Changing Small Group Interaction through Visual Reflections of Social Behavior

by

Joan Morris DiMicco

Sc.B. Applied Mathematics
Brown University
Providence, RI, 1995

S.M. Media Arts and Sciences
Massachusetts Institute of Technology
Cambridge, MA, 2001

Submitted to the Program in Media Arts and Sciences,
School of Architecture and Planning,
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY
at the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
June 2005

Author _____

Joan Morris DiMicco
Program in Media Arts and Sciences
April 2005

Certified by _____

Walter Bender
Executive Director & Senior Research Scientist
MIT Media Laboratory

Accepted by _____

Dr. Andrew B. Lippman
Chair, Departmental Committee on Graduate Students
Program in Media Arts and Sciences

Changing Small Group Interaction through Visual Reflections of Social Behavior

by

Joan Morris DiMicco

Submitted to the Program in Media Arts and Sciences,
School of Architecture and Planning,
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Media Arts and Sciences.
April 2005

Abstract:

People collaborating in groups have potential to produce higher-quality output than individuals working alone, due to the pooling of resources, information, and skills. Yet social psychologists have determined that groups rarely harness this potential. This thesis proposes that technology in face-to-face settings can be used to address the social factors that have damaging influence on group decision-making processes.

While there is much work in the area of collaborative software and groupware, this work differentiates itself with its specific aim to influence the way a group shares information without mediating the group's communication. By presenting visualizations to the group of individual levels of participation and turn-taking behavior, the technology aims to augment the group's communication ability, by making it more aware of imbalances.

A series of dynamic displays positioned peripherally to a discussion were developed and used by a variety of groups during face-to-face meetings. Both observational and experimental results indicate that these displays influence individual participation levels and the process of information sharing used during a decision-making discussion. A display revealing real-time participation levels caused those at the highest levels of participation to decrease the amount they spoke. Viewing a visualization of previous turn-taking patterns caused those who spoke the least to increase the amount they spoke in a subsequent discussion; real-time feedback did not produce this change. Additionally, after reviewing their turn-taking patterns, groups altered their information-sharing strategies. For groups that had poor sharing strategies on an initial task, this change improved their ability to share information related to the decision; for those who did not need intervention, feedback on turn-taking was not beneficial for their subsequent information sharing. The central finding of this research is that displays of social information, viewed during or after a meeting, bring about changes in a group's communication style, highlighting the potential for such displays to improve real-world decision-making.

Thesis Supervisor: Walter Bender
Title: Executive Director & Senior Research Scientist

Thesis Committee

Thesis Advisor

Mr. Walter Bender
Executive Director & Senior Research Scientist
MIT Media Laboratory

Thesis Reader

Dr. J. Richard Hackman
Cahners-Rabb Professor of Social and Organizational Psychology
Harvard University

Thesis Reader

Dr. Rosalind W. Picard
Associate Professor of Media Arts and Sciences
MIT Media Laboratory

Thesis Reader

Dr. Candace Sidner
Senior Research Scientist
Mitsubishi Electric Research Laboratories

Acknowledgements

First and foremost I would like to thank Walter Bender, my advisor, for supporting my research from its original, vague conception. He has consistently guided me on how to explore the deep issues while still keeping the broad perspective. I hope that over the years I have acquired his instinctive insight into which problems are the interesting and challenging ones to pursue.

My thesis readers: Richard Hackman has inspired me to believe that groups do have amazing potential and can be improved in innovative ways. Thank you for being so supportive of my research ideas and my future endeavors. Candy Sidner has offered insightful comments from our very first meeting, guiding me on what was technically feasible and what would be beneficial for meetings. Roz Picard has always been an enthusiastic supporter of my ideas, so much so that she offered up both her research group and her pattern recognition class to assist me in my endeavors.

Other mentors: I would like to thank Pattie Maes, my first advisor at the lab, for introducing me to academic research and for inspiring me to pursue the PhD after the Masters. I would also like to thank Dan Ariely. He has influenced my research approach, particularly in his questioning the benefits of groups, and I appreciate how he encouraged me to question my original assumptions. Other mentors at the lab include Henry Lieberman, Judith Donath, Hiroshi Ishii, Chris Schmandt, Cynthia Breazeal, Bruce Blumberg, Ted Selker, and John Maeda. I especially want to thank Judith and Cynthia for being on my general exam committee and steering my original exploration into the areas of social visualization and sensing of human behavior.

Lab community: My research could not have been done without the amazing Media Lab community. I would like to thank each and every person who helped me along the way with a seemingly small technical issue or question. Each of my projects has benefited from others' input and expertise. Anna Pandolfo, Kate Hollenbach and Sunil Vemuri deserve special thanks for their help on my thesis projects.

At this point I suspect that every single person at the lab has been solicited by me to be in an experiment. I want to thank each person who participated in a study, particularly the Electronic Publishing Group, the Student Committee, and the Affective Computing Group who repeatedly welcomed my technology into their meetings.

Outside the lab: There have been many mentors outside the lab and their comments have helped to shape my work and writing. I would like to acknowledge their input: Judy Olson, Paul Maglio, Amy Greenwald, Michael Muller, Eric Horvitz, Kori Inkpen, Elizabeth Churchill, and the members of Richard Hackman's "Group Group."

Finally, I couldn't have done this without the support of my entire family, especially my husband, Mike DiMicco.

Table of Contents

Figures and Tables

Chapter 1: Introduction

"What doomed the Columbia and its crew was not a lack of technology or ability, the [independent review] board concluded, but missed opportunities and a lack of leadership and open-mindedness within NASA management. ... The disaster, the [board's final] report said, was fully rooted in a flawed NASA culture that downplayed risk and suppressed dissent. 'I am convinced that the management practices overseeing the space shuttle program were as much a cause of the accident as the foam that struck the left wing,' the report said."

—"Final Shuttle Report Cites 'Broken Safety Culture' at NASA"
New York Times, 26 August 2003

The contradiction of groups is the realization that the process of exchanging ideas and discussing decision alternatives can actually lead to fundamentally flawed decisions. The NASA Challenger Disaster is frequently cited as an example of such a decision failure (Griffin 1997) and unfortunately, in the summer of 2003, NASA also determined that the Columbia Disaster was more a flaw of group decision-making than of engineering.

The type of decision failure that led to the Columbia Disaster has been documented many times in the very public arena of politics and also in the experimental setting of the laboratory, illustrating that these flaws are not just bad luck, but rather indicate that our decision capabilities in groups are faulty because of a consistent over-reliance on the social cues and behavioral pressures in group settings. These cues subtly encourage groups to conform to the opinion of the authority, suppress alternative viewpoints, and inadequately share information relating to the decision at hand, and these behaviors systematically increase the likelihood of a flawed decision (Janis 1982).

As an alternative to existing technology designed for collaboration, I propose a new type of interface for groups that does not mediate or control communication, but rather visualizes different types of group behaviors, to encourage a more productive process that allows for more information sharing and exchange of viewpoints, without dictating or controlling the group decision-making process. These applications observe group patterns and reflect back the observed patterns for the group to observe and correct extreme skews in its deliberation process.

This thesis presents the design and implementation of collaboration applications that sense group behavior and reflect the behavior patterns back to the group in the form of dynamic visualizations. An essential aspect of this research is to understand the impact of these information displays on group process, through both qualitative and quantitative evaluation. Therefore, groups in both real-world meetings and in the controlled environment of a lab setting have used these applications during their decision-making discussions.

This research thesis had two objectives: (1) to build dynamic visualizations of social behavior designed to improve group processes during face-to-face collaboration and (2) to run qualitative and quantitative evaluations of group collaboration to determine the impact of these information visualizations on group process and individual behavior. The outcome of this research has been a new understanding of how displays of social information are received by collaborating groups, specifically with evidence of how group behavior changes and group decision-making is altered, and a set of guidelines for building applications designed with the intention to change behavior.

I. Theoretical Grounding

Research in social psychology has demonstrated that our ability to make group decisions is frequently flawed because we overly rely on social cues during a group discussion. Regrettably, when technology is introduced as a mediator in our communication, our group decision-making ability further degrades. As supporting background for this research, I present these two issues in more detail to illustrate why collaboration technology can address the issues of group dynamics more effectively within the role of social facilitator, rather than as a process or communication mediator.

A. Social Psychology of Groups

Conversations held by groups for the purpose of making decisions are fraught with complications. Social psychologists have demonstrated that individuals allow the presence of the other people in the group to influence their behavior to such a degree that through the process of exchanging opinions, the group is led to a lower quality decision (as compared to aggregating individual decisions) (Bray 1982, Hackman 1992, Janis 1982, Myers 1971, Whyte 1991).

Imagine a scenario where a small group of employees has been given the task of interviewing and hiring a new employee for the company. The group has located the job candidates, reviewed resumes, checked references, and interviewed each of the individuals. When the group is ready to make the decision, it meets for a face-to-face discussion of the pros and cons of the different candidates. By sharing information and opinions about the job candidates, the group comes to a decision on whom to hire.

Social psychology predicts a dismal outcome for this well-intentioned group. It is highly likely that the members of this group will inadequately share information they know about the job candidates (Stasser 1987, Stewart 1998). The group will likely spend its time discussing the group's initial inclination on who to hire, to the detriment of seriously considering the other candidates (Bray 1982, Hackman 1992, Janis 1982, Myers 1971, Whyte 1991). And through the process of discussing the prevailing viewpoint, the individuals in the group will likely become more strongly committed to their initial inclination than they were when they originally entered the meeting (Brown 1986, Moscovici 1969). Each of these systematic flaws increases the likelihood that this group will make a strong commitment to a flawed decision.

While groups have flawed decision processes, Raven (Raven 1998) describes a well-known experiment that aptly illustrates the difficulty in universally stating that groups hinder decision-making. In the experiment, groups of people were asked a basic math question (about a man buying, selling, buying back, and selling back a horse[1]) and were then required to come to a consensus decision. On average, substantially more individuals got the right answer after discussing it with the group, indicating that the process of discussion encouraged the pooling of skills and perspectives, allowing more individuals to make the correct conclusion. But, it turns out that if the group were leaning towards the wrong decision, then all of the individuals would end up making a wrong decision, persuaded by the majority. And further confounding the question of whether or not the group setting helped in the decision, the groups that reached a *unanimous* decision felt more *satisfied* with their decision than those who did not, even if they were shown to be incorrect in their judgment. This experiment illustrates that individuals rely on the opinion of others as an indicator of the accuracy of their judgments, but this reliance can occasionally lead to an error in judgment. Yet, as a corollary to this, if a criterion of decision success is satisfaction with the outcome, then individuals' use of this heuristic may be beneficial even in cases where their judgment is wrong.

With this understanding of the complexity of our limitations, what can we do to limit the harm and harness the benefits of groups? By altering its decision-making process, a group can avoid the above communication flaws and over-reliance on others. For example, by allowing for minority viewpoints to be freely expressed (Bray 1982), by continually scanning the available options to find new alternatives that may work (Hackman 1987), and by allowing for open dissent of the authority figure's opinion (Janis 1982), the group

[1] A man bought a horse for $60 and sold it for $70. Then he bought it back for $80 and sold it again for $90. How much money did he make? (A) He lost $10. (B) He broke even. (C) He made $10. (D) He made $20. (E) He made $30. The correct answer is (D), he made $20.

making the hiring decision will more thoroughly consider all of the job candidates they interviewed and will more likely make the best possible hiring decision for the group.

Considering the existing research literature, there are three possible areas I believe should be examined to enhance decision making processes with technology. First, determine ways to increase vigilance in considering choice alternatives (Janis 1982); second, work to limit the effects of group polarization (a group's tendency to shift towards risk or caution) (Brown 1986); and third, discover ways to increase the sharing of information between individuals (Stasser 1987).

Researchers in organizational behavior often propose one of two methods to solve these problems: 1) train leaders to facilitate groups more effectively or 2) develop group procedures that will ensure productive group processes. These two methods are not without weaknesses. Janis (Janis 1982) and Maier (Maier 1967) outlined different ways group leaders can have a positive impact on a group and its decision-making ability and while their leadership guidelines are useful, the practicalities of training and the challenges in ensuring leaders successfully employ guidelines leave this approach lacking. Delbecq (Delbecq 1975) and others have presented specific techniques for implementing group procedures that tackle the common errors in group decision-making. These procedures involve anonymous pooling of ideas and formalized critique and evaluation methodologies. But again, their approach is limited in its ability to impact decision making because of the training issue. In their case, the Nominal Group Technique and the Delphi Technique prescribe rigid procedures for a very particular type of decision-making situation, and the rigidity of their technique invites group resistance and has prevented widespread adoption and impact.

B. Potential of Technology

The majority of today's collaborative tools, of which (Olson 1989, Stefik 1987) are good examples, were built without serious consideration of their impact on the significant decision errors outlined above. Thus, there are still ripe opportunities for developing decision-making environments that mitigate the errors in group interaction. As stated above, I believe technology can be of most use to groups if it is built with the intention to 1) encourage vigilance in the discussion, 2) instill a resistance to group polarization, and 3) increase the sharing of information, three methods for avoiding common decision flaws.

For example, vigilance can be encouraged by a system that observes the communication of a group, notes the alternative ideas mentioned, and makes the infrequently mentioned alternatives re-appear within the discussion. Group polarization can be limited with a re-framing of decisions in terms of gains, not losses, and an interface or tool that is able to re-frame questions from different points of view may be able to achieve this, such as Paese did in a non-electronic environment (Paese 1993). Information sharing can be encouraged within a group by allowing for anonymous comments or by encouraging private documenting of relevant information that will later get automatically shared with

the group (akin to the Nominal Group Technique, but done electronically) (Delbecq 1975).

Technology is already a part of our decision-making environments, both as communication tools and information devices. As these tools become more pervasive, there is an increasing opportunity to build applications that enhance our communication abilities, rather than restrict them. If tools can be designed such that the flaws in our decision-making heuristics are limited, then the potential for harnessing the gains of group interaction increases.

C. Mediating Technology

While people are highly skilled at using their verbal and nonverbal communication skills to express themselves, when technology is introduced as a *mediator* of the communication, individuals have more difficulty expressing themselves and the quality and effectiveness of their communication degrades (Hollingshead 1996, Kiesler 1992, Ochsman 1974, Reid 1997).

Communication technology over the decades has enabled us to hold conversations that could not have been held otherwise, by connecting distant physical spaces (Isaacs 1994, Ishii 1993), by allowing for rich asynchronous communication (Davis 2003, Hollan 1992, Olson 2000, Rosenberger 1998), by providing new tools for expression (Brave 1997, DiMicco 2002, Kurlander 1996) and by providing new interpretations of our communication through data analysis and visualization (Sack 2001, Smith 2001).

There are enormous benefits of mediating technology, but because mediating technology fundamentally changes the way we communicate, the technology can create additional group process problems. For example, video-conferencing is commonly used for meetings within companies, yet studies have shown that there is less trust between individuals and it is more difficult to establish common ground over video than when face-to-face (Anderson 1997, Isaacs 1994, O'Malley 1996, Rocco 1998). When communication is text-based, such as with IRC, email or, increasingly, instant-messenger, decision-making tasks take more time and produce lower rates of task accuracy (Kiesler 1992, Ochsman 1974, Siegel 1986). They also cause individuals to make more declarative position statements and less information-based statements (Reid 1997), making productive information-based meetings more difficult to hold. Finally, in the specific areas of group polarization and information sharing, it has been shown that groups experience *more* group polarization (Kiesler 1992, Siegel 1986) and *less* information sharing (Hollingshead 1996) when communicating in a text-based environment.

Social psychologists point to our faulty decision-making processes when our decision quality suffers. Yet, from the perspective of the Media-Richness Theory (Daft 1986, Rice 1992), it may not be our group processes that cause our decision failures, but rather the limitations of our ability to express ourselves in our current media. As illustrated by Reid's study (Reid 1997), when the medium does not support sufficient expressiveness,

we resort to more simplistic ways of expressing our thoughts, and it is this adjustment of conversational style that is detrimental to our decision-making ability.

As designers of technology for decision-making, our goal is to build technology that falls *higher* on the expressiveness curve than the current standard of face-to-face conversations, based on the premise that our face-to-face communication has been shown to be faulty. Current communication technology effectively supports simpler communication tasks, but when communication becomes more complex, such as in a decision that requires a high degree of information and opinion sharing, we need more than our face-to-face environment offers us. We need an environment that is enhanced in such a way that our current ability to express ourselves and interpret each other's meaning is increased. At this point, we may reach our decision-making potential and begin to correct our previous decision errors.

Mediation effects on communication illustrate how the expressiveness of an environment changes our ability to communicate effectively. Unfortunately, even without mediation, we still have difficulties expressing ourselves and, as decision tasks become increasing complex, our communication abilities begin to degrade and our task performance suffers. The lesson to take from this is that we need to focus on the unique expressiveness that technology *can* offer, and focus on building environments that incorporate these new types of expression into our communication.

II. Overview of the Thesis

The focus of this thesis is on building interfaces that assist groups in improving their interaction processes. The goal is to encourage groups to include a more diverse set of viewpoints in their discussions, in order to promote higher quality group decision-making. The interfaces presented in this document strive to reveal information about the ongoing social dynamics within a group's real-time communication by providing analysis and evaluation of this communication. They are designed for a face-to-face setting, to allow individuals to utilize their natural strengths in communicating while providing a display of information that assists the group in reflecting upon its current interaction.

In real-world decision-making situations, there is not always a "right" answer to be found. And even in situations with a right answer, it is often difficult to identify it because people hold different information about the decision that can sway them in opposing directions. Consider the previous example of the group deciding to hire a new employee. Each member of the group may have slightly different information regarding the job applicants derived from interviewing the candidates, reading resumes, and speaking to references. In this situation, the group must determine the best choice by sharing both opinions on the candidates and factual information they have collected.

In this type of decision task, there is a substantial risk that the individuals who hold critical information will not effectively share it to enable the group to make the best choice (Stasser 1987, Stewart 1998). When there is poor information sharing in a discussion to the detriment of exploring new ideas and when groups move to extremes in

their commitment to the prevailing viewpoint, there is an increased likelihood that a group will make a strong commitment to a faulty decision (Hackman 1992, Janis 1982, Myers 1971, Whyte 1991). So while meetings can be frustrating due to inefficiencies, there is a larger issue looming: group-wide dynamics often prevent all viewpoints from being shared and this process decreases our ability to make high quality decisions.

I am interested in using technology to assist in correcting these flawed processes. By analyzing face-to-face communication between individuals in a group and by dynamically presenting information to the group about their interaction, I hypothesize that technologically-enhanced groups will have the ability to make more effective, higher-quality decisions.

The challenge of this work falls into four categories: sensing group processes as they occur, providing interfaces that show groups their interaction patterns, highlighting potential flaws, and the evaluation of how individuals and groups change in response to these visualization tools. This document is organized according to these challenges.

Chapter 2, "Sensing Group Behavior," addresses advances in sensing technology and networked applications that offer new opportunities for interfaces to interpret and assist in our real-time, face-to-face communication. It is now straightforward to capture many aspects of group interaction, such as verbal comments via automated speech recognition (Morgan 2003) and affect signals with basic sensors (Picard 1997). In this chapter, I present related sensing work, my own exploration of the opportunities for sensing group dynamics, and a discussion of how the use of speaking times, detected through the volume level on microphones, can be used as a proxy for measuring group interaction.

Chapter 3, "Visualizing Group Behavior," introduces the visual displays I built that illustrate the data collected during face-to-face meetings. Presenting data to a group in real-time is becoming a realistic standard in collaborative settings. For example, lectures and conferences can incorporate a simultaneous chat room into the discussion with relative ease (Rekimoto 1998) and conference call applications, such as the Jabber project (Kazman 1995), can capture and present context-sensitive information as the group converses over the network. This chapter presents other related projects and the design rationale for the visualizations, and finally a preliminary user study evaluating how the interfaces convey different social messages to an observer.

Motivated by this increased facility to monitor and simultaneously present context-sensitive information to a group, I am exploring ways in which this can improve group-decision processes. By building interfaces for face-to-face collaboration that do not mediate, but rather augment, a discussion, we expect that groups will use their natural abilities to communicate with each other and simultaneously utilize the tools made available to them to make observations and corrections in their behavior as it occurs.

Chapters 4 and 5 evaluate this aspect of the research. Chapter 4, "Quantitative Evaluation of Group Behavior," presents results from two quantitative behavioral studies run with groups in experimental settings. These studies evaluated the specific behavior changes that groups exhibited while using the displays in decision-making situations. Chapter 5,

"Observations and Lessons Learned from Groups," presents my qualitative observations and higher level understanding of how these displays were integrated into group settings and the emotional and social reaction to having this information revealed to a group.

The final chapter, Chapter 6, "Implications and Conclusion," discusses the broader implications for this research and outlines appropriate directions for future inquiry.

Chapter 2: Sensing Group Behavior

Automated sensing of behavior and salient events in a meeting or group context is an emerging field of study. Experts in vision systems, speech recognition, and gesture and facial recognition are currently advancing our ability to recognize different behavior and events in naturalistic settings. As my goal is to sense a group's interaction for the purpose of assisting a group in altering it, part of my research has been to understand the state-of-the-art ability for systems to detect behavior in natural group settings. This chapter is an overview of the different approaches to detecting group behavior, the current abilities of technology, and my use of speech recognition technology and of microphone volume detection, as measures of behavior in group interactions.

I. Codifying Group Interaction

While we can incorporate sensors, cameras and microphones into meetings, it is important to understand the purpose in the detection of behavior and human activity. For my goal of detecting the higher-level inter-personal dynamics a group, it is useful to consider how social psychologists conceptualize and codify group communication.

Broadly speaking, group communication is frequently categorized as either task-oriented or process-oriented; within a group setting, each individual's communicative acts can be coded by either what the group is doing (task-orientation) or how the group is doing it (a process-orientation). For a task-oriented analysis, McGrath has outlined different types of group tasks and researchers in both social psychology and computer-supported cooperative work have used this framework to both describe group activity and to build supportive tools (McGrath 1984).

These tasks are:

- *Planning Tasks*: Generating plans
- *Creativity Tasks:* Generating ideas
- *Intellective Tasks:* Solving problems that have a correct answer
- *Decision-making Tasks:* Tasks for which the preferred answer is the correct one
- *Cognitive Conflict Tasks:* Resolving conflicts of viewpoint
- *Mixed-Motive Tasks:* Resolving conflicts of motive-interest
- *Contests/Battles:* Resolving conflicts of power; competing for victory
- *Performances:* Psychomotor tasks performed against objective standards

Particularly in the field of CSCW, McGrath's tasks are often used as a way of understanding a group's actions and determining the technical requirements for a software application designed to support that group's activity. For the experiments that will be presented in Chapter 4, subjects were asked to perform "decision-making tasks," yet they were actually performing "intellective tasks" in the sense that there was a particular answer expected if the group used a successful process. In Chapter 5, in the discussion of the case studies of real-world groups, the group meetings were mostly "planning tasks" with brief "cognitive conflict tasks" incorporated into the discussion.

The prevailing approach to understanding group process through behavior coding is the SYMLOG system, developed by Robert Bales (Bales 1950, Bales 1999, Bales 1979). To understand a group's process, trained observers code each group member's behaviors according to Bales' "Interaction Process Analysis" guidelines and, by aggregating these observations, produce a descriptive analysis of a group's interaction dynamics that can later be used by the group for learning and reflection. These guidelines code for behavior such as asking or providing opinions, agreeing or disagreeing, showing tension or releasing tension, and asking or providing suggestions. The aggregate report summarizes the group in terms of dominating and submissive personalities, relationships to authority, friendliness, and emotional expressiveness. As will be discussed in Chapter 4, these aggregate reports summarize the personalities of the group members, and some of the personality characteristics SYMLOG codes for are successfully and automatically visualized within the application Second Messenger 2.0's interfaces.

While these task-oriented and process-oriented analysis approaches provide frameworks for those working on automated detection of behavior, there is still a considerable challenge in determining how to automatically measure behaviors in such a way as to aggregate them into these high-level conceptualizations of the group's behavior. When hand-coded by experts, through observation and video analysis, there are still moments of disagreement and ambiguity in the ratings, and many of the small actions of individuals cannot be understood outside of the context of the entire meeting. For example, at the level of determining through direct observation turn-taking patterns and ownership of the floor, moment-to-moment there is substantial ambiguity as to whose turn it is (Parker 1988, Stasser 1991). In short, there is a gap between observing behavior and successfully summarizing it is either a process-oriented or task-oriented level of the group's behavior.

II. Automated Detection of Group Behavior

An important question to ask is "what should be captured in a group interaction?" Both verbal and nonverbal communication reveal information about individual intentions, and at a high-level, these can reveal both a group's task and process. Verbal utterances can be analyzed in terms of each speaker's attitude and communicative intentions (Grosz 1986, Lochbaum 1998). Nonverbal behaviors, such as hand gestures, eye gaze, posture shifts, and head movements, reveal information about the attitudes between individuals and shifts in topic and task (Argyle 1988, McNeill 1992). Of these behaviors, which can be accurately captured and summarized?

There are numerous research initiatives that study augmenting the experience of the face-to-face meeting by offering technological assistance. The original conception for many of these initiatives came from Xerox PARC's CoLab Project (Stefik 1987). Since then, many different labs are building their own meeting capture environments: Stanford's Interactive Workspaces (Johanson 2002), MIT CSAIL's Agent-based Intelligent Reactive Environments (Brooks 1997), the Media Lab's Facilitator Room (Basu 2001), CMU's Interactive System Labs Meeting Room (Burger 2002), Georgia Tech's eClass (Abowd 1999), and IPSI's AMBIENTE Project in Darmstadt, Germany (Prante 2002) each augment a meeting space with cameras, microphones and interactive displays to allow for numerous types of meeting assistance and full capturing of events.

The majority of the projects associated with these initiatives tackle the technical challenges in providing pervasive technical assistance to dynamic meeting contexts and ways to capture all aspects of communication related to the meeting. After capturing a meeting's data, there is then the design and development of applications for the browsing and scanning of content. Ricoh Innovations is one example of a lab working on how to use salient audio and video events to tag a dataset to aid in the browsing of a meeting's recording (Erol 2003).

Returning to the concept of codifying group behavior in terms of task and process, McCowan, et al. have utilized the McGrath framework to build an application to determine the type of meeting (task) a group is holding using both audio localization (where are people talking) and video analysis of body movements (blob detection of heads and arms around the room) (McCowan 2005). By considering each meeting as a series of sequential actions that are representative of a particular type of meeting, either a monologue, presentation, discussion, or a note-taking session, they were able to use Hidden Markov Models to train a system to detect a meeting type with relatively high accuracy. Although their approach is promising, it is a good example of the weaknesses in the current sensing technology: the robustness of their categorization of a meeting is limited to when groups followed rigid scripts according to strict definitions of the different types of meetings (or tasks). In natural meetings settings, where groups flow from one task or meeting style to another, their system had much lower accuracy rates, because the natural overlapping of events that occurred in real meetings were not within their training set or conceptual model.

Despite the limitation of the work on meeting capture in terms of assisting group process, these projects are very useful for understanding the issues in measuring group behavior in natural group settings and in highlighting areas that are not sufficiently developed to pursue my research agenda of altering group behavior.

III. Speech Recognition in Meetings

One fruitful avenue of research has been capturing a meeting's spoken content for post-meeting analysis. Using either off-the-shelf speech recognition software or custom designed recognizers, and requiring each speaker to wear a close-talking, noise-canceling microphone, many researchers have demonstrated success in analyzing an audio collection for verbal content. The current leading work in this area is being done at ISCI in Berkeley, were their large-scale, multi-year project is referred to as the Meeting Recorder Project. As part of this initiative, they have developed algorithms for performing topic tracking and the detection of agreement and disagreement within an audio recording of a meeting (Hillard 2003, Morgan 2003). Another example of promising use of audio content to extract summarizing information is Kristjansson, et al.'s work where they built a meeting-support tool that augmented an audio recording of a meeting with an outline of the meeting's structure derived from a discourse analysis of the conversation (Kristjansson 1999). Capturing and then summarizing the meaningful structure of a meeting indicates that group process or structure may be able to be captured.

Several projects at the MIT Media Lab have also demonstrated the unique ways speech recognition transcripts can be used for high-level analysis of groups, despite inaccuracies in transcription. Sunil Vemuri has provided valuable guidance for detecting verbal content from informal conversational speech as part of his project "What Was I Thinking?" (Vemuri 2004). By using IBM's ViaVoice with additional accuracy filtering, he produced low fidelity transcripts of conversational speech. Another project at the lab that strives towards high-level topic spotting was built by Jebara, et al., (Jebara 2000) where their system performed real-time topic spotting by comparing the voice-recognition transcript to a fixed corpora derived from newsgroups. During a face-to-face conversation, the current topic of conversation (limited to the corpora) was displayed on a shared screen. Eagle, et al., (Eagle 2003) further developed this work by incorporating the Open Mind Common Sense database (Singh 2003) into the analysis. With OMCS, they located the 'common sense' connections between spoken words to eliminate errors in voice-recognition and provide a topic-level summary of a conversation. While this was not a real-time support tool, it offers a promising method for analyzing speech-recognizer output.

IV. A Speech-Driven Meeting Capture Tool

Each of the above initiatives shows promise in determining the content of meetings and because of them, I was motivated to explore using speech recognition as a way of

tracking and detecting social dynamics. The first application built for this research initiative is called "Second Messenger" and it is designed to support a group in a face-to-face meetings using the verbal content to reflect a group's interaction (DiMicco 2004a).

A. Second Messenger 1.0

Second Messenger 1.0 uses a speech-recognition engine as an input method and outputs filtered keywords from the group's conversation onto an interactive display. The goal of the interface is to improve the quality of a group discussion by increasing the visibility of diverse viewpoints. The system uses a combination of speech-recognition technology and semantic analysis to display a real-time text summary of the group's comments. The content on the shared display can be manipulated with a pointing device, making the display a real-time meeting tool that can be used by the group to organize their discussion. See Figures 2.1 and 2.2 for screenshots of the client applet and the shared group display.

