity video and audio would be sufficient to capture that communication. I contend that our non-verbal actions in the physical world are a critical component of body language, and when creating mediated channels we should strive to create new vocabularies of action that enable people to communicate non-verbally. In much the same way that in a shared physical space we can observe people interact with objects around us, so too should people’s actions in mediated systems be visible and part of supporting a sense of presence and awareness. How these action vocabularies are constructed and communicated to people is critical to the success of these sorts of systems.

*Attention*  Creating new opportunities for simultaneous commentary and communication about some shared experiences creates situations where people have to make choices about which if the stages

<table>
<thead>
<tr>
<th>Project</th>
<th>Main Stage</th>
<th>Shared Display</th>
<th>Side Stage Attention</th>
<th>Audience Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>backchan.nl</td>
<td>face to face</td>
<td>yes</td>
<td>infrequent</td>
<td>20-500</td>
</tr>
<tr>
<td>Information Spaces</td>
<td>virtual world</td>
<td>yes</td>
<td>infrequent</td>
<td>10-20</td>
</tr>
<tr>
<td>ROAR</td>
<td>broadcast video</td>
<td>no</td>
<td>frequent</td>
<td>&gt; 1,000</td>
</tr>
<tr>
<td>Tin Can</td>
<td>face to face</td>
<td>no</td>
<td>occasional</td>
<td>10-20</td>
</tr>
</tbody>
</table>

Table 1.1: Comparing the projects covered in this dissertation on the major axes that distinguish them. Projects in bold are major projects discussed in the most depth. Small project variants (e.g. backchan.nl for remote Q&A, Tin Can Meetings, etc.) are not included.
to attend to and which stage to use for their performances. From a design perspective, there are a number of important attention related decisions to make: How is our attention made visible to others, and how does it affect their impressions of us? How do we design displays for occasional attention of users who are shifting their attention between different stages? Understanding how people think about and enact attention in situations with multiple available communication channels is critical to designing appropriate options and understanding the practices that evolve around them.

This work contributes on two levels. First, by creating and deploying interfaces with particular properties, I provide concrete guidance and insight about particular specific design strategies and interfaces. This is valuable for designers and researchers thinking about how they design this variety of communication systems. I also contribute to the broader discourse about the three research themes laid out above. In each case there are both broader theoretical contributions to be made as well as specific findings that contribute to scholarly discussions about these issues.

1.5 Thesis Organization

This thesis starts with a broad background discussion about some of the methodological assumptions implicit in doing design work as research, a tour through some of the high level related work that all of my work relies on, and a more in-depth treatment of the major research themes I introduced here.

Based on this background, I will then discuss four major project areas in each of the following four chapters: Virtual Worlds (which contains two specific sub-projects for meetings and presentations in virtual worlds), backchan.nl (a tool for managing audience feedback during live events), Tin Can (a tablet-based platform for enhancing small group discussions), and ROAR (a platform for very large scale audience interaction during live events). Through each of these chapters I will relate my findings back to the research themes that are laid out in Chapter 2.
2 Background

Although this dissertation discusses a wide variety of work, it all draws on a common foundation. This chapter will lay out that foundation before moving on to discuss specific projects. There are four components to this foundation: design as a research practice, methodology, related work, and major research themes. I will cover these in order.

2.1 Design as Research

Proposing a design space and arguing for its value as an approach to common problems is not, traditionally, the realm of academic research. It is a frequently taken-for-granted assumption at the Media Lab that designing novel technical systems is a natural and defensible way to do research, but outside of that context this approach is rarely accepted without argument. Given that this assumption is fundamental to my work, it seems useful to address this epistemological question from the start.

Design as research is clearly being conducted in a variety of contexts using a variety of methods, yet there is very little discussion or agreement about the fundamental aspects of how that work is conducted and what we can learn from it. Traditionally it is tolerated, provided that the primary argument for its contribution to the wider discourse is based on a justification from another research field. In computer science related fields, efficiency and technical complexity are reliable ways to argue for the quality of work, even work like mine where novel systems and interactions are described. This has a warping effect on the sorts of systems that are built and studied because there needs to either be a technical challenge in the solution or a way to easily judge the efficiency of the solution relative to existing methods.

Although there are ways to frame my work in those terms, I do not view those as particularly effective ways to judge my work. Instead, in this section I would like to make a positive argument for why building and testing systems may be the best approach for cer-
tain kinds of research questions. I will describe the sorts of contributions one can make working this way and contrast this approach with other approaches dominant in the fields of human computer interaction, computer mediated communication, and computer supported cooperative work.

2.1.1 Engineering as Research

It is a gross simplification, but let us separate work in the field of human computer interaction or computer supported cooperative work into three general categories: technology-enabled sociology and psychology, studies of systems in-the-wild, and design of new systems. In the first category, researchers seek to answer the kinds of questions typically of interest to sociologists and psychologists, but deploy technology to allow them to answer questions they have not been able to answer in sufficient detail in the past. This work focuses primarily (but not exclusively) on drawing conclusions about the behavior and experiences of either individuals or collections of people. Studies of systems in-the-wild are, in contrast, more focused on understanding the relationship between the technology and its use, often described as socio-technical systems. Finally, there are researchers who design and build novel systems and then study them. I describe this final category of researchers (of which I consider myself a member) researchers-as-designers.

These last two categories are deceptively similar. After building a novel system, does one simply run a study on that system like a researcher who didn’t build the system themselves? This suggests a critical thought experiment: if the researcher-as-designer could simply imagine a system into existence that looked exactly like the system they wished to study, would that compromise the research in any fundamental way? Put another way, does the actual design and implementation process actually add value to the research or is it just an overhead?

Although provocative, the thought experiment is subtly misframed. Technical artifacts can never really be imagined into existence because their creation is a constant negotiation between the properties of the tools used to create it, the environment in which the design happens, and the reactions the designer has to their own work. In practice, the artifact that comes out of a design process is the result of a lengthy iterative process, even in design processes conceived as iterative. Simply creating part of an artifact and integrating into another part causes a re-evaluation of those parts in a way that causes designs to drift from their original models. The time spent in the design and implementation processes can be seen as critical
for producing a viable design. If we desire to study systems that
don’t already exist, there is simply no way around spending time on
the design itself because our ideas about what the design could or
should be before entering those process cannot become real without
any intermediate steps, and if they could they would be unlikely
to meet any of the original design goals. From this perspective, we
view the development process as a fundamentally necessary cost of
creating any novel system.

An alternative approach is to consider the distinctive values
of the design process as a research process. In some fields, we expect
that the researcher will become deeply embedded in the process
on which their work focuses. In these fields, putting yourself at a
distance and insisting that you can simply observe without being
part of that process is often viewed as naive. Yet when we shift our
focus to creating novel technical artifacts, we prefer to isolate either
the users (as in lab studies) or the treat the artifact itself as stable
(as in studies of the use of existing artifacts in situ). It seems only
natural to say that claims about the design of socio-technical systems
can be easily augmented by a long-term, rich participation in that
exact process. Playing the role of the active participant in the process
grants us credibility and real analytical leverage. I don’t mean to
say that one can’t make arguments about design choices made in a
design process without being a participant in it—you certainly can—but being a participant in that process provides important insights
that we are unlikely to find if we treat the technology as the black
box output of an historical design process conducted by others.

It is difficult to precisely identify the sorts of contributions that
would not be possible without engaging in the design process be-
cause there are few examples of a team doing novel design work
and handing off the result to a separate researcher and comparing
their results. We do have a large body of research on systems con-
ducted by non-designer researchers, but there’s nothing to systemat-
ically compare it to. I hope instead that my work highlights how the
researcher-as-designer can operate effectively in both roles and serve
as a starting point for a broader discussion about why designing
and studying systems is as valuable a research strategy as studying
existing systems.¹

If we accept that conducting design and development are valuable
research processes, we must consider the challenges to this kind
of work. If we hope to avoid the limits of studying design work in
decontextualized lab situations, then we need to find situations were
our system might credibly be used “for real.” The best situations are
ones in which people interacting with and through the system can
do so in the normal contexts in which they might interact with such

¹ Designing and building systems can
take a substantial amount of time, es-
pecially if you hope to deploy those
systems in situ instead of in lab con-
texts. If papers are the main output
metric for a researcher, this approach
is not necessarily an efficient way to
generate papers, since few conferences
will accept papers on the design or
development of a novel system without
an attached ‘study’. This might help
explain the waning popularity of this
sort of research.
Table 2.1: Comparison between the kinds of technical contexts that the researcher-as-designer typically studies compared to the researcher.

<table>
<thead>
<tr>
<th>Researcher-as-Designer</th>
<th>Researcher</th>
</tr>
</thead>
<tbody>
<tr>
<td>using novel technology</td>
<td>using existing technology</td>
</tr>
<tr>
<td>not widely deployed or available</td>
<td>popular/widely used</td>
</tr>
<tr>
<td>smaller, bounded user groups</td>
<td>larger, fluid user groups</td>
</tr>
<tr>
<td>extrinsically motivated users</td>
<td>intrinsically motivated users</td>
</tr>
<tr>
<td>rougher edges</td>
<td>well polished</td>
</tr>
<tr>
<td>consumer-oriented</td>
<td>consumer or professional</td>
</tr>
<tr>
<td>bounded use durations</td>
<td>potentially unbounded use durations</td>
</tr>
<tr>
<td>internal process traces</td>
<td>publicly observable process traces</td>
</tr>
</tbody>
</table>

...
view and access to the data is not contingent on the arguments being made with the data being acceptable to corporate actors. This is a further reason why the researcher-as-designer can find insights that might not be accessible in systems designed by others.

This is not an argument for one approach to replace another; both approaches can easily co-exist. However, because I rely so heavily on this approach in my work, I want to present the reasons why working in this particular way is a valuable complement to dominant research strategies. This is particularly relevant in a Media Lab context where building systems is a taken-for-granted component of our work, but we rarely seek to justify it on its own terms to other researchers. Hopefully this section can serve as a start to a broader conversation about whether and how this research strategy can make contributions that are respected and valued outside our own Lab.

2.2 Methodological Approach

This work is not a single monolithic study and so it does not share a single methodology across all its components. Still, there are some common methodological approaches that I would like to highlight. In particular, I would like to draw attention to the differences between the sorts of design research that I aspire to conduct, and describe how this approach contrasts with more traditional approaches.

Perhaps the best way to contrast my approach to the approach of much of the systems research literature is to describe my approach as in situ to contrast it with what I would characterize as a traditional ex situ approach. Because of its roots in experimental psychology, studies in the HCI and CSCW fields are typically designed around normalizing as many as possible variables to focus on particular independent variables. Studies are typically conducted with subjects recruited independently, who are given a particular scenario with defined roles and tasks to complete. There are usually rewards involved, and sometimes those rewards are tied to performance. McGrath [1984] describes these as “quasi-groups” and claims that they “are the least natural of the set of groups discussed so far” (emphasis original). The basic axes that McGrath uses to separate groups splits them with respect to temporal scope, activity scope, whether the group frequently handles their current task, and how the group is composed. On all of these axes groups composed for lab studies are extraordinary; they share few common features with the groups that my work actually hopes to address. As a result, conducting lab studies would provide little valuable insight into group-oriented systems.

Setting aside the properties of the group, the setting of lab studies
is also quite problematic. Participants are expected to set aside their prior motivations and replace them with a desire to follow instructions and be a good participant. Studies in this space are rarely this extreme, but the Stanford Prison Experiment [Haney et al., 1973] and Milgram’s work on authority [Milgram, 1983] show how powerful this effect can be. Even when the researcher isn’t trying to push subjects, there is nonetheless a passive and subtle pressure to respect the researcher’s work and help them generate positive results. This is exacerbated when there is a prior relationship between the study subjects and the researcher, which is all too common.

Studies conducted in an *ex situ* lab situation also tend to extract individuals from their existing socio-technical infrastructures. Few systems are used in a vacuum. Even a system as simple as a traditional website can be viewed in quite different contexts: phones, tablets, and computers all have quite different interaction properties, and are often configured in distinctly individual ways. Yet in lab studies we tend to insist on participants using a specific configuration that is likely to be foreign to their own experience outside the lab. We also force a certain level of focus on study participants. Although it is common to frequently switch between applications and interaction contexts, researcher expectations in a lab context will significantly discourage this sort of behavior, even though it is natural and common *in situ*.

Lab studies also tend to limit the sorts of systems that can be fruitfully studied in terms of frequency and duration of use. We expect participants to use an interface intensely for a short period of time measured in minutes. Interfaces that are intended to be used occasionally over longer periods of time are not easily evaluated in lab studies and so an over-reliance on lab studies will tend to exclude those classes of systems. This is an example of the most insidious sort of influence that the pressure to conduct *ex situ* experiments has on research: systems that are hard to study in terms the research community will accept will simply be built less frequently. Researchers must constantly make strategic judgements about what sorts of work they do and whether it will be be viewed as a valuable research contribution.

In my work, I try to balance the focus of *ex situ* studies in recent literature with a strong push to create systems that can credibly be studied *in situ*. Mirroring my earlier argument for the researcher-as-designer to complement the researcher, *in situ* research can more effectively address different sorts of research questions at a different scale than *ex situ* research.

In contrast, my work focuses on recruiting people who are performing an existing task (e.g. asking questions of panelists at a con-
ference or participating in a classroom discussion) regardless of their participation in a study or not. Indeed, in the backchan.nl case there was not a formal study at all, simply organizations that requested access to the system. Their desires and goals predated my technical interventions. The system has continued to be used by various audiences to meet an actual practical need. This demonstrates the stability of both the concept and the technology. Although we should not expect that all in situ-style systems should continue beyond the life of the project, their viability should represent a research contribution commensurate with the challenges of building durable and complete systems that operate effectively beyond a tightly controlled ex situ environment.

I am far from the only person conducting work in this manner. There is a long history of in situ work. The vast majority of the work described in the Channels section in related work is conducted this way. Thunderwire, Cruiser, and Portholes are prime examples of this approach. Within a Media Lab context, it seems natural and uncontroversial to work in this fashion. But when looking at research outside the lab it can be easy to feel like this approach isn’t broadly valued. A common criticism of work at the Lab is that it is all about presenting a system as “cool” and not about addressing any issues that research communities are interested in. Making this argument here is my way of trying to leave behind an argument that might be useful for later researchers at the lab struggling to justify their work to outside audiences that prefer empirical approaches or studies that focus on existing successful platforms.

2.3 Research Themes

The projects in this thesis are connected by a series of research themes. All of the projects speak to these themes in different ways. In this section, I will introduce these themes in more detail.

2.3.1 Grounding

Grounding was introduced as a concept by Clark [1989] to better understand discourse processes. Clark and Schaefer sought to model two party conversations, with a focus on how conversants construct a common ground between them. The notion of a common ground that represents the body of understanding shared by conversants was a feature of many discussion models at the time. However, the process by which a concept moved from being understood by one participant in a conversation to both participants was unclear. Many models argued that by simply uttering a sentence, its content were
immediately part of common ground. Using a corpus of real spoken conversations, Clark and Schaefer propose an alternative model where all contributions to common ground must be accepted by the other conversant before they become part of common ground. This acceptance need not be an utterance itself; an utterance that builds on a previous utterance can represent an implicit acceptance.²

In my work, the Clark and Schaefer model has a number of implications for my work. As Clark identifies in later work [Clark and Brennan, 1991], mediated communication channels can inhibit the formation of common ground in a variety of ways. A slight addition of latency, as is common in voice or video conferencing can make turn-taking challenging which leads to longer utterances and rarer acceptance turns. Furthermore, Clark’s models are primarily for two person conversations. As the number of people involved increases (as is true for all my work), the challenges become even more pronounced. Thus systems that can help provide ways to come to group understandings about goals, topics, or process can be quite powerful.

An interest in promoting the construction of common ground is not limited to mediated technical systems. Having a well defined group process can play a major role in accelerating this process, whether it’s through hand signals for signaling approval or disapproval or structures for transitioning between topics. A simple understanding like “if I stay silent in a discussion it represents agreement; if I disagree, I must speak up” can help alleviate some of the challenges of common ground in group work.

A common theme both in my work and other approaches to managing the challenges of creating common ground in a group is exporting group state into the environment. Post-it notes in design processes represent a similar strategy; by moving ideas onto physical objects in the environment, we are not over-burdening the audio channel with all the content of our ideas, while also creating representations that are durable and shared. A drawing that everyone can see may mean more to its author than to everyone else, but is still more a part of common ground than only verbal contributions. My work frequently uses a similar strategy: shared displays that can be seen by all participants and whose content is generated by the group. By using a shared screen, it becomes possible for group interactions to be represented in a way that they can rapidly become part of common ground without requiring verbal utterances. In each of the projects described in this thesis, I will show how this process works (or fails to work in some cases) and argue that communication tools that create side stage spaces can help mitigate the challenges of grounding in mediated communication.

² Backchannels also play an important role here. Nodding or showing obvious confusion can be non-verbal, visual ways to accepting a contribution or rejecting a contribution. Taking away these backchannels can make it very difficult to judge whether common ground is being created or whether your audience is silently rejecting your attempts to add to common ground.
2.3.2 Non-verbal actions

While grounding and common ground are concepts primarily focused on oral and written utterances, non-verbal actions have an important role to play in this process as well. When people lament the lack of “body language” in mediated communication, it is partially a complaint about the difficulties of building common ground in the absence of easy non-verbal acknowledgements of a contribution. Non-verbal actions are much broader than backchannel acceptances, though. They also include actions like nodding, looking attentive, fidgeting, taking notes, hand-raising, and so on. Although the lack of these actions in mediated environments is frustrating, we can’t easily replicate them all in a mediated context. Non-verbal ways of communicating in non-mediated contexts have arisen over a long period of time, and are frequently culturally specific. Deciding which non-verbal actions to support, how to represent those actions, and understanding how these actions are understood are at the heart of this research theme.

There is an instinct in this space to simply recreate the non-verbal actions that serve us so well in un-mediated contexts. In the same way that interface designers often fall back on physical textures to evoke the properties and uses of the physical object in a digital context, designers of social systems can sometimes employ a similar strategy. This kind of social skeumorphism is quite common in the virtual world domain, and I will discuss its failings in more depth in Chapter 3. On the web, however, such skeumorphism is harder. Although many of our non-verbal actions take their names from physical actions like “poke”, “raise hand” or “flag”, we have also seen a wide range of new kinds of actions like “upvote”, “like”, and “retweet.” These actions, like their non-mediated predecessors, have slowly evolved to carry their own novel meanings and connotations based on context.

Although non-verbal actions of this sort are prominent components of mediated social systems, they are rarely called out as an explicit research area in this way. Nevertheless, a review of the systems literature later in this chapter will reveal that much of the work in this space employs carefully chosen non-verbal actions to facilitate their design goals. By calling out this component of a thoughtful design, I hope to call attention to the important role that non-verbal actions can play. In each of the projects I discuss, I will call out the non-verbal actions that I employ and discuss their efficacy.

\footnote{Apple is the highest profile company to use this practice extensively, with their leather-textured calendar application, bookshelf background for ebooks, and brushed metal window effects. This technique is often called “skeumorphism” and is widely derided by user interface designers as a visual crutch that impedes the construction of accurate mental models for an interface, but nevertheless is increasingly common particularly in mobile applications.}
2.3.3 Attention

Attention, for our purposes, is best understood not as an innate characteristic of someone’s behavior, but as a value-laden socially situated performative behavior. “We are always to a certain extent in a state of distraction,” according to sociologist Emile Durkheim. [Durkheim, 1974] Every situation is composed of stimuli that disrupts some fictional conception of undivided attention. Likewise, every situation requires some aspect of performance, as individuals are required to communicate their attentiveness in response to specific social contexts. Although in many situations we might expect “attention” from our audience, what we really expect is an appropriate performance of attention that satisfies social norms. This performance is related to, but not exactly identical to what someone is focused on at specific moment. The performance of attention is a necessary indirection because we cannot ever really know for sure what someone is focusing on, so instead we rely in social situations on proxy performances.

In this thesis, I describe systems that complement another communication channel. In these contexts, I am interested in two main aspects of attention: how do people make choices about what to focus on in their environment and how do representations of attention shift in the presence of these new ways to be involved?

This research theme is partially a response to a frequent criticism of this sort of work. Won’t creating new stages simply devalue attention on the main stage? Because attention is a social construct, part of the goal with each of my projects is to show how creating side stages tightly related to the main stage can make it easier to reconfigure expectations about what attention looks like. In spaces like this, focusing on a mobile device, laptop, or projected image can shift from being a display of inattention to become another way of attending to an important shared experience. This shift is important, because my work shows that many people can feel disenfranchised by main stages in different contexts and new stages with different affordances can help boost overall engagement.

This takes a shift in attitude. Instead of assuming that being in a privileged position on the main stage means you deserve the full focus and performed attention of all audience members, it is more productive to view people’s focus as fluid; capable of shifting often and easily among a variety of targets on a moment-to-moment basis. A well designed side stage is one many possible targets for attention in a space, and in the best situations can contribute to the main stage making a focus on either the main or side stage a way of attending to the same underlying experience. But because attention is

4 Originally from [Harry et al., 2012].
fundamentally performative, this change will require shifting social understandings of different ways of being part of a shared experience.

Across each of my projects in this thesis, I will try to address how attention is understood and constructed, as well as describe the processes by which people make choices about attending to either the main stage or the side stages.

2.4 Related Work

Designing new systems for collaboration and communication, as opposed to studying existing systems, has long been a major stream of HCI and CSCW research. This section will summarize the most salient past work in this area, although little of this work is recent and responsive to the significant shifts in the way people use technology to communicate, collaborate, and play.

For a variety of reasons, there has been somewhat of a shift in interest away from the kind of hybrid face-to-face/mediated experiences I create towards building and studying systems for asynchronous experiences between much larger numbers of people. The advent of research on mass collaboration systems like Wikipedia (e.g. [Kittur et al., 2007]) and “crowd sourcing” (e.g. [Bernstein et al., 2010]) is part of a larger shift away from what was once the center of gravity of systems research. This shift is a natural response to changes in both technology and the common experience of modern collaborative technology users in a web-oriented world where asynchronous interaction became the norm. But the experiences these sorts of studies focus on are predominantly single-channel and asynchronous. This is in stark contrast to my work, which is concerned with the properties of multi-channel synchronous experiences.

In this work, I focus on interactions within groups both small and large, co-located and remote, but always co-temporal. Although interesting new sorts of work and communication structures are evolving in the asynchronous domain, we should think not just about how to marshal large numbers of people, but about how small groups of people who know each other work, recognizing that much of that work happens face-to-face or co-temporally while geographically distant. It is not effective to treat these interactions as a simple increase in tempo on asynchronous interactions. In synchronous systems, it is much more important to understand how we are perceived (and can control those perceptions) by others. In asynchronous systems, these issues are minimized; we experience others through their actions on shared objects like documents.

My survey of related work is organized into three major design
designing complementary communication systems: translucence and awareness, adding new communication channels, and design techniques to help people reflect on their own participation and the participation of others. These design strategies have influenced my own design process and have important findings related to my three main research themes: grounding, actions, and attention. As I discuss each design strategy, I will point out their connections to the main research themes. Table 2 summarizes the major work covered in this section, and its relationship with the three themes. After a discussion of work primarily in the systems literature, I will also cover related theoretical contributions and perspectives that influence and contextualize my work.

2.4.1 Translucence & Awareness

This work owes a clear debt to the work of Erickson and Kellogg [2000] on social translucence. Their work intersects with my action and grounding themes. Social translucence is a design strategy that aims to create “digital systems such that people’s presence and activity, made appropriately perceptible, will create accountability and more easily coordinated action” [Kellogg and Erickson, 2002]. They call their example systems designed for this purpose “social proxies” that use “abstract visual representations ... to portray information, in addition to contextual information provided by the other common traces of user activity in mediated communication environments (e.g. persistent conversation).” In each of their projects (Babble, Loops, Lecture, Auction, etc.; [Erickson and Kellogg, 2003] is a nice overview of these projects), they seek to promote a sense of “collective awareness” where each person using the system has a sense of the actions of others in the system and appreciates that this awareness is mutual.

We share an interest, in my terms, in how we can construct meaningful actions in mediated social spaces and how we can understand how public displays can help ground collaborative and discursive processes. As Erickson and Kellogg point out, this has been a topic of interest both direct and indirect for quite some time in the systems literature. Their work nicely complements work by Gutwin and Greenberg [2002], who present a framework for thinking about the ways the workspace awareness through actions can be constructed and presented. Ackerman and Starr [1995] share this interest, too, but focus on representing overall system activity as an inducement for broader participation. This is an important finding, and one I echo in my work, particularly when it comes to a public, optional system like backchan.nl. Distributed cognition, as described by Hollan et al. [2000] represents another productive way to think about these processes; by fostering a sense of mutual awareness we can support the kinds
Table 2.2: A summary of the major related work to be discussed in this section and its relationship with the main research themes.

<table>
<thead>
<tr>
<th>Social Translucence</th>
<th>Grounding</th>
<th>Actions</th>
<th>Attention</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Loops, Babble, Lecture</strong></td>
<td>● ● -</td>
<td>[Erickson and Kellogg, 2000]</td>
<td></td>
</tr>
<tr>
<td><strong>Group SketchPad</strong></td>
<td>● ● ●</td>
<td>[Gutwin and Greenberg, 2002]</td>
<td></td>
</tr>
<tr>
<td><strong>CafeCK</strong></td>
<td>● ● ●</td>
<td>[Ackerman and Starr, 1995]</td>
<td></td>
</tr>
<tr>
<td><strong>Airplane Cockpits</strong></td>
<td>● ● ●</td>
<td>[Hutchins, 1995]</td>
<td></td>
</tr>
<tr>
<td><strong>ClearBoard</strong></td>
<td>● ● -</td>
<td>[Ishii and Kobayashi, 1992]</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Channels</th>
<th>Grounding</th>
<th>Actions</th>
<th>Attention</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Class Backchannels</strong></td>
<td>· · ●</td>
<td>[Yardi, 2006]</td>
<td></td>
</tr>
<tr>
<td><strong>Conference Backchannels</strong></td>
<td>· · ●</td>
<td>[McCarthy et al., 2004]</td>
<td></td>
</tr>
<tr>
<td><strong>Semi-Public Displays</strong></td>
<td>● ● ●</td>
<td>[Huang and Mynatt, 2003]</td>
<td></td>
</tr>
<tr>
<td><strong>Rendezvous</strong></td>
<td>● ● -</td>
<td>[Kellogg et al., 2006]</td>
<td></td>
</tr>
<tr>
<td><strong>Audio Backchannels</strong></td>
<td>· ● ●</td>
<td>[Yankelovich et al., 2005]</td>
<td></td>
</tr>
<tr>
<td><strong>Fragmented Social Mirror</strong></td>
<td>● ● ●</td>
<td>[Bergstrom and Harris, 2011]</td>
<td></td>
</tr>
<tr>
<td><strong>VideoWindow</strong></td>
<td>● · ●</td>
<td>[Fish et al., 1990]</td>
<td></td>
</tr>
<tr>
<td><strong>Thunderwire</strong></td>
<td>· ● ●</td>
<td>[Hindus et al., 1996]</td>
<td></td>
</tr>
<tr>
<td><strong>iCom</strong></td>
<td>● ● ●</td>
<td>[Agamanolis, 2003]</td>
<td></td>
</tr>
<tr>
<td><strong>Portholes</strong></td>
<td>● · ●</td>
<td>[Dourish and Bly, 1992]</td>
<td></td>
</tr>
<tr>
<td><strong>Cruiser</strong></td>
<td>· ● ●</td>
<td>[Fish et al., 1992]</td>
<td></td>
</tr>
<tr>
<td><strong>GDSS</strong></td>
<td>● ● -</td>
<td>[Nunamaker et al., 1991]</td>
<td></td>
</tr>
<tr>
<td><strong>Cognoter</strong></td>
<td>● ● ●</td>
<td>[Tatar et al., 1991]</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Reflection</th>
<th>Grounding</th>
<th>Actions</th>
<th>Attention</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Second Messenger</strong></td>
<td>● · -</td>
<td>[DiMicco et al., 2007]</td>
<td></td>
</tr>
<tr>
<td><strong>Meeting Mediator</strong></td>
<td>● · -</td>
<td>[Kim et al., 2008]</td>
<td></td>
</tr>
<tr>
<td><strong>Conversation Clusters</strong></td>
<td>● · ●</td>
<td>[Bergstrom and Karahalios, 2009b]</td>
<td></td>
</tr>
<tr>
<td><strong>Conversation Votes</strong></td>
<td>● ● ●</td>
<td>[Bergstrom and Karahalios, 2009a]</td>
<td></td>
</tr>
<tr>
<td><strong>Conversation Clock</strong></td>
<td>● · ●</td>
<td>[Bergstrom and Karahalios, 2007]</td>
<td></td>
</tr>
</tbody>
</table>
of process that Hutchins [1995] describes in flexible communication systems. I see my work as a continuation of these past approaches to representing activity. Although there are many similarities in terms of findings and design strategies, I will focus here on the points of difference as a way to clarify the contributions of my work.