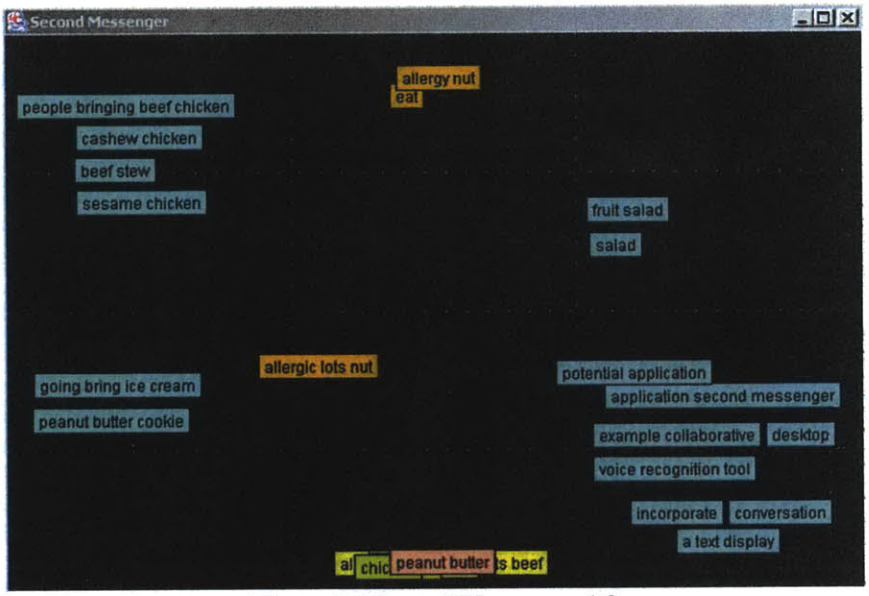


Figure 2.1 Second Messenger 1.0
The spoken phrases fall down the center of the screen. Users can move the phrases around the screen to organize their ideas. Untouched phrases collect at the bottom of the display.

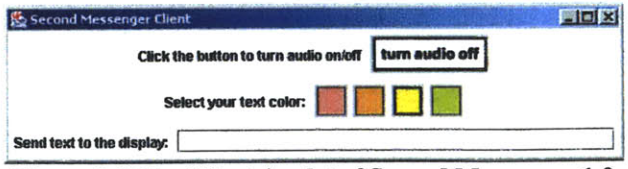


Figure 2.2 The Client Applet of Second Messenger 1.0
Users can turn their audio feed on/off and choose the color of their transcribed text. They also have the option to send text directly to the display.

The core goal of Second Messenger 1.0 is to influence the way a group communicates in a face-to-face setting. Because research has shown that increasing the amount of discussion around minority viewpoints can increase the quality of group decisions

(Nemeth 1989), the application attempts to bring this about through emphasizing the minority viewpoints in the group by selectively filtering which users' keywords appear on the display. By increasing the visibility of group members who speak less frequently, and filtering out the comments of group members who are verbally dominating, the application attempts to amplify the voice of the group's minority members.

B. System Architecture

The following sections explain how the application works: first assembling a transcript of the conversation; and then filtering it according to semantic and social facilitation goals. Figure 2.3 provides an overview of this process.

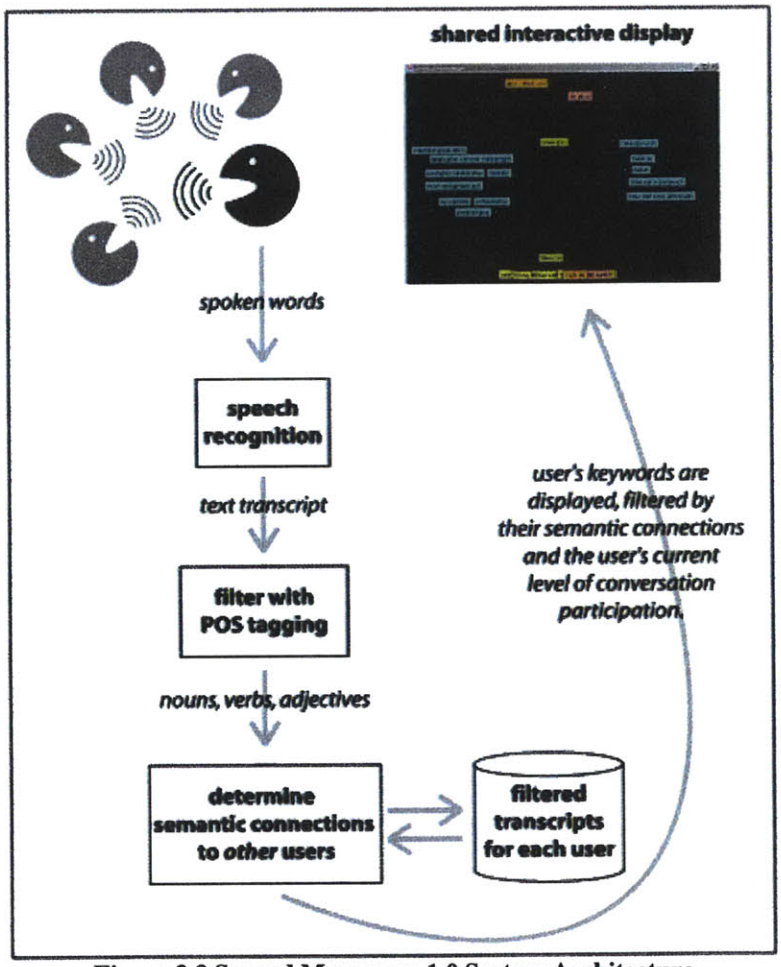


Figure 2.3 Second Messenger 1.0 System Architecture

1. Conversation Transcription

Today's speech-recognition technology is far from ideal, creating a significant challenge to deploying an application that detects informal, non-dictation speech. To achieve the highest accuracy possible and to eliminate the challenge of speaker identification, each user is required to wear an individual microphone that sends his/her speech to an

individually-trained speech model. The system uses IBM's ViaVoice engine to convert each speaker's audio to text and then streams this transcript to a central server for processing.

Second Messenger 1.0 does not need a complete transcript of a conversation, but rather a collection of the relevant keywords. Thus the system performs part-of-speech filtering on each incoming utterance to reduce the text to nouns, verbs, and adjectives. This type of filtering removes the non-verbal utterances (ah's and um's) and many of the uninteresting comments, such as simple affirmations and confirmations.

2. *Semantic Filtering*

The next step is to determine which text to display to the group. My opinion is that the most interesting comments are the ones that haven't been said before, yet have some relation to previous comments made by others. Therefore when the system performs a semantic analysis of the phrases using WordNet, it calculates how close the current phrase is to prior phrases spoken by other members of the group. Specifically, it locates mutual conceptual parents of the spoken words. For example, after semantic filtering, the phrase "I saw some good peanut butter cookies at Star Market yesterday" becomes "peanut butter cookies" if other people have previously used words connected to the parent concept "food." Our objective is to encourage contributions from everyone in the group, therefore the emphasis here is to display the ideas that relate to each other that originated from different people, preventing one person from controlling which text appears on the display.

3. *Social Facilitation*

As stated in the introduction, the primary goal of this project is to encourage equal participation in a meeting for the purpose of increasing the diversity of the discussion, and ultimately the quality of the decision-making. Utilizing the fact that the system knows exactly how many words each user has spoken, Second Messenger is able to calculate if someone is under- or over- contributing to the conversation. Imagine a meeting scenario as described earlier where most of the group is able to state their opinions, yet one member contributes disproportionately less. When Second Messenger detects this type of imbalance in the verbal contributions, rather than displaying the spoken text with the same imbalance in contributions, it visually emphasizes the less-vocal member's contributions. When someone is contributing one-fourth the amount of the other group members, their text phrases are no longer filtered through the semantic filter, so that all of their spoken nouns, verbs, and adjectives are displayed. If someone is dominating the discussion, by saying twice as much as all other group members combined, they will no longer have any of their words displayed on the screen. These contribution ratios were chosen as a baseline and these could be adjusted if desired. I hypothesize that this type of filtering will make the less-vocal members' contributions more apparent and the dominating individuals seek out others' ideas and opinions.

C. Evaluation & Speech Recognition Accuracy

The system was deployed as described and used informally during several weekly research meetings. In a multi-person conversational setting, the voice-recognition results were so poor that they distorted the application's text output to the point of irrelevance to the meeting. To avoid this problem, the system was enhanced to filter ViaVoice's output based on the "phrase score" returned for each word. Although IBM states that the phrase score is not a confidence value, Vemuri, et al. (Vemuri 2004) demonstrated that a word's phrase score is a good predictor of word accuracy. By setting the phrase score threshold high (above zero), one can ensure that the words sent through the semantic filter were actually spoken.

Unfortunately, although this filtering removes errors, in a conversational setting it filters out the majority of the spoken words due to low phrase scores. In use, very few content words appeared on the screen once filtering was put in place. In an informal trial where five people trained voice models and participated in two-person and three-person conversations, the recognition rates were well below a reasonable level. On average, 76.3% of spoken words were assigned a negative phrase score (standard deviation = 7.2%), leaving less than 25% of the spoken words transcribed and available for semantic analysis by the system.

A second observation made during the informal evaluation of the software is that the complexity of the information provided in the display is difficult to incorporate into a face-to-face discussion. The goal of the application was to provide a peripheral or secondary channel of information, yet the content presented requires constant attention if the user is to understand the on-going filtering. Unless the application were to evolve into a tool that a group would directly focus its attention on during a discussion, the current level of semantic filtering and analysis was in excess of what a group needed to support real-time face-to-face interaction.

V. Measuring Speaking Amount

Given this low level of detection and my desire to use this system in real-world meetings to observe behavioral changes, I decided to explore using speaking *amount*, instead of speech content, as a measurement of interaction in a group setting.

To detect speaking time, each user of the system wears a microphone that detects speech on the microphone and sends it over a serial port for processing. To quantify an individual's spoken contribution, the system detects the microphone sound level every 10 milliseconds and when the level is elevated to a participant's natural speaking level (calibrated before the experiment) for 30 milliseconds within a window of 50 milliseconds, the application interprets this as one "speech unit." This calculation eliminates very quick breaths and other isolated non-lexical vocalizations from being interpreted as speech.

I have found that there are several advantages to this basic measurement technique. First, this simple metric does not require a user to train a voice-model or spend time working with the system before use, allowing it be used immediately by anyone. Second, where a speech recognizer fails to accurately evaluate short utterances as such, because of its attempt to place each utterance in context with the person's previous and next comments, a volume detection system accurately captures and treats short utterances as what they often are: non-lexical comments that do not contain verbal content, but reflect involvement and engagement of the listener. Third, because volume detection is computationally simple, there were no longer issues of having a separate machine per user to perform speech recognition or issues of machine latency due to processing speed.

Figure 2.4a shows a picture of custom hardware designed to collect the volume off of eight separate microphone channels. Figure 2.4b shows the box that contained the hardware, enabling the system to be relatively portable and usable by groups in different locations. Appendix A provides the schematic diagrams for the hardware.

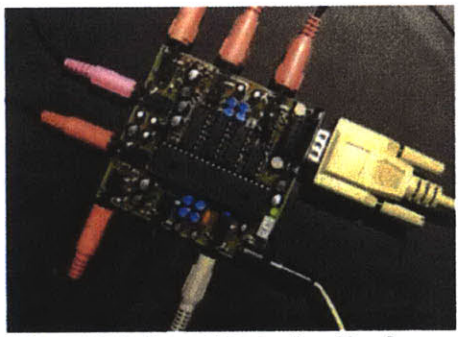


Figure 2.4a Sound Detection Hardware

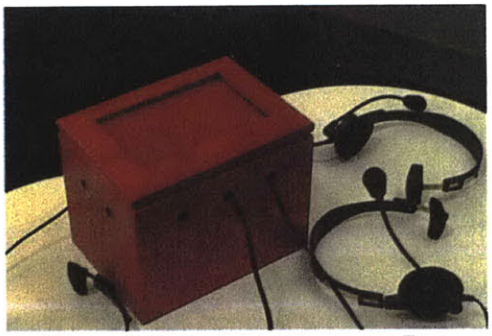


Figure 2.4b Portable Hardware Box

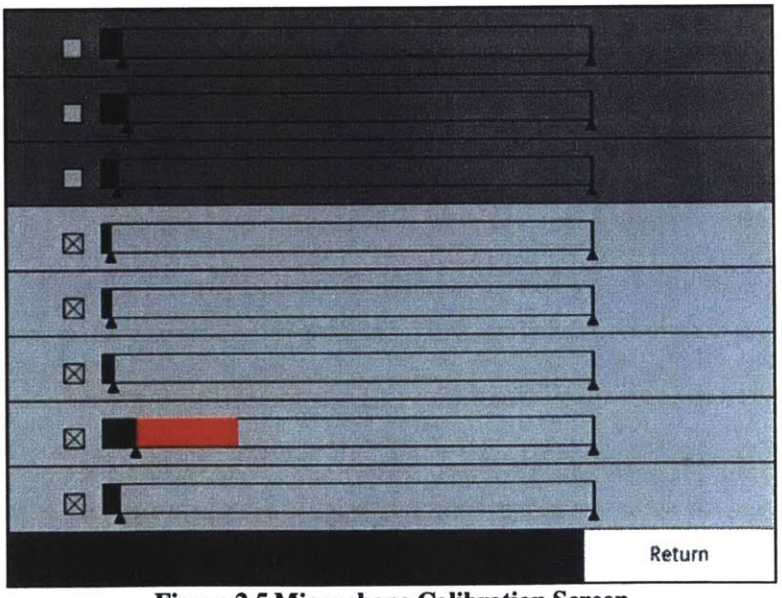


Figure 2.5 Microphone Calibration Screen

Prior to usage, each person's microphone was calibrated to their speaking level, so that when they spoke, and only when they spoke, the application would interpret that as a moment of speech. Figure 2.5 shows the application's screen that indicates the current

volume detection on the hardware's eight microphone channels. Each microphone setting can be adjusted with a slider to accommodate for louder or softer speaking individuals. In this screenshot, the top three microphones are turned off, and hence grayed out, and the seventh microphone is detecting that someone is speaking, indicated by the red bar.

VI. Conclusion

As discussed in Chapter 1, a significant challenge for groups is avoiding the process-based flaws of groupthink and group polarization. Therefore, if technology is to support more productive meetings and face-to-face discussions, a first step is to design technology that can capture and represent a meeting's content based on the process used by the group.

Towards this goal, the first version of Second Messenger uses speech-recognition and semantic-analysis techniques to gather and then display the conversational keywords onto a shared workspace. The application utilizes its knowledge of who said what to assist in finding the relevant keywords and to emphasize the contributions of "minority" group members. In this manner, this application attempts to equalize the contributions of individuals during a meeting in order to increase the diversity of ideas discussed and increase the quality of the decisions made. By working within the process-based construct that participants who contribute less to a conversation may be a minority opinion, the application works to equalize the contributions to the visual display.

The outcome of building this application was a realization that to understand how process feedback has an influence on group process itself, the measurement needed to be robust enough to provide an accurate representation of the group's interaction, even if only a very basic representation. At this time, off-the-shelf speech-recognition technology cannot provide accurate summaries of the content of a face-to-face meeting, and thus, measuring the speaking *quantities* of each individual in the meeting became the method for capturing the dynamics of the group. The next chapter addresses how this information can be used to visualize different patterns and behaviors in a group setting.

Chapter 3: Visualizing Behavior

In building interfaces for group reflection, my design goal has been to present information that encourages a group to increase the breadth of its discussion so that more viewpoints will be considered. As described in the previous chapter, my first work in this area focused on capturing spoken content using voice-recognition software and presenting discussion content to the group within an interactive idea-capturing tool (DiMicco 2004a). This chapter presents tools to visualize speaking amount as an indication of group interaction. The challenge in designing these tools is conveying the appropriate social message to a group with an awareness tool that does not mediate communication or dictate a group process.

This chapter begins with a discussion of related work in the areas of communication tools and social visualization, and then presents two applications that visualize the speaking patterns of who-spoke-when during a meeting, which are the focus of the evaluations in later chapters.

I. Technology that Changes Group Interaction

There are several different types of collaboration applications that provide guidance in how to alter the dynamics of a group's interaction for the purpose of assisting the communication. These include a subset of traditional groupware tools, social visualization tools, and select examples of computer-mediated communication tools that incorporate anonymity and biosensors.

A. Traditional Groupware Tools

For over two decades, researchers have been working on integrating technology into our group decision-making settings. These applications are commonly referred to as "groupware" or "group decision support systems" (for representative examples, see (DeSanctis 1985, Nunamaker 1991, Sproull 1991)). While this work includes many examples of systems designed for meetings, most of these applications focus on collecting and organizing documents and agendas related to the meeting (Mantei 1988, Nunamaker 1991, O'Grady 1994). As such, there are fewer examples of tools that assist in monitoring the social dynamics of a group to assist in communication. The following section highlights some of these applications.

While the "smart meeting room" initiatives described in the previous chapter record the behavior of the individuals in a meeting, none of these projects focused on issues relating to *changing* the behavior of the individuals. A notable exception to this is Project Nick, a meeting-augmentation project presented at the first ACM Conference on Computer-Supported Cooperative Work (Begeman 1986, Cook 1987). The project was a large-scale proposal for supporting group process within meetings through innovative communication tools. It allowed people to send private messages to the meeting's facilitator about their "mood" during the meeting to indicate interest or boredom. It also proposed that summarizing reports that stated how much each person had spoken during the meeting should be generated, although the authors proposed that these calculations would be done by hand. The goals of the project were similar to my own, in attempting improve group decision capabilities, yet the implementation was limited by existing technology and incorporated an assumption that a meeting facilitator would manage the technology.

Another important example is Jabber, a project by Kazman, et al. (Kazman 1995). This application collected information about who was speaking and, with voice recognition, determined some keywords of what was said, during a conference call. By providing a visual summary of the audio stream over the phone line, the tool could be used during a call to monitor who had been speaking and the general words that had been stated. While they did not have the goal of altering the group discussions, Jabber is an example of augmenting verbal communication by enhancing a group's awareness of the interaction.

A more recent project, by Chen (Chen 2003), studied how providing behavior feedback in a classroom setting could assist instructors in "feeling the pulse" of the classroom. In a remote classroom arrangement, the instructor could get visual feedback on who was gesturing and fidgeting during the lecture from camera sensors. As with our work, they used simple visualizations of activity to assist in understanding group interaction to a greater level, although this tool is designed for the group leader (the instructor), not the entire group.

Focusing on small, collocated groups, Huang's IMHere application (along with other projects at IBM Research) provides examples of how a large-scale, shared display can be used to increase a group's awareness of its behavior patterns over time (Huang 2004, Russell 2005). With an awareness of basic actions and movements of co-workers in and

out of the office, groups can coordinate their activities and serendipitous collaborative opportunities.

These groupware tools strive to increase a group's awareness of actions and communicative acts. Whether stated as an explicit goal or not, these tools will have an impact on how a group chooses to communicate and the process by which they share information.

B. Social Visualization

Research in the area of social visualization provides another set of examples of applications that provide individuals and groups insight into their patterns of interaction. My designs have drawn inspiration from previous work that focuses on how to convey social meaning through simple visual forms, as has been done by Donath et al. with the Chat Circles and VisiPhone projects (Donath 2000, Viégas 1999) and Erickson, et al. with the Babble project (Bradner 1998, Erickson 2000). As postulated by Erickson, et al. mutual awareness of actions, or "social translucence," enables more fluid, more expressive interactions between individuals, resulting in a more effective coordination of ideas, opinions and actions. While their visualizations are designed for online communication, this principle of increasing visibility of behavior and interaction patterns is an important design principle in any social software.

C. Biosensor Feedback in Social Settings

While not commonly considered in group collaboration, applications incorporating biosensor feedback are an important emerging area for collaboration (Mandryk 2004). For example, the Galvactivator, developed by the Media Lab's Affective Computing group (Picard 2001) displays a user's current skin-conductivity level on the outside of the glove, providing a public display of normally privately held information. Anecdotally, this glove allows individuals who are less vocal to contribute to a discussion by using the glove's light to indicate to the group that they would like to speak. This use of a display of nonverbal behavior is the sort of interface I am interested in introducing in a more systematic way, for the purpose of increasing the diversity of a group's discussion.

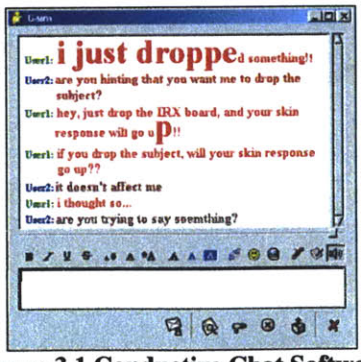


Figure 3.1 Conductive Chat Software

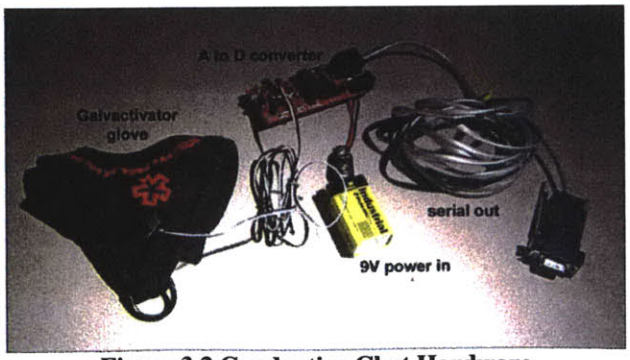


Figure 3.2 Conductive Chat Hardware

In an early research project with co-authors Lakshmipathy and Fiore (DiMicco 2002), I built an instant-messaging client called Conductive Chat that incorporated this skin-conductivity signal into the size and color of the typed text (Figures 3.1 and 3.2). The intention was that by incorporating a new channel of information about emotional state, users could be more effective in their communication.

D. Summary of Related Work

Each of the examples above illustrates different methods for making the behavior of individuals more explicit to a group and, as with my own work, the purpose in doing this is to increase mutual awareness between group members. Tools for reflection and awareness are informative to users, and in this sense are related to the work presented in the remainder of this chapter, but these tools were not designed with the explicit intention to alter the behavior of the individuals for the purpose of improving their collaborative acts, nor were they studied in terms of explicit behavior change.

The majority of the studies examining the impact of technology on decision-making have compared face-to-face groups without communication tools to groups that communicate via text through computer interfaces similar to today's most basic email, chat, and instant-messaging applications. Most of these experimental analyses of group behavior have produced negative results: when communication is *mediated* by technology, information sharing decreases, group polarization increases, decision quality degrades further, and groups report lower levels of satisfaction with the decision (Hollingshead 1996, Kiesler 1992, McLeod 1997, Rocco 1998, Siegel 1986, Sproull 1991).

One of the declared benefits of mediated over face-to-face communication is that mediation results in more equal participation across group members (Dubrovsky 1991, Siegel 1986). These findings have been attributed to a decrease in awareness of social status between individuals afforded by the anonymous interaction. Unfortunately, anonymity is rarely a feature of today's business communication tools and it is unrealistic to assume real-world groups will interact anonymously to make critical decisions. This reality severely qualifies this encouraging finding regarding mediated communication.

First presented as the Media-Richness Theory (Daft 1986) and later expanded by Reid, et al. (Reid 1997), one explanation as to why there are consistently negative results with regard to decision-making and mediated interactions is that the more constrained you are in your conversational abilities, the harder it is for you to express yourself. In the face of communication constraints, people naturally resort to more positional statements as compared to constructive, informational statements. In situations of group decision-making, a decision is easier to evaluate when the group is able to clearly articulate information and persuasive arguments rather than relying on emotionally driven comments.

As discussed earlier, these mediation effects on communication illustrate how the expressiveness of an environment influences our ability to communicate effectively. Yet within the relatively expressive environment of a face-to-face setting, groups still have

difficulties making high-quality decisions. Therefore, the challenge is to focus on how tools that incorporate new types of awareness into our face-to-face setting can enhance our expressiveness and decision-making abilities.

II. A Display of Speaker Participation

To understand how a dynamic display reflecting group behavior could influence group decision-making processes, I designed a simple interface showing fluctuating participation levels during a meeting. By removing the content-rich speech-recognizer input and replacing it with a simple detection of how much each person had spoken, the system could reliably show the quantity of speaking, providing a robust platform for controlled study.

As described in the previous chapter, over the course of a group interaction the application senses how much each person has spoken, and the interface, shown in Figure 3.3, reveals a dynamic display of relative participation. When the application detects that someone is speaking, a color-coded identifying circle moves across the top of the display and the bars on the histogram adjust by their relative percentage participation. Bars are colored and numbered to assist users in identifying their corresponding bar. The dots along the top of the screen show a visual representation of who has spoken over the previous 30 seconds.

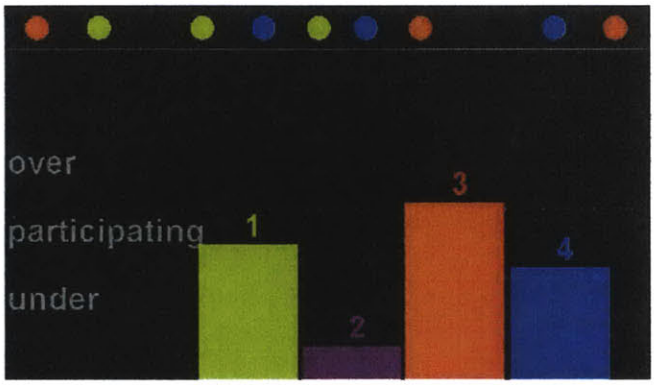


Figure 3.3 Speaker Participation Display.

This interface is designed to be a tool for group reflection, as well as a social interface implying a standard for appropriate group behavior. To make this social message more apparent and to aid the group in interpreting the height of the bars, the display is labeled with the words "under," "participating," and "over." The number of participants determines the relative vertical scaling of the histogram's labels. For example, with four speakers, the word "participating" aligns with 25%, and "over" and "under" are set to plus and minus half that value: 37.5% and 12.5%, respectively. As any display of social information has an inherent social meaning, this choice of words was deliberate to clarify the intended message.

As will be described in the next chapter, this display was used during a controlled experiment to understand how this explicit message would change the behaviors of those

at the extremes of participation. The experiment found that, with the display, over-participators and non-critical information holders tempered their comments, giving others more "air time" in the conversation; however, the display did not encourage under-participators to speak more, meaning that those that spoke more were in the middle-range of participation.

III. Second Messenger 2.0

After establishing that a display of speaker participation placed within a face-to-face setting influenced group participation levels (as will be presented in Chapter 4), the next direction was to build a system that better captures the expressive range of a group's interaction. This application, which supports real-time reflection and post-meeting review of group interaction, is called Second Messenger 2.0. It was built in Java, using the Processing Programming Language developed in the Media Lab's Physical Language Workshop Group (Fry 2004 - 2001). As with version 1.0, this application's role is to augment a discussion by providing a secondary channel of information to a group conversation.

The application reveals the speaking patterns of up to eight participants, either during a meeting or afterwards in "replay" mode. To do this, it detects who is speaking at each moment through the same hardware as used with the simple participation display, as was described in the previous chapter. With the collected data, the system produces five different visualizations, each showing a different perspective of the data. These visualizations attempt to emphasize different patterns in behavior to provide general indicators of group interaction.

When in use, each person in a meeting wears a noise-canceling microphone that detects when he or she is speaking (based on an initial voice calibration). When viewing the interfaces, the group can either view an aggregate of the information over the entire meeting, or watch a sliding window of the previous two or five minutes. When watching in real-time, it reflects the current interaction; in replay mode it plays back the interaction at ten times the original speed of the conversation. Figure 3.5a shows a screenshot of the application window with the playback buttons and viewing options along the bottom.

Figure 3.4 shows two members of a group using the application live during a meeting. The group is using a tabletop monitor in the center of the meeting space to view their participation levels. In other settings, groups use a large shared display off to the side of the meeting to view their behavior. In either situation, the technology is an ambient source of information, allowing the group to check in on their behavior when desired, yet neither disrupting nor dictating the process of the meeting.

Figure 3.4 A Group Using Second Messenger 2.0 with the Group Circle visualization displayed on a tabletop monitor.

A. Why a Replay Display?

Social psychologists suggest that one way a group can improve its interaction, and consequently its productivity, is by having a high-level understanding of its emotional and social interaction (Hackman 1992). Robert Bales introduced the idea that a group can gain insight into its interaction by viewing an aggregate report of its behavior (Bales 1999). Using Bales' SYMLOG system, trained observers can code group members' behaviors and produce a descriptive analysis of a group's interaction dynamics that can later be used by the group for reflection. By producing an automatic summary of the group interaction and allowing a group to replay it, I hypothesize that Second Messenger 2.0 can similarly assist groups in understanding and improving their interactions.

B. The Visualizations

Figure 3.5 shows the different visualizations available of Second Messenger 2.0. In each interface, the group members are identified by unique colors. Yet to encourage groups to use the display as a general indicator of participation, not as an evaluative mechanism for comparing individual behavior, in all of the interfaces, a double-click on any individual's color representation turns the image to an anonymous monochrome. In this way, the group can get a general sense of the imbalance but not know who exactly is the most extreme in his or her participation. The group can also choose to focus on certain individuals, for example the facilitator or designated leader, and not others. An example of this is shown in the histogram in Figure 3.5b.

The Histogram visualization (Figures 3.5a and 3.5b) is a bar graph similar to the display described in the previous section, showing the relative speaking amounts of each person. The explicit labels of "under" and "over" were removed to lessen the social message of equal participation. This straightforward way of comparing participation levels of group members is still relatively explicit; so much so that groups would sometimes request to use this visualization so they could explicitly comment on who was "winning" in participation.

The visualization in Figure 3.5c functions as a timeline that reveals who is speaking at each moment in the meeting. The circles down the left side of the display represent individuals, growing and shrinking in proportion to their contributions. The horizontal lines extending from the circles have vertical blue bars at the moments when that individual spoke. To highlight moments of overlapping speech, transparent vertical red lines are drawn over times where more than one person spoke. As a meeting progresses, the timeline moves right to left, compressing the available data into the available screen space. This timeline view was built as a way of observing turn-taking patterns within a meeting and to determine which moments of overlap indicated a switch in speaking turns. Through pilot testing it was found that this information is too distracting to view in a real-time setting, but is very informative and fodder for further discussion after a group interaction.

In the Fan visualization (Figure 3.5d), a single fan image represents the spread in participation between members, calculated as the standard deviation in participation. As the spread increases or decreases over time, the width of the fan grows and shrinks accordingly. This is designed to explicitly remove any sense of individual identity and treat the group as one unit, where an extreme imbalance in who has been speaking is signaled by an abnormally wide fan.

In the Bouncing Ball visualization (Figures 3.5e and 3.5f), each person is represented by a circle of fixed diameter where the vertical position reflects the individual's relative participation level. When viewed with a limited time window (the previous two or five minutes), this view gives the visual impression of bouncing balls that play off of each other as turn-taking and floor control switches between different people. When viewed in anonymous mode, the circles near the equal participation level are colored white, and then as they move further to the extremes, they turn increasingly brighter red.

The Group Circle visualization (Figures 3.5g and 3.5h) represents each group member as a circle that grows and shrinks in proportion to that person's participation level. The position of the circles is chosen by the group—the group can arrange the circles (with a mouse) to mimic the physical orientation of the group during the meeting. The grey circular shape around the colored circles then adjusts its shape and size to accommodate the physical arrangement. By allowing the group a way to mirror itself, this display emphasizes the intimate nature of a face-to-face meeting. Additionally, while this interface conveys the same information as the Histogram does, it is more difficult to make fine-tuned visual comparisons of circles sizes. Therefore this display is able to emphasize extreme differences in size, rather than encourage a ranking of participation, which groups frequently did when viewing the bars.

The Group Circle display also provides information about the proportions of overlapping speech during the meeting. By clicking on a circle, it reveals who was speaking at the same time this individual was speaking (Figure 3.5h). These pie slices represent moments of agreement, backchannel comments, side comments, laughing, and interruption. Frequently higher levels of overlap indicate synergy and agreement between individuals. The representations of overlap go both ways: e.g., when yellow laughs with purple, it is counted as overlapping speech on both yellow's circle and on purple's circle. Using this

view, it is possible to observe unique relationships between individuals by seeing disproportionately different amounts of overlap for different pairs of people. As with the Timeline, this higher level of information was found to be too much to interpret in a live setting, so was only used when reviewing a meeting's interaction after it occurred.

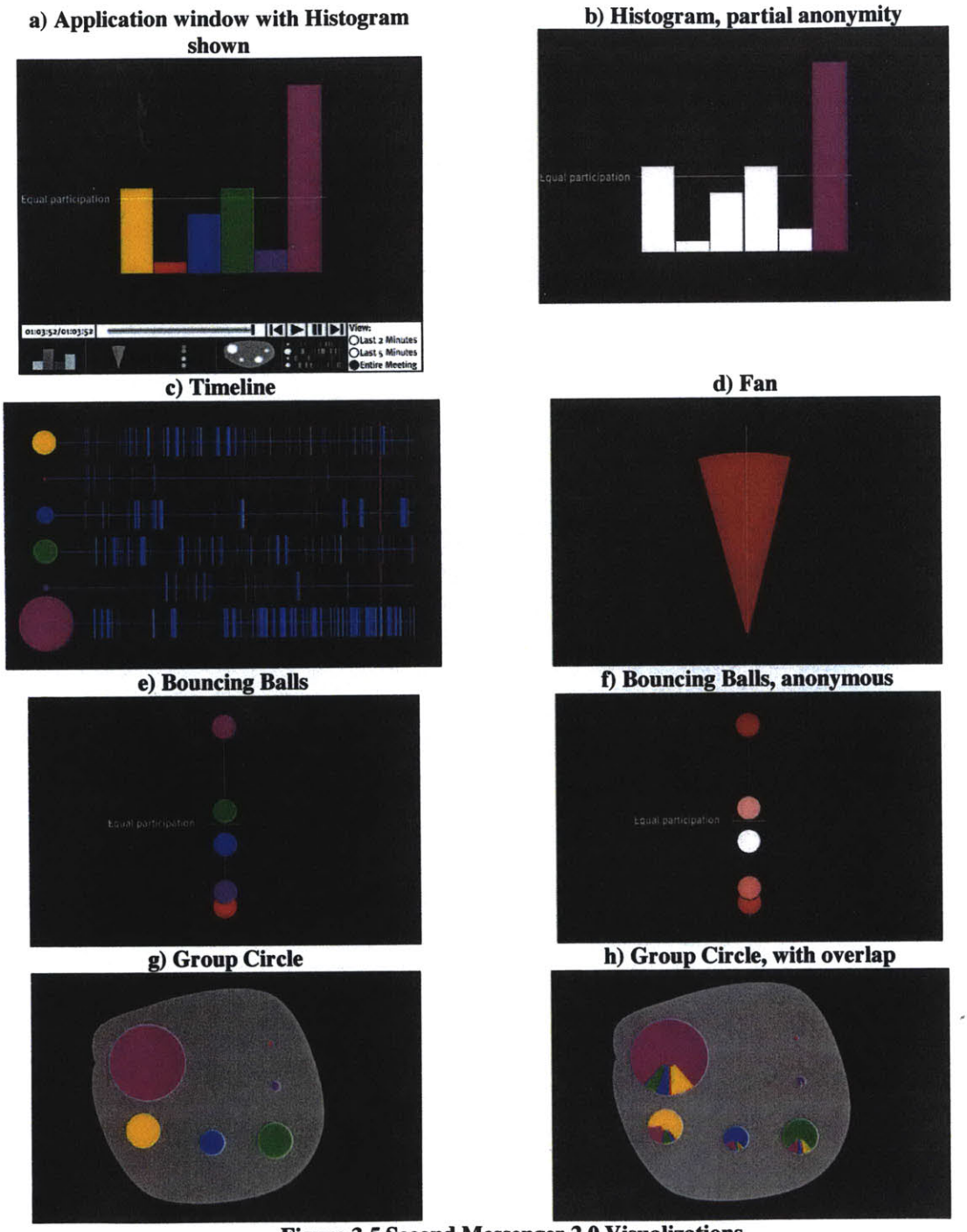


Figure 3.5 Second Messenger 2.0 Visualizations
These screenshots capture the same moment in a conversation of six people.

These different visualizations allow for flexibility to accommodate to a group's individual personality and style. When real-world groups were first introduced the application, it was recommended that they user either the Histogram or the Group Circle view on first usage because they were the simplest to interpret during a meeting. As a group became more accustomed to the displays, they would try different views and even switch back and forth between views during a meeting. A thorough discussion of these groups' interactions with the application is presented in Chapter 5.

IV. Efficacy of the Visualizations

In contrast to when a group is directly observed, these visualizations provide no information about what is being said, the tone of voice used, the relative speaking volumes, or any nonverbal communication such as gesturing, eye gaze and head movements. Without this wealth of information, can anything be conveyed about individual personality and behavior? To determine if Second Messenger 2.0's visualizations effectively provide summarizing information about a group's interaction, I conducted a user study asking subjects to observe two groups with the application and make judgments regarding their behavior.

A total of eight subjects used Second Messenger 2.0 to make qualitative assessments of two groups' interactions. Each subject was given a ten-minute tour of the application, while looking at data of an example group, and then was given as much time as needed to study the two groups with the different visualizations and to fill out a questionnaire about the group dynamics and individual behaviors.

The study's subjects were researchers affiliated with the Media Lab, but none of them had had prior exposure to the application or this research project, nor did they know the individuals in the visualized meetings. Gender was split evenly and the average age was 26 years. We chose to have outside subjects participation (those who had not taken part in the original meetings) because when meeting participants use the application, they make conclusions about what they see based on their memory of the interaction, and we wanted to know in absence of this context what information was conveyed.

The two meetings used in the evaluation had occurred as part of a previous experiment where the groups were videotaped and recorded using our hardware. These meetings had four members each and spoke for an average of 20 minutes. During this time, the groups debated two sides of a defined issue and reasoned back and forth about the different options until they reached a consensus. These two groups were selected out of the pool of available meetings because they had distinct personalities within them: the first group had one animated member who was unfocused in his contributions and a second animated member who facilitated the consensus process; the second group had one controlling member who directed the discussion leading the group to ignore a very quiet member. The goal of our evaluation was to determine if distinct personalities, such as the ones we observed on the videos of these groups, could be distinguished from one another using the visualizations.

A. Evaluation Procedure

Subjects spent on average 30 minutes assessing the two groups, slightly less time than it would have taken to watch the two videos. They were given the option to evaluate the groups in parallel or one after the other and they were encouraged to use any of the visualizations as part of their evaluation.

All of the subjects used each of the visualizations at some point in filling out the evaluations, yet they often chose to use just one or two visualizations for the majority of their answers. Since the Group Circle and Timeline visualizations contain the most detailed information about the speaking patterns, these tended to be the most frequently used interfaces.

We evaluated subject assessments by comparing them with our own assessments of the groups based on watching video recordings of the groups. To perform this assessment, two observers filled out the questionnaires while watching the videos separately and, in cases where the observers disagreed in their answers by more than 2 points (on a 7-point Likert scale), ratings were removed from the results reported below. (We did not have permission to show the videotapes to anyone other than the experimental investigators, thus no one else watched these tapes.)

B. Findings

The findings from the evaluation indicate that certain behavior traits are more effectively conveyed in the visualizations than others. Specifically, personality traits that are expressed with explicit behavior were reported with higher consistency than those traits that imply a positive or negative intention. For example, subjects were better at distinguishing the introverts from extroverts than they were at determining who was cooperative versus aggressive. The study also asked subjects to assess which sides of the debate each person fell on and these ratings were strikingly accurate.

The first part of the evaluation questionnaire asked subjects to evaluate the groups and the individuals along several personality dimensions. These questions were derived from Bales' behavior-trait questions (Bales 1950) and were asked on a 7-point Likert scale. Tables 1 and 2 provide detailed summaries of these responses.

In these two tables, the horizontal bar represents the range of answers provided by the subjects, the "O" marks the mean response, and the "X" marks our assessment of the group based on watching the video. While many of the questions are qualitative judgments and do not have empirically correct answers, as can be seen in the broad range of user responses for some questions, for the purpose of evaluating the application we used our assessment (the X's) as a standard by which to compare the subject responses.

In the group-level evaluation (Table 3.1), subjects were relatively consistent in how they rated the level of interruption and these ratings agreed with the video evaluation. They rated turn-taking as more structured and faster than it appeared in direct observation, and

the equality of contribution, a quantifiable measure, was rated as more balanced than the video revealed. We attribute these inaccurate assessments to the subjects' limited exposure to the application and not yet having a good sense for what average or typical behavior looks like in the visualizations.

Table 3.1 Evaluation of Group Personality
O = Mean subject rating, X = Investigators' rating.
Missing X's indicate investigators did not agree on rating.

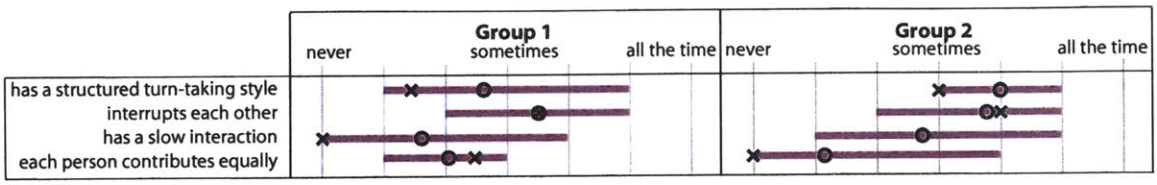
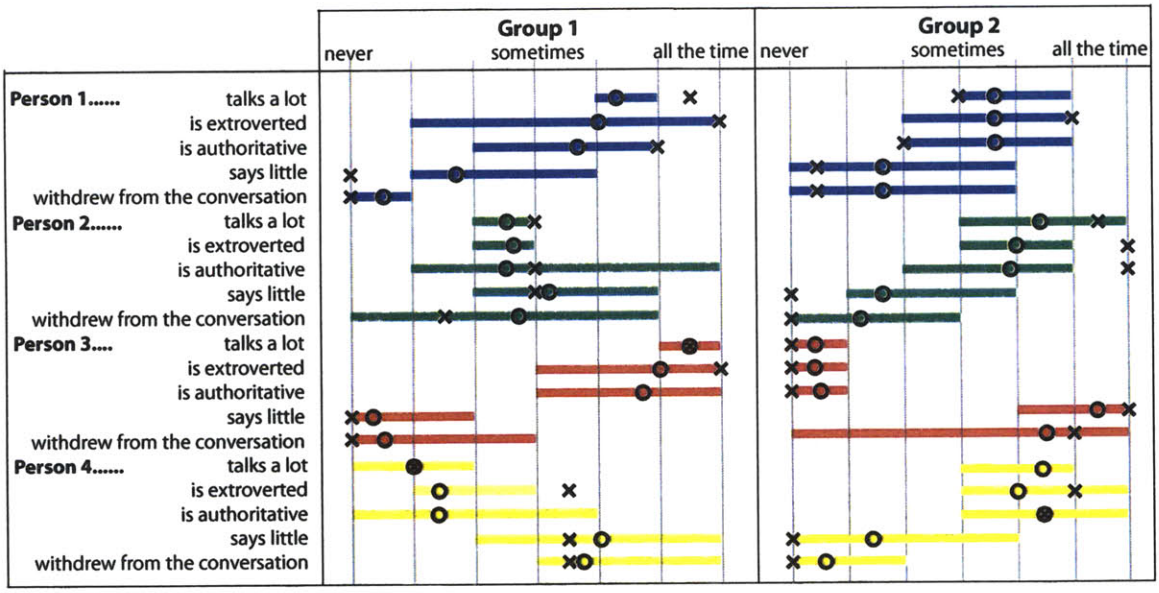


Table 3.2 Evaluation of Individual Personalities
O = Mean subject rating, X = Investigators' rating.



The individual evaluation results (Table 3.2) reveal that subjects perceived the individuals in the meeting to exhibit distinct personality traits. For example, in Group 1, the subjects determined that Persons 1 and 3 spoke more, were more extroverted, and never withdrew from the conversation, as compared to Persons 2 and 4. In Group 2, subjects determined that Person 3 was very different in demeanor than the other three, in speaking less, not being extroverted, not having authority, and withdrawing from the conversation. These were the individuals mentioned earlier that we had hoped the subjects would be able to distinguish from the others in the group.

The second step of the evaluation asked subjects to determine which individuals were on opposing sides of a debate and the results for this were highly accurate (Table 3.3). Based on observing who was exchanging turns, who was overlapping whom and who spoke more or less, subjects could determine which individuals were debating against each

other. In the one case where we could not determine this from the video, neither could the subjects.

Table 3.3 Evaluation of Sides of the Debate
How subjects rated each person in terms of which side of debate they fell on.

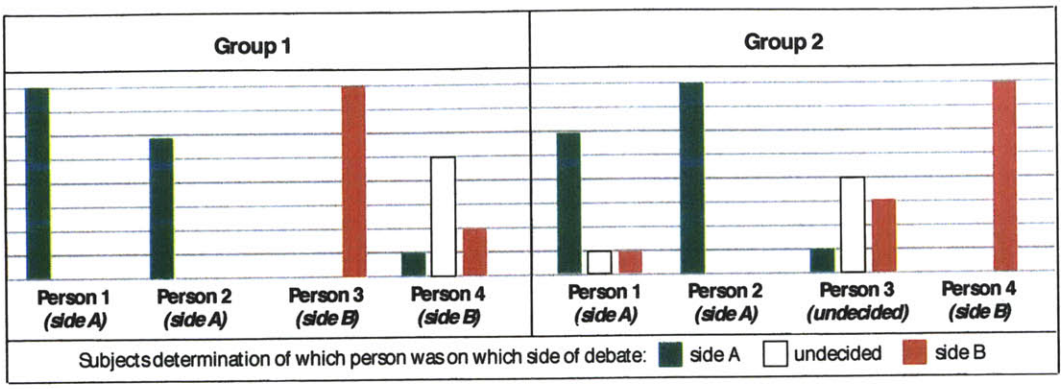


Table 3.4 Adjectives Describing the Groups.
(x) = number of subjects (out of eight) naming this trait.

	Adjective named by both	Only by Subjects	Only by Investigators
Group 1	agreeable (4)	argumentative (4)	active (0)
Person 1	confident (4) interrupting (4)		extroverted (2)
Person 2		respectful (4)	
Person 3	confident (5) extroverted (4)	aggressive (5) interrupting (5) authoritative (4) controlling (4)	facilitating (1) cooperative (1)
Person 4		quiet (7) passive (5)	respectful (0) cooperative (0)
Group 2		active (5)	structured (0)
Person 1	cooperative (5)		confident (2)
Person 2	confident (4) interrupting (3) controlling (3) imposing (3)		authoritative (2) uncooperative (0)
Person 3	quiet (7) passive (3) withdrawn (3) shy (3)		introverted (2)
Person 4	extroverted (3) confident (3)	aggressive (5) interrupting (5) argumentative (4)	outgoing (2) cooperative (1) facilitating (1) agreeable (0)

The final section of the evaluation asked subjects to select adjectives from a list to describe the groups and individuals. Table 3.4 shows which of the pre-determined adjectives were named by both the subjects and ourselves, only by the subjects, and only by ourselves. The traits that subjects accurately named with high frequency included

"*confident*," "*extroverted*," and "*interrupting*." Traits that subjects named with high frequency that were *not* reflected in the video were "*argumentative*," "*aggressive*," "*authoritative*," and "*passive*." Unlike the correctly named traits, these traits imply a negative intention or attitude toward others. This may indicate that the visualizations emphasize or imply aggression when the groups did not exhibit this in person. Traits the subjects did *not* name that were shown on the video were "*active*," "*structured*," and "*uncooperative*." Determining activity level and structure requires knowing what average or normal behavior looks like, something the subjects could not do with a first-time use.

C. Evaluation Conclusion

The user study's findings indicate that the strength of the visualizations is in revealing the extreme differences between the behaviors of individuals in a meeting. Subjects could pick out the extroverts, the quiet individuals, and the controlling personalities. Subjects had more difficulty determining the mood of a group's interaction and were inconsistent in their assessment of individuals on traits that involve personal motivation and intention, such as aggressiveness versus cooperativeness and controlling versus facilitating. Most notable of the results is that based on speaking patterns alone subjects were able to detect which sides of a debate each person was on.

With the limited number of subjects in this study, we can not make definitive or statistically significant statements about these results, yet it appears that while not as complete as a video analysis, the visualizations conveys general impressions to an observer that provide insight into the different personalities in a group. The experimental subjects, using the application for the first time, took the same amount of time to make these assessments as they would have had to take if watching videos of the group. We theorize that with increased familiarity with the application, users could use the system with much greater efficiency, providing a potential benefit over watching videotaped group interactions.

V. Conclusion

This chapter presented visualizations designed to convey information about a group's face-to-face turn-taking and participation patterns. The intention of these displays is to reveal the extent to which a group is balanced in its participation by illustrating the different behavior patterns of each person. With this increased awareness of their interaction, particularly within the context of an on-going group interaction, it is expected that groups will alter their behavior and information sharing patterns.

Chapter 4: Quantitative Evaluation of Group Behavior

As the goal of this research is to *change* group interaction through visual displays reflecting social behavior, a central piece of this work has been to run both observational and quantitative evaluations of groups using the applications presented in the previous chapter. This chapter presents two quantitative studies of groups: the first experiment measures how individual participation levels change in response to the introduction of a bar graph indicating the group's fluctuating participation levels (previously published as: (DiMicco 2004b)); the second experiment studies how group information sharing and individual behaviors are influenced by both the presence of a display of speaker participation as well as by having viewed a visual replay of the group's previous interaction. Chapter 5 will present the qualitative observations made during these experiments and of pre-existing, real-world groups using the applications during their meetings.

As the following sections will describe, the experiments found that a display showing participation levels, imposing a norm of equal participation on groups, causes those at the highest levels of participation to decrease the amount they speak. Reviewing the turn-taking patterns with a visualization causes those who spoke the least to increase the amount they speak in a subsequent discussion; real-time feedback does not produce this change. After reviewing their turn-taking patterns, groups alter their information sharing strategy during a second group decision task. For groups that had a poor sharing strategy, this improves their ability to share information to a level equivalent to the best information-sharing groups. For groups that had already established a successful sharing process, their subsequent change in strategy was harmful to their information sharing. In addition to these results, reactions to the displays and perceptions of the task, self-performance, and group dynamics are presented.

I. Related Technology Studies

The experimental design of both studies is modeled after established tasks and methods from the "hidden profile" literature in social psychology, as will be discussed in the next sections. As such, the most related studies are not from the areas of human-computer interaction or computer-supported cooperative work. Yet there are a few examples worth mentioning of experiments in social psychology that specifically examine how technology changes human-to-human behavior.

There are two recent studies that utilize known psychological phenomena, social loafing and the bystander effect, to understand online group interaction. A project led by Kraut and Resnick at CMU and University of Michigan examined how mechanisms for limiting social loafing— the putting forth of less effort when working in a group versus when working alone—in the offline world could be applied in the design of an online community, demonstrating that the same preventative measures can be effective in the online domain (Beenen 2004). Hudson and Bruckman at Georgia Tech have used the concept of the bystander effect—the tendency to not intervene when others are present as compared to when they are not—to understand and interpret their observation that online classrooms have more equal participation than their offline counterparts (Hudson 2004). This utilization of social psychology as a lens for understanding online interaction illustrates how well-studied observations in our non-digital interactions apply to the digital realm as well.

A recent dissertation at the University of Geneva examined how reflections of behavior, in the form of either a "mirroring tool" or a "metacognitive tool," influence the behavior of groups during a strategy task (Jermann 2004). Their experiments found that a visual representation of the desired state of group interaction was critical for the group's regulation of interaction and was used to alter behavior. Their metacognitive tools provided precise feedback to groups about their progress on a task and these tools assisted groups in their behavior regulation and performance. While their tools cannot be applied to any type group interaction, since they were designed for a specific task, their finding that groups respond and adapt their behavior according to the information provided within a reflection tool corresponds with our own experimental findings.

The one project I am aware of that has given process-based feedback to groups in a computer-mediated environment is a project done several years ago at the EDS Center for Machine Intelligence (Losada 1990). By hand-coding group behavior based on Bales' criteria (Bales 1950), and then providing groups with a full analysis of their interaction dynamics during their task, they found that those communicating through computers increased their amount of socially and emotionally motivated comments, as compared to those without the feedback, who did not. This change in the online groups brought them inline with the ratios typical of face-to-face groups that did not have feedback. These results are interesting because they demonstrate that the feedback had an impact on their group interaction patterns, but these results are difficult to extrapolate from because of the limitations of the technology deployed for the group decision-making tasks (they

contrasted basic text chat with natural face-to-face communication) and the complexity of the interaction feedback provided to the groups (Bales' full SYMLOG output).

II. An Information Sharing Task

In the two experiments described in the remaining sections of this chapter, groups were asked to perform two information-sharing decision tasks. Information-sharing tasks, or "hidden profile tasks" (Stasser 1992), are designed to measure how and when individuals share privately-held information in a group discussion. In terms of McGrath's taxonomy of tasks, these are "intellective tasks" in the sense that there was a correct answer to the decision that a group could only come to through discussion (McGrath 1984). Previous hidden-profile studies reveal that the framing of the group decision greatly influences the amount of information shared and in most cases, groups focus their discussions on the facts that the entire group already knows, rather than talk about the facts that are privately held by only a few individuals (Dennis 1998, Hollingshead 1996, Stasser 1992, Stasser 1987, Stewart 1998).

Information sharing tasks are particularly well suited for studying group participation, as demonstrated by a study run by Stasser and Taylor (Stasser 1991). In their experiment, they directly examined how speaker participation rates related to this issue of information sharing. As one would expect, subjects that spoke the most in the experimental task shared the most information. Somewhat less expected was that they were able to predict the outcome of the group's decision solely from knowing who spoke during the first *ten minutes* of the discussion, because the subjects who spoke the most during these initial minutes had the most influence on the decision. While this finding does not universally apply to all group decisions, it is suggestive that speaker participation in a group setting is a valid metric for quantifying an aspect of information sharing and that information sharing tasks are an appropriate platform for examining participation and decision-making.

The specific decision tasks given to the subjects during the experiments were to select a student to admit into MIT's undergraduate program and to select a neighborhood for a new 24-hour convenience store in the Boston area. For each of these topics, there were three profiles (three students and three store locations) and each profile had a total of fifteen facts. These facts were distributed to the experimental subjects in different ratios so that everyone knew the majority of the facts, but each person had some facts that only they knew. Each subject was provided the same number facts for each profile. In the first experiment, each subject had nine facts, of which seven were public and two were private. In the second experiment, each subject had twelve facts of which only one was private to that subject. The details of these two tasks are in Appendix D.

As will be discussed in the results, subjects in the first experiment discovered the "trick" of sharing information early in the discussion and this meant almost all groups shared the hidden information and came to correct decision. For the second experiment, with much less hidden information, it was more difficult to share the information and thus more

difficult to come to the correct answer. Based on these modifications and other changes to the presentation of the task to the groups, this second experiment was more successful in getting groups to demonstrate poor share of information in their discussion and therefore provide a mechanism for measuring the impact of the displays on information sharing. The modifications to the second experiment were made based on guidelines for what factors hinder information sharing in groups (Wittenbaum 2004).

The topics of student admissions and convenient store locations were designed to be something the subjects were somewhat familiar with, so they would feel comfortable discussing it, but not expert in. This way the information provided during the study would be the most influential to the decision. Each subject was required to be from the university community, so that they would have a similar perspective on the topic and some amount of common ground for discussion.

III. Evaluation Metrics

The following two experiments use information-sharing tasks to evaluate different aspects of a group's interaction. Traditionally, information-sharing experiments measure how many hidden facts are shared to determine task performance and then survey each group on different aspects of the group and the decision after each task. In addition to these measures, the technology used during these experiments captured when each person spoke, and with this data collection, there are several additional measures by which to evaluate group and individual behavior.

The first experiment uses the calculation of individual participation level to examine how much certain types of individuals spoke during the experimental tasks, specifically those that spoke at the extremes of participation ("over-participators" and "under-participators") and those who held critical information within each decision discussion. The questionnaires for this experiment focused on questions of group dynamics, such as inter-group trust, and how each individual felt about the visual display.

The second experiment also used individual participation level calculations to examine how much individuals spoke in the experiment, and then further analyzed speaking patterns to examine whether turn-taking and other patterns in turn switching were influenced by the technology. The questionnaires for this experiment were expanded from the previous study to ask individuals about satisfaction with personal contributions, satisfaction with different aspects of the group decision (performance, efficiency, difficulty, trust) and cultural background.

The following sections present the two experiments in detail, along with their experimental results. The results are organized by experimental hypotheses and evaluate each of the above-mentioned variables in terms of experimental condition.

IV. Changing Individual Participation with a Shared Display

The first experiment examined the impact of a display of speaker participation. The specific display used in the experiment was presented in Section III of the previous chapter.

Participation levels are not a complete measure of successful group dynamics, yet when meetings are focused on collaborative decision-making, particularly when information must be exchanged between group members, extreme imbalances in participation can signal an imbalance in a group's consideration of the different opinions and viewpoints relevant to the decision at hand. So while equal participation is not ideal for all types of meetings, in certain circumstances, such as in information-sharing tasks, it can signal a flawed information-sharing process. Thus, this first experiment was designed to understand if a display of participation levels would have any impact on individual participation levels.

A. Study Design

The experiment had two conditions: control and experimental. In the control condition, the groups discussed the two information-sharing tasks with no shared display. In the experimental condition, the first decision was made without the display and the second was made in the presence of the display. Prior to the second task, the group was given a brief introduction on how the display worked and was told to use the display in any way they felt was appropriate. During both conditions, subjects wore microphones and had their individual speaking times recorded. The task topics of student admissions and store locations were counter-balanced across condition.

The photographs in Figure 4.1 show subjects in the second task of the experimental condition. The histogram display was projected onto a wall and on the opposing wall was a mirror that enabled subjects seated near the display to look into the mirror to see the display rather than turning away from the group to look. Subjects also had a number placed in front of them that corresponded with the number on their histogram bar.

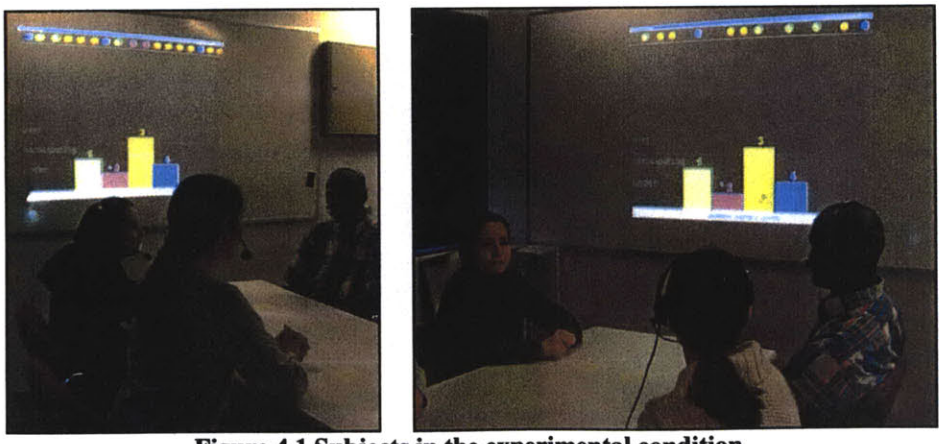


Figure 4.1 Subjects in the experimental condition.

After each task, subjects filled out a questionnaire about both the decision and the group interaction. After the experimental condition, subjects also answered questions about the display.

B. Study Hypotheses

Based on theories on the formation of group norms (Bettenhausen 1985, Hackman 1992) and on the self-regulation of behavior (Carver 1998), any display of information about group performance should cause individuals to be more aware of their own behavior and to attempt to change it, in order to comply with the normative pressure to behave in the same way as the rest of the group.

Therefore, by presenting equal participation as a standard of measurement, it is hypothesized that those participating at the extremes would feel pressure to comply with the group's average behavior and would correspondingly adjust their participation level. Yet, because the groups had already established norms of conduct during the first task, where information sharing and discovery of critical information were important aspects of the discussion, it is predicted individuals would not adjust their participation levels to the point of sacrificing the sharing of relevant information. Also, because of the peripheral nature of the display, it is expected to take a peripheral role in the group interaction, avoiding the negative effects technology has been shown to have on group dynamics, such as decreasing trust and comfort level within the group (Greenspan 2000, Rocco 1998).

These predictions are summarized as three behavioral hypotheses:

> ***H1:*** *In the presence of the display, subjects who over-participated in the first task will speak less and subjects who under-participated in the first task will speak more. Subjects who spoke an average amount will be unaffected by the display.*
>
> ***H2:*** *Subjects who hold critical information relating to the decision will exhibit no changes in the presence in the display, while those without critical information will limit their contributions when seeing the display. This change will result in critical-information holders having a greater relative speaking time.*
>
> ***H3:*** *The display will have minimal-to-no impact on perceptions of the task and on the group's interpersonal dynamics, such as trust formation between group members.*

Beyond these hypotheses, the experiment attempted to gauge the reaction to this type of interface by surveying the subjects on their perception of the display and their perception of their individual performance.

C. Method

1. Experimental Procedure

The study was designed so that four subjects, previously unknown to each other, were given two decision tasks on which to come to consensus. Each decision began with the subjects reading information sheets about the task. They were then told to discuss the task with the group to make a decision. They were explicitly told that most of the facts they were given about the decision had been given to everyone, but that each of them had some information that only they knew, so it was important to share information. The subjects were allowed to consult their information sheets during the discussion so individual memory would not impact the amount a subject shared with the group. In this way, the experiment set up an idealized decision-making scenario where the groups had no pre-existing social norms or status and everyone explicitly knew that the best decision came from sharing information.

Additionally, something not done in the later experiment was that the relative importance of each person's information varied by task. In each task, *one* subject held critical information that was designed to persuade the group towards a certain answer. In the student-selection task, one subject had information about all of the student recommendation letters and, in the convenience-store task, another subject had all the financial information on expected revenue and taxes for each location. In both tasks, this information was crucial enough to the decision that once mentioned would become the focus of the discussion.