Although Erickson and Kellogg are particularly concerned with what is made visible and what is kept private (the difference between, in their terms, transparency and translucence), this is a point of divergence between our work. Although I agree with their analysis of the value of considering what actions should be made visible and what should be concealed, it is not a main focus of my analysis. In my work, reading is essentially always invisible and any other action is visible. This is partially a response to their suggestion that it is “important that participants were aware of the others’ awareness of [the properties of the system]” [Erickson and Kellogg, 2003]. This fits nicely with Clark and Brennan’s [1991] presentation of grounding. Simply being told something by someone is not enough for the conversation to move on - you must accept that presentation of information, and that acceptance needs to be accepted by the original presenter. In this way, grounding plays a role not just in communication itself, but in how we communicate information about who we are and what we’re doing through actions in the system.

This finding also suggests that if you don’t know which actions are public and which are private, it diminishes the value of translucence as a design strategy. In their work, they tend to rely on physical metaphors to communicate the visibility properties of a system. This is a sensible strategy, but I feel this limits the kinds of experiences we can craft. In my work I tend towards not including invisible actions and instead create completely transparent spaces with carefully selected actions that are worth making visible. This is possible partly because the group sizes in my work are smaller than in the main examples they propose, and it’s feasible to show all actions without it being overwhelming.

Erickson and Kellogg a number of specific design findings that complement some of my experiences designing similar systems. They describe three approaches to visualizing activity: realist, mimetic, and abstract. [Erickson and Kellogg, 2003]. I share their interest in abstracted representations, although for different reasons. They argue that realist and mimetic approaches face “substantial pragmatic barriers (e.g. expense, infrastructure, support)”. In the years since this work was originally done (and well before; one might reasonably argue that ClearBoard [Ishii and Kobayashi, 1992] offered an elegant realist approach), many of those pragmatic barriers have fallen. We’ve seen large-scale virtual worlds (like Second Life) that use
mimetic approaches and wide adoption of video conferencing which uses realistic representations. I wish to argue that abstract representations are simply more flexible and better, even given the option of realistic or mimetic approaches.

My work goes into greater depth than Kellogg and Erickson’s does on the issue of “public not personal” displays. While we agree that it is critical that each person’s display doesn’t deviate in the kinds of information it represents, in my work these displays are not monolithic—they are not the only venue for interaction between people. Furthermore, displays in my work are most often themselves public, which reinforces the grounding effect. Indeed, that is the most significant deviation between our work. In all of the social proxy work, the proxy is the primary communication medium; in my work, systems coexist with another primary communication channel, and rarely have any knowledge about the contents of that channel. Public displays also exacerbate issues of attention, which tend not to be major issues for work in the social proxy space (as shown in Table 2.2).

The other major distinction is in the use of metaphor and display techniques. Erickson and Kellogg limit their representations to “a relatively large geometric shape with an inside and an outside and sometimes other features that represent the online situation or context” [Erickson and Kellogg, 2003] with “small colored dots” to represent individual users (similar to [Viegas and Donath, 1999], minus the direct agency). These design strategies are illustrated in figures 2.1 and 2.2. Furthermore, they argue that the best way to represent information is through the use of “relative movement” of the user-dots in a way that has “metaphoric correspondence to the position and movement of people’s bodies in face-to-face analogs of the online situation.” [Erickson and Kellogg, 2003] As I hope my work shows, these limits are not at all necessary to create spaces of meaningful action that facilitate grounded communication and collaboration. Specifically, the need for relying on face-to-face analogs is not a helpful constraint. Instead, my work seeks to create spaces that are easily understood and provide contexts for meaningful action without relying on existing face-to-face metaphors.

2.4.2 Channels

The primary focus of my work is on designing systems that add new communication channels and understanding how those channels operate in contrast to existing channels. In this section, I will present related work that addresses some of these questions.

The work most directly related to these questions comes from
research into so-called “backchannels” in presentation and classroom settings. Yardi [2006] describes how a chat-based backchannel operates over a semester in a classroom, McCarthy et al. [2004] describe a similar approach at a conference. Backchannels can also be considered a potential part of non-event-oriented contexts too, like long-term co-working among small groups. [Huang and Mynatt, 2003] Backchannels are not just focused on co-located groups, however, and Kellogg et al. [2006] (among others, e.g. [Yankelovich et al., 2005]) has addressed how text and audio backchannels can coexist in distributed contexts. Although past work has addressed in general terms the different ways people use backchannels, it has not sufficiently explained the complicated issues around channel selection, attention, distraction, and identity. Furthermore, in my work I try to move beyond just adding new text or audio channels by adding other kinds of non-verbal actions. In terms of my research themes, past work on backchannels has largely focused on characterizing use patterns, with some discussion of attention. More recent work, like [Bergstrom and Harris, 2011], shares an interest in how we can construct actions and how shared displays can be used to help ground the interaction.

Much of the work on creating shared media spaces, driven by experiments at PARC in the late 1980’s and early 1990’s is salient to my work. Although in some cases this work focused on creating new primary channels, researchers quickly became attuned to problems of privacy and attention because such systems always co-exist with face-to-face communication, in much the same way they do in the systems I design. The earliest work at PARC [Olson, 1991] focused on creating flexible video connections between offices and conference rooms. Subsequent work focused less on a phone-call-like model where connections are created and ended and shifted towards creating spaces with different affordances. Sometimes these involved connecting multiple individuals together, as in CAVECAT [Mantei et al., 1991]; other times researchers focused on creating long term persistent video connections in common areas of distributed research groups in the VideoWindow [Fish et al., 1990] project.

Over time, attention shifted more towards taking advantage of the possibilities to do more than just create “being there” experiences. Some researchers experimented with audio-only spaces [Hindus et al., 1996], finding that video was not required to create a sense of connection and space for users, but that the properties of audio did require audio-specific etiquette and coping strategies for the system to be useful. iCom represented a particularly rich design perspective on connecting spaces [Agamanolis, 2003], recognizing that awareness need not be limited to visual awareness, but can extend to informa-
tion awareness which can be productively embedded in a media space. This embodies the “beyond being there” model best of all the work in this research stream: not just trying to create a transparent window between remote spaces, but making something better than a window could be. Furthermore, these projects also focus more on issues of attention, because they are not necessarily always the primary interaction venue for their users.

Serendipity also evolved as an important part of sharing an office environment that was not present with most media space systems. Portholes [Dourish and Bly, 1992] addressed this explicitly by giving people a broader view of remote spaces instead of focusing just on main channel interactions. While my work is not concerned with serendipity, this kind of visual side channel carries important awareness information in much the same way that the side stages in my systems add important contextual information to an interaction. CRUISER [Fish et al., 1992] offered non-verbal ways to signal a desire to emulate some of the office hallway etiquette for signaling a desire to drop in and chat informally, without the explicitness of placing a call. The addition of moves like “cruise”, “glance”, and “visit” are similar in approach to the non-verbal actions at the core of my work like voting in backchan.nl, promoting ideas in Tin Can Classroom, or moving around the field in Information Spaces.

Early media space researchers proposed a distinction between “formal” systems and “informal” systems. [Olson, 1991] While most of the work discussed here (and much of my own work) tends towards the informal side of that continuum, there are some formal elements in my work. This formality manifests most strongly in Group Decision Support Systems research. These systems (exemplified by the work of Nunamaker [Nunamaker et al., 1991]) provide prescriptive systems to support particular brainstorming, decision making, outlining, and voting schemes or policies. In the typical GDSS configuration, each participant has their own computer and interacts with shared structured data in some way, like submitting a new idea or voting on a proposal. In systems like this, the assumption is that having a structured display will ground otherwise informal processes by forcing participants to use the actions the system provides as a set of legitimate conversational moves. Although I tend towards informal systems in my work, the work in this space nonetheless has much to teach us about grounding and actions.

The lack of consistent results in comparative work in this area [Dennis et al., 1988] illustrates the importance of focused design analysis to contextualize findings; it is not useful to view all brainstorming systems as equivalent and comparable in analysis, and I hope that my work will illustrate how the subtleties in interface and
approach can have big impacts on outcomes that help explain some of the contradictory results in past GDSS work. Work in this space also raises serious questions related to attention that is largely unaddressed. In fact, in many situations the GDSS designers advocate for largely shutting down pre-existing primary communication channels to focus on the structured, mediated alternative.

Although somewhat rare in the literature, there are a handful of projects that directly address the kinds of hybrid spaces that I seek to create. *Cognoter* [Tatar et al., 1991] addresses this design space most clearly. Like my work, *Cognoter* created a hybrid space for very small groups (two to five people) that includes both personal and public displays where users can create items and spatially arrange them like on a whiteboard. Textual items can be arranged on a users’ display and that arrangement is mirrored on all other users’ personal displays. The authors characterize *Cognoter*’s model of creating shared text elements as representing a “parcel-post” as opposed to an “interactive” conversational model. Instead of embodying a present/accept process (as described by [Clark, 1989]), they describe their process as being more like literary communication (such as email) where the writer tries to make sure “that the addressees should have been able to understand his meaning in the last utterance” (emphasis mine). This is in contrast to face-to-face interaction, where we can interactively ascertain the extent to which we are being understood (and repair mistakes) before moving on. The authors describe *Cognoter*’s failure to be used effectively by its users as (in part) a conflict between the interactive mode of face-to-face communication and the parcel-post model in *Cognoter*. The *Thoughtswap* project [Dickey-Kurdziolek et al., 2010] also shares the goal of creating a complementary communication system. Like *Cognoter*, the mediated space is used serially with the face-to-face space, while I am interested in creating spaces for legitimate simultaneous performances in mediated and non-mediated spaces. This suggests a major hurdle for my work: can we create systems that use a parcel-post model yet still integrate fluidly with the interactive face-to-face model? *Cognoter* and *Thoughtswap* suggest this is hard, but I will show throughout my work how these barriers can be overcome and suggest ways to explain *Cognoter*’s negative findings.

2.4.3 Reflection

Understanding how we present ourselves to others has been a topic of sociological inquiry for quite some time. Although many of the insights of scholars like *Goffman* [1959] about how we communicate and interpret information about who we are and how we want to
be treated are still relevant, the information that is available about people has changed substantially. In some of the examples in this section, designers have added some new bit of information about people to a face-to-face discussion; in others, we don’t have any of the traditional information we would get from being face to face with someone and rely on new types of signals (like the non-verbal actions I propose) to create a sense of people around us. Part of what sets mediated communication apart is the ability to accumulate behavioral histories and represent and reflect those histories to ourselves and others. The work in this space is not as closely connected to my main research themes, but I include it here primarily because it has served as a source of design inspiration.

My work is substantially inspired by the work of DiMicco et al. [2007] on the Second Messenger project. In this project, participants in a group discussion were presented with a constantly-updating bar-chart visualization representing the relative amount of time they had talked during the discussion. They found that while people who over-participated without a visualization tended to moderate their participation when the visualization was present, people with low participation did not participate more just because others were participating less. Meeting Mediator [Kim et al., 2008] took a similar approach, but focused on situations where groups of two people could see each other and had to interact with another group of two people that they could only hear. Using a different visualization, Kim et al. found that groups were more interactive with the system than without, although there was not a correlation with group performance.

Bergstrom has done a series of projects that adopt a similar design strategy. Conversation Clusters [Bergstrom and Karahalios, 2009b] pulls topics from an audio conversation and presents them in clusters on a table-top display. Conversation Clock [Bergstrom and Karahalios, 2007], like Second Messenger and Meeting Mediator, visualizes conversation participation, but uses a timeline metaphor instead of an aggregative metaphor. Conversation Votes [Bergstrom and Karahalios, 2009a] lets users discreetly vote about the progress of a discussion, and displays anonymous votes on a table-based display. Karahalios and Bergstrom [2009] describes this design space as “social mirrors”.

These are all examples of the accumulate and reflect design strategy, where the system tracks some aspect of behavior: spoken participation in the case of Second Messenger and Meeting Mediator, discussion topics and group attitudes in the case of Bergstrom’s work.
2.4.4 Meeting Structure

Any system designed to support interaction makes a number of implicit assumptions about the nature of the social process it’s trying to support. This is true across design domains. For instance, calendaring systems often assume that the important aspects of someone’s work life can be captured in meetings. Meetings have a time, a place, and other meeting participants. Certainly, this describes some kinds of work life, but plenty of other jobs don’t fit with that metaphorical structure. Imagine a car repair shop; a system to support that sort of work environment would necessarily be concerned with specific tasks that need to be done and which people in the shop were going to do them. A meeting organization system would not be as appropriate for managing that environment. Of course, over time systems influence people’s behavior in such a way that it makes it hard to distinguish between the implicit assumptions about how people would use the system and how their behavior has adjusted to best make use of the systems they have available.

Meetings are rich social experiences that are composed of a variety of deeply interconnected components. I turn to McGrath’s [1984] theoretical framework as a good starting place for understanding the diversity of factors that influence a group’s process. McGrath describes groups as “task performance systems”, recognizing the critical extent to which tasks characterize a group’s interaction, along with the group composition, the properties of the environment, and properties of the individual. He identifies eight types of tasks: planning, creativity, intellective, decision-making, cognitive conflict, mixed-motive, competitive, and performance. In this chapter we are primarily concerned with decision-making tasks.

Various structures for organizing decision-making have been proposed. One interesting structure is IBIS [Kunz and Rittel, 1970], which classifies the structure of arguments in design meetings. In their model, a discussion has Issues (e.g. “Users aren’t using this feature as much as we thought they would”), Positions (“We should include the feature in our tutorial”) and Arguments (“Not many users are following the tutorial either, we need another way to promote this feature.”). These objects are linked by different relationships, like “supports”, “objects-to”, “responds-to” or “specializes”. This creates an abstract argument network that organizes the discussion around concrete issues. Conklin and Begeman [1988] describe a computer-supported system that uses this structure for both synchronous and asynchronous discussions. This basic approach of modeling a particular task network in a formal way is shared by many Group Decision Support Systems of which Nunamaker et al. [1991] is a nice overview.
In McGrath’s terms, though, systems like gIBIS (and many of the systems described in the GDSS literature) are not strictly addressing one task. They’re nominally about making decisions, but a group using gIBIS is also sometimes trying to solve intellective tasks (e.g. solving questions with a single right answer), perform generative tasks, and perhaps resolving conflict tasks. The precise mix of these tasks varies in each group’s process, which creates a significant diversity in system designs. Structures that assume a particular task profile are unlikely to be effective when applied to another task profile.

Task profiles are one example of a formalizing instinct to organize and abstract certain features of human practice in a way that can help us both better understand those practices or (in the case of designers) design tools that support those practices. Not only is it a question of useful or accurate abstractions, it is also a tension between the right amount of structure and too much structure. A model like IBIS or Roberts’ Rules of Order [Robert III et al., 2000] is relatively heavy amount of structure. In contrast, process models like nominal group technique [Bartunek and Murninghan, 1984] operate at an even higher level of abstraction. In general, my instincts lean towards less formality rather than more, but that reflects the sorts of practices I’m creating tools for more than a real argument for either strategy. Regardless of how one decides to abstract, there is still a complex relationship between the selected formalism and the practice it is either modeling or seeking to promote. Berg [1997] describes this interplay this way:

By offering abstracted models of the work and/or by processing input into output, formal tools are attributed central roles in organizing the work within the workplace.

Berg organizes the camps on this issue into two main discourses: naive formalists who believe that formal models are necessary for rational or scientific decision-making and scholars who argue that the texture of human practice defies reduction to formal models. This argument has a political angle to it, as well, arguing that (as Berg describes it) formal tools will “inevitably function in a rigid, impoverished way, thus de-skilling and dehumanizing the work of those who are caught in its cold, instrumental rationality.” These are both extreme views and not widely held at this point. Berg characterizes the distinction between the formal tool and the human practice as being like the distinction between the map and the terrain it describes. Maps are a way of abstracting and describing terrain, while also influencing the ways that people understand and develop that terrain. They are necessarily tightly interrelated.

Although Berg is primarily interested in building theoretical tools for better understanding practices at the boundary between formal
tools and human practices, this perspective is also valuable for designers. Formal model building is a critical step in designing any sort of software system. Indeed, designing abstract representational models for data is a core part of any software design process. As researchers we frequently have our own agendas and are trying to describe future terrains that may not exist yet. Put another way, maps are sometimes guides to exciting new terrain, not just describing a certain perspective on existing terrain. This is more precisely the sort of work that I do.

In the chapters that follow, I will describe a series of projects. For each of these projects, I will attempt to identify the existing practices it seeks to support, as well as the sorts of new practices it aims to encourage through the formalizing effects of software systems.

Hughes provides a clever (if somewhat inaccessible to non-database engineers) analysis of the challenges representing human marriage practices in a database. It illustrates, too, the built in biases of computational systems towards certain sorts of convenient abstractions despite their lack of grounding in practical experience.
Over the past twenty years, the prospect of visiting a virtual world different from our own world has moved from being strictly the domain of imagination to being widely accessible. What started as worlds presented through text-based descriptions and interacted with through typed commands evolved into rich graphical and sensorial experiences that became rapidly accessible to mass audiences. The argument for virtual worlds has always had a revolutionary tone. Virtual worlds might, for example, suppress the prejudices of offline society by hiding identity information, breaking down the boundaries of distance, and making experiences broadly accessible that might not otherwise be feasible. Perhaps we could even build a new virtual society, one that is better than our own. These utopian visions are attractive precisely because the urge to grow beyond the confines of everyday experience is so strong.

This vision was so attractive that in 2007 the market research firm Gartner famously predicted that by the end of 2011, “80 percent of active Internet users (and Fortune 500 enterprises) will have a ‘second life’, but not necessarily in Second Life.” [Ano, 2007b]. This has not come to pass. Linden Lab has stopped publishing public usage data, but it is clear by comparing its cultural impact to systems like Facebook, Twitter, and YouTube that Second Life has been almost completely eclipsed in the public imagination by other types of mediated social experiences. Despite this shift, it is still productive to consider why virtual worlds were viewed as strong potential venues for mediated social interaction. What attracted us to these experiences in the first place? Why did these experiences ultimately fail to catch people’s imaginations? Are there aspects of “virtual” experiences that were left behind in the shift away from a “virtual” approach that might be valuable components of more modern experiences? Why did the reality of accessible virtual worlds fail to live up to their revolutionary, utopian promise?

In this chapter I will address these questions through an analysis of the properties of virtuality and a pair of design projects: Information Spaces and Presentation Spaces. Although these projects do not
exemplify the best sorts of in situ research, they serve as conceptual arguments about how we might think about designing virtual experiences that take advantage of what virtuality offers designers that is missing in other approaches to mediated interaction. The approaches I will describe contrast with much of the interface design work in virtual worlds in this period that focused on replicating the forms of offline experiences and expecting them to have the same properties when transplanted into a virtual context.

Although in its specifics this argument is perhaps a little dated and unlikely to apply directly in the future (barring a resurgence of virtual world design), it can be seen also as an argument for how to do a close reading of the properties of a platform and letting those properties guide a design that is closely adapted to its tools. Considering virtual worlds in this detail helps highlight some of the taken-for-granted assumptions about non-virtual design.

In a larger sense, a deep discussion of virtuality will also demonstrate how the core research themes of this thesis—grounding, non-verbal actions, and attention—operate in an unfamiliar context. Although readers might be unfamiliar with operating in a virtual world environment, this unfamiliarity will help address these research themes without preconceptions.

I open this chapter with some background on my structural approach to thinking about virtual worlds. I will describe how spatiality, dimensionality, representation, and presence interact with the social experiences that a world can effectively support. I will argue that spatiality and presence are the core differentiating factors that define a virtual world, and the two projects that I describe in this chapter try to make maximum use of those factors. Finally, I close the chapter with a reflection on why virtual worlds have failed to gain widespread traction and on whether the work in this chapter might be effective if re-imagined outside a virtual world platform.

### 3.1 Characterizing Virtual Worlds

While there is no widely agreed upon typology of virtual worlds (though there are many examples, e.g. [Koster, 2007] or [Bartle, 2003]), there are a few general categories that scholars of virtual worlds agree are important to properly describe the kind of worlds being discussed.\(^1\)

The two major axes on which virtual worlds can be organized (as proposed by Bartle [2003] are agency (which Bartle calls ‘change’) and persistence. Agency is a broad term to refer to the set of things that someone can do in a world. Agency can be thought of as the interface that you have onto the world. For example, the agency of

\(^1\) This section draws heavily from [Harry, 2008a].
players in a chess game is limited. When it’s their turn they can move their own pieces in certain prescribed ways. Moving pieces is the agency players have in the chess game. In virtual worlds, agency becomes a bit more elaborate and includes movement, avatar customization, communication, object creation, etc. Different worlds have different sets of actions that an avatar can do, and I describe them as having different kinds of agency.

Persistence is the ability of the world to remember events that change something about the world. For example, if you create an object in a persistent virtual world you can expect that the object will be there when you return. This is in contrast to worlds where changes aren’t saved for very long. Multiplayer, first person, shooter games are a good example of this—you interact with a rich three-dimensional world and might change it by dropping weapons, causing explosions that change the appearance of part of it, or by destroying certain objects in the environment. None of those changes will remain in that world the next time you play, though. It will be wiped clean and you’ll have a fresh copy of the original space. These spaces often even revert to their original state after a few minutes: bodies disappear, dropped weapons fade, and explosion marks are removed.

Certainly, these axes are not perfectly distinct (worlds with limited agency often have low persistence as well), but they serve as good organizing principles. For the most part, modern virtual worlds have high levels of agency but relatively low levels of persistence. In World of Warcraft for instance, players can kill monsters in the world, but the monsters will always reappear a few minutes later. There are virtually no actions players can take that modify any aspect of the world apart from killing computer controlled characters. Rich agency exists almost entirely in the relationships between players and the organizations they form. Worlds like this have proven to be commercially successful because they are more resistant to disruptive behavior aimed at degrading other players’ experiences, but they also rule out many of the interesting opportunities that virtual spaces offer over physical spaces.

The best recent example of a world that offers that kind of rich agency and persistence is Second Life, a world developed by Linden Lab. Second Life is a free application that connects to a single monolithic “Grid” of Second Life servers that provide a mostly continuous (flat) virtual space in which avatars can own land and items, build clothing, buildings, or vehicles, and embed behavioral scripts in their creations. The world also provides an economic system with its own currency system (the Linden Dollar, L$), which is exchangeable at a fixed rate for US Dollars on a currency exchange that Linden Lab manages. The community that has arisen around Second Life is
extraordinarily diverse and rich, but an in depth discussion of its
dynamics is beyond the scope of this chapter. There are a number
of books that describe the history and culture of Second Life, which
provide a good introduction to that topic [Au, 2008, Ludlow and Wal-
lace, 2007]. For my purposes, Second Life is critical as an example of
a world in which avatars have a considerable amount of control over
the design of environments, and so Second Life has been instrument-
al in developing an intuition for how virtual spaces influence the
behavior of people in them. It has also been useful as a platform for
exploring the design space of algorithmic architecture. For much of
this chapter, I will turn to Second Life as a source of inspiration, as
well as to draw comparisons between the general model for virtual
architecture that I explore in this thesis and the model that Second Life
embodies.

The following sections lay out the pieces of what might make
something feel like a world. There is not a clear set of necessary and
sufficient conditions for world-ness. Instead, these sections introduce
a series of concepts that all lend-themselves to fostering a sense of
world-ness. These concepts are Representation, Spatiality, and Pres-
ence.

3.1.1 Appearance and Function

Specific features of both physical spaces and virtual spaces can be
thought of as serving symbolic and functional purposes. As virtual
spaces were first being conceived it was not obvious that they would
draw their symbolism heavily from physical spaces [Novak, 1991].
Novak would no doubt be surprised to see how familiar the design
elements became when the general public had the tools to create their
own spaces. In Second Life, for instance, there are countless recre-
ations of both specific architectural landmarks as well as buildings
that mimic familiar architectural styles. Those buildings in turn are
filled with rooms furnished in a way that would not stand out in the
least from their real world analogs. Why is it that, free of the nat-
ural laws of the physical world, so much of virtual world design is
concerned with recreating familiar physical spaces? 2

This focus on familiar representations serves a number of impor-
tant social roles. First, it functions much like the identity signals do
in physical fashions. Although choosing and furnishing a virtual
house is substantially less costly than its physical analog, it is still
a strong demonstration of taste that helps visitors to virtual spaces
understand something about the person who assembled them, much
like a personal homepage or profile page might on the Web. And
though the price of virtual items may be substantially less than the

4 This phenomenon can be thought of
as a sort of three dimensional skeueu-
ismorphism. I discussed earlier how
physical objects are adapted to signal
the properties of two dimensional user
interfaces. The main difference in this
case is the relative frequency of the
practice; skeuemorphism is the excep-
tion in two dimensional visual design
and frequently results in a critical back-
lash. In Second Life, it was far and away
the dominant design strategy. A casual
user would be hard-pressed to find a
truly “native” design strategy.
physical artifact it mimics, virtual economies usually have some sense of relative value that allow these items to also function as signals of wealth.

The second important social role is in building spaces that contextualize behavior. Visitors to virtual worlds are forced to reform their notions of socially acceptable behavior in virtual spaces. It is substantially easier to understand the intent of a virtual space, and to make meaning from the actions of others within that space if it is built to look like something familiar rather than something abstract. In this way, avatars in virtual spaces can reasonably expect that spaces that look like virtual museums, dance clubs, meeting rooms or houses should be used for virtual analogs of what one might do in their offline equivalent. This argument is analogous to Norman’s, with respect to the design of interactions with physical objects [Norman, 2002]. He describes how physical objects use metaphors to demonstrate affordances. Metaphors imply a conceptual model that makes it easier for people to make deductions about what how their interactions with the system will affect it. In a very similar way, literal representations in virtual architecture serve as behavioral affordances. They use architectural metaphors to imply what the social model of
the space should be.

Although literal representational techniques serve effectively as identity signals and behavioral affordances, this does not mean that they are the only way to do effective design work in virtual worlds. Indeed, the work in this chapter will try to demonstrate an alternate approach.

I will describe spaces which not just look different and imply that different behaviors are expected but spaces that actually have different functional affordances that make certain activities more or less effective in a particular space. Functional affordances are the tangible controls that people have over their own spaces or the ways that spaces react to the presence and behavior of people within them. In a physical context, lighting controls, heating/cooling, and security are all concrete examples of functional affordances; each typically exposes interfaces to building inhabitants (in the form of switches, windows, thermostats, and locks) as well as frequently exhibiting a certain amount of reactivity to the presence and absence of people in a space. A major thrust of the work in this chapter is to imagine what form these affordances might take in a virtual context. In line with the “beyond being there” argument, I aim to treat virtual space as a new medium that has its own strengths and weaknesses instead of trying to create the experience. This distinction is best understood through analogy to how symbolism and function interplay in two kinds of physical spaces: cathedrals and nightclubs.

The form of a classical cathedral is rich with religious symbolism that informs the overall structure of the building, detailed adornments, lighting, and scale. It also has certain functional affordances. The space is designed such that a speaker at the podium can be easily seen and heard by the people in the pews. This also means that the whole congregation will easily hear any noises from the pews. This encourages parishioners to be quiet, and re-enforces the power...
dynamics inherent to the church; visitors are not there to interact with each other.

Nightclubs use acoustics and lighting to create a very different kind of space. Although the precise form of clubs varies, the functional aspects of clubs are often quite similar. Loud music makes it hard to hear people far away, which both forces people to be close together to talk and makes it hard to be overheard. In this environment, it is easy to have intimate conversations. Low lighting makes it difficult to see people, hear them, and identify them. This creates a situation in which people must fill in information about each other because the environment makes that information hard to get. Darkness and candlelight in a cathedral would have a different effect—the functional and cultural/symbolic meanings are interwoven. The nightclub is known to be about hedonism and escape; the cathedral, for the believer, about spirituality, solemnity, and perhaps fearful or awe-inspiring experiences.

Lighting and acoustics are two aspects of what I think of as “functional” aspects of space. They operate mostly independently of how a space looks (that is to say two spaces could look the same but have different acoustics) and both respond to people’s actions in the space as well as mold those actions. These two examples demonstrate how the functional side of spaces has a big impact on what kind of activities make sense in them. You would not, for instance, try to hold a business meeting in a nightclub or hold small group discussions in a cathedral. In a virtual world, however, holding a business meeting in a cathedral would have far fewer negative impacts because the space is simply a visual backdrop, and has few functional affordances. The heart of this chapter is to show how building virtual spaces that exhibit some of these functional affordances that differentiate physical spaces so well might create compelling virtual experiences that are competitive with non-virtual interfaces.
3.1.2 Dimensionality and Spatiality

One question we are frequently asked is why use 3D for a collaboration environment? While it might be possible to build a 2D tool with functionality similar to [Project Wonderland], the spatial layout of the 3D world coupled with the immersive audio provides strong cognitive cues that enhance collaboration. For example, the juxtaposition of avatars in the world coupled with the volume and location of the voices allows people to intuit who they can talk to at any given time. The 3D space provides a natural way to organize multiple, simultaneous conversations. Likewise, the arrangement of the objects within the space provides conversational context. If other avatars are gathering near the entrance to a virtual conference room, it is a good guess that they are about to attend a meeting in that space. It is then natural to talk to those people about the content or timing of the meeting, just as you would if attending a physical meeting. In terms of data sharing, looking at objects together is a natural activity. With the 3D spatial cues, each person can get an immediate sense of what the other collaborators can and cannot see. — Sun Labs, describing the features of a virtual world that made it a compelling space for group work.[Yankelovich]

It’s difficult to pin down exactly what makes a virtual world a world. In the previous section we draw the distinctions of agency and persistence from the literature. These are useful for distinguishing between different sorts of world, but are not sufficient to distinguish worlds from non-worlds. Another attractive potential to draw a distinction between worlds and non-worlds might be a certain immersive representation. When we think about immersion, we typically presume that means a three dimensional representation.