2. Participants

A total of 100 subjects were recruited from the university community and randomly assigned to 25 four-person groups. The average age was 25, with about two-thirds of the subjects being students and one-third being members of the larger community. Gender was split approximately in half and 23 of the 25 groups had members of both genders.

Two of the subject groups were eliminated from the data analysis due to unusual behavior. One of the experimental groups spent one minute on the discussion task, rendering the differences in individual participation meaningless. And one of the control groups contained a subject from outside of the university community who elected to only participate in the task topic unrelated to university life, rendering measurements of change in participation irrelevant. As a result, the remaining analysis contains 11 control groups (44 subjects) and 12 experimental groups (48 subjects).

D. Results

The experiment results indicate that the shared display of speaker participation influences the amount subjects spoke relative to how much they had spoken during the first task and what information they held in the second task. Results also indicate that the subjects'

perception of the display's impact on the group was quite low, yet the level of trust between participants may have been adversely impacted by the introduction of the display. As a potential aid in explaining the findings, there are also significant patterns in how subjects perceived their participation and the accuracy of the display.

The following discussion of the results is organized into a general overview of the data results and then an analysis of the data in terms of the three hypotheses.

1. Overview of Results

This section covers the general statistics of the groups: how much time was spent on the tasks, task accuracy, task difficulty, and other general trends.

Time Spent on Tasks

The time groups took on a task ranged from 4 minutes 40 seconds to 28 minutes 39 seconds, with an average time of 13:45 minutes. With this high rate of variance (SD = 6:47 minutes), there were no significant differences between tasks based on order, topic or condition. Within group, there was a strong, positive correlation between the times spent on the two tasks (Pearson correlation coefficient $r=.596$, $p<.005$).

Task Accuracy

Subjects performed very well on the tasks. In the control condition, the eleven groups located the correct answer every time in the first task and ten out of eleven times in the second task. Four of the twelve experimental groups made incorrect decisions: two in the first task and two in the second task. There were no significant differences in accuracy between the two conditions. This overall high task accuracy indicates that not only was information shared sufficiently, but that the instruction to share information with the group was sufficient to ensure it happened in all conditions. (In the second experiment presented in this chapter, the tasks were altered to make them more difficult, so that task accuracy and the amount of information shared could be evaluative measures.)

Task Difficulty

Subjects rated the two task topics as equally challenging. On a 7-point Likert scale of difficulty (with 7 meaning very difficult), control subjects rated the convenience-store and student-selection topics as 3.15 and 3.09 and experimental subjects rated these as 3.09 and 3.04. No significant differences were found between these ratings.

After the second task, subjects rated the task's difficulty "as compared with the first task" where 4 out of 7 meant the second task was the *same* level of difficulty as the first. On average, subjects rated the second task to be more difficult: the average response was 4.52 for control and 4.39 for experimental groups. There was no statistical difference between these ratings. This perception of increased task difficulty can be attributed to an increased effort applied by all groups on a second attempt.

Participation Rates

In groups of four subjects, the average participation across the group will always be 25%, and thus in the experiment the average participation across conditions and tasks was 25%. The participation rates of males were no different from that of females (average male participation: 24.78%, average female participation: 25.20%).

The amount an individual subject spoke was highly correlated between tasks. In the control condition the Pearson correlation coefficient between Task 1 and Task 2 was $r=0.501$ ($N=44$, $p<0.001$) and in the experimental condition was $r=.553$ ($N=48$, $p<0.001$). When examining the change in individual participation between tasks, there were no statistically significant changes in participation rates across conditions (average change in ctrl: 0.01%; exp: 0.07%).

2. *Evaluation of Hypothesis 1*

The first experimental hypothesis was that subjects who over-participated in the first task would decrease the amount they spoke when they saw the display and that those who under-participated in the first task would increase the amount they spoke when they saw the display.

Categorizing Under- and Over-Participators

To test the hypothesis, subjects were divided into three categories: under-participators, middle-participators, and over-participators. This was done in order to understand how these three separate groups naturally varied their behavior between tasks and if they responded in different ways to the presence of the display.

To perform this categorization, subjects were pooled within each condition and categorized based on his/her rate of participation in the initial task. The divisions of under, middle and over were determined by the mean participation rate (25%) and the standard deviation of the distribution of participation in the control and experimental conditions. Under-participators were defined as those participating at a rate lower than the mean minus one SD and over-participators were defined as those participating at a rate higher than the mean plus one SD. This categorization left approximately 68% of subjects as middle-participators, and 16% as under and 16% as over. This division was a subjective decision, so I made a conservative choice by categorizing only those subjects in the extreme tails of the distribution as being outside of the category of "middle-participation."

Exploratory Data Analysis

After performing this split, the change in participation rate was calculated for each subject (participation in Task 2 – participation in Task 1) and these new datasets were examined for violations of normality. Through an inspection of each dataset's kurtosis, skew, and studentized deleted residual metric and with a visual inspection of their

boxplots and normal Q-Q plots, it became apparent that two of the datasets, specifically the under- and over-participators in the experimental group, had problematic outliers.

The boxplot and the studentized deleted residuals revealed that one subject in the over-participator category and another subject in the under-participator category (from two different groups) were skewing the distributions of the datasets to such a significant degree that we could not assume normal distributions. By removing these subjects, along with the other three subjects from their groups, the maximum studentized deleted residual in the under-participator group was reduced from 3.546 to 1.949 and in the over-participator group from 6.465 to 2.445 (the rule of thumb is that values near 4.00 and above indicate that the associated datapoint is significantly skewing the population away from normal). Additionally, this data reduction produced boxplots that no longer highlighted outliers. Because of these changes, the analysis of under, middle, and over-participators was conducted without these two problematic groups. There is not a high-level explanation of why the two individuals in these groups behaved abnormally. The individuals greatly increased the amount they spoke in the second task, unlike the other subjects in their participation category, and were both critical information holders in the second task, but they were not the only individuals in the experiment to match this description.

Data Analysis

After this step of data reduction, the experimental condition was left with 40 subjects in total: five subjects in each of the over and under categories and 30 in the middle category. The control condition had 44 subjects in total: eight in each of the over and under categories and 28 in the middle category. As stated previously, these categorizations were determined by a split of the subject pool based on the mean and plus/minus the standard deviation of the participation rates in Task 1. (In the control condition this was 25±8.26%; in the experimental condition this was 25±8.90%.)

To offer some perspective on these data in terms of the actual groups, of the ten experimental groups, six had members that fell into either the over or under categorizations; of the eleven control groups, nine had members in either the over or under categorization. Thus the majority of groups contained individuals with extreme participation levels, and only six of the 21 groups had neither under- nor over-participators.

To determine the significance of the change in participation for each of these groups, a paired t-test was performed to first determine if there were significant changes within each group. Then a t-test of independent samples compared the changes in participation across the two conditions. These results are shown in Table 4.1 and graphically in Figure 4.2. The statistically significant differences are noted with red asterisks in the Figure.

Table 4.1 Participation Rates for Each Participation Category and Condition with statistics for significance of change and difference.

Participation & Condition		N	Task 1	Task 2	Average Change	Standard Error	Paired t-test	Independent t-test
under	ctrl	8	12.9	21.4	8.43	1.8	t(7)=–4.546 **p<.005**	t(11)=2.804 **p<.05**
	exp	5	11.5	12.3	0.84	1.7	t(4)=–.507 p=.639	
middle	ctrl	28	25.3	25.0	–0.28	1.5	t(27)=.191 p=.850	t(56)=–1.23 p=.223
	exp	30	24.6	26.9	2.27	1.5	t(29)=–1.558 p=.130	
over	ctrl	8	36.0	28.6	–7.45	2.1	t(7)=3.597 **p<.01**	t(11)=2.453 **p<.05**
	exp	5	40.9	26.4	–14.48	1.4	t(4)=10.512 **p<.001**	

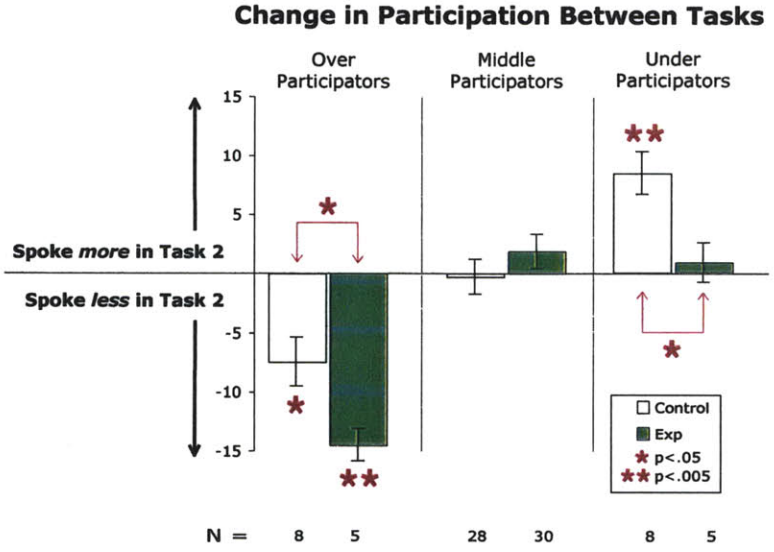


Figure 4.2 Average Change in Participation Rates by participation category and condition. [2]

Over-participators in both conditions had a significant decrease in the amount they spoke during a second task as compared with a first task (t(7)=3.597, p<.01; t(4)=10.512, p<.001; paired t-tests). The decrease in participation was significantly greater in the experimental condition (t(11)=2.453,p<.05, t-test of independent samples). Middle-participators showed no signs of change in behavior between tasks and across conditions. Under-participators in the control condition increased the amount they spoke to a significant degree (t(7)=–4.546, p<.005, paired t-test), while the under-participators in the experimental condition did *not* change their level of participation (t(4)= –.507, p=.635,

[2] The error bars found in all of the charts in this document express the standard error of the mean.

paired t-test). This discrepancy in observed changes was found to be statistically significant across the conditions ($t(11)=2.804$, $p<.05$, t-test of independent samples).

The findings indicate partial support of Hypothesis 1: in the presence of the display, over-participators decreased the amount they spoke in the second task to a significant level, and this level was significantly lower than the corresponding change in the control condition. In partial rejection of the hypothesis, under-participators naturally increased the amount that they speak in a second task in the control condition and the introduction of the display for a second task did not induce them to raise their participation at all, even to the level that they would have without the display. This implies that the presence of the display may have hurt the under-participators by preventing them from increasing their level of participation in the second task.

As stated in the hypotheses, it was expected that the display would cause those individuals at the extremes of participation to become more aware of their behavior and thus motivated to alter it to comply with the group norm. It appears that over-participators responded in the expected way, while under-participators did not. One explanation of these apparently opposing results is that it is easier to say less than it is to say more. So the instruction given to the under-participators to speak up was more difficult to perform than the task given to the over-participators. A second possible explanation of these findings is that under-participators may have responded to the information about themselves differently than the over-participators did. This conjecture will be revisited in Sub-Section 5.

3. Evaluation of Hypothesis 2

As mentioned previously, for each decision topic there was one subject who held critical information that assisted the group in making its final decision. During the task discussion, the group would become aware of which subject's information seemed more critical to the decision and would frequently focus the discussion around having this subject share proportionately more information.

The second behavioral hypothesis was that subjects who held non-critical information in the presence of the display would decrease the amount they spoke, while those with critical information would not behave differently than those in the control condition, because of an awareness of the relative importance of their information. To evaluate this, subjects were pooled across groups and divided them into three new categories: those holding critical information in Task 1, those holding critical information in Task 2, and those holding non-critical information in both tasks. These datasets were then examined for systematic changes in participation between tasks and across conditions.

As detailed in Table 4.2, on average, subjects who held critical information spoke more during that task than the preceding or proceeding task. Yet, in a t-test of independent samples, there were no significant differences in the means of the change in participation rates between the control and experimental subjects who held critical information in the first task, subjects who held critical information in the second task, and those who held

non-critical information in both tasks. When the change between Task 1 and Task 2 was examined within condition though, a significant decrease in participation was found in experimental subjects ($t(11)=2.859$, $p<.05$, paired t-test) that was not found in the corresponding control subjects ($t(10)=.408$, $p=.692$, paired t-test). Additionally, when the change in participation of experimental subjects holding information in Task 1 was compared to those holding critical information in Task 2, there was a significant difference between them ($t(22)=-2.729$, $p<.05$, t-test of independent samples). This corresponding change was not found in the control group ($t(20)=-.806$, $p=.430$, t-test of independent samples).

Table 4.2 Participation Rates for Each Information Role and Condition with statistics for significance of change and difference.

Information Held & Condition		N	Task 1	Task 2	Average Change	Standard Error	Paired t-test	Independent t-test
Critical info in Task 1	ctrl	11	25.1	23.9	–1.17	2.87	t(10)=.408 p=.692	t(21)=1.356 p=0.189
	exp	12	27.1	21.2	–5.91	2.07	t(11)=2.859 **p<.05**	
Critical info in Task 2	ctrl	11	26.2	28.4	2.14	2.95	t(10)=–.728 p=.483	t(21)=–.462 p=0.649
	exp	12	25.3	29.4	4.10	3.03	t(11)=–1.354 p=.203	
Non-critical info-holders	ctrl	22	24.3	23.8	–.49	1.57	t(21)=.310 p=.759	t(44)=–.601 p=0.551
	exp	24	23.8	24.7	0.90	1.68	t(23)=–.539 p=.595	

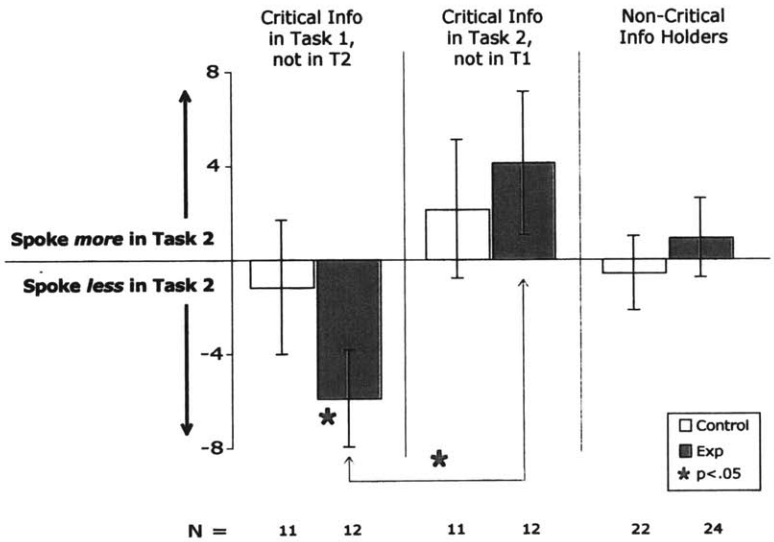


Figure 4.3 Average Change in Participation Rates by information held and condition.

While a significant difference was not found between control and experimental subjects, the subjects in the experimental condition had a significant difference in their changes in

participation depending on when they held critical information. Shown graphically in Figure 4.3, those who held critical information in the previous task significantly decreased their participation (the left-most colored bar) and that the change in behavior of those holding critical information and those not was significantly different (the left-most colored bar vs. the central colored bar). The significant differences are marked with red asterisks in the Figure. While this is weak support for Hypothesis 2, it is a potential indicator that the display encouraged modulation in participation based on the information held by the subject. I speculate that the significant decrease in participation of subjects who held critical information in Task 1 may have been caused by them realizing the information they held was less pivotal than in the previous task. With this realization, and with a visual display of their elevated participation, they may have deliberately attempted to contribute less.

4. Evaluation of Hypothesis 3

Hypothesis 3 predicted that individuals would not perceive an increase in task effort or report interpersonal complications traditionally associated with incorporating technology into a collaborative task. Because these variables were not directly manipulated, the evaluation of this hypothesis is based on the subjects' ratings of task difficulty and group dynamics and the experimental group's evaluation of the display.

First, based on subject responses, there is no indication that the display directly impacted the group process or the cognitive load associated with the task. As stated earlier, there were no differences in the difficulty ratings across conditions. And when the experimental groups were asked about the display directly, on average they claimed they did not look at it and did not find it particularly useful or distracting. They also did not believe it had changed their own or others' behavior. These ratings are summarized in Table 4.3.

To measure the impact of the display on group dynamics, subjects rated the group interaction along several dimensions after each task. In ratings of task performance, group satisfaction, strength of consensus, and task efficiency, there were no differences found between tasks or across conditions. In measures of group interaction, subject ratings of comfort level, perception of honesty of the group, and perception of group listening were also unchanged between tasks and across conditions.

There were three parameters of group interaction that subjects reported as having a significant increase between Task 1 and Task 2. In the control group, ratings on the richness of the group interaction, the perceived helpfulness of the group, and the subjects' level of trust of the group all increased. Table 4.4 contains these average values and the significance of their increases found via paired t-tests.

For these same questions with the experimental subjects, there were corresponding increases in the categories of richness of interaction and perceived helpfulness, but on the question of trust, there was no increase in the level of trust between the first and second

tasks. Across conditions though, there was no significant difference in the changes in trust ($t(90)=.423$, $p=.67$, t-test of independent samples).

In conclusion of this examination of Hypothesis 3, the only indication the display impacted group dynamics was that while all subjects perceived a significant increase in the richness of the interaction and group helpfulness, the subjects working with the display did not report a corresponding increase in group trust. This comparable lack of increase may indicate that the display disrupted the development of group trust, perhaps because it revealed socially sensitive information midway into the group's interaction. Other than this finding, there are no other indications that the display influenced task performance, the group's ability to focus on the task, or the development of healthy group interaction.

Table 4.3 Average Responses to Questions about the Display on a 7-point Likert scale (1=low agreement, 7=high agreement). (N=40 because of 4 missing values.)

Questions about the Display	N	Average Response	Standard Error
I looked at display	40	2.95	.222
I found display informative	40	3.23	.275
I found it useful	40	2.84	.249
I found it distracting	40	2.55	.274
Changed my participation	40	2.98	.289
Change other's participation	40	3.09	.281

Table 4.4 Changes between Post-task Questionnaire Responses on a 7-point Likert scale (1=low, 7=high).

Question & Condition		N	Task 1	Task 2	Paired t-test
Richness of group interaction	ctrl	44	5.52	5.98	$t(43)=-3.346$ **$p<.005$**
	exp	48	5.48	5.88	$t(47)=-2.729$ **$p<.01$**
Helpfulness of group	ctrl	44	6.04	6.32	$t(43)=-2.386$ **$p<.05$**
	exp	48	6.06	6.40	$t(47)=-3.483$ **$p<.001$**
How much you trust group	ctrl	44	6.18	6.41	$t(43)=-2.493$ **$p<.05$**
	exp	48	6.52	6.35	$t(47)=.405$ $p=.132$

5. Perception vs. Reality

To quantify subjects' understanding of their participation during the tasks, subjects in the experimental condition rated the accuracy of the display and asked subjects in the control condition to estimate their own participation rates in relation to the group. The responses to these questions have provided insight into how well individuals perceive their own

participation, and this in turn has assisted us in explaining the observed behavior of the under- and over-participators in the experiment.

Accuracy of the Display

Subjects in the experimental condition were asked to rate how accurately the display reflected the group's conversation. On average, subjects found the display to be more accurate than not (4.63 on a 7-point scale of accuracy), which is a positive finding, since the display was an accurate representation of how much time each person spent talking. There was also a positive correlation between the amount someone spoke and how they rated the display's accuracy (Pearson correlation $r=.301$, $p<.05$). This correlation indicates that those who participated at the highest rates found the display to be more accurate than those who participated at lower rates. To analyze this trend further, I divided the subjects by their participation in Task 2 (the task where they saw the display) into under, middle and over participation categories (based on the mean and standard deviation), and found that the over-participators rated the display as significantly more accurate than the under-participators did ($t(8)=-2.324$, $p<.05$, t-test of independent samples). Table 4.5 contains these results.

This finding indicates that subjects who spoke at a higher rate had a different reaction to the display, perhaps because they were more aware of their own participation level than others. The perception of under-participating subjects that the display was less accurate could have been influenced by an inherent reaction to the technology or, perhaps more likely, their equating of the quality of their contributions with the amount of time they spoke and assessing that they had spoken at a higher rate than displayed.

Table 4.5 Rating of the Display's Accuracy by Participation Level on a 7-point Likert scale (1=low, 7=high). (Total N=37 because of 3 missing values.)

Participation Category	N	Accuracy Rating	Standard Error	Independent t-test
under	5	3.40	.600	t(8)=–2.324
over	5	5.20	.490	**p<0.05**
middle	27	4.41	.359	—

Table 4.6 Self-rated vs. Actual Participation of subjects in control condition, on a 7-point Likert scale.

Control (N=44)	Participation	Average Response	Standard Error	Paired t-test
Task 1	self-rated	4.48	.140	t(43)=3.357
	actual rate	3.89	.179	**p<.005**
Task 2	self-rated	4.57	.136	t(43)=2.684
	actual rate	4.00	.190	**p<.01**

Perception of Participation

To understand this discrepancy between perception of participation and actual participation, in the control condition, there was an additional question not asked during

the experimental condition: "How much did you speak in relation to the other group members?" Responses to this question indicated that there was a skew in subjects' perception of their participation in the upward direction. To correlate this self-rated participation with actual participation, the subjects' actual rate of participation for the two tasks was transformed into seven equally-divided bins that correlated with the bins of the 7-point Likert scale on the questionnaire.

Table 4.6 shows that the average self-rated participation level was above 4.00 (defined on the questionnaire as "equally participating") and that this is significantly higher than the subjects' actual participation rates, as found by a paired t-tests ($t(43)=3.357$, $p<.005$; $t(43)=2.684$, $p<.01$). These results provide further evidence that under-participators are unaware of their lower level of participation, while over-participators are overly aware of their higher participation.

This finding may explain why under-participators did not increase their participation in response to seeing the display. We originally hypothesized that subjects would respond to the display by becoming aware of their own participation and then seeking ways to comply to the group's standard of speaking an average amount. This is how Carver and Scheier (Carver 1998) explain how individuals self-regulate their behavior and this is what we believe we have observed with the over-participators in the study. But Carver and Scheier point out that sometimes heightened focus on personal behavior can lead to an opposite result because of an interaction with one's expectations of ability to succeed. In the case of the display, success was measured by one's ability to change the display to reflect that one was participating at an average level. If a subject did not believe the display was accurately reflecting his/her level of participation, then his/her confidence in being able to change the display would be very low. Furthermore, if the subject did not feel he/she had much to contribute and the display reflected this self-perception then they may have decided to stop trying to contribute more to the discussion. Carver and Scheier found in their own studies that when subjects "expected to do poorly, [...] self-focus led them to avoid items for which norms were available. In this way, they were showing evidence of disengaging themselves from the goal of performing well compared to other people." (p. 182) (Carver 1998). What I believe this means is that the under-participators responded to the display by rejecting it as a standard of behavior and withdrawing from an attempt to comply with the pressure to speak at a higher level of participation. Although this cannot be confirmed post-experiment, this conjecture would be supported if quiet individuals reported negative feelings associated with the display, such as rejecting the labels on the display of "under" and "over" or exhibiting a feeling of being singled out as a lesser contributor.

E. Experiment Conclusions

The results show that over-participators responded to the display by significantly decreasing the amount they spoke, to a degree not observed in the control condition. Under-participators responded by *not* increasing the amount they spoke, in contrast with under-participators who did increase their participation in the control condition. Subjects who held critical information during the first task but not the second significantly

decreased the amount they spoke during the second task, only in the presence of the display. The peripheral display did not disrupt task process or group interaction, although the introduction of the display disrupted the trend of increasing group trust that was found in the control condition. Lastly, subjects over-estimated their level of participation in a conversation and under-participators rated the display as a less accurate reflection of their behavior than over-participators did.

This last finding assists in explaining why under-participators did not behave as expected. When subjects did not find the display to reflect their internal understanding of their behavior, they withdrew from the act of comparing themselves to others based on the display, and this resulted in a reduced effort to comply to this group norm. In a related study done by Losada, et al., where technology was introduced into a face-to-face setting to provide feedback on group interaction, subjects also withdrew from some unexpected activities (Losada 1990). The authors also concluded that a discrepancy between one's mental state and the feedback standard explained these observations.

Taken together, the results indicate two overall trends: (1) subjects who spoke more than those around them were aware of it and were able to use the display in conjunction with the information they held to decreasing their participation in a conversation; and (2) subjects who spoke less than those around them were less aware that they were doing so and did not find the display to be an informative reflection of their behavior. This resulted in a withdrawal from comparison with the display and a corresponding lack of change in participation.

To add a different perspective to these findings, below is an example of how one group responded to the display. This text is taken from the post-task questionnaires (and each subject's participation rate is noted in parentheses).

> ***S1 (27%):*** *"It impacted how much I spoke at the beginning. Knowing that I am generally outspoken is different from having a quantitative measure of my outspokenness. In other words, I was slightly influenced to speak less out of concern that I didn't want to be the highest pick on the chart."*
>
> ***S2 (29%):*** *"We had one participant that spoke less than the rest of us, and I personally (and I think that the others did too) made more of an effort to ask her what she thought since I could see that her bar was so much lower."*
>
> ***S3 (17%):*** *"Almost never saw the display, I forgot it during the discussion. But when I saw it I tried to speak more. I think that is very good."*
>
> ***S4 (27%):*** *"... I think [the task] was overly engaging to the point where the overhead/participating chart was not a factor."*

These comments highlight several of the informal observations of subjects and combine well with the numerical findings. A first observation is that the specific impact of the display on a group is a combination of each subject's internal interpretation of the display and personal judgment as to how it should be used. This can vary greatly between

subjects and between groups. Second, those that perceived themselves to be over-participators seemed to be most aware of the display and its message. Third, while under-participators may observe that they are under-participating and may "try" to improve, they express less control over their level of participation. Lastly, many subjects stated that they chose to focus on the task discussion and not on the display, reflecting its peripheral role in the discussion.

In summary, in the behavioral study, over-participators responded to the display by restricting their comments, while under-participators did not increase their participation levels as much as when the display was absent. Critical information holders were not adversely impacted by the display, while subjects holding non-critical information for that task significantly decreased the amount they spoke. The findings indicate that introducing the display midway through the experiment hindered trust development between group members. The discovery that subjects had inflated self-ratings of their participation helps illuminate why under-participators did not respond as expected to the display.

In the next experiment, these negative findings of under-participator behavior and trust development are addressed in the design of the experiment and in the use of a "replay display," which explores whether or not a more reflective use of the display can assist quiet individuals in accepting the display's content and using it to change their participation in future discussions. Beyond the scope of this thesis, there are several other directions that could address this issue. One approach would be to introduce some "pre-work" to the group that would establish the expertise and value brought by each person and incorporate casual socializing that could ease some of the social tension of an initial interaction. Another approach is to consider private displays that eliminate the social-consciousness of the displays and focus on how a display can be a private reminder to alter behavior.

V. Changing Group Process with a Visual Replay

The second quantitative experiment of this thesis was an examination of how group process, specifically group information sharing, was influenced by these information displays. After the previous experiment, which demonstrated that a peripheral, real-time display changes individual behavior in a group setting, there were several remaining questions. First, could information sharing, a group-level process integral to decision-making, be influenced by a feedback display? Second, could a more informative display that reveals patterns in group dynamics such as turn-taking and interruption be informative to a group in their decision-making? Third, because the previous experiment revealed that trust formation could be hindered by a display and that the quietest members did not speak more when a display was present, there was a desire to try different means of introducing the display and incorporating it into a group setting to address these concerns.

A. Study Design

The experiment had a 2×2 factorial design with two independent variables. The first variable was the presence of the Group Circle display, as described in Chapter 3, Section IV, which provides real-time participation feedback. In the experiment, it was shown on a tabletop monitor located on a table in the center of the group of four subjects. The second variable was the use of the replay feature of Second Messenger 2.0, to watch the Timeline display of the turn-taking patterns (also described in Chapter 3, Section IV). Groups using the replay tool would watch it on a monitor off to the side of the group in between the two decision-tasks. Figure 4.3 shows screenshots of the tools' displays and pictures of subjects in the conditions with the tools present.

As in the previous experiment, groups of four subjects performed two information-sharing tasks. To test the different hypotheses about information sharing, trust, and participation, there were several changes made to the design of this experiment. First, before the experiment began, each group was given a full explanation of the tool (or tools) they would use. It was theorized that trust development between group members would not be hindered by the tool if groups knew, from their very first interaction with each other, that feedback tools would be present.

Second, also prior to the experiment, groups were given a brief statement explaining that "It may be the case that some people have information that others do not have. Research has shown that, when this is the case, teams perform better when each team member has an opportunity to contribute to the discussion." So each group was told the importance of information sharing, but was not told that there would necessarily be information to share. This made the task of information sharing less obvious to the groups than it had been in the previous experiment. Additionally, all groups, including those in the control condition who did not see any displays, were provided black-and-white images of the interfaces where one person was excluded from the conversation (Figure 4.4). In this manner, every subject was introduced to the idea of imbalance and exclusion from the

conversation. Both of these procedural changes were designed to prevent our experimental results from demonstrating that just the fact of telling a group to avoid ignoring individuals and to try to all share information would cause groups to behave differently.