The role of dimensionality in virtual worlds is subtle. Over the course of my work, I have shifted from focusing on three dimensional representations to two dimensional representations. Although this shift leads to a radical change in how people view a space (few people look at a two dimensional web application and say “this is a virtual world!”), it is perhaps not as fundamental a shift in metaphor as you might initially suspect. This section draws a distinction between the dimensionality of a system (which is primarily a representational quality) and the spatial properties of a system (which is primarily an active, functional quality). Although often conflated, these are separate qualities. Both can promote a sense of world-ness, but it is quite possible to have a feeling of occupying a world without a three dimensional representation, provided there is some sort of spatial structure.

The dimensionality of a world is an aspect both of its display and its abstract data representation. Objects in a three-dimensional world have a position in three-dimensional space, a solid volume, and an orientation in three rotational axes. In an intuitive sense, a
three-dimensional virtual world looks a lot like the three-dimensional physical world we are used to. The programmatic representation of the world need not be bound to its visual representation, however. It is possible to build a three-dimensional representation of a fundamentally two-dimensional world. The MASSIVE system demonstrates this nicely; they had both a three-dimensional visual client and a text based two-dimensional client. Both clients could almost completely represent the world state. As a world, MASSIVE had no vertical dimension, so its three-dimensional representation was simply a skin on a two dimensional world. [Greenhalgh and Benford, 1995] What’s important, though, is drawing a distinction between the dimensionality of the world itself and the dimensionality of its visual representation—they need not necessarily be the same thing.

The perspective on a world is an important related aspect of the world that is related to its dimensionality. Three-dimensional worlds have two major perspective options: first person and third person. In a first person perspective, the user views the world as if the camera was placed where the eyes of the avatar would be. If they see their own body at all, it is usually only their hands or feet. In this mode, the user literally inhabits the body of avatar and becomes that character in a significant way. In a third person perspective, the user sees their character from a camera that is usually behind them looking down. This provides a better sense of the world around their character, but can sacrifice some immersion by showing the avatar animating itself or looking different than a user’s own vision of them. The choice of perspective is primarily one of immersion: first person views are more immersive than third person views. In two-dimensional worlds, third person views are essentially the only option. A first-person two-dimensional view would be Flatland [Abbot, 1899], with all of the challenges of navigation and interaction explored in that book.

Spatiality is harder to precisely describe because almost without exception virtual worlds are all spatial in some way or another. MUDs offer perhaps the best starting point as one of the least spatial examples of a virtual world. In a MUD, players occupy discrete rooms. Each room can contain many players and objects, and is connected to other rooms through a series of nominally spatial relationships. For instance, from a given room, you might direct your character to move north which would move your character into the room that the system thinks is north of the room you were in. Although this model is spatial in the sense that you can be closer or farther from people, avatars in early MUDs had little agency or perception of events anywhere but their current room. In a given room, there is no functional spatiality; all players and objects occupy a sort of indis-
tinct space where they could hear and interact with each other, but have no finer position than the room itself. As a result, there was no context for using the kinds of spatial language that make spatiality so useful. A player couldn’t describe an object as being “the thing on your right.”

Furthermore, the connections between rooms themselves were not reliably spatial in any particular way. Although they were ostensibly arranged in cardinal directions, there is no enforcement of “normal” spatiality. A series of rooms could easily fold back on itself such that moving north a few times would return you to the room you started in. Different routes out of a single room might all go to the same room. Paths might even behave differently in different directions; moving north from one room to another, and then south to try to get back again might not necessarily take you back to where you started. In this way, even an ostensibly spatial metaphor breaks down and fails to convey the contextual and perceptual benefits of true spatiality. In contrast, a world where objects and avatars have distinct discrete locations immediately confers these benefits. Avatars can indicate group membership by avatar proximity, can have a shared visual reference point, and so can communicate about objects behind or to the right of other avatars.

Returning to the quote about Project Wonderland that introduced this section, I argue that it conflates notions of dimensionality with spatiality. All of the beneficial features that are ascribed to a “3D world” are more properly ascribed to a world with rich spatiality. A two-dimensional virtual conference room can have an entrance where avatars congregate just as easily as a three-dimensional world. Two-dimensional worlds confer the same benefits regarding shared gaze, too. An avatar in a two dimensional world can infer another avatar’s view on that world in the same way they might in a three-dimensional world. These are all properties of a world’s spatiality and not its dimensionality. Although the work I describe in this chapter is two dimensional, I have maintained spatiality wherever possible. This maintains many of the benefits described in the quote while avoiding the many challenges of working in three dimensions. Translating this design approach into three dimensions is certainly possible, although maintaining spatiality in three dimensions requires a certain vigilance. Second Life is an instructive example here. Avatars in Second Life freely move their cameras with no external representation of current location. As a result, an avatar’s position in the space has little to do with their current view, and so spatial language isn’t necessarily that useful. Chapter 4 demonstrates how three-dimensional spaces can be used with this same approach, and similar analogs could be built for essentially all the zones and ideas
While I believe that worlds that are fundamentally two-dimensional are not inherently less spatial than their three-dimensional counterparts, there is something to be said for the representational language (as opposed to the functional or algorithmic language) of three-dimensional spaces. As discussed in the previous section, three-dimensional spaces tend to offer more legible spaces because they use a representational language that is familiar. Meeting rooms represented in three dimensions (or even some sort of isometric view) may be more obviously meeting rooms than the very abstract vector-graphics style rooms I show in this chapter. For my purposes, though, the dimensionality is not particularly important for demonstrating the design space. Instead, it is spatiality that is critically important for creating a sense of being in a world.

### 3.1.3 Representation and Presence

The most visible distinction between world-like experiences and non-world like experiences is the way that people are represented. In a system like Second Life, people are represented as quasi-realistic avatars. Not only do they look like people, they move and interact like people: they shift stance, walk, run, wave, and dance in more or less realistic ways. When these representations are combined with literally-designed spaces (that is, spaces that look like non-virtual spaces) it all starts to feel quite familiar and world-like.

This is clearly an attractive approach; it seems natural to expect that a software world that looks like the real world that is populated by people who move look and move like real people that it will be valuable. We’ll defer a deeper discussion of why this is a bad assumption. Instead, I seek to place avatars on a continuum of representation strategies, any of which can support a sense of a sense of presence. This argument is analogous to the dimensionality and spatiality argument made in the previous section; spatiality is where the world-ness comes from, and you can create spatial experiences with a variety of dimensional choices. The same is true for representing people. What’s important is that you foster a sense of presence, not that a particular representation is inherently superior.

There are a wide variety of representational strategies that focus on different aspects of someone’s identity. Pseudonyms are widely used for their simplicity, but are nevertheless quite expressive. A name alone can paint an evocative picture of a person, although it might not mirror their offline identity presentation. [Jacobson, 1996] More elaborate representations can accumulate behavior in some social context like number of posts, number of friends, age in a
community, or metrics representing the quality of one’s contributions to the community. Image-based avatars like those on Facebook and Twitter add a more flexible visual vocabulary for representation that easily supports both photos that leverage offline representational strategies as well as more abstract and conceptual representations. Finally, pure self-description is quite common as well. Whether the third-person description style of MUDs, a one line “bio” on Twitter, or the full suite of different tools that make up a profile page on Facebook, online representations frequently employ text in a variety of ways. The properties of each of these strategies vary, but they all provide non-literal ways of creating a sense of identity online.

Presence, however, is a different matter and is decoupled from representational strategy. We would probably not describe Facebook as a virtual world, not because its representations of people are a combination of text and photos, but because there is little sense of being in a space with someone. We are aware only of someone’s actions, not their presence. Most asynchronous-oriented online experiences share this character. The space is well represented and we see the traces of someone’s actions after they happen, but we don’t see the presence of non-acting people in the space. Simply indicating in some way that other people are viewing the same document (as in Google Docs) can be enough to create a sense of presence, even in the absence of a world more complicated than a simple text document.

Although we can create a sense of presence with or without a realistic approach to representation, it is important not to understate the impact of representational choices. In a world that uses human-like, realistic representations and encourages people to identify with them through customized appearances, we have to sacrifice some of our flexibility in terms of the ways a designer can manipulate that representation. While it would be acceptable to resize, distort, re-color, and move a non-realistic representation of someone (like a pseudonym, simple shape, or profile picture) in service of a design goal, those sorts of manipulations of carefully controlled human-like representations would be viewed as highly invasive. By way of analogy, consider the way players of a board game interact with pieces representing other players. Within a set of rules, certain manipulations can be handled by any player. In contrast, situations where we are permitted to actually manipulate the bodies and appearances of others are quite rare. Even in a sporting context with rules that mandate actions and appearance, we rely on sanctions and penalties, not physical coercion because we recognize a distinct level of autonomy of the bodies of others. The obvious exception to this is high contact sports, but most sports have significant limits to the ways that bodies can directly interact. Although this autonomy is not quite as high in
virtual spaces, it still exists to a large extent and constraints the sorts of experiences we can create in virtual environments.

There are some notable potential counter-examples to this argument. Many games represent a player as a human character and allow other players to impact that representation, usually through violence. In this case, players tend to view their avatars less as extensions of themselves and more as pieces on a game board. Although being harmed by others is not desirable, it is still within a clear framework of rules. In non-game worlds like Second Life, the social framework of body manipulation looks much more like that of high-contact physical sports. Contact is acceptable but only in particular ways and for particular reasons.

3.2 Information Spaces

As a first step in exploring the potential of virtual architecture, I developed a space in Second Life that focuses on the social meaning that an avatar’s position in a space can have and how that meaning can be augmented using some of the aggregative properties of virtual space. [Harry, 2008b] This particular design is focused on meeting situations. In meetings with more than a few people, it can be challenging to understand other people’s feelings about an issue, reach consensus, and influence others. In McGrath’s [1984] terms, I’m focused on decision-making tasks, with an added interest in supporting some sorts of planning tasks like staying on an agenda and distributing tasks. The design addresses these collaboration challenges by creating a meeting space focused on non-verbal signaling using avatar positions. The main stage in this space is a chat or audio conversation, while the side stage is the positions of avatars and associated visualizations. The goal of this side stage is to help manage the decision-making process by creating visualizations that make magnify and make persistent avatar movement within the meeting space in a way that encourages participants to use their avatar’s position in a more performative fashion. To support this use, I built a range of social utilities—systems that make visible properties of avatars’ social behavior in the space. These visualizations are all controlled by a centralized dashboard, such that the nature of the space can be controlled by the moderator, much like a meeting organizer might set up the chairs and projector in a physical meeting room depending on the kind of meeting they were having. An overview of the meeting space can be found in Figure 3.5.

Beyond the space itself, I also developed a number of meeting support widgets that can augment a virtual meeting room. These widgets are not strictly architectural, but they show how different kinds

3 This autonomy is embedded in the design of Second Life itself. Objects attached to an avatar (usually taking the visual form of clothing or jewelry) have very limited ways to change the appearance of the avatar they’re attached to (beyond their own form) and have very limited ways to act on that avatar’s behalf. Furthermore, objects that aren’t attached to someone’s avatar have essentially no way to impact that avatar.

4 This section draws heavily from [Harry, 2008a].
of tools could create virtual spaces that are more or less appropriate for a certain kind of meeting.

Core to the aesthetics of this space is a belief that virtual worlds have distinctly different properties from the physical world. In this design, I show new ways for people to represent their attitudes and emotions in a mode that is easily effected and understood by others. At the same time, I seek to remind participants on the space of its differences from traditional virtual meeting spaces through a visual design that is strikingly abstract. The presentation and the function combine to guide the interactions of meeting participants with contextual cues that are otherwise missing in a virtual world context.

3.2.1 Space Design

The space is divided into four major zones. The main area is like a traditional sports field with end zones labeled “agree” and “dis-
agree”. It provides a space for people to position their avatars on a continuum to show their attitudes about the issue under discussion. The fluid self-arrangement of people within this space based on their opinions provides a literal basis for seeing where someone is coming from and the status of the group’s attempt to reach consensus. A single-axis continuum is used so that multiple people can easily be at the same continuum position without displacing each other.

Of course, not everyone always wants to reveal their opinion about the issue at hand. Surrounding the main agree/disagree field is an area for people to stand who want to participate in the discussion without putting their avatar on the continuum. Still further from the field is an observation area for people who want to be present, but not participating. Finally, there is a platform for the moderator with controls to manage properties about the space itself. This layout is shown in figure 3.8. For the sake of simplicity, most examples in this chapter will focus on the Agree/Disagree continuum, but the implications of other floor types (for instance a process oriented Keep Talking/Move On field) will be discussed later. This spatial approach is a powerful organizational metaphor because it both relies on our knowledge of the meanings of position relative to other people in physical spaces [Yee et al., 2007] and the metaphors inherent to interactions in a spatial environment. [Lakoff and Johnson, 1980]

The Agree/Disagree floor is only one potential floor design. Although it makes for a good thought experiment and it will be used for much of this section, its actual utility is limited. Few meetings are sufficiently organized for this approach to be effective. Even with a strong moderator who makes clear exactly what issue avatar’s positions are agreeing or disagreeing with, it can at times be ambiguous. There are a number of other options for floor designs that are more generally useful.
Figure 3.5: An isometric overview of the Information Space with all the visualization components turned on.
The primary alternative is a Keep Talking/Move On floor. This floor addresses the need for meeting participants to be able to express when they think the group as a whole should change to a new discussion topic. In much the same way that Robert’s Rules of Order [Robert III et al., 2000] is a process that is focused on how time is spent, using the Keep Talking/Move On floor is a way of actively expressing your desire to move on or keep talking about an issue and for others to passively judge the interests of the group, even if a more formal structure is not in place. This is a generic issue, and is one that can be particularly hard to resolve in medium-sized groups. Participants might not want to speak up and say that they want to move on, but if they had a non-verbal way to express their desire to move on (and gauge others’ reactions) meetings might spend less time on topics that they didn’t need to.

Although these are the only two floor designs that are currently implemented, there is potential in other kinds of floors. Floors with a few multiple choice options make it easy to hold straw polls. A calendar display could be used to show preferred dates for an event. Floors could be used for setting up speaking queues, as well, dividing the audience between those who are waiting to say something, those who are speaking, and those who don’t have anything to say. Multi-variate floors are also possible, although there isn’t a nice way to let multiple avatars stand at the same point in the space in the same way that you can in single variable floors like Agree/Disagree.

Because the space is divided into socially meaningful regions, the most important information we can visualize about someone’s position is how long they’ve been somewhere and where they last came from. When an avatar pauses for a while, a transparent column will slowly rise out of the ground. I call these columns “dwell indicators.” If they move, the column will slowly shrink and eventually disappear. In this way, avatars leave a temporary mark on the space with their presence, and other people can use this signal to better understand what their position means. Someone who has been standing on the agree side of the field for the entire discussion is quite different from someone who just arrived there; this approach distinguishes those meanings visually. A dwell indicator is shown in figure 3.9.

As an avatar moves around this space, their path is drawn behind them. This helps meeting participants understand how an avatar arrived at their current position. In particular, it fits well with the dwell indicators; when an avatar who has been standing somewhere for a while (as shown by the dwell indicator) moves to a new position, their old dwell indicator will start to shrink and a line is drawn between their old position to their new position. This connects the
two dwell indicators and shows how long the avatar was in their previous position, and how recently they moved. In the case of the agree/disagree floor, this would show that someone who perhaps had long been against a particular proposal had changed their mind. This is the kind of behavior that this space seeks to visually emphasize because it’s socially meaningful but is otherwise hard to convey non-verbally. A position history trace is shown in figure 3.10.

The space also records typed contributions in text boxes that appear above the speaker’s head and then rise slowly. This creates a visualization of chat over the course of the meeting, displaying what was said, when it was said, and where in the room the avatar was when they said it [DiMicco et al., 2007]. In aggregate, it also shows the patterns of conversation of the course of the meeting. Much like the conversation visualization work discussed in earlier chapters, this allows participants to be self-reflective about the dynamics of the meeting as it occurs. By making objects appear in the space itself, it foregrounds these issues and makes imbalances in participation harder to avoid. Because the boxes contain the text itself, it is also possible for participants to zoom in on text messages and view the conversation within its spatial context—which is lost in a text-only transcript of the meeting.

Finally, the floor of the agree/disagree field displays the current average vote (by moving side to side along the floor), as well as its deviation (by growing wider or thinner). Like chat messages, a representation of the group’s collective view also floats up into the sky. These bars provide more context about the overall feeling of the avatars in the space over time, showing aggregate views of avatar movement. Furthermore, the history of the average position separates historical chat messages based on which side of the average the message came from. This helps contextualize the chat messages as well. For instance, it is easy to tell if a talkative participant was way out of line relative to the group consensus.

3.2.2 Discussion Tools

During the process of developing the Agree/Disagree space, it became clear that there were a number of other common events in meetings that could be easily supported with applications. From an application design perspective, the main benefit of working in a virtual world is that it’s easy to export information into shared conceptual space of the meeting room. This is a common practice in face-to-face meetings in the form of slideshows, keeping live agendas on a projected screen. It’s particularly prevalent in group design practice, in which a fluid and grounded conversation about ideas is
virtual spaces It’s easy to facilitate this process by creating objects in the virtual world that hold information. The kinds of information that are easily represented in Second Life and the challenges of working without any familiar UI widgets or metaphors limit the overall efficacy of these applications, but they are interesting first steps in building specific objects that offer functionality that can further customize the meeting room experience. Unlike in physical rooms, which are largely limited by the number of projectors available to augment the space with information, virtual spaces have many more opportunities to customize the experience.

Often, meetings involve some sort of distribution of tasks. This process is usually an elaborate (and sometimes unspoken) negotiation between who has time to do a task, is interested in getting it done, and the skills necessary to complete it. This application addresses availability. When a task is going to be assigned, the meeting moderator can press a button on meeting dashboard and a small pyramid will appear. (Figure 3.13) Text can be stored in the pyramid by typing ‘/todo’ followed by the text of the task. Anyone can then click on the task to claim it. Claimed tasks spin slowly above the head of the claimant. As an avatar with claimed tasks moves around, the tasks follow above their head. Tasks also have buttons on them, allowing the owner to release them (so they can be claimed by someone else) or delete them outright. The vision for this particular application is that tasks would also be exportable from within Second Life to non-virtual world tools where people tend to manage their task lists.

The task objects serve as visual reminders of who in the group has already accepted tasks and who hasn’t. Much like visualizing chat can encourage participation from people who are participating below average, this can serve a similar role. Task objects also operate metaphorically - by staying above an avatar’s head, they rely on the metaphor of tasks “hanging over us.” [Lakoff and Johnson, 1980] This kind of allusion doesn’t work nearly as well face to face. Although you could hand out note cards with tasks on them, it would not have the same metaphorical resonance as the virtual approach does.

Having a representation of the agenda in a meeting space can be a helpful way to both remind people of the what the current status of the meeting is. (Figure 3.14) This tool takes a note card object in Second Life and represents it as a hierarchical listing in the meeting room. The moderator can move a pointer between agenda items to indicate which one is currently under discussion. The system also has a voting mode, in which agenda items also act as buttons that accumulate votes. By default, the voting system operates using an

valuable. [Dwyer and Suthers, 2005]
approval voting in which a single avatar can vote for as many options as they want, but cannot vote for a single option more than once. As votes accumulate, the color of the item changes to quickly show which items are the most popular. In the spirit of configurable spaces, there are a number of ways the voting system can be reconfigured. Votes can either accumulate secretly until the end of the vote, or be shown as they are received. Voting can be approval or traditional first-past-the-post. Ballots can either be public or private. Depending on the kind of vote being held, different configurations would be appropriate. Although this doesn’t offer nearly the flexibility of a system like Selectricity [Hill, 2006], embedding the voting in the virtual space serves as a visual reminder to past votes.

The dashboard is the heart of the space (Figure 3.15), and contains a range of controls that customize the social experience of being in the space. Each of the social utilities described above can be turned on or off from the dashboard. The texture of the floor itself can also be changed from Agree/Disagree to an other design. The applications are also sometimes connected to the dashboard. The podium that contains the dashboard is also elevated from the floor itself, which reinforces using spatial metaphors the relative roles of the participants—the avatar in that position is understood to have more control over the space.

In a virtual space where avatars can easily act at a distance, the dashboard stands out as an anomaly; only avatars standing within the gray podium (see Figure 3.5) can push buttons on the podium. This ensures clear attribution to changes in the state of the meeting space. There will never be any question who turned off chat archiving—it had to be the person standing in the podium. From our experiences in physical spaces, we are used to people needing to be proximate to, for instance, light switches to change the lighting in a room. Although this breaks many of the expectations of Second Life users, I believe that enforcing avatar presence near the buttons is an important social cue that helps the legibility of roles and abilities in the space.

3.2.3 Deployments

We have conducted evaluations of this approach. We worked with two existing discussion groups in Second Life to arrange meetings inside an early prototype design. The first group had 7 participants and the second group had 12 participants. Each group was given a list of controversial topics and asked to vote on which one they were most interested in discussing. After choosing a specific topic, they were given an editorial on that topic and then asked to move
their avatar to represent whether they agreed or disagreed with the author’s argument and to hold a discussion about their opinions on the topic. After the discussion, feedback about the meeting space was collected informally through both individual conversations with participants and group comments.
Figure 3.8: A top-down view of the space, to illustrate its different zones.
Figure 3.9: An avatar with a dwell indicator growing slowly at its feet.
Figure 3.10: An avatar after moving from the agree side to the disagree side. The position history shows the avatar’s path, and is colored according to the area in the space they were in.
Figure 3.11: Chat messages rising above the head of the speaker.
Figure 3.12: Left: A view of the floor with a line showing the current average vote. Right: In the space above the meeting room, long bars float slowly up to represent the history of the average vote over time.

Figure 3.13: A todo pyramid floating above an avatar’s head. Buttons below the pyramid provide a mechanism for deleting the todo or releasing it so someone else can claim it.
Figure 3.14: The agenda system, with buttons for manipulating all items as well as deleting individual items. The text for the agenda is pulled from note cards dropped on the agenda object.

Figure 3.15: The dashboard. Left: A button for making new todo objects. Right: Buttons for configuring the space.
Participants were, on the whole, very excited about the potential of the design space. Many of them focused on the social implications of this arrangement, thinking out loud about how they would respond to, for example, their boss moving to “agree.” They also appreciated the way the space aggregates history, because they were often pulled away from Second Life briefly and lost track of what was going on.

In these tests, though, participants rarely moved their avatars around the space to take advantage of the many visualization features. This was primarily due to the modality of the Second Life interface. When we ran these studies, Second Life did not support voice chat (it has since been added), and avatar movement is controlled by the keyboard. Switching between movement mode and typing mode is a little bit confusing. This makes it hard for avatars to both move and talk at the same time.

The other major theme from these experiences was that even relatively sophisticated Second Life users make infrequent use of the elaborate camera controls. As designers, we assumed that all users would be comfortable detaching the camera from their own avatar and inspecting objects in the environment—the conversation archive, in particular. This was not the case with our test groups. Many reported having forgotten about the above-their-heads information because they didn’t normally change their camera view. They also reported that their avatar movement was often used to change their camera view, and so making that movement socially significant was sometimes problematic. Instead of visualizing their feelings about the topic, sometimes the space ended up visualizing their attempts to get a better view of the space without using the advanced camera controls.

As with all domains, applications that challenge users’ ideas about what’s possible can be hard to understand. In the case of Second Life, this is compounded by user interface issues. This leaves us with a number of possible explanations for users’ excitement about the ideas behind Information Spaces but general lack of engagement with specific features. I think the most probable is that because this project reflects both a shift in how to interact with spaces in Second Life as well as forcing users to use the interface in ways they might not be familiar, it would require a longer term study to see effects than we were able to conduct. Perhaps with a user group that was even more familiar with Second Life and could spend more time acclimating to the environment we could better address aspects of the design in more detail.
3.2.4 Analysis

I propose that there are quite a few different ways in which this design influences the behavior of avatars. As in any meeting situation, the space itself is only part of the picture, and so while all these influences have not necessarily been observed experimentally, it is worth thinking through what the range of possible impacts could be.

One theme common to many of the features of Information Spaces is an emphasis on encouraging more equal participation by showing relative participation levels of participants. Over-talkative meeting participants are reminded of their activity, and people participating less can find space to get their say. This can work on a group level, too. It may be clear by the current average vote in the room that the consensus is squarely on one side, but a vocal minority is out-talking a quiet majority. By color-coding chat messages based on where in the spectrum they come from, the for/against balance can be equalized. Compared to the approach employed by DiMicco et al. [2007], using shared displays to show relative participation, our approach has a number of benefits. First, mediated participation is far easier to instrument to generate the sorts of graphs in *Second Messenger*. Second, virtual worlds make it easy to render the data in-place, rather than leaving it to users to locate themselves in a displaced representation and map it back onto people. We cannot make a strong argument that this would change the findings from DiMicco, but it is conceptually appealing.

There is an inherent bias in this approach towards meetings in which participants are trying to find better ways to non-verbally express their attitudes. The whole space is focused on making these signals more legible and visible. This is an approach shared by most side stage designs. If the main stage is focused on one speaker at a time, a side stage that supports multiple simultaneous performances can be particularly useful. This does not preclude main stage performances, of course. For participants who are comfortable breaking in on the main stage, it can be a more efficient way to make a point or represent your feelings on an issue. Yet many participants are not always empowered to perform on the main stage, or don’t want to wait to take a main stage turn when they can perform immediately on the side stage.

The challenge with this approach is that we are assuming that there is an interest or support for this sort of non-verbal communication. In contexts where the sorts of performances the side stage is designed for are not valued contributions (e.g. in situations where there is a stark power differential within the group), it may be that making these signals more visible is in fact a reason for a participant
to be even less open about their feelings. This highlights a limitation in *Second Life*—any system that relies on an avatar’s location makes it difficult to anonymize their participation. If side stage performances could be anonymized, they could be a powerful way to balance otherwise imbalanced power relationships. Anonymized participation is a common feature in GDSS systems that is difficult to replicate in a user-friendly way in *Second Life*.

If there is not group interest in equal participation, it may be that this kind of space would reinforce hierarchies instead of subverting them. While in a face-to-face meeting, substantial effort might be spent trying to figure out what an important decision maker feels about an issue, in this environment they may just move right to that position and everyone else could follow them there, obviating any real discussion of the issues. It may be that if this is really how meetings work, then quickly acknowledging that everyone wants to follow the boss might not be a bad outcome. This effect is limited by the informality of the space. No one is required to indicate a position at any time; instead changes in position (from undecided to decided, for instance) can be done strategically. The system is not designed to be coercive, simply to create a new stage for performances if people want to use it.

By making visible someone’s ongoing attitude about something, meeting participants might be discouraged from moving because dwelling somewhere for a while becomes seen as a status symbol. I suspect that in this situation, the dwell indicator is only making visible pre-existing attitudes about changing ones opinion. In a meeting environment where opinion changes are valued, the bias could just as well be the opposite; staying somewhere too long looks like inactivity and disconnection because the group expectation is that participants will be actively representing and changing their opinion. Participants’ use of avatar movement in place of camera control exacerbates this issue; a user that wants to look around will have to give up their dwell indicator, which might cause users who become attached to their dwell indicators to simply not move at all, limiting the impact of the other visualization systems spread around the space.

### 3.3 Presentation Spaces

For many people, virtual worlds like *Second Life* offered a chance to be part of a broader intellectual life that might not be available to them offline. Although it has faded somewhat, there was a period where lectures and presentations in *Second Life* were reasonably common and provided a platform for people to find interesting social and intellectual contexts without leaving their house. In many

---

5 Of course, adoption in tools like this is always a challenge and often requires the buy-in of high status members of an organization. Tools that aim to disempower or level status among a group can be viewed as threatening. This can lead to tools not being introduced at all, or their introduction being systematically undermined. [Orlikowski, 1992]

6 This presumes that an avatar’s identity is what matters. If the mapping between avatar and keyboardist is not made clear, anonymity is manageable. But fluid systems where you can flexibly claim anonymity sometimes but not others (like *Quora*) are difficult to create.
ways, this was always the strongest argument for virtual worlds. While they ultimately had a difficult time competing with enterprise collaboration software, Second Life in particular had a vibrant community around these sorts of events. Like meeting rooms, the spaces in which these events took place were overwhelmingly literal in their design. The speaker typically gave a presentation either through streaming video or (later) used the in-world voice system. Slides were sometimes used, either in a video stream or embedded in the world and managed using in-world tools.

In this section, I describe Presentation Spaces—a system to support these sorts of presentation activities in a virtual world. I take as my starting place the set of challenges laid out by Yankelovich and Kaplan [2008] (with some adjustments):

**Static Presenters** Managing your avatar in a way that’s convincing and demonstrates engagement while also handling the usual challenges of presenting is difficult.

**Dynamic Audiences** In face-to-face presentations, our movement is influenced by social norms about having to walk through other people, create noise, etc. In virtual worlds, these tend not to be meaningful limits so people move around much more in ways that is not socially significant but can be quite distracting.

**Communication Confusion** How can audience members talk to each other? How does that scale up to lots of people? How can the audience communicate to the speaker in a meaningful/useful

---

7 The only real concession to the virtuality of the environment was that each “sim” in Second Life had a population limit, so the best event spaces were located on the corner where four sims met, to maximize the population. You can see across sim boundaries, but not talk.
way?

Usability Virtual worlds make it easy to see people, but building applications in them can be very difficult; managing HUDs/UIs/cameras is typically quite difficult.

Many of these problems are shared by non-virtual mediated presentations as well. Tools like WebEx suffer from many of these issues, too. Although it might be nice to have better audience/presenter interactions, a virtual world context forces these issues. Watching an un-moving avatar representing someone giving a presentation as an audience member is far more disconcerting than watching a slide deck in WebEx because there is no ostentatiously awkward representation of that person that they’re attending to. The same is true of the audience. In a traditional conferencing context the audience is essentially invisible, so their absence or activity doesn’t get in the way of anyone else’s experience. Representing that audience graphically raises the stakes and means we need to either develop ways to incentivize certain sorts of behaviors and discourage others, or think about alternative representations of their activities.

Presentation Spaces seeks to use spatiality to address these issues in an internally consistent way. Instead of assuming a static space, we use the slides of the presentation itself as the primary organizing feature of the space. In an otherwise flat and featureless space, the slides are spread out linearly. The slides form a space that is defined by the slides, and people’s positions relative to the slides can become socially meaningful. As in the Information Spaces project, there are a series of independent tools that help manage issues like communication confusion, audience movement, and audience/presenter communication.

Using the stages terminology, the configuration here is quite similar to Information Spaces. The main stage includes all the components of the space that are primarily the domain of the presenter: audio content, the slides themselves, and the visual field at the front of the space. The side stage includes all the tools that support non-verbal communication on the part of the audience. Although the semantics of an avatar’s position in the space is still a core component of side stage performances, the way those positions are interpreted is quite different from how position and movement were interpreted in Information Spaces. The sections that follow will show how the structures in Presentation Spaces create a different framework. But the projects share an interest in creating predominantly non-verbal side stages that can enhance the main stage by creating feedback loops in contexts where verbal feedback would be unwieldy.

This system was built on the Wonderland platform [Kaplan and
Yankelovich, 2011] originally developed at Sun Labs. Although there are significant technical differences between Wonderland and Second Life, they are broadly similar: they present people with avatars, they support spatialized voice communication, and objects in the world can respond to user input in different ways. The two primary features of Wonderland that differentiate it from Second Life are its federated, open source server model (compared to Second Life’s monolithic, closed source server model) and its broader programmatic access for system developers to more of the internal behaviors of the world (which Second Life eschews for security reasons).

3.3.1 Design

Presentation Spaces has four major components: slide spreader, a moving platform, chat zones, and thought bubbles. A new presentation space is created by dropping a PDF into a Wonderland world. Slides can be spread in a variety of different ways, and the size and spacing of the slides is controllable. The slides form the backbone of the space, giving it its dominant visual feature. By making all the slides in the presentation immediately visible to all audience members, it becomes instantly clear how long the presentation is going to be and gives an avatar at a certain point in the presentation a sense of context about recent and future slides in a natural way. This sort of use of space is not feasible in Second Life, where land is sold by the square meter and this sort of use of space would be extraordinarily indulgent. 8

The implicit model with this sort of layout is that the presenter and audience will move from slide to slide over the course of the presentation. Instead of the audience staying stationary and slides changing in front of them, it is the slides that stay stationary and the audience which moves. Think of it like a gallery tour, instead of a slideshow. One of the challenges with organizing a presentation space in this way is the increased navigational burden on audience members. If every slide change required the entire audience to move itself from slide to slide, it would turn changing the slide into an elaborate procedure. You would, however, always know who was present and paying attention because inattentive viewers would rapidly be left behind as the group moves on. But in an experience that is already demanding quite a bit of extra attention from both audience member and presenter, this would likely be overwhelming.

To help automate the movement process, there is a sliding platform that can move audience members from slide to slide. The platform is controlled by the presenter, so functionally it works just like changing slides. 9 Audience members on the platform will move nat-

8 This is one of the many ways that design decisions about a world can have a major impact on architectural and social practices; builders in Second Life frequently build tall instead of wide to manage real estate prices, which exacerbates people’s issues navigating and moving in complicated three-dimensional spaces with weak camera controls.

9 We could, of course, move the slides and keep the audience stationary. This is analogous to the camera/document distinction in Pad++ [Bederson, 1998]. In this case, the disconcerting effect of moving graphically enormous and solid slides is a strong reason to move the audience instead. Furthermore, moving the slides would make it harder to move among the slides when not on the platform—the slides could move out from under you in a moment.
urally from slide to slide, maintaining their view of the current slide, with a glimpse of the next and previous slides. The presenter maintains her perspective too, using an automatic custom camera angle that maintains a view of the audience and of the current, next, and previous slides. Audience members can freely move on and off the platform to more closely inspect past or future slides, or simply to signal that they’ve stepped away from their computers for a moment as in the border area in Information Spaces.

One of the benefits of face-to-face interaction is the ability to carefully moderate your own speaking volume to easily address groups of different sizes. Being able to whisper to a friend sitting next to you at a talk is part of what motivates us to attend events together and sit together. In virtual spaces, this flexibility is difficult to foster. Chat is often spatialized (e.g. people near you can see your chat message, but people farther away cannot), but it’s quite difficult to have any intuition about whether someone will be able to hear you or not just looking at their relative position. In physical spaces this is mitigated by being able to see that someone is talking but not be able to hear them. This helps us make judgements about the audience for our own spoken comments. ChatCircles handled this quite elegantly by representing speaking activity of others at a distance without making the actual content of the communication available. [Viegas and Donath, 1999]

In Presentation Spaces, we take a cue from ChatCircles, but because of some limitations of the system we render the “chat circles” in a much more literal fashion: as actual circles that avatars can step into
and out of. This mimics an approach first seen in the game world *Puzzle Pirates*, shown in Figure 3.23. Using a button on the *Presentation Spaces* HUD interface (seen in the upper right hand corner of Figure 3.20), any user can create a chat zone underneath their avatar. A new chat zone starts small, but increases its radius whenever anyone walks into it. All zones are public, and can be arranged hierarchically (e.g. zones within zones) or in a disjoint fashion (e.g. partially overlapping). Stepping into a chat zone adds a new zone-specific chat channel to the main chat system. Stepping out of a zone doesn’t remove the tab, but it does remove your ability to listen to new messages sent to that group since you left it.

This sort of interaction is possible without relying on a spatial organization. A simple instant messenger-like experience with invite-only group chat makes it possible to functionally group people and route chat messages in the same way. But this approach suffers from discoverability problems: how can you tell what conversations are going on if they take place invisibly? By linking the distribution of messages to a spatial property like avatars’ positions, we make the existence of side conversations visible and discoverable. Using position also organizes the audience in a more sensible way. Instead of avatar position in the audience being arbitrary, an audience using chat zones will be a more legible social space with chat zone boundaries illustrating the broad strokes of relationships in the audience.

Chat is a useful mechanism for small groups to interact, but it has two major limitations. It does not persist in a useful way (nor would you want it to, since chat messages are written for a particular audience, not for anybody who comes to the space later), and it is too high frequency for a presenter to keep track of. The thought bubbles system is a way to write messages explicitly to be left behind in a specific space. These messages can be thought of a little more like a tweet in *Twitter* and less like an instant message or chat message. Binding them to a space is another way of using the slides to create contexts for conversation asynchronously. The spread out slides serve as a site for ongoing conversation, both synchronous through chat and asynchronous through thought bubbles. Because thought bubbles are composed in a different way than chat and are a lower-frequency channel, they also become a viable way for different chat zones to interact across zone boundaries as well as for the audience and presenter to interact. Although not fully fleshed out, there’s the potential for systems like backchan.nl (described in Chapter 4) to be re-invented in a virtual world context using spatiality and motion to convey the audience’s activity in a more dynamic way.
3.4 Grounding, Actions, and Attention

For most of the projects described in this thesis, there is a necessary transposition between the visualization and the world. We are usually limited, in physical spaces, to using projections or screens that require translation between the digital representations of people and any sort of visualization of their behavior. If we take for granted that people identify with the avatars of themselves and other as essentially extensions of their physical body, then being able to do in-place visualization seems like a powerful tool. In grounding terms, the appearance and behavior of others is one of the strongest taken-for-granted assumptions in face-to-face conversations; it automatically meets the criteria of being mutually and recursively understood. We may not necessarily parse its content in precisely the same way, but it is clearly a part of someone’s appearance and an automatic part of common ground.

In our deployments of both Information Spaces and Presentation Spaces we saw little evidence that these representations were a subject of conversation. By analogy, though, it is rare in a group situation to make more than a brief passing comment about someone else’s appearance. It may be distinctive and meaningful, but being part of the common ground doesn’t mean it’s the constant subject of conversation. In fact, the opposite is more likely true; the lack of conversation about people’s locations the visualizations based on those locations could also be read as evidence that it didn’t need to be discussed. It was mutually understood that you had been standing somewhere for a certain amount of time or held a certain opinion. A conversation about these states signals confusion about their meaning or significance.

Although these sorts of visualization strategies are difficult in physical spaces, they are not impossible. The artist Lozano-Hemmer’s project Subtitled Public [Lozano-Hemmer, 2005] projects phrases on people’s bodies, and allows people to transfer these phrases between them by touching other people. This is made possible through an elaborate technical infrastructure of cameras tracking people in the square and a network of powerful projectors. One of the major benefits of virtual world work is that it is possible to conceptualize and experiment with embodied interfaces in a much more rapid and low-cost way. If we accept that this sort of technique is useful, it would be interesting to see how these approaches could be extended in physical spaces. Moving around is clearly more problematic, but providing non-verbal ways to represent your attitudes might be feasible, particularly in already-mediated heterogeneous remote meetings with some local participants and some remote participants.
Given that the primary mode of experiencing virtual worlds is moving around, movement seemed like a natural fit for the primary non-verbal action. Users could be expected to already understand the movement controls. Furthermore, relative avatar movements and locations are meaningful offline social signals. Although avatars may maintain interpersonal distances per [Yee et al., 2007], that does not mean that movement through the world operates precisely the same way. With our bodies, we can naturally stay in one place while looking around us. This is possible in Second Life, but not an interaction that many people found natural. Instead, most residents would simply spin their entire avatar around to look at something behind them and treat their position in the world primarily as a way of setting up their view on the world in a particular way. These two competing desires—to manage their view through their position and to express their attitude about the topic under discussion—were not easily reconciled. By trying to map a pre-existing action onto some new set of meanings, the new meanings faltered and older behaviors tended to take over. This made interpreting movements extremely challenging. Did someone actually change their mind or were they trying to look at one of the active visualizations? This was alleviated somewhat in Presentation Spaces with the more elaborate camera controls available in Wonderland, but it is also difficult from a usability perspective to be shifting camera angles and expecting people to adapt.

In hindsight, this conflict is essentially irresolvable barring widespread education about advanced camera controls. And yet, if we are to give up on position as a primary non-verbal action, there are few other options. The “touch” action in Second Life is widely used, but suffers from a complete lack of action attribution because avatars can invisibly touch any object in a large radius around themselves. This can be useful for creating opportunities for nearly-anonymous behavior, but is otherwise a sort of thin experience. Wonderland lacks any sort of standardized non-verbal action, and instead leaves it to each object to create its own interface.

These approaches mimic the behavior of traditional applications where each user has a set of inputs and some actions are made visible to other users and some aren’t. This is a fine (and clearly effective) model, but doesn’t take advantage of the embodiment of a virtual world. If avatars can simply stand in arbitrary locations and manipulate their agree/disagree state, the scene is drained of its social content.

Understanding attention in virtual worlds is tremendously difficult. A conversational participant is not simply attentive or not, attention is something that we use a variety of signals to communicate to others. These signals are almost entirely absent in Second Life.
Although an avatar’s position does suggest the likely view of that user, it is possible for that user’s camera to be looking somewhere else entirely. Visualizations of users’ movement that expects people to manipulate their cameras independent of their avatar requires this to be the case at least some of the time. This places these two values in contrast: being able to understand the visual fields of others by looking at their avatars and being able to look at visualizations in the environment without moving your avatar.

This makes it difficult to actively perform attention in the ways we do offline. Performing attention offline involves more than just gaze direction. It involves our entire bodies, our pose, and our face. These are difficult to control in Second Life, which makes the performance of attention essentially intractable. In particular, the difference between attention and complete inaction is quite hard to distinguish. Avatars whose users have not touched the keyboard or mouse in quite some time will slouch, but that too is a mixed signal that may just as well represent rapt attention as inattention. Inattention is actually somewhat easier to communicate in the form of manic action. In much the same way it’s difficult to imagine that an avatar running around the space is paying any attention. But this is a particularly extreme form of inattention, and more moderate forms of attention and inattention are quite difficult to express.

The ramifications of this lack of expressivity for attention varies somewhat depending on the context. In discussion contexts, conversational participation is a decent (if very low frequency) proxy for attention. We would not expect completely inattentive participants to be able to produce credible contributions to a conversation. But again, a lack of participation is not reliably understood as inattention. These problems are exacerbated substantially in situations where most people are not expected to participate at all, as in a presentation context. Issues with presentations in a virtual world will be addressed in more depth in the next section.

3.5 Past and Future Virtuality

This chapter opened with the breathless claims of a market research firm about the bright future for virtual worlds. This was, at the time, a widely quoted report. There was a large population of skeptics of course, but there were just as many magazines and pundits making similarly bold predictions. When this work was done, I sat somewhere between the skeptics and the optimists. I was interested primarily in virtual worlds as attractive conceptual spaces. Although there was some work that truly pushed the boundaries of how interactions in a virtual space might be different than in offline spaces, the
The vast majority of the work was focused on mimicking offline designs for social spaces. This made virtual worlds an attractive space to do design-based research. For most of the period when I was working in virtual worlds, I was essentially agnostic about the question of whether or not they would be broadly successful. That was a question on which someone in my position had no leverage.

Despite my disinterest in prognosticating about virtual worlds, their decline subsequent to this work poses a challenge. Are there ways that this work speaks to non-virtual designs? Are there fundamental conceptual problems with virtual worlds as an experiences that other design approaches that solve similar problems don’t suffer from? Are those problems likely to have future technical solutions or are virtual worlds simply a seductive approach that is not broadly applicable?

In this section I aim to provide a series of potential explanations for the failure of this vision to come to pass. Trying to understand a negative result like this in hindsight is an inherently fraught process, but based on my work and the challenges I encountered, I will propose some potential reasons why virtual worlds did not have broader success in the particular contexts I was designing for. I will also try to tease out some lessons from this work that are applicable for non-virtual world work.

The virtual world experience is deeply seductive. It has played a major role in science fiction since True Names [Vinge, 1987]. Virtual worlds have always represented freedom of various sorts. Freedom from geographic constraints, freedom from scarcity, freedom from your past, and freedom from physical bodily constraints. The virtual would be the place where you could live in a penthouse apartment, be ravishingly beautiful, and have adventures in exotic places with anyone in the world.

In fiction, virtual worlds are treated primarily as direct analogs of the offline world. Certain physical laws are relaxed and representations are flexible, but speech, body language, and vision are usually assumed to operate identically. Furthermore, the world itself tends to look and behave like the offline world. This vision elided the many practical interface challenges to creating a virtual world that was functionally indistinguishable from the offline world. Instead, there was a tremendous focus on creating something that looked “real”, not a place that supported “real” social interaction.

This is not to say that interactions in virtual worlds were not or could not be socially meaningful. They can and are. But they don’t have the same dynamics as offline interactions and skills in one domain don’t transfer easily to the other. People new to a virtual world interface have little intuition about how it operates as a social

---

11 Other virtual world novels of note include Neuromancer [Gibson, 1986], Snow Crash [Stephenson, 2000], Rainbows End [Vinge, 2007], and the Otherland series [Williams, 1998]. A longer discussion of the competing visions of virtuality can be found in [Harry, 2008a].

12 The one major exception to this is Neuromancer [Gibson, 1986], in which the world is purely abstract. Gibson’s “cyberspace” was almost purely a way to interact with data, not with other people. It was geometric and abstract, while almost all other worlds looked “real” and have human actors featured prominently.
space, which blunts its efficacy. Just because something looks familiar doesn’t mean we can reason about it in the same way at all.

The distance between the real world and virtual worlds is clearest at the interface layer. The ways that our bodies interact with the physical world are elaborate and rich. Virtual world interactions are far more limited. We move using (typically) four keys, or by clicking on where we want our avatar to stand. We move our eyes by spinning our avatar around, or by using elaborate three dimensional camera controls. We animate our avatar with a series of recorded animations identified by names like “wave” and “dance.” Compared to the expressive capabilities of our bodies, it should not be surprising that having a handful of iconic actions would not make it possible to translate social practices without significant changes. Not just do our bodies appear low fidelity, but our movements are even lower fidelity.

This causes a wide range of insidious problems if we’re expecting virtual worlds to work like the physical world we’re used to. There are a number of assumptions that start to break down. In the physical world, we assume we know what someone else is seeing based on the angle of their head. In the virtual world, camera view and avatar position are frequently correlated, but not always. In the physical world, we have detailed control over how our bodies are situated and move. This is the foundation for what we think of colloquially as “body language.” In the virtual world, avatars tend to be iconic, demonstrating a single named pose at a time that is usually shared by all avatars. Other times, avatars idly animate to make them seem more alive, even if those movements are unrelated to any user input. Communication is radically different, too. As discussed in Presentation Spaces, we can build models for who will hear something we say in physical spaces. Online, these models are both difficult to build and hard to come to trust. All of these issues are potentially surmountable and there is interesting research in each of these areas. But given where we’re at now, these inconsistencies between the physical and virtual worlds mean we have to treat the virtual world’s “realistic” rendering as a sort of user interface oddity. Instead of making the interface fall away and be a shortcut to rapidly learning how the world works because of its similarities to the physical world, it is its differences that stand out the most. Virtual worlds are not close enough to the physical world in affordances to make the distinction blurry. Instead, they are much closer to any other mediated communication system, and we should judge their value as a design approach relative to traditional graphical user interfaces. On these terms they seem idiosyncratic and a bit strange.

As a thought experiment, it is useful to consider how the systems

---

13 It is useful to consider what sport looks like in a virtual context; how do we render the most elaborate ways we use our offline bodies? There exist a variety of video games played professionally in a sport-like way, but none mimic offline sports. They focus on hands instead of full bodies, and have a much more elaborate rule sets than physical sports. They adapt to input techniques available instead of assuming they can meaningfully recreate a physical experience with limited inputs.

---

Figure 3.24: A rough mockup of what a two dimensional version of Information Spaces might look like.
described in this chapter would (or would not) work as non-virtual world designs. Moving to a 2D GUI approach doesn’t necessarily mean throwing away the core design values of spatiality and presence. These principles are essentially “free” in a design sense when operating in a virtual context, but as discussed earlier, spatiality doesn’t mean three dimensional and presence doesn’t mean avatars. Those are simply one technique for creating those experiences. Information Spaces could be re-imagined as an addition to a WebEx-like experience. A 2D Agree/Disagree continuum could occupy part of the screen, with people’s names both creating a sense of presence (who is present) and their showing their current position on an issue. An even less-spatial version might be viable: use the agree/disagree continuum primarily as an input only, and show only the average position. Icons or colors next to people’s usernames in the users-present list could provide a person-by-person view. By treating all the inputs of the virtual world as simply actions in an interface, much of the experience can be recreated without any of the elaborate heft of a world.

A similar approach is possible for Presentation Spaces. The screen is dominated by the current slide, but viewers can desynchronize themselves from the current slide and explore past and future slides. An indicator shows how many people are currently looking at each slide. Questions or comments can be posted on past/future slides, and can be voted up or down (ala backchan.nl). Chat zones would be a little difficult to arrange, but something like an IRC approach with named rooms (similar to ROAR, a project I will describe in Chapter 6) would be effective.

I don’t think this sort of adaptation would work for all virtual world experiences. There is something special about a monolithic world like Second Life where people socialize, meet new people, play games, and visit notable places all in a continuous, synchronous world. But taken as a venue for traditional mediated experiences like meetings and presentations, there is simply nothing special about it that we couldn’t replicate in a traditional interface. Traditional interfaces have the benefit of being much more familiar, not requiring a powerful computer to handle elaborate 3D environments, and already having rich ecosystems of other tools for these sorts of interactions. This sort of flexibility is even more important now with our plurality of devices; an interface that scales easily to a tablet or phone is easier to integrate into your life than one that requires a powerful computer and large screen.

This non-virtual-world thought experiment is a way of arguing for the value of this work, despite it taking place originally in a virtual world context. The principles of spatiality (e.g. creating a sense of
place and shared audience) and presence (e.g. representing people, not just the side effects of their actions) can be useful techniques in a wide range of social experiences. These principles are at the heart of the work in this chapter, and translate well into other domains.
In this chapter, we describe backchan.nl, a web-based system that focuses on providing greater audience participation during question and answer sessions. The system allows audience members to propose questions and comments, and to vote on the questions of others. Top rated submissions are projected into the presentation space where audience members, moderators, and panelists can see them. We discuss the results of deploying this system at many different kinds of conferences and relate those results to the particular design of our system, demonstrating how systems for audience participation can be more than just shared chat rooms. From our experience with this work, we discuss the broader implications of configurable mediated social spaces and how subtle design decisions can influence user experience.

4 backchan.nl

4.1 Introduction

The utility of computer mediated communication techniques to provide a sort of “backchannel” to some other conversation has been demonstrated in a few different venues, most commonly in conferences [McCarthy et al., 2004, Rekimoto et al., 1998] and classrooms [Cogdill et al., 2001, Yardi, 2006]. Most backchannels have been text based chat or instant messaging systems that support a dialog between people who are co-present in a real world space having some sort of shared experience. The challenge with backchannels is their covert nature. The backchannels in conferences and classroom contexts are usually not accessible to the presenter, and instead focus on creating conversation within the audience. We think of this configuration not as a front channel and a back channel, but as a main stage (the audio and visual components of the presentation) and a side stage we designed to support intra audience interaction and audience-presenter interaction.

In this project, we propose a design that, rather than relying on chat, lets participants pose questions or comments to presenters, moderators, or participants in a public discussion, which can then

1 The project name backchan.nl is rather confusing in light of my refutation of the front channel/back channel distinction. This work pre-dated that framing, and when this project was named I still viewed my goal as creating a useful backchannel experience. Despite this, this project represented progress towards the development of the notion of stages, and so I have updated its language and analysis to reflect this shift.
be voted on by other audience members. The top posts are projected on a screen to the side of the presenter’s table, visible by the audience. This focus on posting specific questions and not on supporting dialog creates a focused environment that is less about connecting audience members with each other, and much more about making sure popular questions get asked of the presenters in the often limited question-and-answer period. We see this as a valuable alternative to traditional question asking procedures, which favor those audience members who are most vocal or who happen to be seated near a microphone (or those familiar with the moderator, panelists or organization hosting the event). Furthermore, should the format of the event allow, this system enables moderators and panelists to address key concerns as they occur without the potential interruption of taking questions throughout. Finally, this project demonstrates audience interactions during live events need not be limited to chat nor take place in separate spaces from the main discussion; we can imagine a wide variety of tools that support interaction between co-present people focused on different goals.

4.2 Related Work

The phrase “backchannel” has historically referred not to mediated communication at all, but the verbal and non-verbal cues that non-speakers give a speaker during a conversation. Non-lexical utterances such as “uh-huh” and “sure”, or body language cues such as shaking your head or averting your gaze all provide meaningful and important feedback to a speaker without necessarily trying to take a turn speaking. Audible and non-audible signals have both been shown to be important for maintaining communication efficiency by Krauss et al. [1977]. Use of “backchannel” to describe these actions suggests developing non-primary communication channels between speakers and listeners can be a powerful way to create more effective conversation spaces. This metaphor has been extended in the literature to include any system in which there is an additional mediated channel separate from the primary channel of communication (which may itself be mediated). Such systems often serve to connect groups of listeners to a single speaker, though other configurations are certainly possible. Although this term represents the chat-oriented, audience-only systems discussed in this section, we view our work as being in a separate category of systems that aim instead to create a side stage that complements the main stage experience.

There are two main approaches to understanding backchannel use in past projects. The first focuses on existing chat or instant messaging tools, and examine the type of communication that takes place on
these channels and the relationship between this communication, the users of the backchannel, and the presentation they are participating in.

Based on IRC logs recorded during a conference, McCarthy et al. [2004] explores the kinds of conversations that took place and the relative involvement of different users, to propose a general taxonomy of the kinds of conversations held. This approach is similar to Yardi [2006], who describes the use of an IRC backchannel in a higher education environment. Both projects build models that describe the kinds of messages that happen on backchannels. Yardi in particular focuses on the ways that over time, participants in the backchannel develop both fluency with the tool and community standards for its use.

Ratto et al. [2003] discuss their experience deploying the Active-Classroom tool, which (like ours) lets students post and vote on questions. Their interaction is PDA based, and is not projected in the space—instead giving all moderation control to the teacher. Their analysis focuses on one semester long class, and presents little quantitative usage data to drive their analysis.

Golub [2005] discusses the tension between the utility of backchannels (and internet access in general) during a presentation and provides observations of what audience-members use laptops for during presentations.

Backchannels have also been proposed in the audio domain, in which members of an audio conference can create sub-conferences separate from the main shared audio channel. [Yankelovich et al., 2005]

Our work contrasts with these examples because we focus on posts instead of a live chat space and making the contents of that space visible to presenters and the audience, our work shows how the contents of a side stage-style system can be more effectively integrated into physical presentation spaces.

The second approach examines backchannels as an augmentation of physical spaces to create new social environments. These projects seek to connect multiple physical spaces [Karahalios and Donath, 2004] or alter the interpersonal dynamics in a space by visualizing aspects of the conversation [DiMicco et al., 2007, Donath et al., 1999, Bergstrom and Karahalios, 2007], which start to point in the direction of a side stage approach. These systems are focused on one-to-one or small group interactions instead of connecting large audiences with a small number of presenters. At this scale of interaction, the value of a side stage diminishes because participation on the main stage is easier and more equitable.

The work of Rekimoto et al. [1998] combines elements of these two
approaches by making the contents of a backchannel chat channel visible on a screen to the side of the presentation slides, though the rate of chat messages makes it hard for the audience to engage with its contents.

At the 1988 Junior Summit, participants were given two-way pagers that could send messages to a scrolling LED text screen in an auditorium. Much like in our system, while the students found the system empowering and engaging, the organizers struggled with identity, moderation, and accountability issues. Unfortunately, there is little data available about this particular deployment. [Chesnais]

A very similar tool has been produced by the Berkman Center at Harvard University for classroom use. Although we were not aware of it during development, our designs are broadly similar. Source code for this tool is available online. [Ano] Question Tool’s primary difference is that it is not designed to play a major visual role in the classroom. Students can use it as a sort of guided discussion space and teaching assistants are often present to answer some questions, but it doesn’t have a distinct public projection view. The view that students use to interact with it can be projected (and sometimes is), but the system is not designed specifically with that use in mind. As a result, the tool is generally effective at creating a side stage experience but has trouble merging that experience with the main stage in some situations.

4.3 Design

Backchan.nl is a web-based system for posting text items (nominally questions, but any text could be posted) and voting on other people’s submissions. Audience members participate by visiting the backchan.nl website on their laptops. Posts can be voted either up or down, and are ranked using a formula that rewards positive votes, a high volume of votes, and recent votes. The current top eight posts are displayed in the presentation space on three different screens: a large projection screen facing the audience, a monitor for panelists/presenters, and a monitor for the moderator. Text on the large projection screen is sized such that it is visible even in the rear of the room. An auditorium with backchan.nl can be seen in Figure 4.1.
Although the public displays show the top eight posts, the web interface maintains a chronological listing of all the posts that have been submitted during the session. As posts receive positive votes, they tend to rise in the ranking and will eventually reach the top eight and are projected on the main screen. The web interface from which votes are cast and items are submitted can be seen in Figure 4.3.