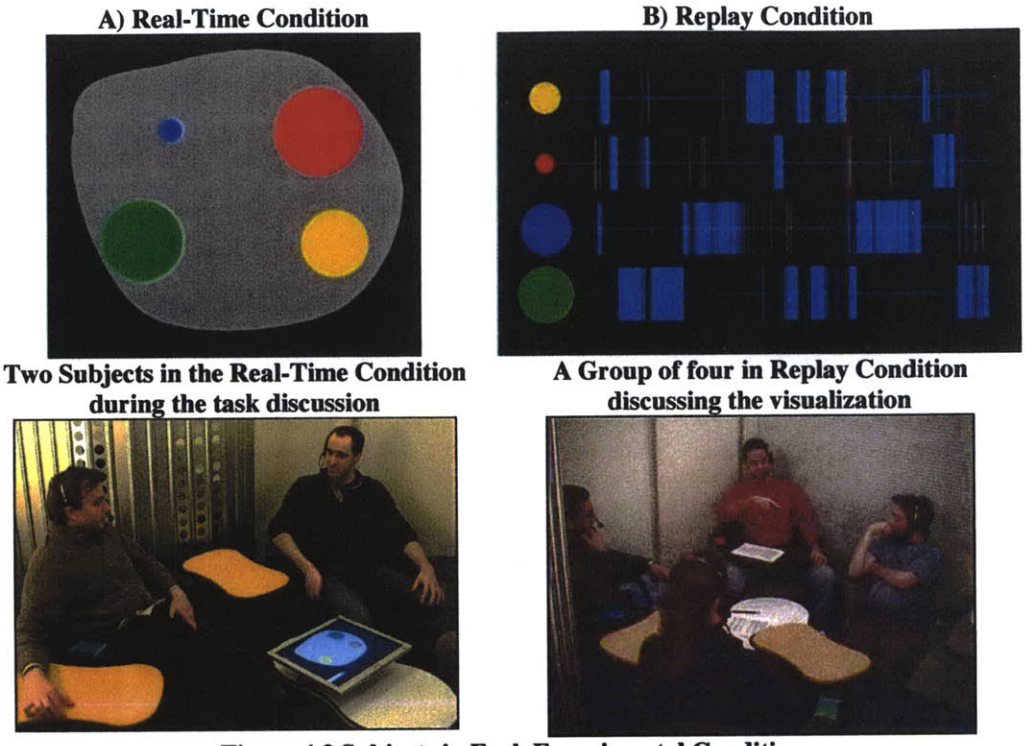


Figure 4.3 Subjects in Each Experimental Condition with screen captures of each display

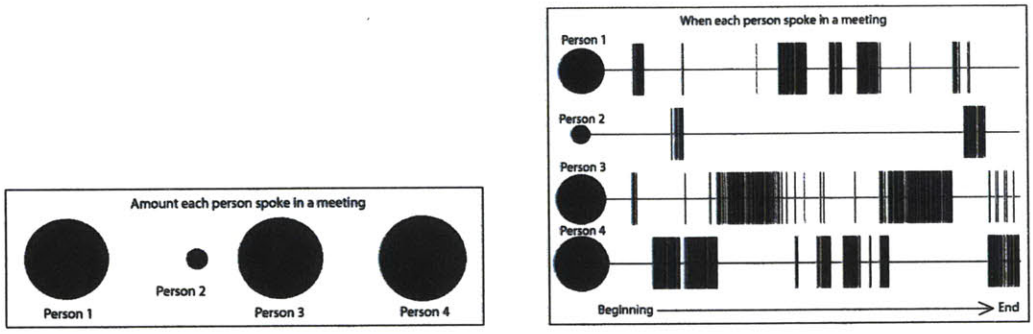


Figure 4.4 Images of the Displays Provided to All Experiment Groups

Another change from the previous experiment was that groups in the real-time display condition had the display visible during both decision tasks, so unlike in the previous experiment there was no baseline measurement of the group's participation before the display was present, nor was there a period of working together without this information present.

Between the two decision tasks, all groups were given up to two minutes to discuss how they thought they had done on the first task and how they would like to approach the

second task. This strategy session was incorporated into the experiment at the midpoint because it has been shown that a midpoint reflection on process can be beneficial to groups (Gersick 1990). According to Gersick's theory, this strategy discussion should result in an increase in group performance and if any dramatic changes in group process were to occur, this midpoint would be moment of change.

While all groups were given this midpoint reflection, only the groups with the replay display were shown the replay before this discussion. When shown the replay, they first saw a visualization of the group's entire interaction and then were shown a replay of the interaction with a two-minute sliding temporal window. (In replay mode, the display shows the interaction at ten times the original conversation, so this never took more than a minute.) These groups were told they could use this replay as part of their strategy discussion.

These experimental procedures, as they were explained to the subjects before the experiment, are found Appendix B.

B. Study Hypotheses

As in the first experiment, it was expected that by setting a norm of equal participation and by instructing groups to listen to each others' opinions, that all groups would be congenial and attempt to listen to everyone (Bettenhausen 1985). Therefore it was expected that, with the privately held information sufficiently buried in the task descriptions, only those groups that actively sought the information would succeed in sharing information. It was anticipated that both of the tools would assist in this information-seeking task. It was not known if the combination of the two tools would be distracting or if the centrally placed real-time display would be distracting.

In terms of the findings from the previous experiment and the changes made in the design to this experiment, it was anticipated that the quietest group members would speak more due to the full disclosure of the use and purpose of tools (unlike in the previous experiment) and that the most verbal members would speak less (as in the previous experiment, they would comply to normative pressures).

Furthermore, we anticipated observing systematic changes in the turn-taking styles of the groups, such as longer turns and a smaller amount of variation in the number of turns per person.

The experiment hypotheses can be summarized as:

> ***H1:*** *With the feedback tools, information sharing will improve.*
>
> ***H2:*** *With the feedback tools, under-participators will increase their rate of participation and over-participators will decrease their rate of participation.*

H3: With the feedback tools, groups will exhibit different turn-taking styles, including longer turns and lower levels of dispersion in number of turns and in speaking time.

In addition to these hypotheses, the groups were asked about how well they thought they did on the task, how much they thought they had each contributed, and for their reactions to the displays.

C. Method

1. Experimental Procedure

As in the previous experiment, this study was designed so that four subjects, previously unknown to each other, were given two decision tasks on which to come to consensus. Each decision began with the subjects reading information sheets about the task. They were then told to discuss the task with the group to make a decision. If they were in the real-time feedback conditions, they would see the display's visualization of participation levels as they spoke. Those without the real-time display were not provided any information on their participation levels. All subjects wore microphones and had their individual speaking times recorded. The task topics of student admissions and store locations were counter-balanced across condition.

After the first task, groups were given a few minutes to discuss how they think they did on the task and were told the next task was "a different topic, but similar in structure." Depending on a group's style of interaction, this interaction varied a great deal: some groups launched into full discussions of how to approach the next task, others discussed more inter-personal issues of listening and feeling like they had the opportunity to speak, and other groups decided they did not have anything to discuss during this time.

After both tasks, groups filled out questionnaires about their perceptions of the task, their own behavior, and the behavior of each group member. If they had used a tool, they answered additional questions about their perceptions of the display's information.

Unlike in the previous experiment, there was no critical information holder in the group of four because of the new structure of privately-held information. In each group, each person was given 36 facts (twelve facts on each of the three decision options), of which only three were private to them. Over the group of four, if all privately held facts were shared (twelve in total), the task was designed to reveal the best choice. Because of the extreme imbalance between public and private information, overall, the groups did a very poor job of sharing this information, and this is in line with previous studies of information sharing (Wittenbaum 2004).

2. Participants

A total of 123 subjects were recruited from the MIT community (via mailing lists and posters) and then randomly assigned to 27 four-person groups and 5 three-person groups.

The average subject age was 27, with about three-fourths of the subjects being students and one-forth being members of the larger community. Gender was split approximately in half (55% women, 45% men) and gender was not controlled within groups, resulting in 27 of the 32 groups comprised of both genders.

Two categories of groups were eliminated from the group-level analysis of the experiment. Although the protocol required four subjects for each group, five groups had subjects not show up for the experiment appointment, resulting in three-person groups. It was found that there was a significant effect of group size on the percentage of information shared in a group's second task discussion ($t(26)=2.229$, $p<0.05$). Therefore these groups of three were removed from the analysis of information sharing and change in participation. The second category of problematic groups were those groups that agreed on what the group's final decision should be prior to any discussion. When a group agreed prior to discussion it meant that the group did not need to discuss any of the related information in order to come to consensus, and therefore no privately held information had an opportunity to be shared. There were five additional groups that fell into this category. For the group-level analysis, this leaves 22 groups for analysis. The 10 removed groups were kept in for the analysis of individual-level behavior. Table 4.7 presents the number of groups within each condition in the experiment:

Table 4.7 Number of Subject Groups in Each Condition

	No Real-time Display	**Real-time Display**	**Totals**
No Replay Display	6	7	13 with no replay display
Replay Display	5	4	9 with replay display
Totals	11 with no real-time display	11 with real-time display	22 groups total

D. Results

In the experiment, during each of the task discussions, the amount of information shared during the task was logged and the amount of time each person spoke was collected. After each task, groups filled out questionnaires about the task, their own performance, and their perceptions of each other person in the group. With this data, the results focus on addressing the three behavioral hypotheses and examining how people perceived the task, trusted each other, and perceived their own contributions. The following sections first provide an overview of the data collected, then address each of the hypotheses (regarding information sharing, participation levels, and turn-taking patterns), and ends with addressing the issues of how the technology influenced perception of task, trust, and self.

1. *Overview of Data*

This section covers the general statistics of the groups: how much time was spent on the tasks, task accuracy, task difficulty, and other general trends.

Time Spent on Tasks

From the previous experiment, it was known that the tasks took an average of thirteen minutes of discussion, and, because of the limitation of an hour total on the experiment, it was decided to provide groups with fifteen minutes for each task. If they had not decided after fifteen minutes of discussion, they were given an extra minute and then the discussion was stopped, regardless of whether or not they had come to consensus. The average discussion time was twelve minutes for both the first task and the second task.

As an aside, if a group had not made a decision after fifteen minutes, they were typically beyond the point of information sharing and had moved to a phase involving emotional debate and thus no hidden task information was shared in the later minutes.

Task Accuracy

In terms of information sharing, groups did not perform very well on the tasks. The average number of hidden facts shared on the first task was 40% (approximately 5 of the 12 facts) and on the second task was 52% (approximately 6 of the 12 facts). This increase in information sharing between tasks was approaching the level of significance (paired t-test, $t(21)=-1.911$, $p=0.070$). Only one group shared all of the hidden information (in their second discussion) and this group also gave the correct answer on this decision.

As for the group arriving at the correct decision after discussion, 8 of the 22 groups got the right answer on the first task and 3 groups got it right on the second task. But, there was no difference in the amount of information shared between groups that got the right answer versus groups that got the wrong answer (Task 1: $t(20)=1.520$, $p=0.144$; Task 2: $t(20)=0.553$, $p=0.587$). The likelihood of arriving at the correct choice was dependent on one of the group members initially choosing the correct answer based on his/her information and then arguing for the choice during the discussion, rather than on the aggregation of hidden information revealing it as the best option. Therefore the decision outcome was not an indicator of successful information sharing and will not be used as a criterion for evaluating the groups' decision processes.

Task Topic and Order

After each task, subjects rated the task's difficulty. Unlike in the previous experiment, it was found that one of the task topics was more difficult than the other. On a 7-point Likert scale, the store task was rated as more difficult (Groups with store first: Task 1 difficulty=5.00 (SD=1.34), Task 2 difficulty=3.30 (SD=1.72), paired t-test, $t(62)=6.834$, $p<0.001$; Groups with school first: Task 1 difficulty=4.04 (SD=1.44), Task 2 difficulty=4.96 (SD=1.71), paired t-test, $t(51)=-3.003$, $p<0.005$). Furthermore, the first

task was rated as more difficult than the second task (Task 1 difficulty = 4.57, Task 2 difficulty = 4.05, paired t-test, $t(114)=2.242$, $p<0.05$).

As will be discussed in later sections, subjects found the first task to be more difficult when they anticipated seeing a replay of their turn-taking. Other than this finding, there were no significant effects of the experimental conditions on task difficulty.

Culture and Gender Effects on Participation

In each group of four, the average participation rate was 25%. There was a positive correlation in how much someone spoke on each task ($r=0.507$, $p<0.001$). This corresponds to the previous experiment's findings.

There were no significant differences in participation rates by gender, in line with the previous experiment's findings. Task 1 average participation rate for men was 24.5% (SD 8.9), for women 25.4% (SD 9.2). Task 2 average participation rate for men was 24.0% (SD 9.9), for women 25.7% (SD 10.1).

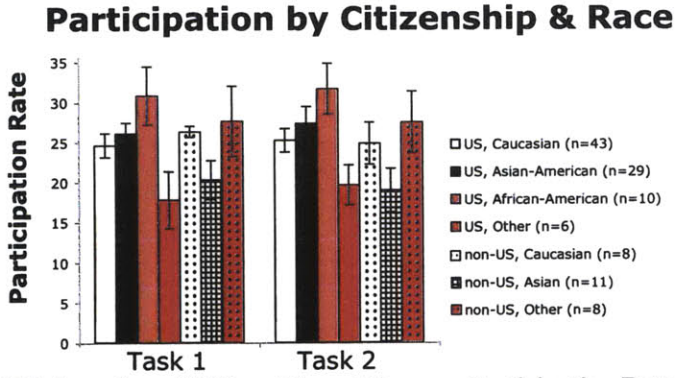


Figure 4.5 Interaction of Citizenship and Race on Participation Rate

Culture, something not asked about in the previous experiment, did influence how much someone spoke. The ANOVA showed that citizenship and race had a significant interaction effect on the participation rates in the second task ($F(2,74)=3.158$, $p<0.05$) and an almost significant effect in the first task ($F(2,74)=2.783$, $p=0.068$). Post-hoc comparisons using Tukey's test showed that "US-citizen African-Americans" spoke significantly more than "US-citizen Other" in Task 1 ($p<0.05$) and significantly more than "non-US-citizen Asians" in both Task 1 and 2 ($p<0.05$ for both). These effects can be seen in Figure 4.5.

Participation and Self-Perceptions

A frequently received comment regarding the previous experiment is that participation rates do not reflect the contribution a person has made to a discussion. While it is certainly true that participation rates do not capture what was said or how the group received it, in this experiment there were several significant correlations between individuals' perceptions of what they contributed and how much they spoke during the task. The correlations show that the more someone spoke in either task, the more satisfied

they were with both the *amount* they spoke and the *quality* of what they had said. In the second task, the more someone spoke, the more they felt they had contributed to the *decision.* While participation levels do not capture the content of what someone has said, there is a relationship between how much one speaks and one's perception of significance of their contributions to the meeting.

In addition to those correlations, it was also found that in the first task, the older the subject, the more they spoke. And in terms of the real-time display, those that spoke the least were more distracted by it in the second task (supporting the previous experiment's conjecture that quiet subjects are more disturbed by real-time feedback). Participation rates did not correlate with any questionnaire responses relating to the task or any additional questions about the displays. The significant correlations are listed in Table 4.6.

Table 4.6 Correlations with Participation Rate
The Pearson correlation coefficient is listed for cases where the correlation was significant ($p<0.05$).

Survey Question:	Task 1	Task 2
Satisfaction with amount I spoke	0.390	0.400
Satisfaction with quality I contributed	0.370	0.405
Wish I spoke more	–0.439	–0.343
How much I spoke in relation to others	0.691	0.686
Age	0.297	No correlation
My contribution to the decision	No correlation	0.215
How distracting I found the real-time display	No correlation	–0.362

2. *Evaluation of Hypothesis 1: Information Sharing*

The first hypothesis was that the feedback tools would influence the amount of information sharing. To measure how much information was shared, during each experimental task, a coder (myself) marked on a checklist each hidden fact as it was brought up in the discussion. For a fact to be counted as "shared," it had to be mentioned and had be acknowledged in some way, either by another person stating that was a new fact to them or by the fact being incorporating into the discussion.

Figure 4.6 shows the average percentage of information shared by condition. "A" indicates the real-time display was present; "B" indicates the replay display was present; "CTRL" was the condition with no tools present.

On average, across all groups, information sharing went up during the second task, but this was not a significant increase (paired t-test, $t(21)=-1.911$, $p=0.070$). As mentioned earlier, on average, groups did a poor job of sharing the hidden information, but more importantly, there was a wide variation amongst groups in their inclination to share information. In the first task, some groups shared as many as ten of the twelve hidden facts, while others shared none; in the second tasks, the best group shared all twelve facts and the worst group shared only three. With this wide variation in group performance, the feedback tools themselves did not have a significant influence on the amount of

information sharing (in a two-way ANOVA, all p-values > 0.3). As Figure 4.6 illustrates, information sharing increased in Task 2 for all groups except those in the replay condition (B). This increase was only significant for the real-time condition (A).

The one potential trend though is that groups with the real-time display in Task 1 shared less than those without, as illustrated in Figure 4. Related to this, in a pair-wise comparison of information sharing by condition between tasks, the groups in Condition A demonstrated a significant increase (paired t-test, $t(6)=-3.79$, $p<0.01$). This change is marked with a red asterisk in the figure. Therefore, if the real-time display did hinder the amount of information shared in the first task, these groups rebounded from that hit in performance during the second task.

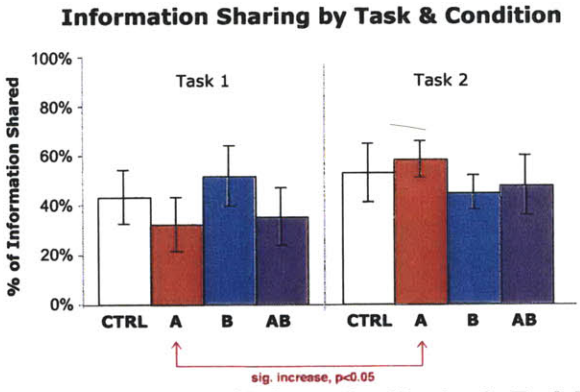


Figure 4.6 Comparison of Information Sharing in Each Task
CTRL = no displays; A = Real-time display shown; B = Replay display shown

Table 4.7 Information Shared in Each Condition and Task

Condition	N	% Shared in Task 1	Std Error	% Shared in Task 2	Std Error	Paired t-test
No displays (CTRL)	6	43.1%	10.8	52.8%	11.7	t(5)=–1.929 p=0.112
Real-time (A)	7	32.1%	11.0	58.3%	7.3	t(6)=–3.79 **p<0.01**
Replay (B)	5	51.7%	12.2	45.0%	6.8	t(4)=0.416 p=.699
Both displays (AB)	4	35.4%	11.5	47.9%	12.0	t(3)=–0.550 p=0.621

Based on the wide distribution in information sharing amongst the groups, as shown in the histogram in Figure 4.7, and the observation that some of the best information-sharing groups shared less on the second task, I divided the groups into three categories: "low," "medium," or "high" in terms of how much information they shared on the first task, in order to examine whether or not the displays had a different effect on these different types of groups. The divisions for high, medium, and low are highlighted in the histogram. There were 9 low, 8 medium, and 5 high performing groups.

ANOVA analysis comparing the real-time feedback, the replay feedback, and the three levels of performance in Task 1 reveals a significant interaction between the replay

feedback and the Task 1 performance on how well the group shared information in Task 2 ($F(2,11)=6.643$, $p<0.05$). This interaction effect is illustrated in Figure 4.8, which plots the high and low performers, divided by whether or not they saw the replay display. The data associated with this ANOVA is in Table 4.8.

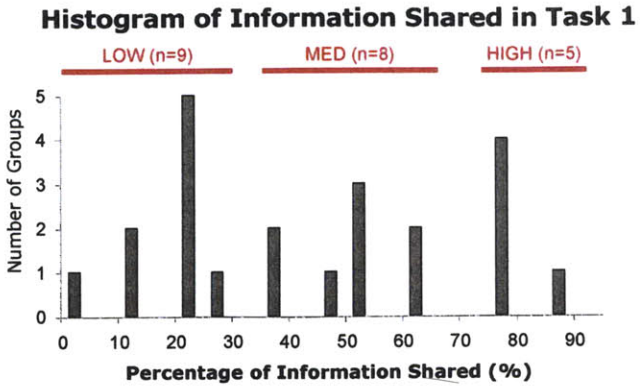


Figure 4.7 Distribution of Information Sharing in Task 1

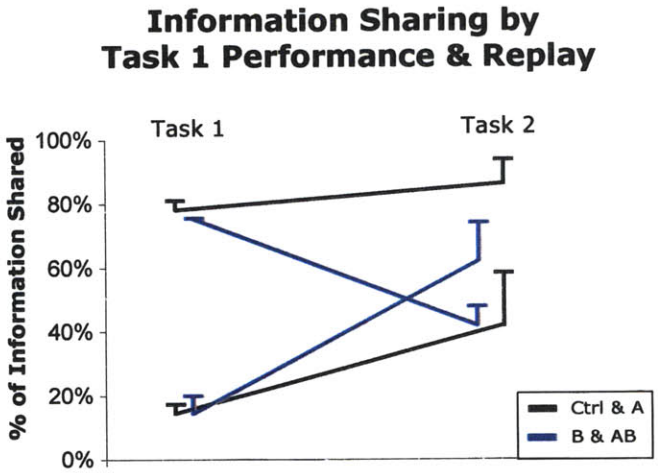


Figure 4.8 Interaction of Task 1 Sharing and the Replay Display
CTRL = no displays; A = Real-time display; B = Replay display

Table 4.8 Information Shared in Each Task by High/Med/Low Performers

Condition	Task 1 Performance	N	% Shared in Task 1	Standard Error	% Shared in Task 2	Standard Error	Paired t-test
No Replay (CTRL&A)	low	6	14.2%	0.03	41.7%	0.06	t(5)=–3.91 **p<0.05**
	medium	4	41.5%	0.06	54.3%	0.10	t(3)=–1.57 p=0.22
	high	3	77.7%	0.03	86.0%	0.07	t(2)=–1.00 p=0.42
Replay Display (B&AB)	low	3	13.7%	0.06	61.0%	0.12	t(2)=–2.82 p=0.11
	medium	4	52.0%	0.02	37.3%	0.04	t(3)=2.80 p=0.07
	high	2	75.0%	0.00	41.5%	0.17	t(1)=2.03 p=0.29

Table 4. 9 Post-hoc Tukey Results
Comparing Information Shared in Task 2 for Each Condition

Condition		No Replay			Replay Display		
		Low	Med	High	Low	Med	High
No Replay	Low	—					
	Med	**<0.05**	—				
	High	0.9984	**<0.05**	—			
Replay Display	Low	0.8274	0.1521	0.6873	—		
	Med	0.5427	0.4323	0.4224	0.9919	—	
	High	1.0000	**0.0711**	0.9996	0.9411	0.7665	—

These results illustrate that without the replay display, groups had an increase in information sharing in Task 2 (although this was only a significant increase in the low performing groups). With the replay display, those that did the best on Task 1, decreased their information sharing, so that they are now indistinguishable in performance from the low-performing control groups. Those that did poorly on Task 1, do better than their control condition counterparts, to the point that they are indistinguishable from the high-performing control groups (they are statistically not different than any of the groups).

In summary, the groups that had access to the replay display demonstrated a change in information-sharing strategy, as compared to the groups without access to the display. For those that had demonstrated a poor information sharing strategy, this change was beneficial to them, pulling them up to the level of the best information sharers who did not have the display. For those who had done very well at information sharing in the first task, this change in strategy was harmful to them, resulting in them performing significantly worse than the equivalent groups who had not seen the display, putting them at the same level as the poorest information sharers. It is often the case that when a group (or individual) performs at an extreme level, high or low, on a subsequent task that group will regress towards the mean. This was not shown in the control condition; groups that performed well in the control condition continued to perform equally well on the next task. It was the introduction of the replay display that caused these groups to change their performance in the second task.

An important aspect of these findings is in the acknowledgement that none of the groups were given feedback on how well they had shared information. Presumably if groups had been told that they had performed well or poorly in terms of information sharing that would have influenced how they used the display as feedback about their turn-taking patterns. This also highlights that there are downsides to providing groups feedback: they may use it to incorrectly adjust their behavior.

In Chapter 5, we will present the stories of some of these groups, discussing in more detail what was observed in the groups that dramatically changed their information sharing strategy, either by increasing or decreasing the amount they shared.

3. *Evaluation of Hypothesis 2: Participation Levels*

The second hypothesis of the experiment was that those people who spoke the most and the least would bring their participation in towards the middle when they had access to the feedback displays.

To perform this analysis, the same categorization of over, middle, and under participation was done to these groups as was done in the previous experiment's analysis. By calculating the standard deviation of the participation rates across all twenty-two groups (9.2%), pooling the subjects together, and labeling those subjects who spoke mean plus the standard deviation as over-participators (spoke > 34.2%) and those who spoke mean less the standard deviation as under-participators (spoke < 15.8%), the remainder were left as middle-participators.

Examining just the over-participators and under-participators, the change in participation between tasks was calculated for each subject (Task 2 participation – Task 1 participation). Figure 4.9 shows a graph of the change in participation rates for the subjects within each condition and Table 4.10 provides the participation rates per category.

ANOVA analysis found no significant main or interaction effects of the conditions on participation for either the over-participators ($F(3,12)=1.22$, $p=0.346$) or the under-participators ($F(3,12)=0.22$, $p=0.884$). In a pair-wise comparison of Task 1 and Task 2 participation rates in different conditions, the over-participator groups who saw the replay (B), had an almost significant decrease (paired t-test, $t(4)=2.189$, $p=0.09$) and the under-participator groups that saw the replay (B) had a significant increase in the amount they spoke (paired t-test, $t(5)=-4.96$, $p<0.01$). Examining the groups that saw the replay and had the real-time display (B & AB), the under-participators exhibited a significant increase in participation (paired t-test, $t(9)=-3.767$, $p<0.01$). Although the control condition group appears to exhibit changes (decreasing in the over-participators and increasing in the under-participators), these changes are not significant. Figure 4.9 illustrates these significant changes, highlighting the significant changes, and Table 4.10 provides the average values for each condition along with the t-statistics.

The data did not reveal a change in participation due to the real-time display, but this is to be expected because the real-time display was present during both tasks (unlike in the previous experiment), thus the over and under participators in these conditions were able to monitor (and change) their behavior during the first task, as well as the second.

The findings to be described below are not as strong as the previous experiment's because the statistics only show a change *within* condition, not between the conditions. Another caveat of these findings is that with the extremely low number of samples, the variance within each group is very high. It must be acknowledged that with a higher number of samples, these results could change. Yet, the results, as they are, reveal interesting effects of the replay display:

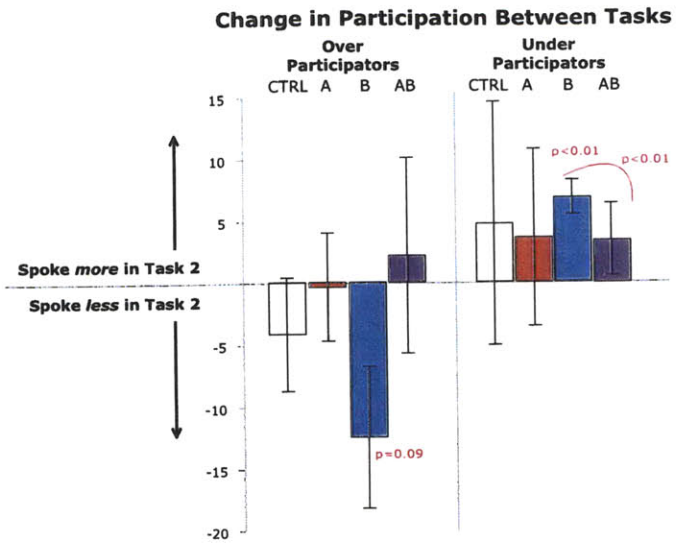


Figure 4.9 Over- and Under-Participators' Change in Participation

Table 4.10 Over- and Under-Participation Rates and Significant Changes Between Tasks

Participation Type	Condition	N	Part in Task 1	Std Error	Part in Task 2	Std Error	Paired t-test	Paired t-test (combining cond)
Over-Participators	No Displays (CTRL)	4	36.61%	0.77	32.42%	5.13	t(3)=0.918 p=0.427	t(7)=0.757 p=0.474
	Real-time (A)	4	37.05%	1.34	36.67%	4.84	t(3)=0.086 p=0.937	
	Replay (B)	5	42.06%	3.15	29.55%	4.22	t(4)=2.189 **p=0.094**	t(7)=1.387 p=0.208
	Both Displays (AB)	3	39.95%	3.23	42.11%	4.83	t(2)=–0.274 p=0.810	
Under-Participators	No Displays (CTRL)	3	13.99%	0.83	18.71%	10.48	t(2)=–0.481 p=0.678	t(5)=–0.764 p=0.479
	Real-time (A)	3	11.92%	1.05	15.51%	6.13	t(2)=–0.501 p=0.666	
	Replay (B)	6	12.20%	0.62	19.03%	1.36	t(5)=–4.96 **p<0.01**	t(9)=–3.767 **p<0.01**
	Both Displays (AB)	4	11.32%	1.26	14.72%	3.83	t(3)=–1.156 p=0.331	

The under-participators who watched a replay of their turn-taking interaction spoke significantly more during the next task, and those that did not see this replay did not have a corresponding increase (the increases in the control group and the real-time groups were not significant, as shown in Table 4.10). In the previous experiment, the under-participators did not increase their contributions when the display was present, and it was theorized that this was a combination of not believing the information on the display and feeling less able to contribute after getting negative feedback. I propose that the *replay* feedback in this experiment was more contemplative and allowed the under-participators to reflect and plan how to speak more during the upcoming task, making the task of speaking up a more achievable act than when behavior feedback was provided real-time.

The over-participators who watched a replay of the turn-taking had a decrease in the amount that they spoke, but it was not to a significant level ($p=0.07$). In the previous experiment, they decreased the amount they spoke to a very significant level ($p<0.01$). Thus it appears that a replay interaction produces a less dramatic change in the over-participators, although this may be found because of the low sample sizes. Furthermore, those over-participators who saw a replay *and* had the real-time display, actually had an average increase in the amount that they spoke, so their response to the display was not to speak less or change their previous behavior. A possible explanation of this is that because the groups could discuss and contemplate their turn-taking, the over-participators may have felt that the group was accepting of their high level of talking and the display imposed less pressure on them to conform that it would have without the discussion.

One explanation for these results is "regression towards the mean." This can be seen in the under- and over-participators in the control condition of both this experiment and the previous one. While there is a natural tendency for extremes to move towards the mean on a second task, this was the desired effect of the displays in this experiment. Therefore, the challenge in finding systematic, significant changes in the individuals who had the displays present is in discovering a *stronger* effect than regression toward the mean. Thus, the findings presented here should be considered preliminary because the low sample sizes have a high variance and the rule-of-thumb that extremes move towards the mean with or without feedback weakens their claims.

Minimum Participation Rate

To further examine the issue of under-participators though, there is another way to approach the question of whether or not the displays encouraged quieter group members to speak more. For each group discussion, there is one person who participated at the minimum and this was not necessarily the same person in each task. In fact, in nine of the twenty-two groups, the quietest person was different between the two tasks. If the displays encouraged more equal participation, this minimum value should be higher in the groups with the displays.

The average values for the minimum participation in each condition were not significantly different from each other within each task, but the *change* that was exhibited by the groups between tasks was influenced by condition. An ANOVA analysis found a main effect of the replay display on the change in minimum participation between tasks, calculated as Task 2 minus Task 1 ($F(1,14)=9.67$, $p<0.01$). Additionally, there was a significant interaction between the replay and real-time displays on the change in minimum participation ($F(1,14)=5.22$, $p<0.05$). Figure 4.10 and Table 4.12 show this interaction: those groups with the replay (B & AB) had an increase in the minimum participation level and those without the replay (CTRL & A) had a decrease in the minimum participation level. It appears that the real-time display (A & AB) tempered the increase in the AB Group and tempered the decrease in the A Group. A post-hoc Tukey test indicates that there are no significant differences in the minimum participation rates within each task, but the change in the minimum rate for the control group (a decrease) was significantly different than the change in the replay condition (an increase) with $p<0.05$ significance.