When a user first loads the site, they are asked to identify themselves with a name and affiliation. This information is included with any posts that a user made. Votes are publicly anonymous, but are tracked internally with the voter’s name to prevent multiple-voting. Identity is easily changed and no formal account registration system is included. The implications of how identity is handled in this system are discussed later in the paper.

4.4 Observations

Although the backchan.nl system was designed to be used at one specific event, the success of the tool at that event led to many sub-
sequent inquiries for future deployments. Over the life of the project thus far, backchan.nl has been used at hundreds of events. We’ve had over 14,000 unique users (though this number is artificially high; see discussions of identity issues in a later section), over 20,000 items posted, and more than 60,000 votes cast. The events have included large events with hundreds of audience members and smaller scale events with audiences of fewer than 100. Most of these events have been panel discussions of some sort, though we have used it for presentation sessions as well.

For our analysis, we will focus on two of our early deployments that share many characteristics but resulted in significantly different experiences. The first event was the Futures of Entertainment conference. This event was our original deployment target, so its structure informed many of our design decisions. The conference was single track and each session was a two and a half hour panel discussion. Panelists gave brief personal introductions, after which discussion amongst panelists was managed by a moderator. The first two thirds of a session were predominantly guided by pre-prepared questions from the moderator, while during the final third, questions were solicited from the audience. We knew from previous years that the conference audience was likely to have large numbers of laptop users and wireless Internet access in the conference venue was known to be excellent. The audience was largely entertainment industry professionals interested in exploring industry issues from an academic perspective.

The second venue was ROFLcon, a two-day conference exploring Internet culture with panel discussions involving significant figures from the Internet community. The convention had multiple tracks, and backchan.nl was only used in the biggest presentation space on the second day of the conference. ROFLcon’s attendees tended to be much younger than the Futures of Entertainment audience. The ROFLcon audience included many more students and was generally rowdier and more exuberant. It was not uncommon for audience members to interrupt discussion by shouting something at panelists. Anecdotally, this was many attendees’ first conference experience. More general descriptions of the atmosphere at ROFLcon can be found in popular media coverage of the event. [Raftery, 2008]

In terms of their overall format, these two conferences were quite similar; both focused on panel discussions, took place in similarly sized auditorium spaces, and had technically savvy audiences likely to have laptops for accessing the system. Despite this, the audience of each conference used the system in very different ways. In this section, we present and compare usage data between these two conferences.
4.4.1 Usage Data—Futures of Entertainment

Over the course of the five sessions in the conference, there were 125 distinct users. Most names and affiliations were reasonable, with only a handful of obviously anonymous names chosen. The use of pseudonyms or nicknames was generally rare.

These users posted a total of 224 items, with a mean of 37.6 questions per panel and a mean of 20 questions per each morning introductory session. Across all sessions, we observed a rate of question submission of 0.26 posts per minute and 1.8 posts per registered user over the course of the entire event. We didn’t have accurate overall attendance data, but the main hall for the conference has a seating capacity of 190 and was rarely full, so we can conclude that a significant fraction of attendees used the system at some point during the conference.
Figure 4.3: A screenshot of a backchan.nl event in action.
There were a total of 676 votes, with the average post receiving 2 votes. The vast majority of votes were positive: there were 568 positive votes compared to 108 negative votes. The distribution of final scores for posts is shown in Figure 4.4. Votes were cast at a rate of 0.78 per minute.

4.4.2 Usage Data–ROFLcon

ROFLcon had five primary sessions that used backchan.nl. We observed 450 different usernames, although there are a very large number of pseudonymous names and many duplicate IP addresses recorded. Over the course of those sessions 420 items were posted, at a rate of 1.08 per minute. There were 1667 distinct votes, recorded at a rate of 4.27 per minute. Of those votes, 1142 were positive, 525 were negative.

4.4.3 Post Contents–Futures of Entertainment

Over the course of the conference, we observed a number of different categories of posts. Early on in the first panel (and in every subsequent panel), alternative options for audience interaction set up by members of the audience and advertised on the backchan.nl system. The first two posts in each panel advertised a Skype public chat and an IRC channel. Private chat was not an inherent feature of our tool, and we were surprised that the first submitted items were advertisements more than questions. In subsequent panels, these postings occurred very quickly, just as the conversation was commencing. Neither channel sustained the same level of involvement as the backchan.nl system itself, but its position as a screen visible to all participants in the conference empowered audience-members to create and publicize their own alternative channels. We view the opportunity for the audience to co-opt the channel for competing tools as a demonstration of our commitment to an uncensored channel. Furthermore, such postings highlight the utility of the backchan.nl system for audiences to self-organize, and to mediate their own conference experience. These kinds of informational posts didn’t score well, and were often the first posts pushed outside of the top eight as soon as questions targeted at the presenters started to be submitted.

During panels, the bulk of popular postings were questions targeted at the panelists. These questions varied in specificity from follow-up questions to panelists (“I would love to hear more about buzz marketing - how it actually works, and how clients want it to integrate it with more traditional methods.”) to general synthetic
questions ("What’s the role of Social Media in advertising and Convergence Culture?").

The other most popular posts represented public sentiment in some way. Near the beginning of the third session, someone posted “Can we make sure some more questions from the board get answered this time? xthxbai.” This was the sixth most voted on item in the entire conference. Later in the conference, someone else asked the second most popular question: “So, is NOW the time the panel should turn some attention to these excellent user-generated questions?” There was also a complaint about temperature in the auditorium.

Among questions that failed to attract significant attention and votes, there were a number of common themes. Posts that didn’t feel sufficiently question-like tended to get passed over. The same was predominantly true for funny and snarky comments.

4.4.4 Post Contents–ROFLcon

For the most part, items on backchan.nl at ROFLcon fall into similar categories as those at Futures of Entertainment. The biggest difference was the balance. There were fewer questions for panelists, but some were certainly generated that were subsequently asked of the panelists, e.g. “Moot, what is your favorite 4chan meme?” As at Futures of Entertainment, these items tended to be well received. Most of the top posts in each session were a question for the panelists of some sort.

A significant majority of the posts, however, were not questions. There was a constant flood of jokey posts, for instance: “WAKE UP SHEEPLE, ALEXIS DID 9/11”, which combines a number of popular Internet memes about a notable community figure (Alexis), a satirical exclamation from the site Alexis runs (“wake up sheeple!”) and 9/11 conspiracies. The success of this kind of post varied widely. Sometimes they were wildly successful, but the vast majority of them languished in obscurity and never made it to the top eight. ROFLcon also had many more announcement type posts like “::abuses backchannel:: Someone lost a Lumix camera yesterday. Find Susannah on the ROFLTeam to describe it/pics on it.” As at Futures of Entertainment, messages like these were never highly rated, but did get visibility at the start of sessions. They were rarely submitted later in a particular session. This indicated an understanding that there were phases in a session when different kinds of posts were more or less appropriate.

ROFLcon’s use of backchan.nl was much more playful than at Futures of Entertainment 2. This audience was quite familiar with manipulating social tools like this and so pushed the system to its limits.
In one session, users engaged in a wide-spread coordinated attempt to rig the item rankings. Eight posts were made containing parts of the lyrics to Rick Astley’s “Never Gonna Give You Up.” [Ano, 2007a] Users then voted these items up and down to make them appear in order in the top eight, attempting to “rickroll” the audience. “Rickrolling” — forcing an unsuspecting public to watch/listen to Astley’s song — was a relatively popular Internet meme at the time of the conference. The same user name can’t vote more than once on an item, so users participating in this process quickly switched between pseudonyms to trick backchan.nl into letting them vote again. When the lyrics were finally in order, someone in the audience yelled “WE DID IT” and there was spontaneous applause for their achievement. In this way, users demonstrated a clear internalization of the system dynamics and co-opted the system for their own playful ends. We believe this shows the power of the system that users only play with systems that provide a meaningful stage. Pictures of this happening are available in [Chillag, 2008].

4.4.5 Voting and Posting Patterns

Over the course of each session, voting and posting patterns emerged that depended quite a bit on the organization of the conference sessions themselves. As seen in Figure 4.5, most of the activity of both posting and voting occurs around the same times. Overall, participation decreases substantially in the second half of every session at Futures of Entertainment 2. This has two possible explanations. Official question time for each panel started between 90 and 120 minutes into each session. Audience members might have chosen to ask questions themselves rather than rely on backchan.nl system, given the opportunity. The decline might be a result of users misunderstanding the ranking system and assuming that new posts were unlikely to reach the top eight with so many highly rated posts already submitted. We
saw a similar falloff in ROFLcon, though it wasn’t as significant. This is closely related to our discussion of tempo later in this paper—it might be the case that this is simply a function of a poor selection of the time constant for these long sessions.

What is notable about the ROFLcon participation rates is the high number of questions very early on in a session. This is likely because the ROFLcon sessions included public figures the audience was largely already familiar with, and so audience members had questions prepared before the panelists said anything.

The voting timelines of a number of different posts are displayed in Figure 4.6. In most cases, the voting in the first five minutes of a post’s lifetime indicates whether or not it is going to become popular. Posts were rarely contentious—although many posts had some negative votes, they were usually predominantly positive or predominantly negative. These patterns are not unlike those observed in work studying Digg and Reddit, internet-wide systems with a similar voting design. [Lerman, 2006]

### 4.4.6 Main Stage/Side Stage Integration

Over the course of the Futures of Entertainment 2 conference, backchan.nl was frequently integrated into the foreground conversation. Although the moderators came to each session with a set of prepared questions, most of them quickly went off the script and integrated the posts from the audience into their questions. Often, the moderator would combine a few different questions into a single broader theme and put that question to the panelists. This process was almost always explicit. The moderator would verbally acknowledge the source of the questions, which ones were being combined, and the audience members who asked the original questions. The audience began to expect this kind of behavior and complained when they felt moderators weren’t integrating backchannel questions enough. While this also occurred at ROFLcon, the increased non-question “noise” meant that moderators had fewer options to choose from. They almost exclusively ignored the funny posts, which really didn’t need to be addressed explicitly.

After the Futures of Entertainment event, it became clear that we needed a way to dismiss posts that had been addressed by the moderator. For subsequent events (including ROFLcon) we had some moderation tools that allowed us both to mark a post as answered (reduce its points to zero but still display it in the upcoming list with a checkmark) and “remove” a post (remove all user-visible history of the post). The challenges inherent in moderating a system like this are described later in this paper.
Panelists could also see backchan.nl, and would sometimes pull questions from it into their responses. This worked particularly well because the panels at both events were very discussion oriented and opened-ended.

This underlines the distinction between traditional “backchannel” designs and the main/side stage distinction we aim to create here. With a backchannel approach, integration with the front channel is difficult because the main broadcasters on the front channel have a hard time keeping track of a high frequency backchannel. However with backchan.nl, integration was frequent and fluid. This is the goal with side stages, and backchan.nl demonstrated it nicely.

4.5 User Responses

In general, users and presenters were very positive about the backchan.nl experience. At Futures of Entertainment, conference volunteers conducted interviews with participants after the conference finished, and some participants had comments about backchan.nl in particular. One audience member thought “the ability of people to vote for what they were interested in was great.” Another participant particularly appreciated that the system “gave [audience members] opportunities to participate in direct ways.” Informal chatter at ROFLcon was similar. Moderators tended also to enjoy the system, although they reported having to rethink their moderating approach significantly in light of backchan.nl. Over the lifetime of the tool, moderators who have used it more than once are quite positive about its role in panel events, and the organizers of both ROFLcon and Futures of Entertainment requested backchan.nl at their next events.

Users and presenters in sessions involving backchan.nl expressed a number of common concerns about the system. We collect and address those issues here.

4.5.1 Distraction

The most common concern from presenters was distraction. Some panelists didn’t want to use backchan.nl because they thought it would draw attention away from their own comments. We believe that in a space with wireless access, there are plenty of ways for bored audience-members to distract themselves. Indeed, when we watched the screens of audience members, they were rarely staring at the backchan.nl system for long periods of time. They tended instead to bounce between it and many other different websites and applications. This finding is very similar to Golub’s observations in [Golub, 2005]. The projected display itself is intended not to be flashy
and attract the attention of people who aren’t interested in tracking its content. Still, this is a fair criticism. Although we think this system displays much less information than, say, a modern cable news program, it is still more distracting than non-backchan.nl equipped panels.

In general, users and presenters were very positive about the backchan.nl experience. At Futures of Entertainment, conference volunteers conducted interviews with participants after the conference finished, and some participants had comments about backchan.nl system in particular. One audience member thought “the ability of people to vote for what they were interested in was great.” Another participant particularly appreciated that the system “gave [audience members] opportunities to participate in direct ways.” Informal chatter at ROFLcon was similar. Moderators tended also to enjoy the system, although they reported having to rethink their moderating approach significantly in light of backchan.nl. Over the lifetime of the tool, moderators who have used it more than once are quite positive about its role in panel events, and the organizers of both ROFLcon and Futures of Entertainment requested backchan.nl at their next events. Users and presenters in sessions involving backchan.nl expressed a number of common concerns about the system. We collect and address those issues here.

4.5.2 Presentation Preemption

Presenters were also concerned that questions would be posted to backchan.nl during their presentation that they would answer later in their presentation. Some moderators at Futures of Entertainment 2 expressed concerns that the system placed pressure on them and the panelists to address the audience’s concerns first. This concern was also voiced at the handful of non-panel presentations we have run with backchan.nl. In these situations, we would typically turn the projection off during the presentation itself and then back on during the question period. During the projection black-out, users could still participate, but the audience as a whole could focus on the presentation materials. Despite presenters’ fears, we did not see many instances of the audience pre-empting presentation material with questions.

4.5.3 Replies

Many users lamented the lack of an explicit reply structure in backchan.nl. Audience members frequently made explicit comparisons between backchan.nl and services like Twitter and IRC, which both have been used in conference situations to provide backchannel conversation.
Figure 4.7 shows a common configuration from a Media Lab event. In some deployments these conversations took place over backchan.nl itself and became quite contentious. The core complaint in most instances had to do with the lack of a way to distinguish the timing of posts or to indicate that a post was a response to an earlier posting. Some users tried to adopt the “@username” format of Twitter, but often these posts would get voted into the top eight without the post it was replying to, decontextualizing the response and confusing audience members who weren’t using the tool on a laptop. Over the course of an event, users have almost always moved beyond their initial attempts to use backchan.nl in ways it wasn’t designed for and settle into a pattern of posting items that aren’t explicitly intended to generate responses from the audience.

In future versions, we would be interested in better affording responses. From a design perspective, we find integrating replies a challenging and interesting problem. Given our limited projection screen space and the contextual nature of discussions, we’re hesitant to display elaborate conversations on the screen. Although filtering moderation systems like Slashdot’s have been shown to be appreciated by users [Ratto et al., 2003], we have a non-interactive, low density group display and would have to make a global decision for all users about what threshold is appropriate.

We also don’t want presenters to feel (accurately or not) that the backchan.nl is home to significant conversations that they don’t have any visibility into. This is at odds with our desire to have an uncluttered main projection screen. A successful design will have to accommodate both users’ desires and presenters’ concerns.

4.6 Analysis

What sets backchan.nl apart from traditional backchannel approaches is the presence of side stage conversation in the physical space. This approach had a few categories of effects, relative to work on non-physically situated backchannels.

4.6.1 Content

The content in backchan.nl’s side stage is markedly different from the content observed in chat-based backchannel implementations [Yardi, 2006, Cogdill et al., 2001, Golub, 2005, Rekimoto et al., 1998]. While the vast majority of posts in this system could be classified as “Work” messages in McCarthy’s categories, they were usually focused on a specific audience — the panelists. While panelists have sometimes been involved in backchannels in other systems, they knew that the
backchannel was largely inaccessible to the main participants on the front channel. Showing backchan.nl in the conference space itself makes it clear who is seeing what’s posted, and creates the integration that is the signature of a side stage approach. This increases the stakes of communication, as it enjoys an audience that extends to those not actively using the interface itself.

This played out in different ways in the two conferences. For Futures of Entertainment 2, the increased stakes meant users mostly took the system seriously and there were few snarky or critical comments. This is no doubt influenced by the fact many in the audience were attending for professional purposes, either as representatives of their businesses, for networking, or to learn more about their industry. The more playful approach of the ROFLcon audience reflects that community’s interests and normal mode of interaction. Higher stakes made users likely to use the system, but content was similar to what they might have normally posted in an online-only discussion. In this way the implementations were not so different; both offered material representations of the different values of the communities present. This echoes the behavior of the children in [Chesnais] who, even in the face of adult disapproval and public attribution of their messages continued to submit messages that the organizers found objectionable.

4.6.2 Adoption

Adoption is often a serious problem for social systems like backchan.nl [Orlikowski, 1992]. Having a physical representation of audience interaction in the space serves as a constant reminder about how to get involved in the side stage and what is currently happening on it. This provides an effective hook to get new users involved. In particular, the appearance of posts that an audience member thinks is great (or terrible) and the promise that they can help promote or demote it is a powerful incentive.

The moderators played a key role in backchan.nl adoption. After a moderator effectively demonstrates how backchan.nl can be used, users quickly build an expectation about its use in future sessions. As with the adoption of other social technologies, it was important for the audience to see that the tool was being taken seriously and that their interactions across it were meaningful to the conference organizers.[Orlikowski, 1992] This was most clear when the audience co-opted the system to complain about the lack of attention the board was receiving. Indeed, the moderators explicitly responded to each major complaint that appeared on the projection during Futures of Entertainment 2. The chaotic nature of the items at ROFLcon some-
times precluded acknowledgement, though issues with sound were frequently addressed in response to audience comments.

The physical arrangement in the room further underscores this organizational support. Although people in the audience were able to set up their own backchannels using other software, having a projector and a screen in the room demonstrated the material commitment of the conference organizers to this tool. Although it’s possible to imagine situations in which backchan.nl might be marginalized by user-organized backchannels, this didn’t happen in any of our deployments and we suspect a large part of that is the physical-ness of the design. It’s hard to simply ignore a display in the presentation space in favor of a virtual-only conversation space.

4.7 Side Stage Configuration

Based on our experiences deploying the backchan.nl system, it has become clear to us that there are a number of ways in which a side stage like backchan.nl can be configured that would make this tool both more broadly applicable and more finely tuned for the needs of specific groups of users. In this section, we discuss two major design questions: how should post scores change over time and how should users’ identities be represented? In both cases, we argue that a mediated social space like backchan.nl would substantially benefit from having configuration options that adapt it to different situations. In much the same way that you might arrange the chairs in a room differently for a lecture versus a group discussion, so too should mediated social spaces have a range of options that foster different social situations.

4.7.1 Tempo and Time

There are a number of modern systems that provide some sort of “top” list of items. Digg and Reddit are perhaps two of the most notable (and straightforward) examples of this service; users submit URLs, which can be voted on by other users and by some metric the items are sorted and top items are displayed prominently. These systems (like our system) face an interesting algorithmic challenge: how do you keep the top items changing fluidly such that popular items rise to the top but previously popular items don’t linger too long? Certainly, a rank ordering of items by the number of votes is not generally going to be sustainable; older items will tend to accumulate the most votes and make it hard for newly posted items to graduate to the “top” list of items. In designing a technique for ranking posts, we had a number of goals in mind:
1. Highly positively voted posts should get a high ranking.

2. New posts should eventually displace older posts.

3. Contentious posts should be rewarded somewhat, but less than uniformly well received posts.

4. Posts that continue to receive attention over time should stay highly ranked.

The score is comprised of two factors: an age factor and a vote factor. The vote factor (where $U_p$ and $D_p$ are the up and down votes of a given post $p$, and where $(U + D)$ is the average of total votes across all items in the session) is:

$$voteFactor(p) = \frac{k U_p}{U_p + D_p} + \frac{U_p + D_p}{(U + D)}$$ \hspace{1cm} (4.1)

The first term rewards posts with many positive votes, while the second term promotes items that have a high vote total relative to all posts in the system. In general, the first term rewards positive votes while the second term rewards negative votes. The balance between those terms is set by the constant $k$.

The age factor (where $t_{p,v}$ is the timestamp of a specific vote on item $p$ and $t_{now}$ is the current timestamp) is:

$$ageFactor(t_{p,v}, t_{now}) = \frac{(t_{now} - t_{p,v})}{\tau}$$ \hspace{1cm} (4.2)

The average time difference was computed for the most recent (up to) five votes on each item. The time constant $\tau$ varied in our experiments, but something on the order of $10^4$ was usually effective.

The two factors are multiplied together to generate the final score. The main difference between our ranking system and those discussed in [Ano, 8] is that ours measures time not on the basis of the initial posting of an item, but on a moving average of the ages of its votes. In this way, an old item that receives new attention from voters is rewarded.

One of the downsides to this approach is that the points used for internal rankings are not typically made visible to users. This leads to many situations in which an item with few votes is more highly ranked than an older item with many more votes. Although this seems dissonant, users have very rarely commented on it when using the system. Perhaps popular adoption of similar systems has made familiar the notion that voting helps promote something, but that rankings are not strictly tied to votes.

The time constant in the system plays an important role in configuring the space. The time constant sets a sort of tempo for the
system; if items decay slowly the top ten is a sort of low-pass filter, showing items that have been of enduring interest to the audience over a long period of a time. This might be useful in a session where the backchan.nl’s role is to accumulate questions when there are only a handful of panelists and each panelist speaks for quite a while before there are questions. Conversely, if items decay quickly, the top listing will turn over quickly, exposing a very high-pass view of the audience’s interests. This might work well with a panel where the topic is shifting quickly or there are many panelists speaking for a short time. Which kind of time constant is appropriate also has to do with the number of participants. If there are more voters, faster decay makes more sense because more votes will offset the time decay pressure of the ranking equation.

Returning to the goals for a ranking system that we proposed, it is important to note that these goals are specific to the conference situations we were initially designing for. In line with our longer-term interest in describing the ways that a side stage can be configured for different situations, it is easy to imagine conference situations when our goals don’t necessarily make sense and we might design a different scoring system. For instance, in a paper-oriented conference like CHI, our goal of having a “live” list of questions receiving attention might not be appropriate. Instead, a naive ranking mode in which questions accumulate during someone’s talk and are ranked purely by their positive and negative votes would be a useful way of identifying the best questions for the end of the talk. This is biased against questions that might be raised towards the end of the talk, but that might actually be valuable. The side stage could act as a counterpoint to the existing structure in which questions often focus on the end of the talk because it’s freshest in the audience’s mind.

4.7.2 Identity

As in all mediated social systems, how identity is represented in the system can have profound impacts on the behavior in that system. In our design, identity is handled in a very informal way. Users can enter a name and affiliation, but it is trivially easy to change one’s personal information. This keeps the threshold for involvement quite low, and we built the system this way to encourage more use. In general, the tradeoff in building identity systems is between low thresholds that encourage use (like our model) and systems that have higher costs to join, but also provide more reliable signals about identity across users. An extreme example of a higher cost identity system might require credentials from a trusted source, but such a system would keep out people who didn’t have such credentials
or didn’t want to expose that much information about themselves. Given our initial venue, we built a system that was biased towards a lightweight identity.

We found (as discussed earlier) that in our first deployment at the Futures of Entertainment 2 conference, there was very little identity play, and our decision promoted productive discussion. This is in stark contrast to ROFLcon, where playful and funny posts outnumbered traditional questions. This playful attitude towards identity is evident in the usage data. There were a much larger number of unique users at ROFLcon of which the vast majority were pseudonymous.

From a design perspective this comparison makes clear that different identity structures make sense for different audiences and presentation structures. When considering the design of future side stage systems, we believe a range of identity options should be offered. Beyond the extremes of no registration and trusted certificates, the system could require email verification or track IP addresses, which would increase the costs to changing identity. In this middle ground, identity in the system is still easily gained, but changing identities is more problematic.

Archiving also plays a role in identity. By changing what kinds of behavior are stored in the system, low-cost identity forms can accrue more costs to changing. If users see other users with rich histories in the system, they will tend to be rewarded for their history by other committed users. [Resnick and Zeckhauser, 2002] In this context, users with a history of submitting useful questions can easily be distinguished from someone with a throwaway account. The costs of this kind of pseudonymity have also been explored from a theoretical perspective by[Friedman and Resnick, 2001]. Indeed, what about users’ behavior is archived and made publicly available is another important axis along which a side stage might be configured.

4.7.3 Democracy and Moderation

This system encodes a basic democratic principle: the best items will rise to the top based on the aggregate will of the audience. There are limits to this principle, though. As mentioned earlier we quickly discovered that we needed some sort of moderation vocabulary. We settled on two basic actions: “answered” which sets an item’s points to zero and “remove” which removes all visible record of the item. Because moderators who used early versions of this system had requested this feature, we hoped that they would also take responsibility for making the decision to demote or remove a post. In practice, this was too much of a burden on them, especially in situations (like
ROFLcon) where participants were regularly submitting offensive items, and moderation didn’t just involve marking “answered” items.

In practice, one of the authors would sit in the audience and take responsibility for moderating posts. We moderated with a very light hand, marking as answered only those posts that were explicitly mentioned by the moderator and that panelists seemed to answer and removing only those posts that were broadly offensive. Of course, there are substantial grey areas in these criteria. Questions were sometimes posed to panelists and subsequently evaded. Presumably the author of that question would like to see it remain in the top eight, even though the panelist ostensibly responded to it.

ROFLcon offered other kinds of posts that we struggled to respond to. In the process of arranging song lyrics in the top eight slots, all other submissions were pushed out. After the ordering was successfully achieved and publicly acknowledged, should we remove them? Indeed, humorous posts in general tended to clog up the top eight because they couldn’t really be “answered” and so there was no clear contract between the moderator and the audience about how they should be handled. Furthermore, the sheer achievement resonated with the spirit of the conference itself. In the case of the song lyrics, we left them on the board hoping that the audience would subsequently down-vote them to clear the board for new content. This turned out not to be the case, and backchan.nl was largely useless for the rest of that panel.

These tensions between keeping the tool effective and letting some version of a democratic process run its course pervade social tools like this. In this case, our approach was to at least maintain an open approach to moderation. Posts that were demoted were clearly noted with icons, and all their original votes were still shown on the website view. In this way, the moderator is at least accountable for her actions. In the same way that the audience spoke out against panel moderators who were clearly ignoring backchan.nl questions, members could co-opt the tool to protest moderation decisions they disagreed with.

As discussed with respect to tempo and identity, we argue that configuration has an important role to play here. Because the standards and desires of different communities can vary widely, it makes sense in some situations to devolve the moderation controls to the users. Digg, for instance, has systems that allow users to “bury” a post they think is inappropriate for some reason. Of course, the risk of tools like this is that they can be easily abused by organized groups of users that want to suppress certain points of view. An effective compromise is “flagging” systems used notably by Craigslist and Metafilter. In this model, users could flag items as being “offen-
sive” or “answered.” Moderators can use these flags as a proxy for the audience’s attitude about specific posts. Having a person making an interpretive decision from this data makes it much harder for groups of users to manipulate an automatic moderation system. There are roles for both of these moderation strategies, as well as the simple benevolent-moderator model we used during our testing. This is the final axis that we propose should be available for configuring side stages.

4.8 Conclusions and Future Work

In this chapter, we have demonstrated how integrating a side stage into the physical space of a conference can create effective new ways for the audience to interact with panelists, moderators, and other audience members. This approach is in contrast to traditional backchannel systems which focus instead on creating separate audience-only interaction spaces that are somewhat exclusionary to presenters and create a covert character for audience interaction. This shows why chat-based backchannels are viewed as effective for the audience, but rarely integrate well with the front channel. Our posting approach was both a legible interaction metaphor and made backchan.nl more appropriate for public projection.

In terms of in situ research practice, backchan.nl has been a significant success. From its inception, backchan.nl was a system whose uses was driven by collaborations with organizations running events with needs specific to those events. After much of this research work was done, the system was extended in such a way that third party organizations could deploy their own backchan.nl instances without intervention from anyone on the backchan.nl team. The system has seen significant independent use: over 2000 distinct sessions, over 20,000 posts with over 60,000 votes from over 14,000 unique users. This speaks to the extent to which backchan.nl accurately addressed a real need in people’s lives had a significant impact beyond a lab context. Although not all in situ research work should be aimed for broader adoption, we can nonetheless safely argue that broader adoption underlines our claim that we have built something that works in varied social and physical contexts.

There are a number of incremental changes that could be made to improve this system. While the system is currently adapted to posting snippets of text, it could easily be diversified to include many more types of objects. Links to web resources, polls, and discussion threads could all be promoted to first-class objects in the system. An embedded chat interface would also be helpful, providing live discussion opportunities. We are also interested in including some
sort of reply mechanism for web-users that does not adversely impact
the projected non-participant view.

In the current implementation, the view that is made available to
the panelists and moderator is the same as that projected for the au-
dience. Although in this design we have strived to make the display
relatively accessible to the panelist by limiting update frequency, us-
ing large font sizes, and aiming for visual simplicity. Still, in many
contexts a panelist, moderator, or presenter might desire an even
more abstract view. One compelling direction might be a display
that uses very little text, but represents general activity levels of the
audience over time to judge engagement. We might also look at in-
terfaces that use a panelist’s gaze to adjust levels of detail. For quick
glances, the abstract view is appropriate. But a longer stare could
transition the display to a more detail-oriented view that shows cur-
cent top lists. These changes are particularly compelling for the solo
presenter mode, which (as discussed earlier) is a use case the current
backchan.nl design struggles somewhat to support.