This finding indicates that not only did the control groups have quieter members in the second task (down to 10.8% from 15.9%), but those groups that had the replay display had less quiet members on the second task (the average quietest person spoke 17%, up from 14.3%). This supports the previous findings about under-participators in the replay display conditions exhibited the greatest increases in participation.

Table 4.12 Minimum Participation Rate in Each Condition

Condition	N	Task 1 Participation	Standard Error	Task 2 Participation	Standard Error
No Displays (CTRL)	6	15.9%	1.0	10.8%	1.5
Real-time Display (A)	7	15.6%	1.5	14.0%	2.1
Replay Display (B)	5	14.3%	2.6	17.0%	1.2
Both Real-time and Replay Displays (AB)	4	13.1%	2.6	13.9%	3.4

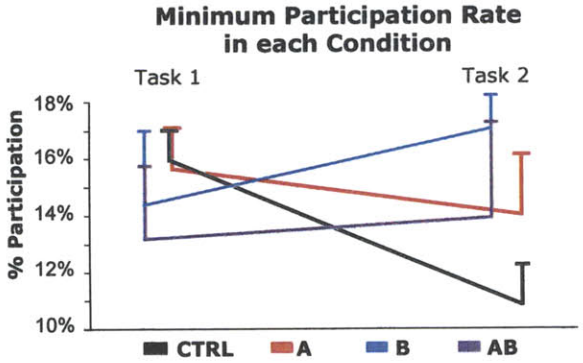


Figure 4.10 Change in Minimum Participation Rate between Tasks

Table 4.11 Turn-Taking Variables Calculated for Each Group

Variable	Description
dispersion in participation	distance calculation[3] between group members' participation rates
dispersion in participation during the first five minutes	distance calculation between group members' participation rates measured after five minutes of discussion
number of turns	total number of turns taken by all group members
turns/min	total number of turns in the discussion divided by the length of the discussion
dispersion in number of turns	distance calculation between the group members' individual number of turns
dispersion in the % of turns	similar to the above calculation, but measures the distance between each individual's number of turns, taken as a percentage of the total number of turns in the discussion.
average turn length	average of the average turn length of each group member
dispersion in turn length	distance calculate between the individuals' average turn lengths
percentage back-talking	the percentage of spoken utterances that were made during other people's turns

[3] "Dispersion" was calculated for a group of four members as |A–B|+|A–C|+|A–D|+|B–C|+|B–D|+|C–D|. This value expresses the spread between the four values.

4. *Evaluation of Hypothesis 3: Turn-Taking Patterns*

Beyond the study of the over- and under-participators, the speaking data was analyzed along several group-level variables to understand how the displays may have influenced the group turn-taking patterns. To calculate many of these variables, the data was examined in terms of the turns taken by each person, rather than using just the participation rate. The calculation of turns was based on previous research on turn-taking in face-to-face conversations (Choudhury 2003). Each second of the conversation was assigned as being during the turn of the person who had spoken the most in the previous three-second window. From this second-by-second assignment, the number of turns each person had, the length of these turns, and who spoke during someone else's turns were computed. Table 4.11 lists the group-level variables that were analyzed in an ANOVA analysis.

The ANOVA analysis revealed that the displays influenced the groups' dispersion in participation rates during the first five minutes of discussion and the average turn length over the discussion. The remaining variables were either unaffected by the experimental conditions or the interaction effects of the variables did not produce significant differences in group behavior.

Dispersion in Participation in First Five Minutes

Dispersion in the participation rates is a measure of the spread across the group's participation levels. The higher the value, the greater the difference between the individual participation rates; a value of zero indicates exact equal participation within a group.

There were no significant effects on the dispersion in participation rates over the entire task, but there were when looking at just the first five minutes. The real-time display becomes less dynamic over time because it aggregated more and more data into the display as the meeting progressed, therefore I hypothesized that the first few minutes of discussion may result in different behavior when comparing groups that had real-time feedback versus not. Furthermore, the beginning few minutes may be the time when the groups are most committed to and demonstrating a certain turn-taking process.

The ANOVA shows that the replay display has a main effect on the change in participation rate dispersion during the first five minutes of discussion ($F(1,14)=8.12$, $p<0.05$). As shown in Figure 4.11, there is less change in dispersion between tasks for groups with the replay in the first five minutes, as compared to those without the replay. Those without the replay increased the amount of participation dispersion in their first five minutes of discussion. This finding was unexpected, since it was theorized that the real-time feedback would influence the dispersion more so than the replay.

The majority of information sharing generally occurred during the first five minutes of group discussion. (The tasks ranged in length from 5 minutes 30 seconds to 17 minutes, so each group talked for more than this five-minute window of analysis.) Groups that changed strategies between tasks would often decide to read the information sheets to

each other at the beginning of the task. Depending on how long this took, this could result in large amounts of dispersion in participation, as one person would spend several minutes reading aloud.

Table 4.13 Dispersion in Participation Rates for Each Task Comparing Groups With and Without the Replay display

Condition	N	Dispersion in First 5min of Task 1	Standard Error	Dispersion in First 5min of Task 2	Standard Error
No Replay Display (CTRL & A)	13	69.9	11.6	129.7	12.2
Replay Display (B & AB)	9	113.1	14.8	99.1	13.0

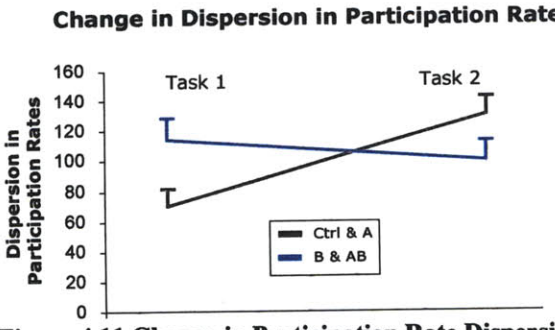


Figure 4.11 Change in Participation Rate Dispersion during the First Five Minutes of Discussion

Average Turn Length

Average turn length measures the average number of seconds each person in the group spoke, on average, during their turns. A high average turn length indicates that each person spoke for longer periods of time, while a shorter value reflects a faster turn-taking style. It could be expected that as groups get to know each other better, they will manage turn switching more efficiently and have shorter turns as a result. Or as they become more animated and excited about a discussion, turn length will decrease.

There was a main effect of the real-time display on the *change* in the average turn length ($F(1,14)=5.916$, $p=0.029$). As is seen in the figure, when there was real-time feedback during both tasks, there was less change in the average length of turns between turns. When there was no feedback, the length of turns became shorter. This indicates that when groups have a tool for monitoring their interaction real-time, they are more consistent in their behaviors across tasks; those without feedback, have a change in average turn length, perhaps because the tool was not available for monitoring. Although, a qualification of this finding is that the average turn length in Task 1 and Task 2 do not differ from each other by condition.

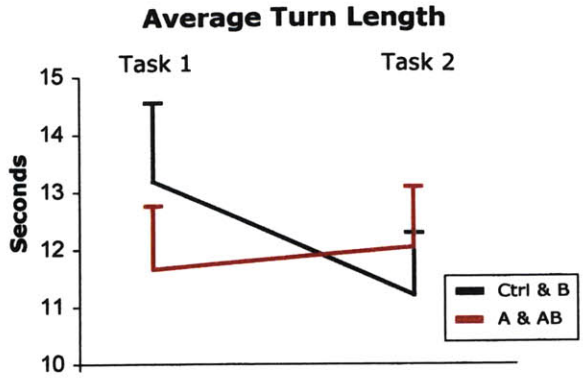


Figure 4.12 Change in Turn Length Between Tasks by Condition

Table 4.14 Impact of Real-time Display on Average Turn Length

Condition	N	Task 1 Turn Length	Standard Error	Task 2 Turn Length	Standard Error
No Real-time Display (CTRL & B)	11	13.16 sec	1.38	11.18 sec	1.09
Real-time Display (A & AB)	11	11.64 sec	1.09	12.02 sec	1.05

5. *Impact of the Replay on Perceptions of Task*

As the previous sections have illustrated, the replay display influenced how people shared information during a task and led to an increase in the amount the quietest members spoke. Also, groups with the replay did not show an increase in the dispersion in the participation levels that those without the replay showed in the first minutes of the second task. In addition to these behavioral changes, the replay display also had an impact on the individuals' perceptions of the task, in terms of difficulty, efficiency, and satisfaction. This section presents these findings.

Unlike the previous data analysis that examined group-level responses, the results here present individual-level responses during the experiment. These data include the groups that were removed from the group-level analysis, so there are 32 groups with up to 123 individuals answering each question. Instances of lower sample sizes are due to missing values or unanswered questions.

The first finding related to individual perceptions of the task is that individuals who viewed the replay display (Conditions B & AB) found the *first* task to be more difficult than those who did not (Conditions CTRL & A) (ANOVA, $F(1,111)=5.902$, $p<0.05$). This is unexpected because the groups answered this question before seeing the replay, and up until this point, there were no procedural differences between the groups. Their perception of greater difficulty must be associated with their anticipation of seeing the replay and anxiety about seeing their "performance" visualized. Table 4.15 provides a summary of average perceived difficulty for these conditions.

Groups using the replay also thought they had done significantly better on the first task than those who did not have the anticipation of seeing a replay (ANOVA, $F(1, 111)=4.605$, $p<0.05$). Again, it is surprising that the anticipation of seeing their interaction influenced their confidence in their ability to perform on the task. On the second task, there were no differences in perception of task performance across conditions. Table 4.16 provides this data, divided by groups seeing versus not seeing the replay.

In terms of satisfaction with the group's final decision, those who had the replay were more satisfied with the final answer in both the first and the second task as compared to those who did not see the replay (Task 1: ANOVA, $F(1,111)=6.351$, $p<0.05$; Task 2: ANOVA, $F(1,111)=7.851$, $p<0.01$). Table 4.17 provides these values.

Table 4.15 Ratings of Task Difficulty
On a 7-point Likert scale

Condition	N	Task 1 Difficulty	Standard Error	Task 2 Difficulty	Standard Error
No Replay Display	55	4.2	0.22	3.8	0.24
Replay Display	60	4.9	0.16	4.3	0.26

Table 4.16 Perception of How Well Group Performed
On a 7-point Likert scale

Condition	N	Task 1 Difficulty	Standard Error	Task 2 Difficulty	Standard Error
No Replay Display	55	5.55	0.16	5.47	0.19
Replay Display	60	6.00	0.11	5.77	0.15

Table 4.17 Satisfaction with Group's Final Answer
On a 7-point Likert scale

Condition	N	Task 1 Difficulty	Standard Error	Task 2 Difficulty	Standard Error
No Replay Display	55	5.62	0.20	5.35	0.23
Replay Display	60	6.18	0.10	6.12	0.15

Table 4.18 Ratings of Task Efficiency
On a 7-point Likert scale

Condition	N	Task 1 Efficiency	Standard Error	Task 2 Efficiency	Standard Error	Paired t-test
No Replay Display (CTRL & A)	56	5.33	0.20	4.91	0.21	$t(55)=1.780$ $p=0.081$
Replay Display (B & AB)	60	5.48	0.17	5.53	0.19	$t(59)=-0.22$ $p=0.842$

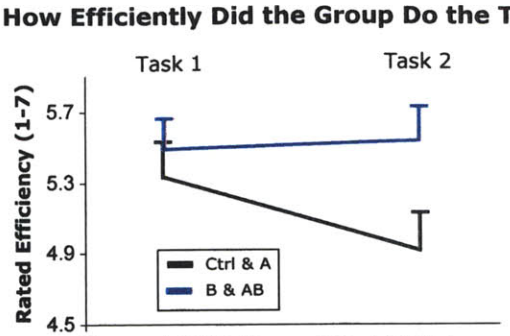


Figure 4.13 Ratings of Task Efficiency
Groups without the replay display perceived they were less efficient on the second task.

When asked how efficiently the group completed the tasks, members of groups with the replay felt they had done the second task as efficiently as the first task, and this was a significantly higher response than those who had not see the replay. These latter individuals felt they had been less efficient on the second task (ANOVA, $F(1,111)=4.041$, $p<0.05$). This result is shown in Figure 4.13 and Table 4.18.

These four results indicate that the replay had a greater impact on the perception than the real-time feedback display. The anticipation of seeing their turn-taking patterns resulted in group members perceiving the task to be more challenging and their confidence in their performance were higher than those who did not anticipate seeing a replay. In both tasks, the members of replay groups had a higher level of satisfaction with the group's final answer and in the second task they felt they had come to consensus as efficiently as they had in the first task, where as members of groups without the replay display perceived they were less efficient in the second task.

6. Trust

As in the previous experiment, subjects were asked how much they trusted the other group members between tasks. For this question, they rated how much they trusted each individual person in the group, and for the analysis those values were averaged into one rating of trust toward the group as a whole.

There were no significant changes in trust between the two tasks. There is a significant interaction effect though on the levels of trust in the second task (ANOVA, $F(1,81)=5.309$, $p<0.05$) and the level of trust is significantly higher in the replay condition compared to the control condition (Tukey, $p<0.05$). It appears that having both tools present dampened the amount of trust as compared to having just one tool. This finding differs from the previous experiment, indicating that the method by which a tool is introduced to a group has an impact on how it is accepted and how it may or may not influence how people feel about each other. See Table 4.19 for the average trust values.

Table 4.19 Ratings of Trust
On a 7-point Likert scale

Condition	N	Task 1 Trust	Standard Error	Task 2 Trust	Standard Error
No Displays	28	5.22	0.195	5.20	0.18
Real-time Display	28	5.45	0.12	5.51	0.14
Replay Display	33	5.47	0.13	5.66	0.12
Both Real-time and Replay Displays	24	5.295	0.15	5.37	0.14

7. *Ratings of One's Own Contributions*

As mentioned earlier in the discussion of participation, the amount that someone spoke during a task correlated with their satisfaction in the amount they had contributed to the discussion, in terms of both quantity and quality. Examining these responses in more detail, there are interesting effects from the displays. First, between tasks, there is a significant decrease in satisfaction over all subjects in terms of quantity and quality of contribution (paired t-tests: quantity of contribution, $t(115)=2.932$, $p<0.01$; quality of contribution, $t(115)=3.298$, $p<0.005$). This was the only question about self-performance that exhibited a significant change between tasks.

In terms of the displays, there were no discernable differences in satisfaction on the first task, yet on the second task, there was a significant interaction effect on perceptions of both quantity and quality from the presence of the real-time and replay displays (ANOVA, quantity of contribution, $F(1,81)=6.562$, $p<0.05$; ANOVA on quality of contribution, $F(1,81)=10.408$, $p<0.005$). Those who had the replay display were the most satisfied with their contributions, followed by those who had the real-time display, then the control group, and lastly, the least satisfied with their contributions were those that had both tools available to them. The difference between one's satisfaction with quantity of contributions in B versus AB is approaching significance (Tukey, $p=0.065$) and the difference between satisfaction with quality of contributions in B versus AB *is* significant (Tukey, $p<0.05$). Table 4.20 shows the data for both the first and second tasks and Figure 4.14 charts the satisfaction with quantity and quality of contribution for the second task.

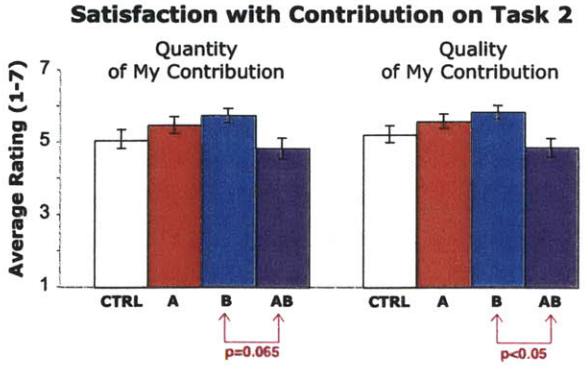


Figure 4.14 Satisfaction with Contribution for each Condition, for Task 2

Table 4.20 Satisfaction with Quantity and Quality of Contributions by Condition, on a 7-point Likert scale

Condition	N	Task 1 Satisfaction	Standard Error	Task 2 Satisfaction	Standard Error
Satisfaction with Quantity of Contribution					
No Displays	28	5.79	0.19	5.07	0.26
Real-time Display	28	5.68	0.22	5.46	0.23
Replay Display	33	5.73	0.19	5.73	0.19
Both Real-time and Replay Displays	24	5.54	0.20	4.83	0.27
Satisfaction with Quality of Contribution					
No Displays	28	5.86	0.16	5.21	0.23
Real-time Display	28	5.89	0.17	5.57	0.20
Replay Display	33	5.82	0.16	5.85	0.16
Both Real-time and Replay Displays	24	5.46	0.20	4.86	0.23

Table 4.21 Correlations with Amount of Information Shared with the Ratings on How Informative and Useful

Question	Correlation	Significance
How *informative* was the real-time display during Task 1?	r=0.529	p<0.05
How *useful* was the replay display after Task 1?	r=0.549	p<0.05
How *informative* was the replay display after Task 1?	r=0.566	p<0.05
Would you want to use these displays in other types of meetings?	r=0.416	p<0.05

8. Responses to the Displays

The final analysis of this experiment is to examine the responses subjects gave on how useful and informative the displays were to them. First, there were significant, positive correlations between several questions and the amount of information a group had shared on the first task. These correlations, listed in Table 4.21, indicate that the groups that were the most skilled at sharing information, so arguably need assistive tools less so than others, found the tools to be the most informative and useful to them. There were no significant correlations between information sharing and these questions after the second task.

Individuals who had both the real-time display and the replay display found the replay to be less "useful" than those who only had the replay display ($t(65)=2.314$, $p<0.05$) and found the real-time display on the second task to be less "useful" than those who only had the real-time display available ($t(51)=1.997$, $p=0.051$). These groups did not find the displays to be less "informative" than groups that only had one tool. These findings are not surprising because of the redundancy of the information provided in the two displays.

E. Experiment Conclusions

The results found that a replay tool, allowing groups to observe their turn-taking patterns in a previous task, led to several systematic changes in behavior and perceptions. With the replay tool, groups changed their information sharing strategy, significantly benefiting those groups that had previously exhibited a poor strategy. Those groups who were already successful were hurt by this change in strategy. This apparent conflict will be discussed in the next chapter, where groups from each of these categories will be profiled in detail.

Along with this cognitive-level change in group behavior, the replay tool appears to also increase the participation of the quietest members. The quietest subjects across all groups in the first task, the extreme "under-participators," increased their participation to a significant level in the second task after seeing a replay. Furthermore, the quietest member in the discussion was less quiet during the second task in the groups that had viewed the replay. These two observations build a case for the replay tool being an effective method for increasing the participation of quietest members of a group, although a higher number of samples will be necessary to definitively state quiet members in the replay condition speak more over the individuals in the control condition due to regression toward the mean.

The groups viewing the replay were more satisfied with their group decision and felt it was made more efficiently than those who did not have the replay. There are also indications that just the anticipation of having the replay was enough to make the task seem more difficult and to increase the confidence of the group in coming to the right choice and in satisfaction with their decision.

The real-time tool did not have as many effects on the groups. It may have caused a slight decrease in information sharing in the first task, which was corrected for by the second task. An unanticipated finding was that the tool did not influence the under- and over-participators or the turn-taking patterns during the discussion. The experiment design was different from the previous experiment in that the real-time tool was present during both tasks. This prevents us from knowing whether or not people changed in response to the display because there is no baseline measurement of their behavior without the tool (as there had been in the previous experiment). This design also enabled individuals to monitor their behavior the entire experiment, perhaps encouraging them to not change their behavior over the course of the experiment. In support of this is the finding that the average turn length for the group did not go down during the second task as it did for groups without the real-time display.

VI. Conclusions

At the outset of this chapter it was unknown whether or not a display of group behavior would have any influence on a group's interaction. As was shown with two behavioral experiments, individual behavior changes in response to real-time feedback as well as to an offline review with a replay feedback tool. The replay tool also influences individual perceptions of the task and of one's performance. And most importantly, the replay tool influenced the process of information sharing in a group setting.

The tools in both of the experiments captured and visualized how much each person spoke in a meeting, and one of the goals of these experiments was to examine whether or not such a tool would cause those at the extremes of participation to change their participation levels. It appears that the initial introduction of a real-time display of speaking levels causes those speaking the most to speak less, while a more reflective tool that is observed after a task causes those who spoke the least to increase the amount that they speak. Both of these findings are encouraging, and will be further supported in the next chapter in the discussion of real-world groups and their initial and long-term responses to participation feedback.

In real-world decision-making, outside of the laboratory setting, groups rarely know if their decisions are right or wrong. More often than not, groups have to rely on other cues to evaluate a decision outcome. These include their satisfaction with their contributions to the decision, their sense of how efficiently the groups worked toward the decision and their satisfaction with the final outcome. These three metrics were all higher when groups viewed a replay of their interaction and then made a second decision.

The paradox in this result is that groups evaluate their decisions based on their understanding of the process and the outcome. The feedback tool provided information about the process and this conveyed information that caused groups to be more satisfied with their decision outcome. This is only a good thing if the decision outcomes of the groups had been consistently better because of the replay tool. While the worst performing groups did greatly benefit from using the tool, the replay hurt the best information sharing groups. Viewing their previous interaction caused them to share much less information than they would have if they had not viewed the replay. The challenge in this then is to create visualization and review tools that accurately reflect the success of a group's decision, so a group can appropriately evaluate and change their behavior to improve it. The visualizations had multiple, positive effects on the groups' perceptions of the interaction but, because turn-taking and participation levels do not fully capture the essence of a group's decision-making process, they did not lead to universally better group processes.

Chapter 5: Observations and Lessons Learned from Groups

The previous chapter presented quantitative studies of group behavior finding that displays of participation and turn-taking influence participation and information sharing. To elaborate on these results, this chapter begins by telling the stories of a select number of the experimental groups to highlight different reactions to the displays. The chapter then presents a detailed description of three real-world groups that used the applications during their regular meetings, with an account of their use and reaction to the displays. These observations support the experimental findings as well as provide a broader context in which to understand them.

Every group is unique, with distinct character and personality. Similarly, every group's response to an awareness or reflection tool will also be unique. The following sections illustrate the response of several groups and will conclude by drawing together the similar themes across them.

I. What Happened During the Experiments?

The experimental results in the previous chapter reported that groups that watched a replay of their first interaction changed how well they shared information in a second interaction. Groups that were poor at sharing information benefited by sharing more, while those that had a successful sharing process were hurt, by sharing less. So, what happened? What process were these groups using to share information in a first task that either worked or didn't work, what happened when they watched the replay and strategized for a second task, and then what happened during the second task? This section will answer these questions by describing four groups and their stories.

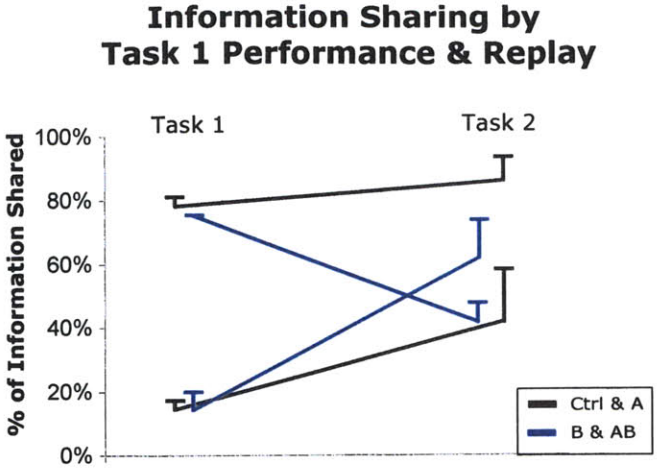


Figure 5.1 Four Types of Groups

1. Control Groups

Figure 5.1 illustrates the interaction between performance on the first task and of watching the replay on the sharing of information in the second task. The black lines represent control groups that had a slight improvement in sharing in the second task. To provide an understanding of how groups interact without a tool, it is helpful to hear the context in which two groups did the tasks: one that was naturally adept at sharing information and one that was particularly poor at sharing information.

A High Information-Sharing Control Group

"Group 20" was in the control condition, so it did not have any of the displays present during the experiment. This group was very successful in sharing information during the first task, sharing 75% of the hidden facts, and extremely successful during the second task, sharing 100% of the hidden facts. They were the only group out of the thirty-two in the experiment to share all of the hidden information in a task.

The best way to characterize how this group differed from others is that they came out of the gate running. They had spent a couple minutes chatting before the experiment started (due to unusual logistics), so they had established a friendly rapport. From the very beginning of the first task, they were enthusiastic and inquiring and figured out that the key to the decision was to exchange information. During the task, the participation was not completely equal, but not as extreme as many groups (Subject1: 22%, Subject2: 33%, Subject3: 15%, Subject4: 30%). Subject2 emerged as a leader and facilitator of the discussion and Subject4 performed the critical task of integrating information and performing a math calculation for the group.

During the strategy session, the group was told to talk about how well they thought they had done on the first task and how they might like to change for the second task. They were not shown any information about their participation or turn taking. During this

session, Subject4 mentioned that it was "important" that they all share information during the discussion. Subject3 pointed out that some information didn't appear important until it was shared. Subject1 commented that it seemed like his information was the least important. He also thought it was important that they "trust" each other's information and "share" the information. In summary, this discussion focused on the critical importance of sharing the information and in trusting each other. This group understood the task they had just performed and had generalized what they did into a strategy they could apply again.

During the second task, they were successful in applying their strategy and sharing all of the hidden information. In terms of participation, the quietest person in the previous task was now the most talkative and took a stronger role in the conversation. Participation was as unequal as it had been in the previous task (Subject1: 11%, Subject2: 20%, Subject3: 39%, Subject4: 29%). They were efficient and congenial in their discussion and once all the information was shared they quickly came to a decision.

A Low Information-Sharing Control Group

"Group 27" was very different in character from the previous group. This group was still congenial, but more awkward in getting things started and in making progress in the discussions. They shared just two facts in the first task (17%) and then just three facts (25%) in the second task. Another unfortunate outcome is that the group was unable to come to consensus within the allotted 15 minutes on either decision. This group desperately needed some help!

During the first task, Subject1 dominated the discussion by attempting to dictate a methodology for making a group decision. Subject2 attempted to help in leading the discussion by supporting him. These patterns are reflected in the participation levels of the group (Subject1: 38%, Subject2: 29%, Subject3: 17%, Subject4: 16%). The process did not work very well, as they got caught up in discussing the first point mentioned and didn't move through the information available very quickly. They shared two facts very late in their discussion and then time ran out without a decision made.

In the strategy discussion, the group talked about how they had not come to consensus and Subject1 emphasized the need to have a better "process" and that they needed "prioritize" items in their discussion in order to come to consensus faster. There was no discussion of information sharing or of listening to others' points of views.

The discussion in the second task followed a very similar pattern as the first. Subject2 attempted, more adamantly this time, to impose a process to their discussion where they prioritized each point. New points (relating to new information) were sidelined from the discussion because they were not relevant in the current "priorities." The participation pattern changed a bit in that Subject4 contributed much more in this discussion, but Subject1 remained the dominating character and Subect3 spoke very little (Subject1: 36%, Subject2: 20%, Subject3: 8%, Subject4: 36%). They shared one more fact than in the previous discussion, but the outcome was the same: they were unable to make a consensus decision in the allotted time.

2. *The Replay Groups*

In Figure 5.1, the blue lines represent the groups that viewed the replay: the ones who shared information well and then did poorly on a second task and the ones who did poorly on the first task and improved dramatically on the second task. In the control group, the top-performing group did not need help in developing a strategy to share information, while the low-performing group knew they were having trouble but continued to flounder in their second discussion. Here are the accounts of two similar groups that watched a visual replay of their first interaction before discussing their strategies for the second task.

A High Information-Sharing Replay Group

"Group 25" had a slow start in sharing information in the first task, but eventually shared 75% of the hidden information. After watching a replay of their turn taking and discussing how they could do better on the next task, they went on to share only 25% of the facts in the second task.

During the first task, Subject3 was the dominating voice in the discussion (Subject1: 12%, Subject2: 23%, Subject3: 43%, Subect4: 20%). After she repeated a hidden fact three separate times, Subject1 finally asked "do we all have the same information?" After this point, the group clarified that they all must have some private information and began to exchange notes. They ended the discussion by sharing 75% of the facts and coming to a decision.

After watching a replay of the discussion, where Subject3's verbal dominance became evident, Subject3 declared that she needed to be quiet next time. Subject2 and Subject4 both told her it was okay that she had spoken so much because she was the "facilitator" and it helped them stay "organized." In terms of strategizing for the next task, Subject3 thought they hadn't questioned "why" enough and that they should all listen to the quietest member more (Subject1). These are all reasonable suggestions that are further articulated in the questionnaires they filled out at this point in the experiment:

Subject1: *"I thought I had spoken more than I actually did speak. I will try to speak more in the next task."*

Subject2: *"[We should do more] explaining why we made our choices and what different information we had, to get it out sooner."*

Subject3: *"I hope I'll be less dominate and we'll question each other more." "Although I'm embarrassed to have spoken so much, I'm glad the group didn't find it crushing, but rather 'facilitating.'"*

Subject4: *"Some of us should speak more."*

During the second task, the quietest member (Subject1) spoke more and Subject3 spoke a little less (Subject1: 21%, Subject2: 15%, Subject3: 40%, Subject4: 23%). In this discussion, Subject3 attempted to read her information sheets aloud, a totally new

strategy that had not been discussed by the group. She was cut off in doing this by the others, who moved the discussion on to other points. This unorganized and changed strategy did not work well for this group: they shared only 25% of the facts.

With the exception of Subject2's written comment, their reflection focused on who spoke the most and the least, rather than on how they should organize the discussion. As a result, the group made changes to how they made decisions that were unnecessary and not beneficial to the group's information sharing. This highlights the drawback of providing a visualization of turn taking that emphasizes equality in speaking amount: groups use this measurement to evaluate their behavior and in instances where inequality did not signal a failure of information sharing, they alter their behavior in harmful ways.

A Low Information-Sharing Replay Group

"Group 6" needed some help after the first task. They shared just one hidden fact (8% of the facts provided) during their discussion and it wasn't clear that they even knew there were hidden facts as part of the task. After watching their replay and discussing their group dynamics, they had a marked increase in sharing, by exchanging seven facts (58%).

This group would be described as laid back and easy going. During the first discussion, Subject2 and Subject4 assumed equal leadership roles where they traded off leading the decision-making process, but did so without being overly organized or dominating. The participation was unbalanced in that the two leaders spoke a lot and the other two spoke much less (Subject1:10%, Subject2: 39%, Subject3: 14%, Subject4: 37%). The one hidden fact was shared by Subject4.