More broadly, we’re interested in the ways that mediated interac-
tion spaces can be configured for supporting different kinds of social
situations. We discuss the ways that tempo and identity function
in this particular system as a way of laying out a broader structure
for the relationship between these kinds of design decisions and the
kinds of interactions they create.

Finally, we’re interested in the potential of side stage systems like
this one as a way to involve remote participants with co-located par-
ticipants. Remote participants are often marginalized and forced to
rely on a local advocate to interrupt the flow of conversation and
check for questions from remote users. Our side stage approach
offers a way to more fluidly involve remote participants by encour-
aging both local and remote users to interact through the same me-
diated system. This blurs the lines between local and remote partic-
ipants and could counteract some of the disadvantages of being a
remote participant. Although for the most part the deployments we
discuss did not have significant numbers of virtual participants, our
future work in this area will explore the implications of this system
on remote participants.
5 Tin Can

In the introduction, I argued that taking a “beyond being there” approach to designing collaborative interactions can yield powerful results. I have shown in the past two chapters how designs that aim not to create the experience of “being there” but instead to imagine new kinds of relationships and interactions can be both meaningful to users and improve on traditional “being there” approaches to mediated interaction. In this chapter, I will introduce Tin Can, a tablet-based application to collaboratively track discussion topics and ideas in a seminar-style discussion classroom. Based on our study of this system, I will argue that we have created a system that remains quite useful even when everyone using the system is face-to-face and could eschew the system entirely if it wasn’t useful to them. Looking forward, Hollan and Stornetta [1992] argue in 1992 that “no one seems to be asking the question, “what would happen if we were to develop communication tools with a higher information richness than face-to-face?” I view Tin Can as one of many potential answers to this question.

In a classroom using Tin Can, each student uses his or her own tablet to share text ideas in a synchronized, visual environment. The system is designed to promote diverse participation and increase engagement. Using this platform, we observed twelve class sessions and conducted interviews with the participating students. Instead of simply introducing an additional text-based communication channel into the classroom, we find that the system creates a new “stage” (in the Goffman sense) on which students could perform in ways that the main spoken stage could not support. This stage coexists with spoken communication, and augments how students attend to the material and each other. We conclude that spoken participation alone poses barriers for some participants and the addition of a non-oral, text-based stage can help establish more equitable, diverse, and engaging discussions in the class.
5.1 Introduction

The physically co-located small group discussion is often viewed as the gold standard for effective collaboration and communication. It can provide a space for participants to voice their opinions and can readily lead to deliberation and collective problem solving [Burkhalter and Gastil, 2002]. Not surprisingly, it is often the case that designers seek to virtually reproduce the characteristics and norms of the small group discussion in technologically mediated communication media. Hollan and Stornetta [1992] provide a valuable counterpoint to this approach arguing against viewing experiences mediated by the “physically proximate” reality as necessarily superior to those mediated by technology. We have adopted the following challenge: instead of assuming that the small group discussion is good enough and the only appropriate design consideration is its preservation and replication, we seek to appropriately apply the unique properties of a technological system to the established affordances of a small group discussion. We would not deny that face-to-face interaction offers many substantial benefits when compared to interactions mediated by, for instance, a video conferencing system; nevertheless, we argue that there is room to improve the physically proximate small group discussion by intervening in the assumed normal frameworks of turn-taking and attention.

Much of the contents of this chapter are drawn from [Harry et al., 2012], which represents collaboration between the author of this dissertation and Eric Gordon.

In this paper, we describe the design and enacted use of a tablet-based system for a discussion-based graduate seminar. Although this is not a common educational venue for intervention (lecture classes are a more traditional venue, e.g. [Kam et al., 2005] or [Bergstrom and Harris, 2011]), it is one in which we identified a number of potential problems with pure face-to-face discussions that a tablet-based system might effectively address. We had two major goals for this work: first, create a class discussion context that encouraged more diverse participation in class; and second, to help students feel engaged and connected to the learning environment.

To meet these goals, we sought to expand notions of legitimate participation beyond speaking, using the affordances of a text-based communication system. Our system creates an alternate communication space within the learning environment. Typically, in group communication contexts, spoken participation is viewed as the primary or dominant interaction medium, one that is often the target of modification as in, for example, Second Messenger [DiMicco et al., 2007] or Meeting Mediator [Kim et al., 2008]. Like the Cognoter [Tatar et al., 1991] system, our system placed emphasis on the combinatory possibilities of text-based and oral participation in a co-located group communication environment. Unlike Cognoter, however, we

\[1\] Much of the contents of this chapter are drawn from [Harry et al., 2012], which represents collaboration between the author of this dissertation and Eric Gordon.
focus on enhancing the performative space of group interaction. Consequently, our goal was not to create alternative communication channels; instead, it was to expand the space of performance. As part of this conceptual shift, we argue for moving from the metaphor of a “front”, spoken channel and a “back” channel to the metaphor of a “main” performance stage and “side” performance stage.

This study seeks to understand how having simultaneously accessible stages in the context of a group discussion affects the methods and outcomes of participant engagement. We start by describing how our system, called Tin Can, is related to existing systems that similarly augment face-to-face communication. We introduce the idea of stages and contrast it to previous models of channels. We then describe in detail the critical design elements of the Tin Can application and the class context in which it was implemented. Then, we present the results of the study, based on class observations, process traces, and interviews. In our discussion, we return to the concept of stages and describe how this formulation of participation can be productive for thinking through how people can interact with additional stages introduced into the dominant context of face-to-face communication. Finally, we discuss some specific insights about the tablet as a platform and describe the extent to which the design met our initial goals and how our results compare to those from past findings in the literature.

5.2 Related Work

There is a rich field of research on the topic of augmenting co-present group communication with socio-technical systems. One significant area of investigation concerns how systems can “level the playing field” of face-to-face communication through reflecting information about a group’s behavior back on itself. Karahalios and Bergstrom [2009] refer to this as a “social mirror”; a real-time visualization of social dynamics that is shared by the whole group and can cause changes in group dynamics. They suggest that “social mirrors become another channel for interaction (or a back channel) and, in the process, become a signal that influences interaction.” Their exemplar social mirrors measure behavior in an audio channel and visualize different aspects of it on shared Displays.² This strategy is shared by Second Messenger [DiMicco et al., 2007] and Meeting Mediator [Kim et al., 2008]. In these systems, presenting real-time participation visualizations tended to close the gap between over-participating group members and under-participating members, although in most cases this effect was primarily from over-participating members decreasing their participation. This work demonstrates how visualizing main

² There are a series of projects in this stream, including table-top interfaces for visualizing machine-recognized topics [Bergstrom and Karahalios, 2009a], reaching consensus through discreet voting [Bergstrom and Karahalios, 2009a], or balancing relative participation in group conversations [Bergstrom and Karahalios, 2007].
stage spoken participation in different ways can impact relative participation rates by encouraging individuals to censor or otherwise alter the nature of their communication to correspond with perceived group norms and group behavior.

Another strain of this work is less interested in altering the oral channel of group communication, and more focused on creating separate, productive backchannels. There are a variety of contexts where people have added communication channels. Yardi [2006] describes how a chat-based backchannel operates over a semester in a class, McCarthy et al. [2004] describe a similar approach at a conference. This early work focuses on characterizing the kinds of use that occur in backchannels using existing systems like chat, but do not engage specifically with design issues in backchannels. Harry et al. [2009] propose a new design for projecting question-oriented backchannels in panel presentations. Yankelovich et al. [2005] discuss oral backchannels during remote, audio-conference-based meetings, and the “social translucence” research stream (Rendezvous [Kellogg et al., 2006] is most closely related to this work) explores the design of systems to represent engagement in different kinds of mediated social situations.

Tin Can is designed specifically for use in a class, and is thus influenced by the systems designed for this specific context. Like much of the backchannel work described above these systems are typically concerned with creating new channels for communication in, for example, a large lecture hall. Bergstrom’s lecture class system [Bergstrom and Harris, 2011] supports question-asking and commenting and the Livenotes project [Kam et al., 2005] supports taking shared notes on lecture presentation slides. The ActiveClass project [Ratto et al., 2003] creates a channel between students and instructors for asking anonymous questions during a lecture from PDAs. Each of these systems seeks to increase participation in very large group settings by establishing separate channels for participation. Work in this space is typically not focused on directly influencing spoken participation because the expectation is that there is none; the lecturer is (except for question-asking) the only legitimate participant.

Classroom 2000 [Abowd and Brotherton, 1998] created a ubiquitous computing environment to create rich records of a lecture. Tin Can shares Classroom 2000’s interest in generating a record of the class, but takes a decidedly low-sophistication approach to it, relying on members of the class to generate the appropriate metadata instead of a broad technical infrastructure. Although questions play a role in our system as well, they are in-the-moment guides to discussion and not primarily aimed at the professor (indeed, the professor asked many of the questions in the system). Classroom 2000 is focused pri-
marily on lecture situations, which have very different interaction dynamics than the discussions Tin Can supports. Classroom 2000 focused on adding value (in our terminology) to the main stage performances of the professor, and are not trying to create credible side stages for student interaction.

In Tin Can, we take a different approach. We seek to expand the stage of participation by diversifying the sites of performance. In other words, we are not interested in creating better oral or better text-based channels; instead, through their correspondence, we seek to create a rich environment for participation composed of multiple, simultaneous stages. Perhaps the earliest research to explore this sort of approach is Tatar et al.’s [1991] work on Cognoter. They pointed to some interesting problems with creating stages, namely the limits of the “parcel-post” model of communication—where a message is sent and subsequently received and interpreted. Although this model works well for written correspondence, they found that within Cognoter, because the written contributions were designed to be interspersed with verbal dialogue, it was difficult for users to understand them “within the time frame of the actual communication” unless the oral conversation paused while written contributions were processed. In other words, in actual practice, users resorted to channel switching in order to accommodate the written or the oral modes. Another example is the Thoughtswap [Dickey-Kurdziolek et al., 2010] project which takes a much more structured approach by interspersing periods of engagement with the system with freeform discussion. In Tin Can, we employed a similar model of written communication to Cognoter and Thoughtswap, but we found that we were able to successfully create simultaneous stages for participation instead of stages one at a time. We will discuss the reasons for this disparity at more length in the conclusions. Work in this space on how alternate communication channels are selected and used owes a clear debt to Ochsman and Chapanis’s [1974] early work on mediated collaboration.

5.3 Setting the Stage

The research questions and goals for the students’ experience using Tin Can are captured in the main stage / side stage model. We wanted to create a side stage experience where participation was viewed as a legitimate part of class discourse and had a clear impact on the oral discussion. This is in some ways a radical strategy: why should we add technology to a classroom discussion if we want people to be more involved and attentive?

Traditionally, educators have accepted the quality and sufficiency
of the main stage in small groups and eschewed the addition of other communication channels because they might be distracting. A seminar class already adheres to the gold standard of face-to-face communication. It is often assumed that the pressures of the performance are exactly what we want them to be—students talk and the professor evaluates. But there are two faulty assumptions here. The first is that engagement in a co-present discussion can be manifested only in established methods of performance, for example, speech. And the second is that all students are equally capable of convincing performances. Social psychology research suggests that introverts rely more heavily on written communication to express themselves [Ross et al., 2009, Wilson et al., 2010]. But when there is only one legitimate kind of performance in a class, when there is only one way to perform on the front stage, the structure of the learning environment may not be as equitable as it could be, and it may not even be as productive as
possible, even for extroverts.

Consequently, when designing for the seminar, we sought to intervene in the established norms of the front stage by adding a well-crafted additional stage. The goal was to create a context for the legitimate performance of the back stage without having to go away from the front stage. As such, we move away from the front/back distinction, preferring the notion of a main stage with a side stage. We sought to design a system where performers could be on both stages at once, where performances were simultaneous, not alternating. Furthermore, the front and back stage as Goffman uses it implies different audiences for front and back performances. Moving to main and side stages reinforces the shared audience of the two stages, which has a big influence on how people perform on each stage and makes it easier to integrate those performances in a meaningful way. We use the terms main stage (face-to-face, spoken conversation) and
side stage (text input) to explain the context of performance created as a result of using Tin Can.

5.4 System Design

Tin Can is a tablet-based application to support class discussions. It provides a synchronous environment shared by the students and professor and each user has his or her own tablet. Students are physically co-located with the professor. Students arranged their tablet in different ways. Some keep them on their laps, some on the table in front of them. All users (including the professor and researcher) have the same capabilities in the system. The system serves as a visualization of the current state of the group discussion. It focuses on three main parts of a class’s process: topics, time, and ideas. Figure 5.1 shows the interface in action.\(^3\)

5.4.1 Topics

The topics pane in the UI collects past, current, and potential future discussion topics. These topics can be added using the “Add Topic” button at the bottom of the pane. The current topic is highlighted in
a topic-specific color. All topics have a short text description. Past and current topics show a kind of clock pie chart, illustrating the start and end times of a discussion topic (or the current time in the case of ongoing topics). The total duration in minutes of past and current topics is also shown as part of the topic text. Topics can be tapped to bring up an interface for changing their state: starting future topics, stopping current topics, and restarting past topics.

5.4.2 Time

The clock in the center of the screen serves as a reminder both of the current time as well as a concise visual representation of the history of discussion topics covered in the class. The time spent on each topic is swept out radially on the clock such that large blocks of color represent topics that occupied a longer period of the class. When an hour of time has passed, the central area in the clock is cleared and the colored record of the previous hour appears at the edge of the clock. Up to four hours can be easily represented in this way. The clock is non-interactive.

5.4.3 Ideas

The ideas pane contains a time-sorted list of ideas. An idea is simply a text contribution. Although we had presumptions about what would be posted here (as indicated by the terminology we used in the interface), ideas evolved to include statements, questions, recording main stage discussion themes, and a simple Twitter-like reply syntax. When entering an idea, the author of the idea could do one of two things: “add idea” or “add idea to group.” The former option would store the idea in the user’s “personal” idea drawer. The latter option would immediately put the idea at the top of the group idea timeline, as well as adding it to their personal drawer. Users tap and hold to “like” an idea. The idea flashes and a “+X” notation appears in its text, where X is the number of likes that the idea has received in total. Ideas in the group timeline have their author’s name displayed in parentheses after the text of the idea. Ideas are colored based to match the color of the current topic.

5.4.4 Users

Each user logged in to the system is displayed on a tab around the edge of the screen. The arrangement is essentially random. Tapping a user extends that user’s idea drawer. This drawer contains all ideas created by the user, whether shared or not. These ideas are differentiated in the list by “(shared)” being appended to ideas that have been
shared. Any unshared idea in the idea drawer can be dragged from the drawer to the group idea area, even if the user didn’t originally author the idea. Ideas dragged by other people are attributed differently in the main timeline. For example, an idea created by Alice and shared by Bob would say “(Alice, shared by Bob)”. By design, personal folders are not private. They are semi-public spaces meant to give users some choice in how their contributions are read by the group.

5.4.5 Archive

All ideas and topics are recorded on the server. At the end of each class session, the server emails everyone who attended the class with a list of their personal ideas and a link to a shared Web page that include a list of all student ideas sorted by topic and by user.

An example of such an archive is shown in figure 5.4.

5.5 Research Context and Methods

We deployed the Tin Can system in two sections of a graduate seminar on media and social theory taught by one of the authors at a liberal arts college. One section met in the morning, the other in the afternoon, twice a week. Class assignments were reading-based. Each discussion class was usually lead by a student or pair of students. While what it meant to lead class changed somewhat over the course of the study, the pre-study norm was to prepare a slideshow and accompanying media (images and video were quite common) and present it to the class. The morning session had eight regular students and the afternoon session had eleven regular students. In total, thirteen students were male, six were female. There were five non-native English speakers in both classes.

Our study lasted for six weeks and utilized mixed methods, including classroom observations, capture of text inputs, and semi-structured interviews.

Throughout the deployment, user interactions were captured. This did not, notably, include live recording of class audio, only text-based interactions with the Tin Can system itself. We made the decision not to record audio because we felt that this would make students self-conscious and would be too disruptive to main stage interactions. In lieu of audio recordings, for most class sessions, a researcher other than the professor was present to observe the class. We employed a form of direct observation known as continuous monitoring, where the researcher documented everything he saw throughout the study period, including the description of the environment and partici-
Figure 5.4: A screenshot of the post-class archive.
pant actions, as well as inferences about their meaning [Montgomery and Duck, 1991]. The researcher’s observations were not prescribed before the study, because of the exploratory nature of this first deployment. We did not know what to expect, so the observations were designed to be generative and not conclusive. We documented patterns of student attentiveness to peers and professor; interactions with tablets (i.e. position of tablets on the table or in laps); and correlations between speaking and writing. The students were aware that their use of Tin Can was being studied and they were aware of the presence of the researcher. Because the students were invested in the use of Tin Can, the presence of the researcher was not disruptive, but instead added to the excitement they had about testing a new system. The field notes were recorded by hand and subsequently transcribed and shared with the teaching researcher.

All text inputs into Tin Can were recorded over 22 hours of usage across twelve class sessions. Each class was about two hours long, but classes often had a non-discussion logistical content from the professor at the beginning of sessions. The average Tin Can session was 105 minutes long. After the deployment, the inputs were categorized into types, including topics and ideas, shared and non-shared.

Finally, at the conclusion of the discussion-based component of the class, the researchers conducted semi-structured individual interviews with fifteen of the nineteen students (79%). Interviews were conducted by the non-teaching researcher to alleviate student concerns about sharing judgements about the teaching researcher, although there were no sections of any of the interviews that students did not want to be shared with the teaching researcher. All interviews were recorded and transcribed and entered into Dedoose, a qualitative analysis tool. Because we view this work as generative, we iteratively coded the transcribed interviews, letting themes (and codes) arise organically as we reviewed the interview data, observational data, and process traces from the class. This strategy closely resembles Glaser and Strauss’s [1967] grounded theory approach. All names mentioned in interviews or shown in screenshots have been obscured or changed to pseudonyms to protect the identities of those participating in the research.

In the results section to follow we will analyze the kinds of content entered into Tin Can. The discussion section will then connect the interviews and observations to these findings.

5.6 Process Traces

Over the course of the deployment of Tin Can, 839 ideas and 119 topics were created. The majority of ideas created were shared: 72% of
Figure 5.5: Distribution of ideas per participant per idea type: personal, promoted, and shared.

Figure 5.6: Timeline of topics and ideas posted in two example class sessions. Ideas posted by students are in black, ideas from the professor are in red. The height of the line indicates the length (in characters) of the idea. Topic names are shown above their corresponding section of the class.
ideas were shared on creation. Another 5% were personal ideas that were turned into public ideas by being dragged by another user from a user’s personal idea drawer to the public idea area. The balance, 23% of ideas, were never shared. The distribution of these different idea types on a per participant basis can be found in figure 5.5.

Over the course of the study, 119 topics were created in total. Of these, 79 were actually discussed and the remainder were raised as potential topics but never actually used. The average class had 6.5 started topics, with a standard deviation of 2.3. Topic duration had a much wider variation: the average topic lasted 851 seconds with a deviation of 673 seconds. This is skewed largely because of very short topics.

A deep analysis of temporal patterns within the data is beyond the scope of this paper, but to provide a rough sense of what topic and idea posting activity looked like we have provided two example timelines in figure 5.6. The most striking thing about the data from this perspective is that the temporal distribution is very uneven. In Kleinberg’s terms, we see a bursty structure [Kleinberg, 2003] in idea posting. This is a common structure in communication systems and it is no surprise that it appeared in our study results.

5.6.1 Ideas

Over the course of the semester, ideas were adapted for many different purposes. Based on a review of the captured ideas, we organized them into a set of broad categories.

Statements were the most common type of idea. Statements capture an argument or idea, like “Talking about sex is a means of controlling it.” These contributions are similar to what someone might say if they had a speaking turn. Alternatively, they sometimes represented note-taking behavior. Students were also particularly fond of “X vs Y” dualisms, which we describe as Theme ideas. Like statements, these often had a dual purpose of either proposing a useful dualism or capturing the nature of the current discussion.

Questions were about half as common as statements. They usually took a rhetorical form, like “Where do you draw the line between not being sexually repressive and being excessively open about it?”, or were framed as less forceful variants of statements like “Is identity a sign? Or is it that which is signified?”

Early in the study, students developed techniques to address their ideas at specific other students. Using “@” syntax, as in “@dan, I don’t think the two are mutually exclusive.” they could respond to other students’ ideas non-orally.
5.7 Discussion

Much of the existing literature on backchannels focuses on situations in which the audience has one stage to perform on and the presenter has a separate one. [Yardi, 2006, McCarthy et al., 2004] Tin Can’s symmetry, where anyone can use either stage at any time, is the focus of our analysis. In this section, we use the stages metaphor introduced earlier to describe how students managed their attention to stages. Then, we discuss how students made decisions about when and how to perform on the available stages.

5.7.1 Attention

As described in Chapter 2, attention is best understood as a socially situated performative behavior. In a lecture, attentiveness is typically demonstrated by looking at the presenter and potentially taking notes. In a small group discussion, attentiveness might be expanded to include looking through class reading material or looking at other people.

Traditional approaches to managing attention in education tend to take as their starting point the desire to maximize audience attention on the presenter through the physical architecture of lecture spaces, presentation media, and rhetorical strategies. Gordon and Bogen [2009] argue that attention and distraction are best understood as being “hand-in-hand. The very same new technologies and landscapes that cultivate a state of distraction are themselves directed simultaneously toward the cultivation of attention.” Educators tend to look to technology (broadly construed) to help manage overwhelming sensory inputs while simultaneously blaming the lack of attention of students on that same technology.

Of course, it is not simply a definitional matter to decide what constitutes attention in a class using Tin Can. To understand how students and professor understood attention in this context we can look to how they talked about attention and distraction. Students predominantly viewed participation as obligatory. Speaking to his motivation to use Tin Can, one student said “to be part of the class I had to use it.” Students were never admonished for interacting with tablets, and didn’t report feeling like they needed to minimize their performance on the side stage to avoid negative perceptions of themselves by the professor or others, except to the extent that students felt like over-participation on either stage might crowd out other students. A student who was particularly active on the side stage worried that “I take so much space that people that are shy...have more problems ... standing up when they have personal
To the extent that students were concerned with attention, the most common problem was not being viewed as inattentive, but struggling to track performances on both stages simultaneously. Although students were not concerned with annoying the professor, they were worried about offending their peers who were presenting that day: “It is a little tough to keep your attention on both [stages], and sometimes you get a running conversation on Tin Can, which can be interesting but it is maybe a little unfair to the presenter.”

Although this perspective represents a pull towards enacting traditional models of attention, it wasn’t enough to significantly diminish involvement (either as a performer or an audience member) on the side stage. Student presenters often used Tin Can as a way to gauge interest in future discussion topics and to decide on whom to call.

In resolving the conflict between compelling simultaneous performances, students could fall back on the persistence of performances on the side stage. In making a choice to attend to the main stage, they could, in Josh’s words, “have a comfort that you’re not going to miss anything because you can always go back and see other people’s posts whenever.” Still, there seemed to be a difference between browsing posts later and being part of the live conversation. This came up most frequently when students expressed frustration with text entry on the tablet and missing the right moment to post something: “I didn’t get it out as fast as I’d hoped and it was already passed and it wasn’t worth typing it anymore.”

Deciding between stages was really only a problem when both stages were compelling. If only the main stage was compelling, students could freely attend to that. The reverse was also common, and students frequently reported attending to the side stage as an escape from an un-engaging main stage, as in this quote from Quinn:

“I can remember a particular ... presentation that he was doing a lot of PowerPoint, I think he was completely oblivious to the Tin Can conversation and [the Tin Can conversation] ended up going in a very good direction ... as a result, I do not remember anything he said, because ... the conversation on Tin Can was a little more engaging”

Moments like this highlight the extent to which our characterization of stages as “main” and “side” is itself a product of attention. The presence of a system like Tin Can does not automatically create a side stage, nor does the ability for spoken communication guarantee that such communication will create a main stage. The addition of a mediated communication platform simply creates the possibility of a new stage. Whether or not it becomes a viable stage, and whether the mediated stage is a main stage or a side stage is all the result of people’s attention to and actions in the system. Furthermore, the
The situation Quinn describes is a moment where the main stage ceased, for a little while, to command people’s attention and Tin Can took on some main stage properties. Although such moments were rare, they point out how stages are not created by technology or decree: they are designated and sustain by the collective attention and action of people using them.

The professor’s high level of activity on Tin Can throughout the class can be seen as playing a role in setting the main/side distinctions. His activity was a way of giving students permission to take the side stage seriously, both because it was clear that he was going to notice contributions from students, but also because he was frequently entering ideas himself and not looking at the current speaker. This underlines the extent to which this was an ideal context for testing a system like Tin Can. Had we deployed in a class where the professor was neutral or hostile to people attending to Tin Can, traditional class expectations of attention would more likely have been practiced by students, reinforcing those norms and making a side stage much less viable.

There was a moment towards the end of our study when the professor brought in a colleague over video chat to discuss his work and answer questions from the class. The remote presenter had a very limited view of the room from the professor’s laptop video camera and could see only a few students. Although the Tin Can system was available for this section of the class, it went almost entirely unused. This may simply be because the activity on the main stage was engrossing, but the total lack of side stage performance was still well outside of the bounds of normal disuse during a particularly engaging presentation in class. This suggests to us that the students were concerned with enacting the traditional model of attention for this outsider to the class. He could have viewed intense tablet use (something that was normal and viewed as attentive during normal class sessions) as inattentive or disrespectful and so his presence (even though his view of the classroom was quite limited) triggered a reversion to the more restrictive expectations of attention in a traditional class context.

During class sessions, we also conducted focused observations of student attention. The researcher would pick a student and mark down when they changed what they were looking at. Although this was not comprehensive data for all students, when students weren’t looking at the current speaker they were predominantly concerned with their own bodies and clothing, not the tablet. The tablet hardly dominated their attention. Even among the most active users of Tin Can, their attention was usually on the speaker and shifted to the
tablet during periods of silence or topics they found less interesting. When the professor spoke, students were far more likely to look at him than when other students were speaking.

5.7.2 Performance

The presence of an additional participation stage complicates the experience of being a member of the class. When should you submit an idea on the tablet rather than say it out loud? When is the right time to say something? Should you share an idea or make it a personal idea? The enacted (and self-reported) answers to these questions can provide some insight into the experience of using the system as well as deepen our understanding of the stages metaphor. In many cases, students viewed the side stage as complementing the main stage and valued its presence in situations where a range of problems with the main stage impeded their participation. In this way, Tin Can acted as a kind of escape valve: when the main stage was working for people, they used it; when they felt like they could not use it or did not want to use it, they turned to the side stage and valued its complementary affordances.

Performance on the main stage was widely viewed as more challenging and having higher stakes than side stage performance. Among the students who were reluctant oral participants in class, this was particularly acute. Geoff, a very rare oral participant in class before Tin Can, was particularly frank on this point: “I don’t really talk a lot in class because I’m scared of sounding stupid.” Geoff was a more frequent side stage participant. Although he would still rarely speak up directly in class, he was often called on by others in class to speak about ideas he had posted on Tin Can; he would happily speak in those instances. This change in behavior on his part was frequently brought up by other students as being a major benefit of using Tin Can because they valued the opportunity to hear and see what he was thinking. Irene, a more talkative student, characterized Geoff as a member of a “good chunk of people who I think are thinkers and they would just think and write down what they were thinking” as opposed to speaking on the main stage. This feeling was common among people who were comfortable on the main stage, who acknowledged that “not everyone feels comfortable speaking in class, so I think [Tin Can] definitely allowed for certain ideas to be shared that probably would have been either suppressed or just ignored or forgotten.”

Students’ comfort with the different properties of the main stage and side stage influenced which stage they chose for a performance. Students for whom English was a native language were more com-
fortable in spoken conversation, and when faced with complex ideas preferred to express them orally, turning to Tin Can to express simpler ideas because typing complex ideas was slow. Students who did not speak English natively had the reverse logic, preferring to type complicated ideas so they could, according to Rong, “organize my language a lot before I actually talk because I want my thoughts to be systematical and clear, I want people to get it.” In both groups, though, students viewed the revisability of written contributions as a potential benefit: “[Tin Can] gave me the advantage of thinking it through in a writing sense a little bit before I vocalized the idea.”

A lack of confidence about one’s performances was not the only reason to choose the side stage over the main stage. Students had a clear sense of etiquette surrounding when they could participate on the main stage and in what ways. Because only one person could be talking at once and conversations were fundamentally linear, students often felt like speaking up themselves would be changing the flow of the conversation in an inappropriate way. Instead, students would prefer to write their comments on the side stage instead of “interrupting” on the main stage. This was intertwined with ideas of timeliness. Performances seen as being closely related to the current main stage conversation were more appropriate than performances that might drag the conversation in a significantly different direction. Although similar, these concerns are not precisely the same. The worry about interruption was primarily a desire to not unduly influence the path of the conversation because that was perceived by some students as the role of the professor or presenter, not the role of the individual student. In contrast, ideas that were seen as “not quite as relevant [and not] really [fitting] into the conversation” were not really valid performances on the main stage at all because not only would they move the conversation significantly, they did not necessarily have anything to do with the existing main stage conversation.