After watching their turn taking interaction, they began their strategy discussion by continue talking about the previous task (exchanging anecdotes about high school education). When the conversation turned to how to do the next task, Subect2 said she thought "we had a good process." Subject4 told Subject3 that he wished she had spoken more, to which Subject3 said she would speak up more if she disagreed with the direction the decision was going, but she had agreed with what happened during the previous task. In this way, their discussion did not focus on the visualization as much as on the particulars of the previous decision.

In their questionnaire comments, the subjects wrote:

[Subject1 and Subject2 did not write comments.]

Subject3: *"If the groups were not so friendly, it [the replay] would probably be more useful."*

Subject4: *People should review their private comments."*

In the next task, the participation was more balanced in that both Subject1 and Subject3 increased their contributions (Subject1: 18%, Subject2: 39%, Subject3: 22%, Subject4:

21%). Subject3 shared all of her hidden facts in this discussion (compared to none in the previous discussion), and Subject2 and Subject3 shared two of the three they had. Subject1, the lowest participator, contributed none of his facts. They continued their congenial manner of discussion from the first task and integrated the hidden information, and this resulted in three of the four being persuaded to change their minds and the group coming to consensus.

This group was friendly and got along well, but they needed a push in the direction of sharing information to get them going in the right direction. An interesting aspect of their strategy discussion is that they did not focus very much on the details of the replay, but rather took away one message, that Subject3 could have spoken more, and that overall they had a good process. Subject3 felt comfortable with the group, as indicated by her written comment, feeling the display was not needed for this group. But from observing many groups, I think this is exactly the type of group that the display is ideal for because its information was received without judgment, enabled the group to articulate something about their process they needed to change, and was integrated by the group into their existing style and process.

3. Lessons Learned from the Experimental Groups

These examples from the experiment highlight the potential strength and the potential weakness of having a group watch a visualization of a previous interaction.

The control groups illustrate how, when put into a decision-making setting, groups go with their natural instincts, and when given time to reflect on their interaction with no external support, they reinforce their instinctive group process. For the control group that demonstrated an excellent process, this reflection was a time to articulate it and clarify how they would do it a second time. For the group that struggled with their flawed approach, their reflection was about how they should impose more structure into their discussion, leading to a second frustrating discussion.

For the groups that were shown a replay of their interaction, this insight into their dynamics provided them a new way of considering their interaction. For the group with a healthy group process, they perceived that they had an inappropriately balanced discussion and this insight caused them to change their interaction and lose their effectiveness as a group. But for the group that was unfocused in its deliberation and had not developed a method for sharing information, the replay was the guide that pushed them toward a more productive process of sharing information through more equal participation.

II. Three Case Studies

The controlled experiments provided insights into how groups respond and incorporate the display's information into their specific decision-making situations. But these

experiments do not reveal how these tools would impact groups that occur naturally, in our pre-existing social and organizational contexts.

To complement the experimental approach of the lab studies, Second Messenger 2.0 was used by three pre-existing groups in their regularly scheduled meetings. I directly observed these meetings, collected information on how much each person spoke in them, and asked each group member to comment on their reaction to the software and their opinions on the dynamics and participation levels in the group.

The outcome of these observations indicate that all three groups wish that quieter members would speak more and that this dissatisfaction in the participation originates from both the quiet individuals and the group as a whole. Second, the display did not appear to impose unnecessary disruption or discomfort to the groups, and as such, provided a mechanism for people to monitor their speaking levels, something lower participating individuals have been shown to do inaccurately without a tool (DiMicco 2004b). In self-reports, individuals stated that they altered the amount they spoke in meetings in response to seeing the display. There is also evidence that over extended usage (in this case, ten weeks) individuals at the lower levels of participation moved to a higher level.

Collaborating groups that hold regular weekly or monthly meeting do not always come together for the specific purpose of information sharing or decision-making, the collaboration activities this thesis has focused on thus far. But, it is important to acknowledge that within any meeting, for the communication to be effective, the ideas and opinions of each group member must be communicated to a satisfactory level for both the speakers and the listeners. This does not mean equal participation, but rather participation levels that are appropriate to the group's needs. What was observed by studying these three groups using the application is that the individuals who speak the least are dissatisfied with their own participation and groups wish to hear more from them. Second Messenger 2.0 aids in this situation by making all members more aware of the imbalance in participation. As first postulated by Erickson, et al. (Erickson 2000), this mutual awareness, or "social translucence," enables more fluid, more expressive interactions, resulting in a more effective coordination of ideas, opinions and actions.

1. The Three Groups

The three groups that used the application during meetings agreed to have it running during the meeting, to have their participation recorded, and to provide information after each meeting on their reactions.

The first group is a graduate-student organization that used the application in eight weekly meetings over ten weeks. This group has no formal leadership structure and has a rotating number of topics discussed each week.

The second group is a research group comprised of one professor and six graduate students. They used the application in two meetings, a month apart from one another. This group used its meeting time to exchange research ideas around a single topic: each

meeting had a designated speaker who presented his/her work and then solicited research feedback.

The third group used the application only once, but has unique expertise and interest in the topic of meeting dynamics. They are a group within a human-resources department that designs and teaches courses on group leadership and meeting facilitation. They used the application during their weekly meeting and, like the research group, had a designated topic and leader.

Each of these groups is from within the university community – the first two are within the Media Lab and the third operates at MIT's administrative level.

The following sections present the results from the case-study analysis: first, a descriptive analysis of the observations of each group, including their participation patterns; second, a report of each group's satisfaction with the participation and contributions of each member in the group, illustrating that groups are not satisfied with the levels of participation of all members; lastly, an examination of the impact of the display. I present the reported reactions of the groups to having the display present and their self-reports on their behavior changes in response to the display. In the case of the student organization, that used the application over a ten-week period, there is preliminary evidence that they had more equal participation over time. This section concludes with a discussion of group dynamics and the broader implications of live feedback in general collaborative settings.

B. Analysis of Participation

As all groups do, each of the three groups had unique behavior patterns and social dynamics. The following subsections present an overview of each group's interaction style and how that was reflected in the measured participation levels.

1. The Student Group

As mentioned before, the Student Group is an informal group of graduate students who meet weekly to discuss the status of various projects. There is no designated leader; the only designated role in each meeting is note-taker. Their meetings range in length from fifteen minutes to an hour, depending on the attendees and the number of topics to discuss. The group began the study with six members, gained a new member in the second week, and lost a member by the fourth week, so ended the study with again six members.

The application was used during eight meetings over a ten-week period. Figure 5.2 shows the participation patterns from five of these meetings: Week 1, Week 4, Week 5, Week 8, and Week 10. (The participation data from Week 3 and Week 7 was lost and the application was not used in Week 2 or Week 9.) While these charts show the Histogram for each of these meetings, the group did not always use this particular visualization. Based on their own preference, they used either the Histogram, the Group Circle in its simplest view, or the Bouncing Balls in anonymous mode (each of these visualizations

are discussed in Chapter 3). They also often chose to have the application show the previous five minutes of discussion, so the display would be more dynamic and responsive to the turn taking during the meeting.

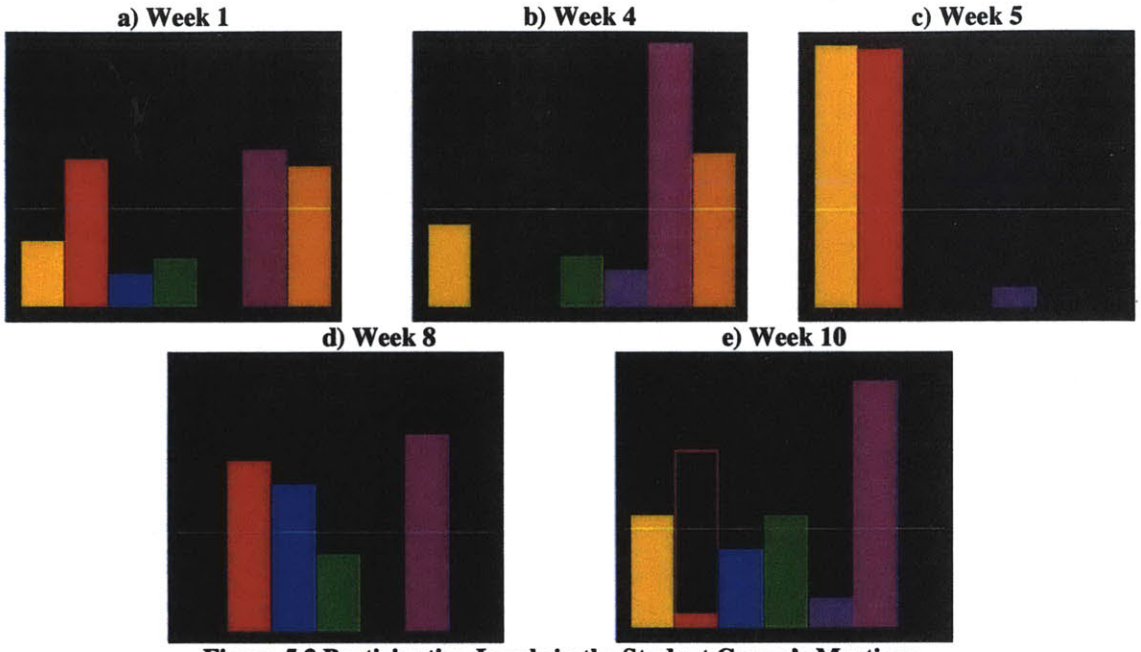


Figure 5.2 Participation Levels in the Student Group's Meetings
There are seven members of the group, with a varying number attending each week. The horizontal line represents where equal participation falls for this size group. In week 10, the red member chose to not speak into a microphone, so the red outline represents the estimation of this person's participation.

As shown in Figure 5.2, there were some trends in participation over the ten weeks, with occasional dramatic fluctuations. When the red, pink and orange individuals (the 2nd, 6th, and 7th bars, from the left), attend a meeting they speak the most. The yellow, blue, green, and purple individuals (the 1st, 3rd, 4th, and 5th bars, from the left) speak the least when the others are present, yet speak at a higher amount when they are not. Because their meetings are not driven by discussion leaders or a manager, these fluctuations reflect individual differences rather than differences in meeting discussion topics.

2. The Research Group

The second group in the study is a research group of seven members. A professor runs the group and the other six members are graduate students. Each of their meetings begins with a brief status-update period, run by the professor, and then one student spends the rest of the meeting presenting his or her research and soliciting feedback from the group.

The group used the interface twice, in monthly meetings. The first time they used the Histogram display and in the second they switched back and forth between the Histogram and the Group Circle displays. The participation levels for these two meeting are shown in Figure 5.3. In the first usage, the discussion leader (the purple bar, 5th from the left)

decided to not wear a microphone during her presentation because she wanted to use the display as a mechanism for measuring how much feedback she got from the group. In the second usage, the discussion leader (the pink bar, furthest right) decided to switch the display when he finished his presentation to show only the previous five minutes of the meeting, so he could then view who was providing him feedback. I encouraged groups to use the display tool in these different ways to accommodate to their existing group culture.

Across these two meetings, the professor (the red bar, 2nd from the left) spoke a high amount and one member (the yellow bar, 1st from the left) spoke very little. The variation amongst the others' behavior can be attributed to the topic of the meetings and who was leading the discussion.

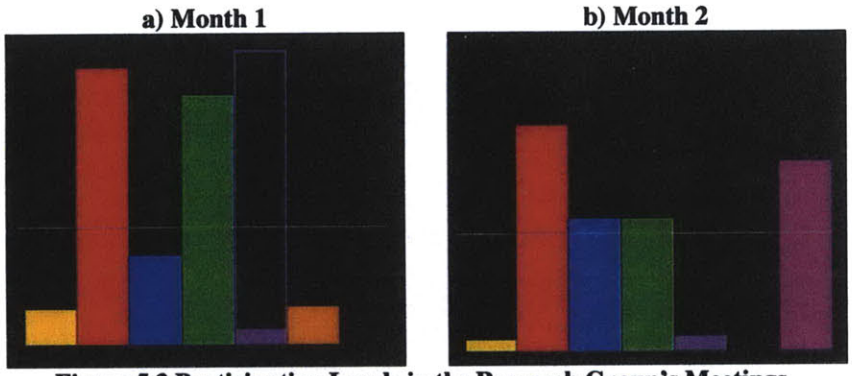


Figure 5.3 Participation Levels in the Research Group's Meetings
There are seven people in the group, with six of the seven attending each meeting. In the Month 1 meeting, the designated presenter (the purple bar) wore a microphone only during the first few minutes of the meeting. The outline of the bar is the estimation of her participation level.

3. The Human Resources Group

The third group in the study is a team of professionals who teach training courses on team leadership and meeting facilitation. Because of their expertise, they are finely attuned to their own group dynamics and participation levels. This group is differs from the other two groups because they are not as familiar with technology. For these reasons, I was particularly interested in gathering their response to the display.

The group used the application during one meeting that had six people in attendance and had a designated topic and discussion leader. In Figure 5.4, the right-most pink bar is the leader and the red bar (2nd from the left), the lowest participator, is the note-taker. I asked the group how typical this meeting was for them in terms of participation and they said typically their meetings have a very different pattern and that was why during the meeting two people specifically asked the red person if she had more to contribute. This group was unusual in that they felt comfortable directly asking people to make comments and were comfortable discussing the group interaction process as it occurred (i.e. pausing the meeting and asking for a "check-in" on how people felt the meeting was going).

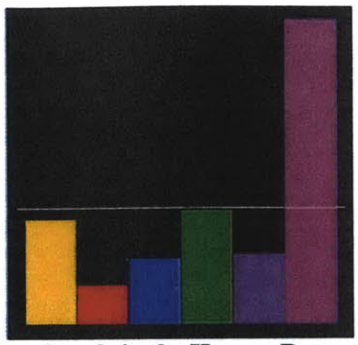


Figure 5.4 Participation Levels in the Human Resources Group's Meeting
Six group members attended this meeting and the highest bar was the designated discussion leader.

Table 5.1 Dispersion in Participation Levels

Meeting	Num participants	Standard Deviation of participation	Dispersion (SD * Num participants)
		Student Group	
Week 1	6	9.1	55.2
Week 4	5	13.9	78.5
Week 5	3	27.1	76.1
Week 8	4	15.8	57.3
Week 10[4]	6	13.8	69.2
		Research Group	
Month 1[5]	6	16.6	82.8
Month 2	6	12.5	75.2
		Human Resources Group	
Initial Use	6	13.7	82.1

As a final comment on participation patterns in these three groups, Table 5.1 lists the standard deviations of participation levels in each of their meetings. To create a metric for comparing participation "dispersion" between meetings, I scaled the standard deviation by multiplying it by the number of people attending the meeting. This dispersion value provides some indication of the spread in participation week to week. While there are very few data points, it appears that the groups with designated discussion leaders (the Research Group and the Human Resources Group) have a higher dispersion in participation than the Student Group. This observation is marginally supported by a t-test comparing the means of the dispersion in these two types of group ($t(6)=-1.939$, $p=0.065$).

C. Analysis of Satisfaction with Group Interaction

Second Messenger 2.0 measures and displays the quantity of speaking in a meeting, not the quality of what is said. There are obvious weaknesses in using quantity as a

[4] One speaker chose not to speak into a microphone, so was not included in the calculation for standard deviation and dispersion.
[5] The designated speaker for this meeting did not wear a microphone, so was not included in the calculation for standard deviation and dispersion.

measurement of group interaction and groups using the system often point this out as a drawback. Therefore I asked groups about their satisfaction with the quantity of their contributions, the quality of their contributions, and if they wished they had spoken more or less. I also asked groups to rate each person in the meeting for how much they had perceived that person had contributed in terms of quantity and quality and if they wished that that person had spoken more or less. These questions were designed to gain insight into how quality and quantity relate to satisfaction and participation in meetings.

When asked directly if they wished they spoke more, the individuals who spoke the least in the Student Group and the Research Group stated that they wished they had spoken more. When asked about their satisfaction with the quality of their contributions, there was a wide range of satisfaction. The Human Resources Group was unusual in that they all expressed extreme satisfaction with their individual contributions and had no desires for changes in participation.

Table 5.2 Correlations between Quantity and Quality

Significant Correlations	Student Group	Research Group	Human Resources Group
Satisfaction with my *quantity* of contribution vs. satisfaction with my *quality* of contributions	0.943 p≤0.05	0.923 p≤0.01	0.939 p≤0.05
Quantity others perceive you to be speaking vs. *quality* they perceive in what you say	–	0.821 p≤0.05	–
Quantity others perceive you to be speaking vs. *how much* they would like you to speak more	-0.912 p≤0.05	-0.955 p≤0.01	-0.846 p≤0.05
Quality others perceive of your speaking vs. *how much* they would like you to speak more	–	-0.901 p≤0.05	–

Table 5.2 lists the significant correlations on the questionnaire responses asking about quantity and quality of contributions. Across all three groups, there was a strong, positive correlation between the satisfaction individuals had with the amount they contributed and with the quality of what they said. Also across all three groups, there was a negative correlation between how much the group perceived someone to be speaking and how much they wish that person would change the amount they spoke, meaning that when someone was not talking very much, the group desired to have them speak more.

In the Research Group, there were further significant correlations involving quality. In this group, the professor, who was rated as having the highest-quality comments, spoke a large amount (the second-highest amount). The quietest member of this group, who did not make any comments beyond agreement with what was being said, was rated as having the lowest-quality comments. These two extreme values explain why this group has a positive correlation between perceived quantity and quality of comments and a negative correlation between quality spoken and desire to have that individual speak more. This group wishes the quiet member would speak up, and feels that the person speaking the most is contributing high-quality comments.

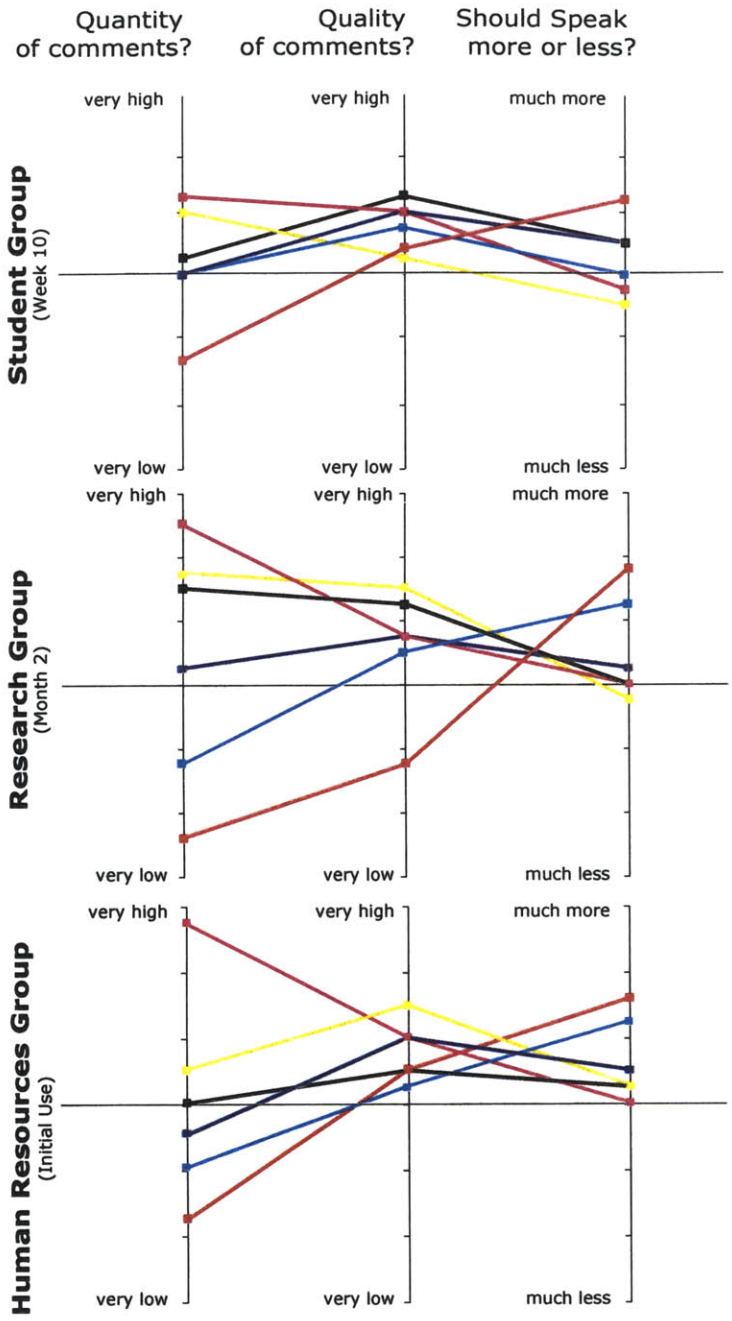


Figure 5.5 Each Group's Average Ratings of Each Member's Contributions
Each line represents a member of the group and the value is averaged across all of the other member's ratings of that individual, expressing the perceived quantity and quality of comments and if they wish that member had spoken more or less. The discussion leaders in the Research Group and the Human Resources Group meetings are colored pink.[6]

[6] Note: The colors in these graphs do not correspond with the Histograms in Figures 3, 4 and 5.

Figure 5.5 diagrams how each group rated each of its members in terms of quantity, quality and desire to have the person speak more or less. The designated leaders can be distinguished in the Research Group and the Human Resources Group because they spoke the most and were rated as such: the leaders are the highest points on the left-most axis (representing quantity of speaking). With the exception of the quiet member of the Research Group mentioned above, every person in every group was rated as making comments of above-average quality, indicating that the groups appreciate the input of each member. The Human Resources Group indicated that they did not want anyone to speak less than they did – this is consistent with their self-reports in which no one stated that they personally wished to change the amount that they had spoken. In the Student Group, in which no person had a role of greater or lesser importance, the distribution in perceived speaking quantity was smaller than in the other two groups, yet there is a stronger desire for the most talkative members to speak less. From an overall perspective, this diagram highlights that groups desire to hear more from the quiet members.

One possible issue with these measures of personal satisfaction and the desire to hear more or less from different group members was that the display itself was causing this dissatisfaction. So I asked the Human Resources Group, prior to using the application, to comment on their perceptions of their average group participation levels and if they wished they spoke more (or less) and if they wished others would speak more (or less). They revealed the same patterns illustrated across Figure 5.5: the group desired to hear more from those that usually spoke the least and less from those that usually spoke the most. An interesting observation though is that when the group used the software, they exhibited different participation patterns than they had described in the pre-usage survey, so while the trends in their ratings stayed the same, the individuals that they referred to in their ratings were different.

Although the amount one speaks in a meeting is not a measure of the quality of one's contributions, this study found that individuals in groups have varying levels of satisfaction with the amount they are speaking, with those at the lowest levels expressing the desire to speak more. Furthermore, groups wish to hear more from these same members. Because effective communication requires that individuals express their ideas and opinions to a satisfactory level for both the individual and the group, it appears that when some people speak significantly less than others in meetings, there is deficient group communication.

D. Analysis of the Impact of the Display

The previous sections described the unique characteristics and patterns in participation of the three groups, yet all expressed desire to hear more from quieter members during their meetings. So the question is, does the introduction of the display have an impact on this? Can increasing the awareness of participation correct an extreme imbalance? Do groups accept the technology into their interaction? And when used over time, are there any longer-term changes in behavior?

The experimental studies of groups using this software found that, when first using a display of participation, those at the highest levels of participation spoke less and the lowest participators did not speak more (DiMicco 2004b). The second set of experimental findings indicates that after reflection on participation levels through a replay, the quietest members speak more. While there are many challenges to studying real-world groups over long periods of time, the artificial nature of laboratory groups (in structure, relationship, and task) limits the ability to generalize findings to the real world. Therefore, although qualitative in nature, I am optimistic that the findings presented here can be generalized to other types of groups.

This section discusses the impact of the display from three perspectives: first, the initial responses to the display reported by the Research Group and the Human Resources Group, to answer questions about accepting the technology into the meeting environment; second, a summary of how the Student Group and the Human Resources Group reported that they were changing their behavior when they saw the display; and last, the change observed in the participation levels in the Student Group over the ten weeks of usage.

1. Initial Reactions

After the Research Group and the Human Resources Group used the application for the first time, they rated their experience and interaction with it. The goal was to gauge if the application was useful or disruptive and how comfortable they were with having this behavior information revealed publicly. A summary of these ratings is in Table 5.2 and Figure 5.6.

The groups found the display to be marginally useful and informative. Although not statistically different, the Human Resources Group rated the usefulness of the display to be much lower than the Research Group did (2.6 versus 4.0). I attribute this to their expertise in gauging meeting progress and interaction and feeling that this tool was redundant.

Overall, individuals reported looking at the display during the meeting and finding the information shown to be accurate. Those in the Research Group who questioned the accuracy of the display (2 of the 6) were familiar with voice and speech recognition technology so naturally took a critical approach in evaluating measurement accuracy. There was not a similar skepticism in the Human Resources Group.

Both groups reported being very comfortable using the technology, rating their own comfort and the perceived comfort level of others to be quite high. There were no differences between the two groups, despite the difference in the groups' experiences with sensing and collaboration technology. It should be noted though that each group contained one member that rated their comfort level far below the others in the group, so the finding does not indicate that all individuals welcome this type of feedback. In one of the three groups, this person was one of the quieter members; in the other two groups these were individuals who spoke an average or above-average amount.

Table 5.2 Reactions After First Use
(Average group ratings, on a 1 to 7-point scale.)

	Research Group	Human Resources Group
Was it useful during meeting?	4.0 (SD, 1.1)	2.6 (SD, 1.2)
Was it informative?	4.2 (SD, 0.8)	4.0 (SD, 2.1)
Did you look at the display?	5 out of 6 people	6 out of 6 people
How accurate was the information?	2 out of 6 people questioned accuracy	6.4 (SD, 0.8)
Comfortable with info shown?	5.3 (SD, 1.9)	6.2 (SD, 1.3)
Think others are comfortable?	5.2 (SD, 0.8)	5.4 (SD, 1.1)
Willing to use in other meetings?	4.7 (SD, 1.0)	3.0 (SD, 1.2)

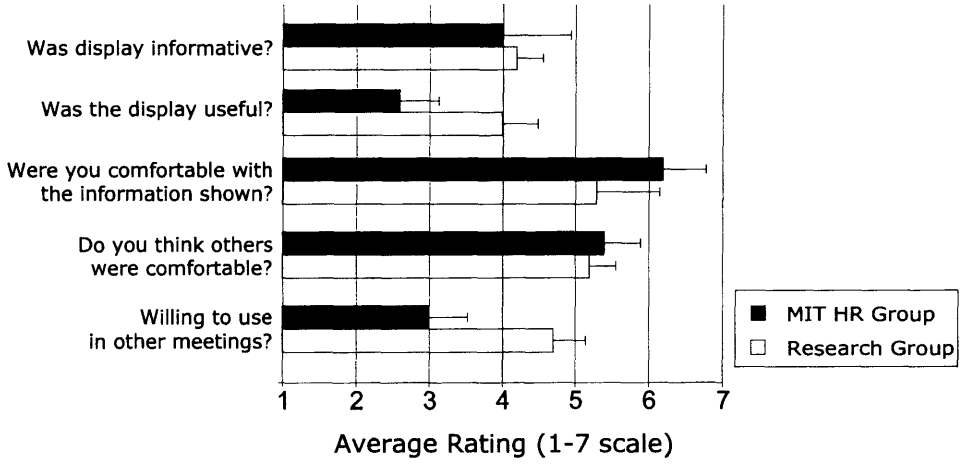


Figure 5.6 Reactions After First Use

The last question on the survey asked if individuals would be willing to use the display in other meetings. The Research Group expressed marginal interest in using the display in other types of meeting (4.7 out of 7) and the Human Resources Group was less interested in using the display in other meetings (3.0 out of 7). This average rating was found to be significantly lower than the Research Group's rating in a t-test of independent samples ($t(9)=2.453$, $p \leq 0.05$). Again, this can be attributed to the culture and expertise differences between these groups. The Human Resources Group has expertise and skills in detecting participation and engagement and did not feel the tool was adding additional information to their meeting.

In the post-interview with the Human Resources Group, I asked when they would find the displays useful, because they were enthusiastic about the visualizations of their interaction and were knowledgeable about the issues of group intervention. They suggested that a facilitator could use the system to check-in to see if he or she had been

following a particular group process that he or she desired to use at different points in the meeting. As an example of this, the leader of their meeting said that she was reassured that she had conveyed her information to the group when she saw her circle grow very large, and then as it shrank later in the meeting, she was assured that she was listening and getting the feedback she desired from everyone else. They did not suggest the display should be made private, but felt that someone with facilitation expertise should use it in conjunction with their other techniques, rather than providing it to everyone to use in any type of meeting.

They also recommended that the tool be used with groups that had specific participation issues, such as one dominating member or one very quiet member, but that it should always be accompanied by "pre-work" done with a consultant to assist the group in being ready to receive the display's information, so that it would inspire the appropriate change. Because their expertise is in consulting and working one-on-one with groups to alter their processes, their recommendations were to use the tool as an aid to those techniques.

In summary, the groups all paid attention to the display and welcomed it into their meeting, but did not rate it as extremely useful to them and did not wish to incorporate it in more of their meetings. This is an issue in terms of the prospects for long-term adoption and deployment of such an application and indicates that the Human Resources Group's suggestion to use the tool in conjunction with human facilitation might be the most effective way to incorporate these tools into meetings.

2. Behavior Change: Reported

To answer the question of whether or not people changed their participation during a meeting, one approach was to ask them directly. In the previous laboratory experiments people changed their participation levels without consciously realizing they are doing so, yet because I was not controlling other factors in these meeting, there was no other means for measuring change.

In the Human Resources Group, one person said she spoke up because she wanted to prove that she was not a low participator. Another person said she spoke less when she saw how big her circle was on the display. The remaining four said they did not change the amount they spoke.

In the Student Group, in Week 1, the three highest participators all responded that they decreased the amount they spoke in response to the display. They also discussed it openly with the group during the meeting, talking about how they were trying to be the "second-highest" participator not the highest. The three lowest participators reported that they did not try to change how much they spoke. This finding is in support of the previous controlled study that found that over-participators decreased their participation on a first time usage, but under-participators do not change. In Week 4 though, the reverse was reported. The highest participators now said they had not changed their behavior and two of the lowest participators stated that they tried to speak more. This finding is in line with the experimental finding that suggested reflection on past participation led under-

participators to speak more. This observation of the Student Group suggests that over time under-participators may use the display to alter their behavior, more so than those who speak at higher levels.

3. *Behavior Change: Observed*

One of the challenges in measuring long-term changes in groups is that membership changes and the topics discussed week to week vary, both of which greatly influence how much each person speaks. Although there is support for the hypothesis that over long-term usage a group changes its participation patterns, it must be acknowledged that other factors may have influenced this result.