Both of these worries, though, led to the same thing: increased use of the side stage. Because turn-taking was not a concern on the side stage, it easily supported ideas going in different directions simultaneously. To the extent that those directions were interesting to other people, they could serve as the basis for future ideas. If they were not, it was not seen as problematic to have put them there in the first place. When an idea did not seem to lead to any future ideas, students “didn’t think anything of it. Not all ideas are great.”

5.7.3 Sharing and Promotion

Key to our argument about stages is moving from a model where we view people as “tuning in” to a single channel to one where we
recognize that computer-mediated communication systems offer new simultaneous stages on which we can perform and be observed. It is critical, then, that we describe how performances shifted between stages, influencing what students said and how they said it. One common pattern was the positive reception to ideas on the side stage encouraging those ideas to be performed on the main stage. This same process happened even within Tin Can, when personal ideas were dragged by someone (usually the professor; 57% of ideas promoted from personal to shared were promoted by the professor) to the public idea timeline. Students also viewed “likes” and replies as good indicators of interest in their ideas. Geoff, the quiet student discussed earlier, captured the impact of these promotions nicely: “At first I started just putting them in my box [i.e. making them personal ideas] without even sharing with the class. Then I saw that [the professor] started dragging them out and putting them in discussions so afterwards I was more open to sharing my ideas within the class discussion.”

Activity on the side stage was sometimes explicitly moved onto the main stage. In most cases, the professor or student presenter called on someone based on something they had said on Tin Can and asked them to re-perform the idea on the main stage. The professor might say, for instance “Olivia, you had a nice point here on Tin Can, do you want to expand on it?” and Olivia could elect to take a speaking turn (and nearly always did). The other common strategy was for a speaker (particularly a student presenter) to use an idea recorded on the side stage as a starting point for a comment of their own or to introduce a topic known to be of interest to students based on side stage activity. Promotion moves by the professor were valued over those by other students, but both were appreciated and clearly remembered by students.

Part of our goal with the system was to use participation on Tin Can as the basis for an archive. Students were aware of this goal, and it was reinforced by the emails they received after class with a link to the shared class record of topics and ideas generated in class. Students occasionally talked sometimes about their ideas as an attempt to “take notes”, but more often they viewed that as the professor’s role. Not, perhaps, because it was a natural role they a priori expected of him, but because it captured (in their view) his observed usage.

In the stages framing, we can understand note-taking as moving performances from the main (oral) stage to the side (text) stage. For example, these were ideas entered into Tin Can: “Play is no longer having fun, it is work”, “Question of self-efficacy in public sector” or “Consumption leads to feeling good about yourself.” When posted
by students, ideas of this form were frequently attempts to move the discussion, but when they were posted by the professor they were seen as records of the main stage. Students characterized the professor’s role in this process as “the note taker person so if ... the presenter said something [the professor] would summarize what they just said.” This was seen as a valuable contribution by the students that showed interest in reviewing the archives after class: “I liked the way [the professor] used it. Because that also meant that I didn’t need to take notes ... because he posted it in Tin Can and I could get access to that later.”

As is evident from figure 5.5, the professor was a significant outlier in terms of his performance on Tin Can and his participation clearly had a big impact on how students understood and used the tool. Beyond his role as a note-taker, students also viewed his performances as oriented towards trying to guide the main stage conversation in particular ways. Students characterized this use pattern as, variously, “guiding”, “influencing”, or “driving.” He was particularly interested in “[initiating] conversation”, primarily by posting thought-provoking questions like “Why do we feel responsible for a corp’s feelings?” or “What is the role of god in modernism?” Most students avoided starting or stopping topics (or proposing them at all), arguing that it was the professor’s job to do that, although some students took more active roles in administrating topics when they were in the presentation role. In total, 53% of topic-related state changes were done by the professor.

On the main stage, the professor was also a frequent promoter of side stage activity. Sam characterized the professor’s role in a particularly evocative way:

“I feel like [the professor] would be a speaker for people who couldn’t speak, you know. The fact that he was really into Tin Can, so he would read something that [a student] had written and be like oh, I want to quote this or talk about it and [act] as a spokesman for people who aren’t really comfortable speaking”

This underlines the professor’s role as a bidirectional bridge between the stages. By taking notes on main stage performances, he reinforced Tin Can’s note-taking role. By speaking out about side stage performances and drawing people into the discussion based on written ideas, he legitimated their side stage performances. It is very hard to imagine Tin Can being as well integrated into the class as it was without the extensive involvement of the professor. This does not, in our minds, diminish the contributions of this work. Although we cannot speak to how a skeptical professor might react to the system, having a fertile classroom situation gives us an opportunity to make important insights into the potential for this design space that
we might not have otherwise been able to access.

5.7.4 Hardware

While one could easily imagine Tin Can working on a laptop, its deployment on tablets substantially affected use and outcomes in a variety of ways. First, there is a simple visual benefit to using tablets. Unlike laptops, which can create strong visual boundaries between people, tablets lie flat (or nearly flat) on the table or in people’s laps. When organized around a rectangular seminar table, tablets do not disrupt sight lines between people. In general, laptops give people something to hide behind while tablets more strongly signal availability.

Accordingly, tablets offer less privacy than laptops. Participants can easily see when other participants are using the system, and typing is easily distinguishable from browsing other people’s ideas. Surprisingly, we frequently saw students looking at other students’ tablets while they interacted with them, even though everyone’s view of the space was the same. Students seemed to be interested in knowing how other people were using the system.

Because of the way the tablet program was administrated at the school, students did not have any particular ownership over a specific tablet. This inhibited any sense of ownership; students talked about the tablets as being essentially disposable, for example “sometimes the [tablet] would run out of battery and kick you off and you’d have to get a new one.” The benefit of this lack of ownership was that it limited the tablets’ non-Tin Can uses. Unlike a laptop, on which the Web and communication tools were a click away, the tablets were not personalized. Even using Web tools was tedious, because they had to log in to each one which was both slow and obvious to people around them.

The biggest challenge with tablets is data entry. The most frequent complaint about the system was how slow and difficult they found accurate text entry to be. Students complained about slow typing speed making it hard to post timely ideas (“I didn’t get [an idea] out as fast as I’d hoped and it was already passed and it wasn’t worth typing it anymore”) and distracting them from the main stage (“it takes time to type on the [tablet] and so probably it takes you away from the presentation sometimes”). We also saw a number of ideas correcting typos and autocorrect mistakes in previous ideas. These problems mitigate the system’s utility as a conversational stage and seemed to depress overall use.
5.8 Conclusions

In deploying Tin Can, we had two major goals: increase the diversity of participation and increase engagement. We feel that we were successful in each of these goals.

When judging participation, we consider activity on each stage. In terms of the main stage, we saw some evidence that people who might not have spoken up in class were prompted to speak by Tin Can. Most often, this came from the promotion processes described earlier. This moderately increased the diversity of participation on the main stage. Side stage participation was viewed by both students and professor as a legitimate way to be a class participant, and we saw much broader participation in Tin Can than we saw on the main stage. The distribution of side stage participation was relatively flat, setting aside the professor, especially when compared to the steep power law reported in a chat backchannel [Yardi, 2006]. Based on our discussion of how and when students chose to participate, it is clear that the distinct affordances of the main stage and side stage meant that each captured kinds of participation that would not have been effective on the other. It is not the case that adding Tin Can detracted from the main stage and that there is a simple conservation of participation across all formats; we saw a more subtle case in which having a communication outlet with different properties drew out contributions that otherwise would not have happened at all.

We found little evidence of students making use of the archival records generated by Tin Can. The presence of the system did not, as we hoped, encourage students to participate on Tin Can because it generated a shared record of the class. In practice, students referred to the archive infrequently, and rarely reported that it influenced the kinds of ideas they wrote or how they wrote them. Although the archive was not frequently used, it did prove useful to some students. One student, Lien, valued the presence of the shared archive because “it [took] the pressure off of me. I don’t have to write down all of the notes.” Frequently, students said they had stopped taking written notes in favor of relying on the archive instead (even if they subsequently did not access the archive). Perhaps the lack of access should be seen as unsurprising because students may have been just as unlikely to review their own notes. Also, it may be because the course was a humanities-style theory course with a research paper and no final exam, students did not view the material as additive. If it was deployed in a skills-based course with more incentive for review, the archive may have played a more central role. Still, we were disappointed that it did not seem to motivate participation or reflection outside class.
This system was consciously designed for students who were less likely to participate orally in class. We were surprised at the extent to which attitudes about the system aligned along active oral participant / reluctant oral participant lines. Active oral participants tended to be indifferent about how Tin Can affected their personal participation in class. However, they acknowledged its effect on less active participants. Members of this group almost always commented on the increased diversity of involvement that Tin Can promoted, with observations like: “it tends to be a certain group of people who would talk and a certain group of people who were thinking but not talking. So I would like to see what they were up to.” Philip, an active oral participant with essentially zero Tin Can participation noted of reluctant oral participants “maybe [reluctant participants] would have something to say but maybe there’s like some sort of reluctance to actually to speak the thing aloud. So it gave another sort of channel to express ideas.” Knowing that reluctant participants had a place to participate made active participants feel less guilty about their own participation on the main stage.

Reluctant oral participants broadly relished the opportunity to participate in new ways with which they were more comfortable. Students described the system as “more efficient”, it “gave more people a chance to say things that they wouldn’t say”, and it helped students “feel more connected to the other students.” Olivia poignantly described the system as “something that was on my side, so to speak. You know what I mean? ... Like it was a resource.”

This dynamic between reluctant and active main stage participants suggests a new view on DiMicco et al.’s [2007] findings. They found that although visualizing participation decreased participation among over-participators, it did not boost participation of under-participators. In contrast, we found that although Tin Can did not decrease oral participation among active oral participants, it did boost oral participation among reluctant participants by letting them try out potential comments in a less intimidating medium and gather support for those ideas before speaking about them to a wider audience. Furthermore, if we include non-spoken participation, reluctant participants increased their participation substantially. This suggests that a lack of participation is not simply an issue of under-participators not finding conversational space to jump in, but can represent low conversational confidence that needs to be specifically addressed to boost participation.

We can also compare this boost in participation to Bergstrom and Karahalios’s [2009a] finding that under-participators on the oral stage were also under-participators in voting. Our findings suggest that if participation rates are strongly correlated, perhaps the votes do not
represent a different stage. This would fit with Tatar et al.’s [1991] findings with Cognoter; although Cognoter had more communication opportunities than Conversation Votes, participation in Cognoter nonetheless frequently stalled audio conversation while discussants processed the contribution. Simply providing another communication venue does not necessarily create a stage.

Our findings are also surprising in light of Tatar et al.’s [1991] analysis of the “parcel-post” style of communication. Although it would be fair to describe ideas in Tin Can as parcel-post, we did not observe any of the breaks in main stage participation resulting from submitted ideas that were observed in Cognoter’s use. Students frequently talked about waiting to read ideas when there was downtime on the main stage, something Tatar et al. view as a central problem with the parcel-post model in a face-to-face environment. We suspect that the main difference is group size. At small group sizes (like Bergstrom’s table-based work and the Cognoter studies) it is quite difficult to maintain separate stages because participation on the side stage is so conspicuous and attracts immediate attention. In Goffman’s terms, it is a venue where it is difficult to “get away with going away”. At larger group sizes like those we observe in this work, it is difficult to constantly participate on the side stage (either as a reader or a writer), so immediate awareness or consideration of all side stage contributions is simply not feasible. In fact, this lack of obvious immediate attention on each contribution could be a big part of why reluctant participants were more comfortable making side stage contributions.

Based on our study, we conclude that the introduction of a tablet-based system into a seminar classroom can have positive effects on the engagement of students and diversity of their participation without sacrificing the primary mode of interaction: spoken conversation. Unmediated face-to-face group discussions can be high pressure situations. Participating in face-to-face discussions requires participants to perform “live” (performances cannot easily be composed in advance); they also typically come with predetermined notions of etiquette to which one must adhere in order to preserve the integrity of the performance. We wanted to design for this context in order to take advantage of the rich texture of participant attention and performative attributes the seminar setting presents. Our study demonstrated that Tin Can gave a voice to students who otherwise were not confident enough to perform on the main stage or who needed to “test out” ideas before sharing them on the main stage. It also demonstrated a need for a new model of thinking about how computer mediated communication systems can work in face-to-face discussion settings. Stages are not tied to particular technologies, nor
are they predictive of specific behaviors. They are a way of conceiving of modes of communication within situations that are flexible and that emerge from participant performance.

From a design perspective, we believe this work offers several important lessons. First, it shows how the model of a collaboratively managed discussion dashboard can be an effective way both to generate a record of a face-to-face experience, but also to improve the quality of the experience while in progress. Furthermore, we demonstrated how tablets can be an effective platform for these kinds of interventions into otherwise un-mediated face-to-face experiences. Second, our work reinforces the “beyond being there” argument [Hollan and Stornetta, 1992] by demonstrating a system that improves on the experience of “being there” in a meaningful way. Productive interventions in face-to-face interaction will be useful not just in guiding the design of similar future systems (which we think this work shows can be a fertile design space), but also will suggest design techniques and strategies that can be effective in computer mediated communication tools for spatially distributed users. A system that is useful even when its users are face-to-face and could simply set the system aside and continue to interact without it is one that likely has something more general to teach us about the limitations of face-to-face interaction and the design of systems that aspire to replace it.
The projects described thus far: *Information Spaces*, *Presentation Spaces*, backchan.nl, and *Tin Can* are all interested in supporting audiences who are not merely passive receivers of information, but active participants in an experience. My interest in non-verbal actions as a design technique is the main way I accomplish this. This is an important and effective strategy because verbal participation becomes more challenging as the number of potential participants grows.

As discussed in Chapter 1, the constraint that only one person can talk at a time and the switching costs with synchronous verbal communication impose stiff costs on the engagement level of a group as that group size increases. I have sought in my designs to alleviate these costs by creating other ways to participate that don’t have the constraint of seriality, which in turn frees us from the costs of negotiating turn-taking. I have shown the various ways this can create experiences where people feel more engaged with the process, more connected to the audience, and feel like they have an impact on group process in ways that traditional “being there” approaches to mediated interaction have trouble providing.

Still, these approaches have limits. As audiences scale up beyond a few hundred (in the case of the largest backchan.nl events we’ve observed), the design approaches we’ve proposed start to break down.¹ Size also represents some intriguing research questions: How do we represent an online crowd? How do we create opportunities for online crowds to interact in some of the same ways that physical crowds interact? As always, my interest is in trying to find analogous new interactions, not to attempt a direct translation.

The challenges of scale manifested in some of my past projects, which played a role in my thinking on how to handle large scale contexts. With backchan.nl specifically, the first failure mode is that the “recent” posts section becomes overwhelmed and hard to keep track of. This causes many users to simply opt out of voting on new posts because they appear faster than they can practically judge them. Larger audiences also bring increased odds of abusive behavior like spamming and mass-voting for low-quality posts. One common ap-

¹ This is not specific to mediated interactions; interactions between unmediated communities change dramatically as the size of the community goes up, too. [Dunbar, 2010]
approach to handling increasingly large groups is to segment them into manageable chunks where our traditional techniques work well. This works, but it’s a bit of a dodge. This chapter will argue that there are ways to simultaneously create compelling small-scale experiences that provide synchronous text-based interaction (e.g. chat) while also providing a series of mechanisms that help a very large group stay aware of each other’s moods and interests and engage in various forms of collective activity that make it feel like you’re really part of a very large audience. Thinking about mediated crowds in this way brings up compelling conceptual questions like:

1. How do people find groups of people to talk with?

2. Do collective activities like chanting or doing the wave have online crowd analogs?

3. How do you manage antisocial behavior in online crowds?

4. How can you create opportunities for deeper engagement with the event that have an impact on other audience members’ experiences?

To address these questions, I will show how we can create meaningful small scale interactive spaces, identify common behavior across an entire crowd, represent those shared behaviors back to the entire crowd, and create a suite of other opportunities for members of the crowd to have shared communication experiences, all in a context that has some simple ways to manage anti-social behavior.

Constructing a sense of remote viewership is not a new activity. As radio, film, and television removed the physical constraints of audiences and performers being co-located, we were able to create enormous audiences all experiencing something together. Yet there is clearly something important about the presence of the audience. We still go to movie theaters to watch movies together, and TV shows frequently have live audiences (or simulate live audiences with a laugh track) to try to foster a sense of experiencing something with someone else.

Over time, the structure of these events has even evolved to take advantage of technology to create a sense of engagement and involvement. Shows like American Idol use text messages to allow audiences to vote for specific contestants. This is a relatively thin form of engagement: feedback is quite delayed, votes are essentially anonymous, and the pool of votes is huge which makes it hard to feel like you’re making a difference. Nevertheless, this is part of a long-term campaign on the part of broadcasters to try to make it feel like broadcast television isn’t simply receiving data, but trying to bring back the historical experience of being in a crowd with other viewers.
In this chapter, I will describe a system called ROAR that tries to develop the design techniques described and studied in past chapters towards audiences of extremely large scale. I will talk about related work in the social TV space as well as discuss the other sorts of tools that people use to create similar sorts of experiences. Finally I will describe the major components of ROAR’s design: sections, pulse, shouts, and feedback.

Unlike the main projects in this thesis, backchan.nl and Tin Can, I have not done any deployments of ROAR. This is primarily for practical reasons. It is quite difficult to reach audiences of the sizes for which ROAR’s design is specialized. Organizations with audiences of this size tend not to be interested in trying prototype code during a large scale event. Still, these sorts of explorations can provide valuable guidance about a design space that we might otherwise overlook because it is prohibitively difficult to deploy prototype code. As a capstone to a series of more elaborately studied systems, I view this chapter as a forward-looking description of future work that can show how the principles developed in earlier chapters could grow to address the needs and interests of larger public audiences.

With this shift towards larger and public audiences, I’ve also chosen to design for more practical technical infrastructures. Unlike with virtual worlds in Chapter 3, extra projectors and screens in Chapter 4, or per-person tablets in Chapter 5, the systems described in this chapter don’t expect elaborate pre-existing technical infrastructure and are instead designed to work with whatever platforms already have available. Practically, this means designing for desktop web browsers, mobile phone browsers, and tablet browsers. This also entails a movement away from more speculative design approaches that expect significant changes in behavior to designs that instead aim to emulate and extend existing practices in less radical ways.

### 6.1 Related Work

In the academic literature, social interaction around television has been a research topic for years. As with the broad interest in CSCW systems work of building systems to support experiences that had historically been co-present, the focus of this research stream has been on recreating the experience of sitting around a television talking about what’s on screen. This is the primary goal for essentially all of the systems discussed in this section.

There are a number of common threads across all the social TV systems discussed here:

- Content is viewed on a television
• The system is generalized, i.e. for essentially all kinds of TV content and supports awareness between programs

• A buddy-list system provides awareness of others’ TV experiences

• Co-watching is conceived as an intimate, interactive experience between small groups

• There is zero awareness of non-friends

Best practices about how to handle each of these basic features is discussed by [Geerts and De Grooff, 2009].

Where the systems tend to diverge is how exactly they try to create a sense of co-presence. The main distinction is the primary mode of interaction: symbolic, chat, or voice. In general, researchers agree that voice is a natural and effective way of promoting a sense of co-presence, but that voice has technical and environmental challenges to adoption. Chat can be effective for users who are used to chatting while doing other activities, but for unacclimated users it can be distracting and frustrating. [Geerts, 2006] This finding is mitigated, though, by other researchers who found that when only chat is offered, users report having satisfying experiences. [Regan and Todd, 2004] Symbolic interaction, like a “thumbs up” button or pre-canned messages alone are “simply insufficient and too impersonal for fostering a feeling of connection.” [Metcalf et al., 2008] One interpretation of these conflicting findings is that Geerts [2006] are less concerned with co-presence because of their interest in asynchronous viewing, so their research focus was not on whether or not people felt like they were watching with their friends, but was instead on simply whether or not they enjoyed marking up video.

Most social TV work responds to the rise of digital video recorders in some way. For systems that are interested in creating a sense of co-presence, the wide-spread availability of devices to both pause shows in real time and record them completely for later viewing poses a major challenge. CollaboraTV [Harrison and Amento, 2007] addresses this most directly, making synchronous and asynchronous viewing essentially identical by providing only a limited vocabulary for interaction to avoid the expectation of actual between-viewer conversation. Other systems that do aspire to supporting synchronous co-present experiences typically make a nod towards potentially time shifting interactions for later viewing by friends.

The final major axis of distinction is around the representation of viewers. Social TV [Harboe et al., 2007], Media Center Buddies [Regan and Todd, 2004], 2BeOn [Abreu et al., 2002], and ConnecTV [Boertjes et al., 2008] all use buddy-list metaphors throughout the experience.

This is mostly an artifact of the research context. It’s hard enough to arrange a group of 5-10 households in a social network willing to install and use the system, let alone a broader network of unconnected users. Few researchers would bet on the value of crowd interactions when assembling a crowd for experimental purposes is so challenging.

Figure 6.3: Screenshot of Media Center Buddies, from [Regan and Todd, 2004]

Figure 6.4: Screenshot of SocialTV, from [Harboe et al., 2008]
Other systems like AmigoTV [Coppens et al., 2004], CollaboraTV [Harrison and Amento, 2007], and Vision Television [Bove] focus on using either real or iconic versions of faces to represent viewers. Most of these support ways of animating those avatars in ways that indicate an emotional state in response to the program.

ROAR sets itself apart from this past work in a number of ways. Due to technical constraints in the period when this work was most popular, almost all social TV systems used the TV screen as site of the user interface. ROAR’s focus on supporting interactions around video watched on a computer is a marked difference in approach that has a number of benefits: text input is easier, the interface can have finer details because of increased resolution and shorter viewing distances and multi-user scenarios are quite rare. This was, of course, an approach that was historically possible, but the recent availability of wide-spread access to streamed versions of live events has made this venue more viable. Furthermore, consensus seems to have emerged that interaction in a traditional TV context is going to be tablet and mobile phone centric, not taking place on the TV screen itself.

The other most important difference between ROAR and past social TV work is its interest in supporting awareness among all viewers, not just between viewers with a pre-existing relationship. There is some work in this space, like Visual Backchannel [Dork et al., 2010] that focuses on providing reflective tools for analyzing and browsing through tweets with a particular hashtag. But there is little work in trying to build interfaces that merge participation with real-time shared visualization. The co-incidence of participation and visualization is a critical component to grounding, as discussed throughout this thesis. If the visualizations of group participation are not part of everyone’s shared display, they lose value as a site for collective action - one of the main goals of ROAR’s crowd visualizations.

Although ROAR is interested in co-presence, it aims to live in somewhat more ambiguous social settings where groups are made up not necessarily just of close friends (where voice might be comfortable) but with somewhat larger groups where 5-10 people might co-watch with different levels of engagement. Although voice is a richer experience, it has relatively hard limits in terms of numbers of active participants. Furthermore, voice tends to be a higher intimacy channel, which is less useful in contexts where not everyone necessarily knows everyone else. Thus to support co-presence, ROAR focuses on chat and some other associated features like polls and drawing. This focus on co-presence precludes creating in any serious way an asynchronous version of ROAR. Indeed, Harboe et al. [2008] note that among some mixed results, the strongest experiences were in live events like sports.
6.1.1 Other Conversational Tools

Social TV research has faded somewhat in popularity since the mid 2000’s. There is a widely-held assumption that conversations around live events are going on not in TV-specific interfaces like those built by researchers, but instead on existing social networks like Facebook and Twitter. In this model, when someone wishes to interact with friends or strangers watching the same event as them, they simply turn to existing social network channels. If this were true, it would suggest that ROAR’s approach is deeply flawed, and instead our focus should be on building systems that use Facebook or Twitter as a substrate. This approach would look much more like Visual Backchannel [Dork et al., 2010]. However, I will argue in this section that the social spaces created by Facebook and Twitter are not (and barring fundamental changes to their mechanics cannot become) spaces for truly interactive discussions about a live event. Instead, we need a new platform to be the home for these conversations.

As I’ve argued about grounding elsewhere in this thesis, creating social situations with mutual awareness is critical to creating effective conversation spaces. If there is any concern that others can’t see what you’re saying, it breaks the cycle of awareness that is critical for moving a conversation forward. The publish/subscribe model adopted by and Twitter gets in the way of this once conversations scale beyond just one person. If Bob follows Alice, and Charlie follows Alice but not Bob, Charlie will see only parts of a conversation that are not directed at Bob with an “@” reply. 5

The nominal solution for this is hashtags, for example words prefixed with a “#” character. These mark a tweet as part of a larger set of tweets on a theme, and places that tweet in a stream in the so-called “discover” part of the interface. This interface has a similar grounding-inhibiting design. First, it is impossible to tell if others are actively viewing the hashtag in the “discover” mode. Whether or not that audience exists is essentially unknowable to posters. Hashtags are visible amongst follower networks (i.e. you can see hashtags that seem to be popular within your network and be involved in this conversations) but there is little evidence that participations cross existing follower/following relationships.

The problem of an audience for tweets with a hashtag is compounded because tweets with a hashtag do have one guaranteed audience: your existing followers. Every tweet, hashtagged or not, is inserted in the streams of all your followers. When balancing a potential imagined audience and a guaranteed one, people will write for the audience they know they have. For major events like the Superbowl or World Cup, we can safely assume that our audience is 5 I’m going to mostly ignore Facebook for the sake of simplifying the argument. The outlines of the argument are similar, but Facebook’s more integrated reply format avoids some of these problems. This is offset by the lack of reliability of posts being seen by friends; as few as 20% of your friends see any given post [Constantine, 2012].
likely to contain many people who are watching the same event. But even for moderately less popular events, it is far more likely that most of our followers are not co-watching and are unlikely to be deeply interested in a detailed conversation about what’s happening.

We can see that this is true by looking at the sorts of tweets that people write during live events like conferences. Ebner et al. [2010] try to categorize conference tweets into four categories: irrelevant tweets, administrative tweets, topical discussions, and topical tweets. The distinction between the latter two categories seems to be whether a tweet contains an “@” reply or not. The authors argue that without external information like a link or picture, a tweet is too decontextualized to be valuable to a non-attendee. I see their data differently. If we look at their exemplar topical tweets (excluding questions, whose audience is clearly the presenter), most of them are clearly written for an audience that does not share their context. For example, the tweet “nice idea of @estudyskills Aggregation of all student weblogs at Tumblelog - gives overview. #eciothh” is recording the content of a talk for someone who isn’t there. A simple evaluation-oriented tweet like this for a co-present audience would look a lot more like “nice idea of @estudyskills” and leave off the rest of the context. Indeed, this is the core difference between chat messages and posts to Twitter or Facebook: chat messages rarely make sense out of context, but posts are nearly always written with context included to make them comprehensible to a non-present audience.

While looking at how people write tweets versus chat messages is interesting, volume of participation is also interesting. We can look at summary data from Bluefin Labs, a company that captures and analyzes Twitter conversations about TV shows and compare the participation rates in Twitter conversations to similar chat conversations. For each show they analyze, they report the total number of tweets about the show and the total number of people who tweeted about the show. The ratio of these two values is the average number of tweets per tweeter. For the vast majority of shows they track, this value is between 1 and 2; the average viewer tweets at most twice about a single show. This holds true even for sports matches that can last 2-3 hours. It is intuitive that only one or two messages per person is a strong indication that there is not significant conversation going on; conversations simply cannot take place in one or two messages. Nevertheless, it is useful to compare these participation patterns with a social context that looks more like ROAR to argue for ROAR’s design approach.

Although we don’t have data for people using a system like ROAR yet, we can look to comparable experiences that already exist. Many sports fans use platforms like Internet Relay Chat to talk about live
events. Based on logs collected over the course of two weeks, we can calculate a metric similar to that reported by Bluefin Labs for conversations on Twitter. We can’t easily link specific messages to specific events, but we can measure how many messages each person sends per hour in which they send any messages at all; essentially for people who are chatting in a given window, how many messages do they send? This approximates the tweets per unique author per show metric. This measure varies pretty substantially between which chat room you collect data from. Data is reported in Table 6.1.

The most direct comparison to Twitter is the final numeric column on the right. Mean messages per active user-hour represents the average number of chat message a user sends in hours where they send any messages at all. If every user logged on, sent one message, and then stayed silent, this metric would be 1. It would also be 1 if every user sent one message per hour. As a result, the floor of this metric (as with the tweets / unique author metric) is 1. Larger values represent users who tends to send many more messages during hours where they are chatting at all. This metric is artificially depressed relative to the Twitter metric because our data covers not just moments when there are live events going on, but 24 hours a day. This is captured by the variance in the fifth column: the number of active chatters varies widely between active moments and inactive moments. Inactive moments don’t exist in the Twitter dataset because it focuses on only the tweets associated with a television show. Despite this, we see values ranging from 3.4 to 16.6, compared to Twitter’s 1-2. This suggests that chat contexts elicit a 200-1600% increase in per-active-user participation.

The large variance in participation rates seems (in this very preliminary review) to be correlated with the size of the rooms. The rooms with the highest per-user participation were also the rooms with the fewest active users. This supports the ROAR’s core argument that smaller interaction contexts will be more interactive than larger undifferentiated contexts. This might also explain part of the difference between Twitter and a chat-style system: Twitter looks much more like a single large undifferentiated room than a focused small scale social space.

This informal analysis underlines both the strong differences in

<table>
<thead>
<tr>
<th>Channel</th>
<th>Total Messages</th>
<th>Total Users</th>
<th>Mean Hourly Active Users</th>
<th>Mean Messages Per Active User-Hour</th>
</tr>
</thead>
<tbody>
<tr>
<td>#reddit-soccer</td>
<td>20,317</td>
<td>251</td>
<td>7.9 (SD=9.9)</td>
<td>16.6 (SD=25.5)</td>
</tr>
<tr>
<td>#football</td>
<td>11,936</td>
<td>199</td>
<td>8.8 (SD=8.2)</td>
<td>10.1 (SD=11.8)</td>
</tr>
<tr>
<td>#teamliquid</td>
<td>110,910</td>
<td>1,940</td>
<td>29.7 (SD=19.2)</td>
<td>9.5 (SD=18.1)</td>
</tr>
<tr>
<td>#joinDota</td>
<td>93,010</td>
<td>9,022</td>
<td>52.8 (SD=113.3)</td>
<td>3.4 (SD=9.6)</td>
</tr>
</tbody>
</table>

Table 6.1: Comparison of participation rates across different IRC chat rooms.
behavior in post/subscribe models like Twitter or Facebook compared to chat-based designs like ROAR.

6.2 Design

ROAR is a social platform that wraps streaming video of a live event. This section will cover the basic components of a web-based version of ROAR. An overview of the visual layout is shown in Figure 6.8.

6.2.1 Sections

I take the organizing metaphor of a “section” from the physical experience of being in a crowd at, for instance, a sporting arena. Audience members rarely come to a stadium alone: most people come with friends. Existing tools for watching live events with chat tend to simply group all viewers into one anonymous mass. As I argued in the previous section, this approach tends to depress overall participation because too much chat can crowd out deeper and broader conversations. In contrast, ROAR organizes viewers into many smaller sections. Each section can be named or unnamed.
When a viewer joins a ROAR-enabled stream, they are provided with a section browser. If they log in with a Facebook or Twitter account, ROAR can prioritize sections with friends in them. A draft of this interface is shown in Figure 6.9. For viewers who don’t want to log in with Facebook or Twitter, they can search for named rooms on a particular theme, like a school, employer, location, or team they’re a fan of. Conceptually, sections are like tables at a bar or circles of people at a cocktail party. You’re likely to know at least some of the people at a table you want to join, but not necessarily everyone. It’s easy to scan the room to evaluate options. When you’re in a section, most of your attention is focused on the interaction within that section, but there are a number of other systems that try to maintain some awareness of the crowd as a whole.

There are a variety of potential interaction styles within a section. Chat is the dominant one shown in this prototype, but chat depends on easy access to a keyboard. It is easy to imagine versions of ROAR that would make sense on a mobile phone or tablet device. In these contexts, text entry is somewhat more challenging, so having interaction models that require less text entry but that still support a sense of presence would be valuable.

There are two varieties of interaction that the ROAR prototype supports. The first is the “shout.” Shouts are composed less like chat messages, and more like tweets. They are intended to be shared out of context to more people. A shout might be a witty comment, a link to a funny image about the game, a clever insult to a hated player or team, or simply a message that captures the feeling of the crowd at that moment. When a shout is entered, it is shown first to viewers in the section of the author. Unlike chat messages, shouts can be voted for. Shouts that accumulate enough votes within their initial section will spread to other sections. Eventually, widely popular shouts could spread to the entire audience, while some will peter out having only been seen by a small fraction of the crowd as a whole.

The second variety of novel interaction is the creation of graphical signs. Although this doesn’t make much sense with a traditional keyboard and mouse interface, drawing is much more natural on tablet devices. For users in that context, drawing can be more expressive than typing. The creation of signs meshes nicely with existing fan practices around bringing signs to a stadium. Indeed, signs brought to the stadium are explicitly written to attract the attention of camera-operators and commentators with the hope of being broadcast to the remote audience and put on a live display within the stadium. In this way, the practice of sign creation is already implicitly acknowledging the attention and awareness of the remote audience. Empowering remote fans to participate in this practice seems like a

---

Footnote: There are a number of moderation challenges with shouts that will be addressed later. Shouts are also significantly rate-limited, and having previous shouts judged negatively by a moderator will strongly increase this rate-limit.
natural extension, especially on input devices where drawing is easy and accessible. An example of the drawing interface and a sign’s representation in the section stream are shown in Figures 6.10.

As a primary communication medium, non-audio communication has a number of major advantages for ROAR. Chat requires minimal extra hardware, can be done in loud places, is more easily analyzable at large scales, doesn’t require turn-taking, and requires substantially less technical infrastructure to support at scale. Still, it is far less immersive and communicative than voice communication. Although audio becomes problematic for large groups (for reasons discussed throughout this thesis), the potentially small size of sections might mean that for groups of 3-5 audio could be an effective addition.

6.2.2 Predictions, Betting, and Promoting

Although chat is the foundation of creating a sense of presence with other people in a section, this thesis has argued throughout for the value of non-verbal actions that provide lighter-weight senses of presence. These kinds of actions can help remind section-members of the presence and engagement of people who might not have something to say, but are still attending to the section and the event. In the context of ROAR, there are a number of different kinds of non-verbal actions that would be appropriate.

The most simple of these actions is making predictions. An event organizer could pose a question to an audience like “Which team will win the game?” These questions could be section-specific or global. Responses to a prompt like this can be aggregated globally or (more importantly) be represented within the section to catalyze a discussion. A simple version of this is shown in Figure 6.11. These questions could also come from section-members and be scoped just to the section.

These sorts of lightweight predictions make sense across a wide range of the sorts of events that ROAR aims to support. Live competitive performance shows like American Idol or The Voice have occasional moments where the audience could weigh in and change the course of the show. Sports-oriented events could offer an even more elaborate and high-frequency interaction. Instead of the occasional audience-wide poll, fans could place time-sensitive bets on the outcomes of various game events. Baseball provides a particularly effective example. Within a single game, there are a series of interlocking time scales where predictions make sense from the inning to the individual batter, to the individual play. At each of these levels, a viewer could opt to place a bet about the outcome of the event. These bets would be shared at the section level, and effective betters would...
be highlighted on a leader-board. While these kinds of actions don’t make sense for all events, they’re another design element that could be added or subtracted depending on the style of event.

All sorts of section and crowd-level participation can also support voting to guide the promotion of interesting content and the identification of inappropriate content. This serves as an additional stream of non-verbal actions for the engaged but perhaps not verbally inclined viewer. Votes are aggregated for shouts and signs, and used to spread those contributions to a wider audience. This serves to create a stronger relationship within the audience. By participating in identifying the sorts of content you like, it creates a sense of the community’s values represented in the top contributions as judged by viewers. It also supports a sense of ownership. Without voting, the acknowledgement of a crowd contribution on-air (in the way that cable news now frequently takes tweets out of context and discusses them on-camera) seems arbitrary and analogous to shouting requests to a band during the breaks between songs. Whether a message catches the attention of the performer or anchor is essentially arbitrary. But by being able to identify contributions that you see as valuable, you can buy into the success and quality of that contribution and celebrate its broader spread rather than rue the arbitrariness of a random selection process.

One of the major benefits of nonverbal actions like those discussed in this section is their ease of aggregation. As in backchan.nl, where voting serves as a simple low-impact way to participate which guides the system’s decisions about what content to show on the main screen, non-verbal actions like making predictions, placing bets, or voting for other kinds of content can be easily aggregated across all sections and represented back to the crowd as a whole. This creates a continued reminder of the crowd’s presence in the same space. Furthermore, ROAR could leverage the existing infrastructure in the physical spaces where events take place to represent the remote audience back to the co-located audience. Using the example of audience-drawn signs, we can imagine using a jumbotron in a stadium, for instance, as a venue to share remote audience content that has been voted up. This creates a strong incentive for creating and voting for content because, as in backchan.nl, the shared screen carries with it a guaranteed audience for your contribution. Having something you created or promoted appear in the main video stream in a physical space closes the feedback loop, and reminds physical spectators and remote spectators of each others’ presence. An example of this approach is shown in Figure 6.12. As with using shared displays in other projects in this thesis, using existing displays in physical venues can help ground the experience between remote and local audiences.
and create a stronger sense that they are sharing an experience.

6.2.3 Pulse

Section-oriented features focus on creating a small scale sense of community between small groups of viewers. While this is part of an enjoyable spectating experience, an awareness of the crowd as a whole is a critical component to the ROAR vision. More than any other feature in ROAR, the crowd visualization has evolved considerably over the design process. In this section, I will describe each of these designs, and explore the extent to which they satisfied the different sorts of awarenesses you can have of a physical crowd:

Volume  How active is the crowd right now, relative to the recent past?

Topics  What topics is the crowd talking about right now, and what are they saying about those topics?

Collective Action  More speculatively, might we be able to give the crowd opportunities to do some sort of activity analogous to singing or chanting?

Each of the designs described here is supported by a keyword detection algorithm that is essentially unchanged across the visualizations. The algorithm scans all chat messages and applies an adapted version of the Term Frequency, Inverse Document Frequency algorithm to identify words that are globally uncommon but frequent in the last few seconds of chat. TF-IDF depends on a notion of documents, and compares the term frequency within a given document to all other documents within the corpus. For a real time stream of data like chat, this obviously requires some adaptation. ROAR bins all chat into 10 second windows of chat, and treats those as documents as far as TF-IDF is concerned. This makes sense conceptually because we can assume that chat is about the video, and as the contents of the video shifts over time, the conversation about it will as well. This means we can reasonably assume that each window of chat will function similar to documents in a traditional TF-IDF model. Of course, this assumption is not iron-clad, and can lead to quite noisy results.

All the chat shown in screenshots of these visualizations is recorded from Internet Relay Chat rooms organized to discuss various spectator events like video game tournaments or sports matches.

The analysis engine provides a rank-ordered list on a per-window basis (10s) of the words with the highest recent TF-IDF scores. I call this list the “term rankings.” This is the input to each of the visualizations described here. It is quite common for the top words
in subsequent rank ordered lists in the previous analysis window to share very few words with any previous analysis window. This high amount of turnover poses a design challenge that each design deals with in different ways: how much inertia should be added to the visualization to dampen out the inherent noisiness of the term rankings?

The first visualization adopted a word cloud model. When a word appeared in the term rankings list, that word was added to the word cloud. The higher its position in the term rankings, the larger the initial size of the word. Each second, every word currently displayed in the visualization would decay in size slightly. If a word already displayed in the cloud appeared in a subsequent term ranking list, its size would be increased, correlated with its ranking in the list. If a word slipped below a certain size, it would fade out of the cloud. The cloud was rendered on top of the live video, ostensibly in a part of the screen that was not as critical to viewers. This approach is shown in Figure 6.13

This visualization had a number of valuable properties in terms of our three original goals. Volume and topics were both easily comprehensible in this format. In particular, the spatial stability of words (i.e. they stay in more or less the same place on the screen for their lifetime) made it easy to both track the performance over time of an individual topic and not be distracted over-much by constant movement. This approach struggles with the temporal component to the data, though. It combines terms that were historically popular but haven’t yet decayed enough to be removed with up and coming words. While the lack of a time component simplifies the visualization, it also makes it somewhat less effective for collective actions like cheering or chanting because earlier stages of a cheer are conflated with current ones. Perhaps the biggest problem with this approach is its space inefficiency. Responding to a demo video using this visualization style⁸, potential users expressed an intense dislike for any approach that attempted to overlay the video itself. As a result, all fu-

⁸ This visualization can be seen in motion in this demo video: http://www.youtube.com/watch?v=KK13_JE_NGg
Figure 6.14: A direct visualization of the term rankings. Larger words have higher rankings, and the total number of words is correlated to the volume rating at that moment. This list changes completely on each new term ranking list.

Figure 6.15: In this visualization, words appear on the left edge as they appear in term rankings and move smoothly to the right. Larger words move more slowly than smaller words.

Figure 6.16: Similar to the river-style visualization, in this approach words appear at the bottom and rise smoothly to the top. In this screenshot, there is relatively low activity.

ture visualizations are located at the edges of the screen using pixels dedicated for visualization, not attempting to overlay the video.

The other major drawback of the word cloud approach is the lack of semantics to the position of words. The next three visualizations use screen location in a variety of ways. The most direct approach was to simply display the term rankings along the bottom of the screen underneath the main interface bar, shown in Figure 6.14. In high activity periods, words lower down the term rankings were shown. In this way, the total number of words displayed functioned a bit like a volume meter, and the most important words always occupied the same part of the screen. However, the noisiness of the underlying data made updates distracting. A word could easily jump from the top spot to the bottom of the list between subsequent updates, which made it hard to track the life cycle of a term you were tracking.

Using the same part of the screen, our next approach adopted a river metaphor in which terms would appear on the left edge of the screen when they first appeared in the term rankings and move smoothly to the right. This better introduced a sense of time that was missing from earlier approaches. If words continued to appear in the term rankings, they would grow in size. If they were missing from the term rankings for a few cycles they would fade away completely. This ensured that any word still showing had appeared in term rankings at least once in the last 10 seconds. Furthermore, words that made it all the way to the right edge had at least been present in the term rankings for a while, making the right side of the visualization a reliable space to see long-term popular words. This approach is shown in Figure 6.15.
While this did create a nice sense of a time passing, there are a number of challenges to this approach. In order to handle a high word density in a tight space, it proved necessary to vary the velocity of each word slightly so they never permanently overlap in a way that makes them illegible. This subtly breaks the contract implied by the spatial configuration; no longer is a word’s horizontal position tightly correlated with the age of the term. Furthermore, it seemed sensible for words higher on the term ranking list to be larger, and to maintain a physical metaphor it seemed like larger objects should move more slowly and be on screen longer than small objects. Ultimately, this seemed to be enough at odds with the underlying metaphor that time moves left to right and an object’s position on the timeline should represent its age. In a tight visual space, it was too difficult to both satisfy that constraint, make it readable, and manipulate the visibility of words in ways that would properly highlight the highest performing words.

The final design iteration took the basic mechanics from the horizontal, river-like visualization and mapped it onto the vertical axis. Even with the same basic visualization mechanics, words rising up from the bottom instead of moving left to right helped alleviate the sense that position was mapped tightly onto time. Instead of a river, there is a stronger sense of bubbles rising. To reinforce this sense, a simple visualization of the crowd size could be added in the lower right hand corner, from which the words would appear to be rising. The shifted location within the interface also makes it easier to provide an interactive component. In this version, clicking on a word causes a dialog box to slide into place showing the chat messages that triggered that word’s inclusion in the visualization. This provide a convenient way to shift from very high level abstractions on the word level down to individual messages. It might also serve as a way to navigate between sections to spaces where there seem to be interesting conversations happening. This final approach is shown in Figure 6.16 and Figure 6.17.\footnote{A video version of this approach can be found here: http://www.youtube.com/watch?v=Ky9U1v3Bzsk}

Returning to the original values for this visualization, the final visualization approach meets all three goals in an acceptable (if not totally perfect) way. Volume is represented by the bursts of new words at the base of the visualization. This helps show instantaneous increases in participation. Furthermore, it has essentially the same property as the word-cloud approach where the total number of words on screen is loosely correlated with an average of recent activity. In this version, truly popular and long-lasting topics are better distinguished from transient topics, and new topics and easier to differentiate from popular old topics because of the way the words move through the visualization space. This also better supports the
sorts of collective action concepts we have in mind; the words of a chant could appear out of the bottom one at a time and not be confused with prior words in the sequence. The biggest remaining weakness of this visualization is that its heavy reliance on motion could be distracting from the primary event content. To help alleviate this, the visualization could be easily run in a minimized mode where only one or two words at a time are shown and travel very slowly so as not to take away from the action during tense moments.

6.2.4 Moderation

Mechanisms for managing contributions in pseudonymous online spaces are a critical component to creating a compelling, pro-social experience. Although the prototypes have little by way of moderation, there are a number of clear extensions that would fit with the overall model. First, the way sections are configured serves as a powerful moderation mechanic. Anti-social behavior thrives when it has both an audience and the means to cause inconvenience for that audience. If the bulk of the sections around an event are created through friend networks, they are both likely to be relatively small and have significant social connections between participants. This increases the social costs to anti-social behavior and decreases the value of anti-social behavior because of the small potential impact.

For large public rooms, however, significant problems remain. In this environment, two potential approaches can work together to minimize the impact of anti-social behavior. First, enabling users to mute other users or flag individual offensive posts creates a set of non-verbal actions that can be easily aggregated to help human moderators focus on problematic users or utterances to make moderation decisions. Muting is a particularly effective signal because it has a clear functional implication: you cannot see chat messages from someone you’ve muted. This defuses many non-serious uses where users might mute in jest or to attempt to get someone else banned by muting them spuriously. In either case, the user is sacrificing their ability to see what the mute target says, which imposes a cost to the signal unless that person’s messages have no value. This differentiates between authentic uses of mute and inauthentic uses.

Based on the mute signal, it becomes tractable to start to make some automated moderation decisions. Muting can also help guide human moderators identifying problematic users. Without inspecting the content at all, a user who has been muted recently is a potential ban target. By creating a simple moderation dashboard that focuses moderator attention on just messages from recently-muted-users, a small number of moderators can effectively moderate a much larger
body of chatters without having to closely watch all the individual chat rooms.

Moderation in this context would result not in bans, but in silent muting. In a chat system like ROAR, it is easy to make it appear to a user that their messages are being sent and seen by others, but simply stop transmitting them to anyone else. This removes the audience for an anti-social participant and denies them the satisfaction of a response. While they may eventually realize they’ve been muted, it will certainly increase the overhead of anti-social behavior somewhat. Furthermore, chat participation in public rooms could require silent viewing for, say, 20 minutes before chat is unlocked. Watching with friends would not be impacted, and any honest user will rapidly pass that threshold perhaps even without noticing the limitation. But a muted user cycling identities would have significant downtown after getting muted before being able to cause trouble again.

6.3 Conclusion

Although untested, the ROAR platform serves as a sort of extended picture of the future directions that this thesis has explored in its main projects in Information & Presentation Spaces, backchan.nl, and Tin Can. In each of these projects, I had the implicit constraint that it needed to operate effectively as a (potentially) deployable system. The work in this chapter was released from that constraint and serves as a kind of extended “future work” section for the dissertation.

The other primary difference between ROAR and the work described in the previous chapters is its interest in operating within modern constraints. One of the ways that researchers try to differentiate their systems design work from that done in industry is to make assumptions about the availability and wide-spread deployment of certain technologies. This was implicit in the work about virtual worlds; we had to assume that virtual worlds would become a familiar interface for mediated social interaction. When backchan.nl was first designed, smartphones were not yet widely deployed, not were super lightweight laptops or tablets. In the intervening years, mobile devices evolved in a way that made backchan.nl a far more effective tool. Finally, with Tin Can, we assumed that every student in a classroom would have their own tablet. This remains a pretty aggressive assumption, but it seems likely at this moment that tablets will have an increasing role in classrooms in the future. With ROAR, I made no such assumptions and aimed to create something that would make sense without any major shifts in available technology. This shift in assumptions helps show how the major research themes of my thesis can apply in a domain without any special technology.
If we expanded our view to include not just the remote spectator experience, but to the experience at the physical event as well, there are a variety of exciting directions. I suggested ways of having content spread between the online audience to the local audience using existing screens in the space. We can also imagine implicating the physical audience in voting for pictures taken live in the venue and sharing that multiplicity of views back to the remote audience. Identifying friends who are physically present at an event and letting them be part of the chat experience could be another nice way to build links between local and remote audiences. We can imagine providing different sorts of social experiences between fans and, for instance, celebrities or commentators. Fans could subscribe to distinct commentary streams from their favorite personalities to augment their viewing. Finally, we can imagine bringing in some of the lessons learned from backchan.nl into ROAR to provide ways to do crowd-level question and answer interactions between fans, commentators, and players.
A major challenge with a design oriented approach to systems research is identifying the contributions that come from the work as a whole. On a per project basis it’s easier to demonstrate the achievements and drawbacks of a particular design strategy. But what can we learn from assembling a set of projects in a design space that extends beyond the outcomes of each individual project?

One way to think about this design oriented approach is to consider this thesis a kind of map of the broader design space. Part of the contribution is simply to identify a certain class of designs as being part of a set with common traits and themes. This identifies the borders on the map. In my work, each of the projects I have described form the core of the design space that I’ve called complementary communication systems.

Each of the projects within this larger design space represents an in-depth exploration of a particular area. Within each area, my work can be a concrete guide about what is likely to work well and what is likely to be problematic. The presence or absence of a shared display for backchan.nl is one easy example of this kind of guidance. In contexts without a dedicated display, usage of backchan.nl falls dramatically. Chapter 3 has a wide range of specific tradeoffs inherent to designing virtual interfaces that could provide a guide for designers in virtual worlds. The analysis of Tin Can makes a point to show how tablets as a platform operate in a classroom in comparison to laptops. These sorts of findings serve as local directions on the map; areas near the implemented design that we have experimented with and found wanting in various ways.

To try to provide guidance to fellow researchers and designers in the as-yet-unexplored areas of this design space, I have attempted to trace three major design themes through each of the projects. I hope that these can provide broader guidance to others working both in the design space of complementary communication platforms as well on communication systems more generally. Let us now review these themes and the broad conclusions I have drawn regarding each of them.
7.1 Research Themes

Using text-based communication channels to create a sense of grounding in backchan.nl and Tin Can had a mixed effect. In neither case did the system create a true sense of the side channel being an automatic component of common ground for people involved in the event; shared displays are not necessarily an automatic route to group-level common ground. Of course, whether a bit of information is part of common ground or not is not a simple binary judgement. Across all the studies we saw varying extents to which people made assumptions about side stage information being mutually visible and understood by others. One of the ways we judged this was how fluidly speakers moved information from the side stage to the main stage. Situations with single shared displays like optimal backchan.nl setups yielded the smoothest transitions between side stage and main stage, while synchronized but non-shared displays in Tin Can supported transition, but it took somewhat more negotiation. When the transition goes unremarked, we can conclude that the speaker is assuming that the presence of something on the side stage is mutually understood. While none of the projects reached that state, both backchan.nl and Tin Can made significant progress in that direction and longer deployments might build a deep enough sense of mutual awareness to best support group grounding. In contrast, in the virtual context the challenge of attending to the side stage inhibits any strong statements about the visualizations role in supporting group grounding.

Non-verbal actions play a critical role in essentially any design of a mediated communication system. All of the projects in this thesis show different ways of constructing those actions yielding different results. Unsurprisingly, actions that are familiar to people because of their widespread use and offline analogs (like voting, which appears in backchan.nl and Tin Can) are the most successful. The more complex promotion mechanics in Tin Can were also well understood through long-term use, even though there are fewer comparable actions in other mediated experiences. There are also important differences between the explicit voting (like in backchan.nl and promotion in Tin Can) and implicit voting (like muting in ROAR). One way to think about this distinction is that implicit voting mechanics use a user’s behavior in a space to intuit their attitudes about someone else or the contributions of others. ¹ In contrast, explicit voting asks a user directly what they think or gives them an opportunity to register their feelings through rating or voting. This can be more obviously manipulated by users because the costs or rating or voting something are usually effectively zero.

The allure of implicit voting mechanics is that they usually carry

¹ Implicit voting plays a major component in Facebook’s decisions about what content to include in someone’s feed. Posts from someone whose posts you have frequently clicked on or commented on are more likely to be shown in your feed. Whether the frequency of people clicking on someone’s posts influences the frequency of those posts being shown to other people (for example, someone who has never clicked on a post from that person) is an open question, but there is some evidence that for brand pages at least, these mechanics are in use.
costs which can increase the quality of the signal; if you mute a user, you can’t hear their messages anymore. If you click on a link, you have to wait for the page to load. If you leave a comment on something, you cared enough about the content to want to spend the time to say something about it. These costs make these signals more difficult to effectively falsify and so are often more accurate signals of the something’s value. The challenge with implicit votes of this form is that they are not necessarily viewed as signals by users, which can make them difficult for people to reason about. It seems intuitive that behavior that isn’t visible to others in online spaces (like reading or filtering) shouldn’t have an effect on our experience in the space. This hasn’t been true for online experiences for some time, but reminding users of this can be risky.

In general, familiar and explicit mechanics are more quickly understood by users and more complex mechanics take a longer period to become well understood by the group. In both backchan.nl and Tin Can users tended to have a few hours to become acquainted with the outcome, visibility, and implications of their actions within the context of the system. Adding unfamiliar actions to a system has a similar legibility tradeoff to any UI decision; it imposes a learning cost on users with a potential payoff in enabling new sorts of interactions. These costs can be steeper in a social situation because it is not enough to know what your actions will mean to the system, you must also develop an appreciation for how your actions will be presented to others and understood by them. This is analogous to the mutual awareness challenge in grounding; fluency in the actions of a social system rests on mutual appreciation of the mechanical implications. For systems that are intended be used by transient, temporary groups familiar actions are best. For systems with longer-term aspirations, more nuanced action design becomes possible.

One of the main traditional critiques of backchannel systems had been that they split the attention of the audience. In this argument, the backchannel is viewed as an audience-only distraction that is disrespectful to the presenter. All of the projects in this thesis have pushed back on this notion by demonstrating how shifting to a main stage and side stage model recognizes the possibility of fluid transfer of ideas between a text-based side stage and a “being there” main stage. By creating side stages that are designed to influence the main stage in clear, well-defined ways, we can reconfigure the attention problem from being one where attention to anything other than the current speaker is conceived of as inattention, to a situation where attention to the side stage is a different sort of attention that has benefits for the shared experience, broadly understood. In backchan.nl, this is achieved by setting up an expectation that questions and com-
ments should be shared through the system and would transition
to the main stage at specific moments. In Tin Can, the professor’s
presence in the side stage created a clear sense of the system as part
of the classroom space where students could try out speaking turns
in a lower stakes environment but one that nonetheless would have
an impact on the main stage if others liked their contribution. Thus
by building stronger connections between the main and side stages,
mediated systems like those I describe are not competing with tradi-
tional “being there” experiences, but are expanding the ways people
can contribute, creating a more inclusive experience, not one where
the audience is fragmented.

7.2 What’s Next?

Looking forward, I have every reason to believe that this approach to
designing communication platforms that complement other collab-
orative experiences is a growing space. Collaborative applications,
broadly construed, have become part of the daily lives of millions of
people both in their work and personal lives. We are in the midst of a
technical shift as well, that is pushing our traditional hypertext inter-
faces in the direction of fully-fledged, presence-oriented social spaces.
These sorts of systems were challenging to build on traditional web
platforms, but the now-broad availability of real-time, event-driven
technical infrastructure has made large scale implementations of this
vision tractable.

While these spaces don’t look like the visions of cyberspace or
the metaverse that fiction might have led us to expect [Stephenson,
2000, Gibson, 1986, Vinge, 2007], they nevertheless represent a radical
change in the frequency and type of engagements with mediated
social spaces compared to past tools. My work is situated at the cusp
of this transition and seeks to guide future work that attempts to
ask the question: “how can we better support mediated audiences
and groups?” My answer is that by adding thoughtfully designed
communication channels to face-to-face or video/audio systems we
can enfranchise both large audiences and small groups to take a more
active role in the group process. These additional side stages for
interaction need not take away from the main stage. Instead, they
create alternative roles and modes of interaction that can increase
overall engagement and activity in the space.

There are a wide range of domains where these changes can have
a big impact. Chapter 6 shows one kind of approach in an entertain-
ment domain. Online and distance education is another space that
is ripe for the addition of these techniques. Work in that space has
focused primarily on asynchronous content delivery and self-paced
learning. Yet much of the educational value of being in a university setting comes from working in a shared cohort as from attending lectures. *Tin Can* points to how these approaches might operate in a small discussion section, a type of educational space often neglected in online education systems. In a more traditional lecture context, we might look to *backchan.nl* or *ROAR* as starting points for imagining what a socially engaging, interactive remote lecture might be like. I hope that as these platforms become practical, my work can help guide and support the decisions of their designers and enable more effective participant engagement among groups and audiences of many sizes and different contexts.
Bibliography

Question and Answer Tool. URL http://sourceforge.net/projects/questionanswer/.

RickRoll’D, May 2007a. URL http://www.youtube.com/watch?v=oHg5SJYRHA0.


Drew Harry, Eric Gordon, and Chris Schmandt. Setting the stage for interaction: a tablet application to augment group discussion in a seminar class. In *Proceedings of CSCW.* ACM Request Permissions, February 2012.