Table 5.3 Student Group's Participation Levels

Person	Week 1	Week 4	Week 5	Week 8	Week 10
1 (yellow)	*11%*	*14%*	*50%*	-	*15%*
2 (red)	24%	-	43%	28%	21%[7]
3 (blue)	*5%*	-	-	*24%*	*11%*
4 (green)	*8%*	*8%*	-	*12%*	*15%*
5 (purple)	-	6%	3%	-	4%
6 (orange)	26%	44%	-	33%	34%
7 (pink)	23%	26%	-	-	-

Over the ten weeks of usage in the Student Group, the highest participators were still the highest and the lowest participators were still the lowest, but there was a movement in the upward direction in the participation of the three lowest participators that were members of the group from Week 1 onwards. The participation levels of all of the individuals are in Table 5.3. As shown in the table, there is a consistent rise in participation of Persons 1, 3, and 4, the three members who spoke the least in the Week 1. With only three individuals observed, this is not a statistically significant change, but it is encouraging and offers up the possibility that the display produced a long-term change in the group's interaction.

E. Discussion and Conclusion

The primary intention of Second Messenger 2.0 is to convey information during a meeting that enables a group to observe, understand, and change its behavior. These case studies illustrate how the output of the displays provide a way to view the behavior of a group after a meeting and that there is a desire amongst group members to change along these visualized parameters, namely participation. According to self-reports and the one study of long-term use, the display led to the desired changes in group behavior.

---

[7] This value of 21% is an estimate based on observation and above the 5% measured by the hardware. As a result of this adjustment, the other participation measurements were adjusted to maintain the relative percentages. The original measurements for the group were: 18%, 5%, 13%, 18%, 5%, and 41%.

The research described here provides insight into how real-world groups, with their own priorities and meeting agendas, incorporate an awareness tool into their existing work practices. While the groups did not overwhelmingly embrace the technology to the point of requesting it in more meetings, they did respond to the information with an increased awareness of their own behavior and expressed a desire for at least some members of the group to change in response to that information. It is encouraging that self-reports and observation support the conclusion that individuals appear to change their behavior in response to the display, without it causing disruption in the group's interaction or task.

While the temptation was to incorporate multiple layers of social information into each visualization, groups seemed to be less interested in the more nuanced representations of overlapping and backchannel speech (as described in Chapter 3). As a result of this, I recommend to those building tools for increasing real-time group awareness to limit the complexity of the information conveyed, so that it can be incorporated into a real-time interaction.

As a final comment, Erickson and Kellogg describe a simple, yet compelling, example of social translucence in the physical world with their description of a window cut into a hallway door (Erickson 2000). By placing a window into the door instead of placing a sign on the door that says "open door slowly," people approaching the door will see if someone is approaching on the other side, and then will be able to respond appropriately (by not opening the door in that person's face). And they can expect the same from the person that sees them. Second Messenger's visualizations of social dynamics and group behavior propose a similarly simple solution that can assist in the complex problem of improving group processes. Compared to an explicit instruction to a group that it should make sure that each person at the meeting has their time to speak, these interfaces open a window onto the interaction of the group by presenting a new visualization of it. The expectation is that groups will use these new windows as ways of observing and adjusting their behavior in ways that assist them in improving their communication effectiveness.

III. Private Participation Displays

As an extension Second Messenger, which is designed for public display, I built a prototype of a private feedback system for individuals to view during meetings The private displays are tri-color LED lights, each housed within a small box so that the color of the light can only be viewed from one side. Three of these lights are shown in Figure 5.6.

The lights can be changed to any RGB color through a console on a computer, as a means of signaling different messages to individuals. The goal in designing such a simple display was that it should not distract a person away from the ongoing conversation nor should it have so much information in it that it would take time to interpret.

The lights were deployed in a meeting where their color indicated to the individual their current level of participation. At average participation, the light would be a neutral white;

when a person spoke relatively more than others it would become redder and as a person spoke less than others it would become greener.

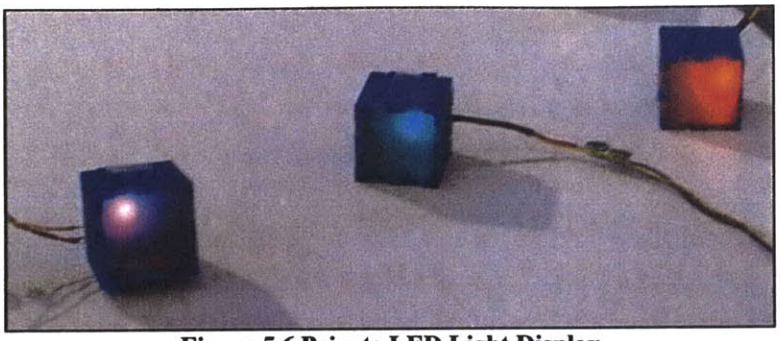


Figure 5.6 Private LED Light Display

Three members of the Student Group who had participated in the previous case studies used the LED lights during one of their meetings. They filled out a similar survey to those filled out in the other case studies and gave a variety of answers. The most talkative member said that he attempted to speak less when his light went red and the other two members indicated that they did not change their participation as their lights had been close to neutral during the entire meeting.

In comparison to the shared displays of Second Messenger, two of the three said they were slightly more comfortable having this private display, yet beyond this, there were few consistencies in their answers about how distracting or helpful it was over the public version of the application.

The most informative comments were that they enjoyed using the private lights and found it easier to interpret than the public display. When looking at a visualization of the whole group, it takes effort to remember which visual object represents you and how it compares to the others. The light filters this information down to the individual perspective for each person, eliminating any need to "interpret" the information real-time. I had predicted that individuals would be adverse to this type of display because it would feel like a computer was instructing each person on exactly how much they should speak, but the response appears to have been quite different, with individuals taking the light as a cue that they could accept or reject. This finding is encouraging for continuing this direction of research.

IV. Overall Conclusions from Experiments

At the conclusion of the evaluation of the impact of the displays, there are several findings across the lab experiments and the real-world observations that suggest some further points to consider.

The first is that reflective feedback, through a replay or over time with extended use, appears to generate the most change in a group. Specifically, after using the display to reflect on participation, the quietest group members speak more. This is a positive

finding, because real-world groups express a desire to hear more from quiet group members and if a group's goal is to share opinions and information, this change in participation patterns can coincide with greater information sharing. These findings illustrate how this tool can bring about positive change in a group without mediating or dictating group communication.

Yet, is equal participation an appropriate goal for a group? As suggested by the Human Resources group and as observed in the high-information-sharing groups in the lab experiment, those who do not need the feedback should not receive it. When provided these tools, groups comply with their own desires for good group process and alter their participation dynamics, yet when the group is functioning very well, this perceived need to change and the subsequent change can be damaging to their process.

The flip side of this argument is that groups that have difficulty sharing information and routinely overlook the opinions of the quiet group members can benefit from this type of tool, particularly when used in a reflective mode. This suggests that it may be best to have a trained expert involved in the process of group intervention to determine when a group needs assistance and to provide preparation and training to the group. Another approach is to consider how different types of tools could emphasize more accurate measures of good group process, such as actually measuring information sharing or task accuracy. Although, without having such tools to evaluate, it is pure speculation they would be received and interpreted appropriately, to universally benefit groups.

The lesson from this is that information technology can influence on our behavior, especially when it provides a guide where there was not one before. The evaluations illustrate the many different ways behavior can change in response to simple feedback visualizations. As some of these changes were anticipated and several others were unexpected, it is important to acknowledge that these technologies must be evaluated and understood prior to universal use.

Chapter 6: Implications and Conclusions

This thesis presented the following findings:

I. In a face-to-face setting, a dynamic display, positioned peripherally to a discussion, influences the behavior of the individuals in the group. When this display reveals information about participation in a meeting, the following was shown to occur:

- *Those at the highest levels of participation decrease the amount they speak.*
- *Individuals who do not have critical information to contribute to the discussion decrease the amount they speak.*
- *Individuals who speak the least do not spontaneously increase their contributions.*

II. When a group watches a visualization of its previous interaction, there are multiple effects on the perceptions of the decision outcome, on individual behavior, and on the group-level decision-making process:

- *Groups that anticipated seeing a replay of their interaction found the task more challenging than others. They also felt they had performed significantly better than those groups that did not anticipate this replay. After viewing the interaction visualization and performing a second task, groups felt they were significantly more efficient during the task than those who had not seen a replay.*
- *Individuals who spoke the least in the visualized interaction demonstrate a significant increase in contributions after viewing the visualization.*

- *Groups performing an information-sharing decision task demonstrate a change in decision-making strategy, which greatly benefits those that were ineffective at information sharing during the visualized task. Those that had a good information-sharing practice were harmed by their subsequent change in strategy.*

III. When pre-existing groups incorporate a real-time display of participation into their regular meetings, they demonstrate consistent patterns week-to-week, and also a desire for change:

- *Groups consistently state that they wish to hear more from the quietest individuals.*

- *The group members who consistently speak the most in meetings initially react to the display by speaking less, but over time express less desire to decrease their contributions.*

- *After having a display of participation present in multiple meetings, the quietest members of real-world groups were observed to speak more. This supports the theory that feedback can produce long-term changes within a group, generated through reflection, that can differ from immediate, short-term effects.*

The implication of these findings is that providing any automated feedback to groups will cause them to adapt their behavior to accommodate to the normative pressure imposed by the feedback. In the case of the Second Messenger interfaces, the implied ideal of balanced participation led to an adjustment of participation behavior. Without mediating the communication and without explicitly instructing groups on how to change, individuals altered their personal behavior and this influenced their perceptions of the task and the group's decision-making strategy.

The challenge put forth by this implication is to find other measures of group behavior that can provide informative feedback to groups about their behavior in collaboration settings. The following section outlines different ways of approaching this challenge.

I. Future Work

There are many different directions to take this research. To further our understanding of how increased mutual awareness caused by technology can our communication patterns and collaboration outcomes, I propose the following subset.

A. Further Analysis of Speaking Patterns

An immediate direction for further work is to explore more rigorously what information can be summarized from an audio stream of who-spoke-when. The fields of discourse analysis and sociology provide frameworks for interpreting interaction patterns (for

example, see (Fay 2000, Parker 1988)). Currently, I have data from over 60 conversations collected and ready to be further analyzed for turn-taking exchanges. A first question to answer is whether the results of Fay can be replicated; these results demonstrate that the most influential person in a large group is the one who speaks the most and in smaller groups is the one who interacts with individuals directly. They quantified group interaction by analyzing turn switches and characterizing a "direct interaction" between individuals to be when one person directly follows another one in a sequence of speaking turns. A second direction to explore would be to use techniques of pattern recognition to determine the social roles in natural conversations, much like (Choudhury 2003) in her dissertation work on detection of social networks.

B. Tools for Personal Reflection

One of the outcomes from the first study with the display of speaker participation was that the introduction of the display disturbed the establishment of trust between group members. Certain subjects also rated the accuracy of the display to be much lower than others did. Both of these findings indicate that a public display may not always be the best way for individuals to receive information about their behavior. A private display of similar information or of personalized, specific suggestions on how to alter behavior could be more useful to individuals in certain circumstances.

As discussed briefly in Chapter 5, private LED lights were built that could signal to individuals information about their personal level of participation. These lights were well received by a trial group, indicating this direction is worth pursuing. Private tools could be used for both real-time feedback and for reflection, over not just single interactions but a collection of meetings, providing a mechanism for individuals to learn about their behavior over time and in different contexts.

C. Social Communication Patterns

In addition to our spoken words, our physical actions convey information about our attitudes and intentions. Eye gaze, head movements and posture provide valuable cues about both the discourse structure of a meeting and the attitudes of the participants. Sensing just one of these behaviors may be enough to enhance a system with knowledge about the attitudes of group members and this information could further be used to provide social feedback.

In terms of sensing bodily movement, head movement and fidget detection may be detectable by placing accelerometers on the users' microphone headsets. For posture-change detection, a collection of pressure sensors could be placed on the seat and back of the user's chair, similar to what has been used by Mota and Picard (Mota 2003). For gaze detection, i.e. when one person is looking at another, head position could be tracked through ceiling mounted cameras or by using IR-beacons on each headset (Selker 2001).

Each of these existing technologies can be implemented in a group setting, opening up many opportunities for new types of interfaces and studies of group interaction dynamics.

D. Discourse Detection

The original vision of my research was to use conversational speech as input to an intelligent system. As speech technology advances and detecting verbal content becomes a feasible reality, the possibilities for analyzing a group's interaction and providing constructive assistance are numerous.

Once a transcript of a real-time meeting is obtained, there are existing techniques from the field of computational discourse analysis to model a group's interaction that can be utilized for building applications that track the content of meeting and the intentions of the individuals in the meeting (Galley 2003).

E. Further Experimental Questions

Outside of developing new applications and computational understandings of face-to-face interactions, there are many experimental questions to explore based on the findings to date. In particular, I am interested in further exploring the issue of subjects over-estimating their own participation and how different displays might assist users in correcting this misconception. Would the use of private displays change the response to the technology? How would viewing the information offline and privately change the resulting individual behavior? Once the social context is removed from the viewing of the information, individuals may still find the information interesting, but not as relevant to their next group interaction. Another area of investigation is in the manipulation of data: if an individual is presented feedback that is a purposefully distorted representation of their behavior, will that individual detect the inaccuracy or will he/she accept it and attempt to change in response to it?

II. Broader Implications

The goal of this thesis was to demonstrate a new way for technology to assist collaborating groups. Previous work in collaboration technology takes a "mediating" stance by controlling and transmitting the messages between people. As discussed, there are many drawbacks to this type of interaction, the main disadvantage being the constraining effect the mediation has on the abilities of individuals to say what they mean.

In an attempt to address existing problems in group dynamics and decision-making, this work produced technology that encouraged group members to change their behavior, with the hope that these changes would lead to improved overall group process. This type of technology can be referred to as "persuasive" (Fogg 2003) because in use the

technology is striving to *persuade* a group of individuals to behave in a collective way that is beneficial to the group.

The goal of a "persuasive collaborative system" is to make a group aware of discrepancies between their decision-making goals and their expressed behaviors. By making groups aware, without direct instruction, the system can persuade groups to thoroughly consider their decision approach and to alter it in productive ways.

The field of human-computer interaction frequently focuses on the usability of tools, rather than considering how every tool is a potential persuasive technology. An important concept interface designers should consider is "decision framing," from Kahneman and Tversky (Tversky 1981). Decision framing is an economic concept that explains why small changes in a description of a decision can result in a dramatic change in the choice individuals make. We make all of our decisions within an existing context and any tool or technology that is present during that decision process is contributing to that context. So using the concept of decision framing, one should consider how collaboration tools can dramatically alter a decision frame.

This research has shown that by introducing a new perspective, or new decision frame, into a group's decision-making context, the group's behaviors and decisions change. And this opens the possibility for creating fundamentally different collaboration tools that, rather than focusing on features, focus on new mechanisms for improving the decision frames that surround us.

Appendices

- Appendix A: Hardware Design
- Appendix B: Experiment Procedures Given to Subjects
- Appendix C: Experiment Questionnaires
- Appendix D: Experiment Tasks

Appendix A: Hardware Design

The follow schematic diagrams were designed and used by Katherine Hollenbach, my summer UROP, to build a custom piece off hardware to perform data collection. The board was designed to measure the voltage (volume) off of eight microphone channels and feed the data through a serial port to a computer for processing.

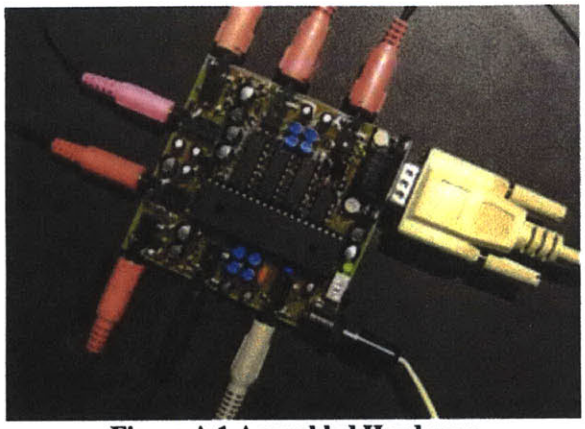


Figure A.1 Assembled Hardware

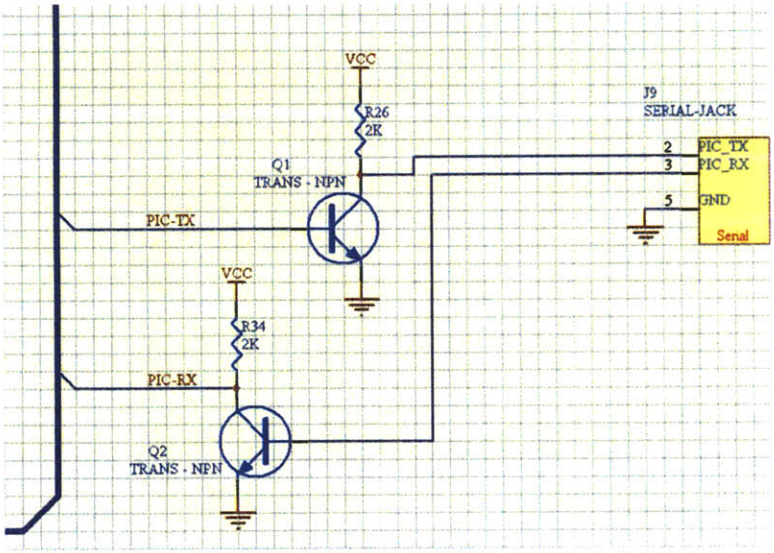


Figure A.2 Serial Schematic

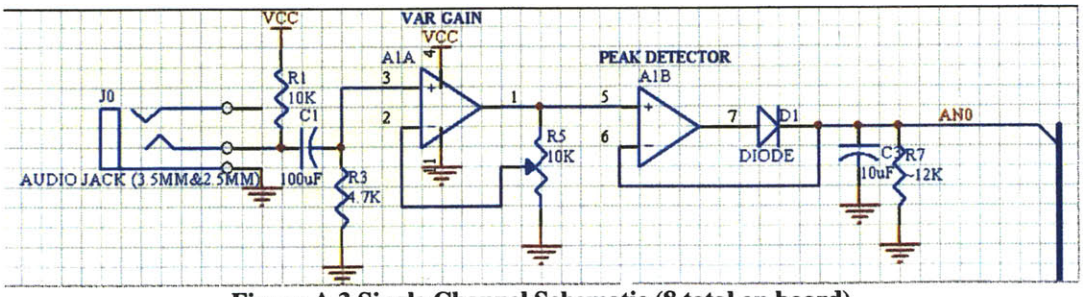


Figure A.3 Single Channel Schematic (8 total on board)

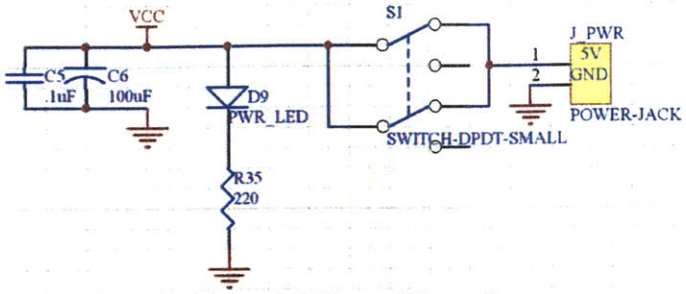


Figure A.4 Power Schematic

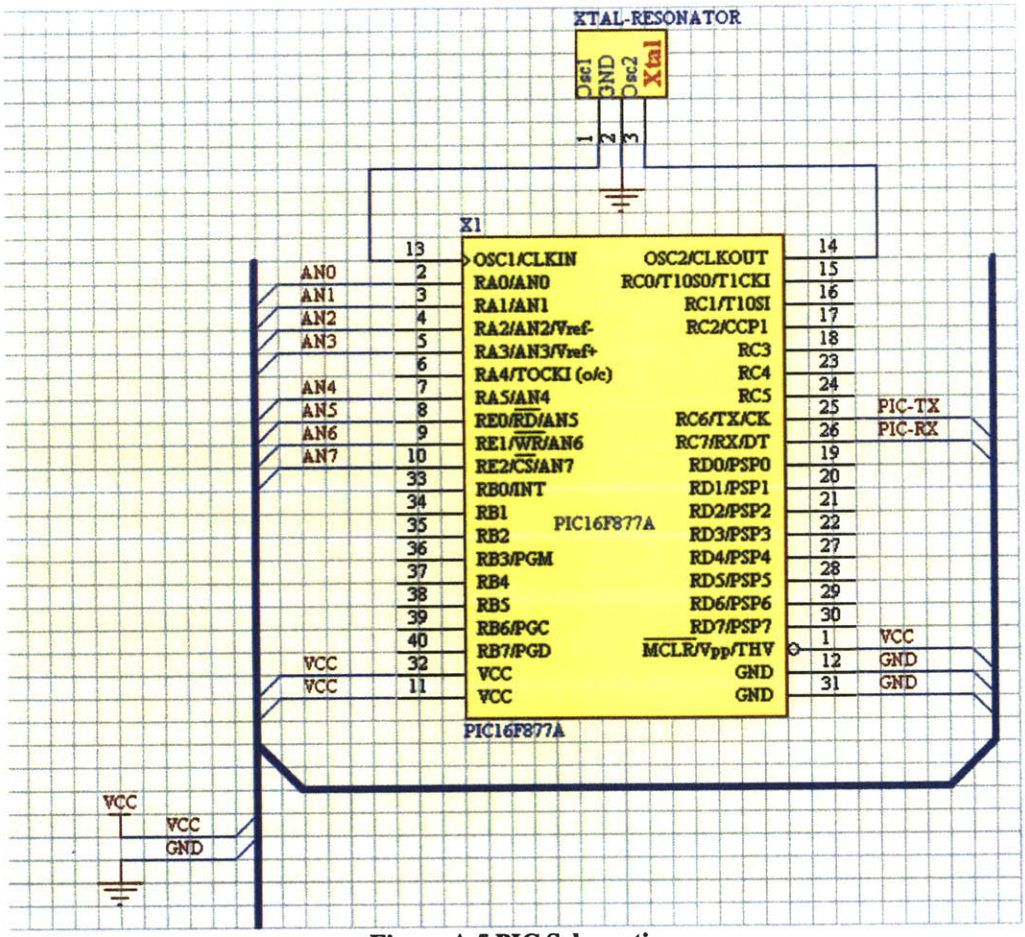


Figure A.5 PIC Schematic

Appendix B: Experimental Procedures Given to Subjects

EXPERIMENTAL PROCEDURES

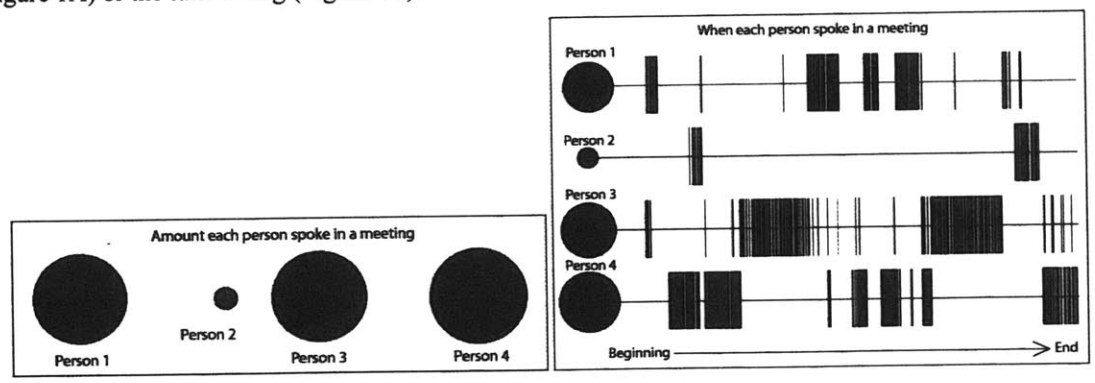
After an initial briefing and calibration, you will be asked to perform two group tasks:

1. You will be given a headset, which you should wear for the entire experiment. This headset is used to detect when you are speaking.
2. You will be given a decision task (Task 1) to complete. Read the task information and make an initial decision on your own; then discuss the decision with the group (for up to 15 minutes) and come to a group consensus.
3. When you have completed Task 1, you will be asked to fill out a questionnaire.
4. You will be given a few minutes to discuss the group's performance on Task 1.
5. You will be given a decision task (Task 2) to complete. Read the task information and make an initial decision on your own; then discuss the decision with the group (for up to 15 minutes) and come to a group consensus.
6. When you have completed Task 2, you will be asked to fill a second questionnaire.
7. You will be debriefed on the goals of the experiment and given time to ask any questions.

Background

For each of the decision tasks, you will be given information about the decision before you discuss it with the group. Before the discussion starts, you should decide which option you feel is the best one. Then you will discuss your decision with the group to make a consensus decision, completing the task. It may be the case that some people have information that others do not have. Research has shown that, when this is the case, teams perform better when each team member has an opportunity to contribute to the discussion.

Data from the headsets can be used to display the relative amount each person speaks in a discussion (Figure 1A) or the turn-taking (Figure 1B).



(A) (B)

Figure 1: (A) the amount each person spoke in a meeting; (B) a timeline of when each person spoke in a meeting

EXPERIMENTAL PROCEDURES (CON'T, CONDITION A)

You will be given access to a tool that will help you monitor whether or not the group has been overlooking someone in the discussion. The tool monitors how much each person has spoken and presents this information in real-time (while the discussion is taking place) in a display (See Figure 2).

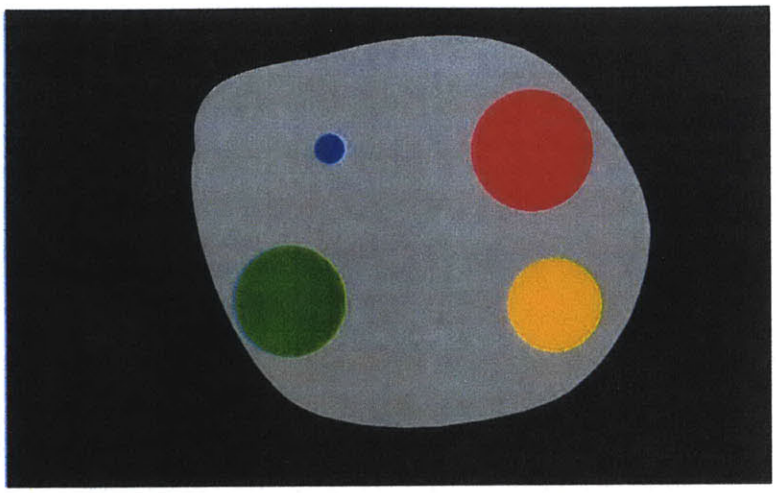


Figure 2: a display of individual participation in a group discussion

Each of you will be assigned a color circle. The circles represent the relative amount each of you has spoken. The larger the circle, the more that person has spoken. The circles grow and shrink in proportion to how much each person has spoken so far in the task.

Before begin the group discussion, we will calibrate the microphones and make sure that each of everyone is clear as to which circle represents each of you on the display.

As stated in the background section above, research has shown that teams perform better when each team member has an opportunity to contribute to the discussion. When one group member is ignored, the team may perform less optimally. For example, the group shown in Figure 3 may have missed out on Person 2's opinion.

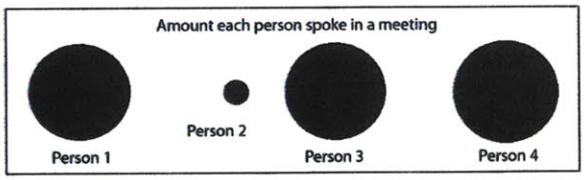


Figure 3: amount each person spoke in a meeting. Note that Person 2 spoke less than the others.

We suggest you use the tool to assist you in making sure each person has had an opportunity to be heard.

EXPERIMENTAL PROCEDURES (CON'T CONDITION, B)

You will be given access to a tool that will help you monitor whether or not the group has been overlooking someone in the discussion. The tool monitors when each person has spoken and then, after the task, you will be given time to replay the turn-taking patterns of the previous conversation (See Figure 2). Before using the display, you will be given some instructions on how to use the interface.

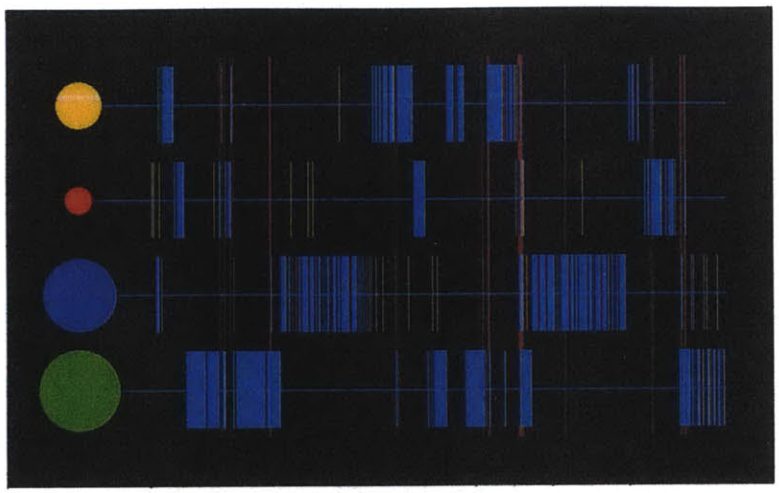


Figure 2: a timeline display of the turn-taking patterns of a group discussion. Each circle represents a person and the vertical blue bars show when that person spoke.

After each task you will be given time to discuss how you believe the group performed and you will be able to use the timeline display to replay the discussion in order to assist you in your review.

As stated in the background section above, research has shown that teams perform better when each team member has an opportunity to contribute to the discussion. When one group member is ignored, the team may perform less optimally. For example, this group shown in Figure 3 may have missed out on Person 2's opinion.

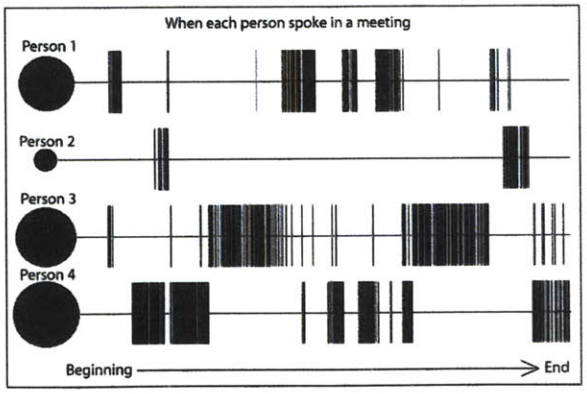


Figure 3: In this timeline, Person 2 has not spoken as much as the other subjects.

We suggest you use the tool to check to see that you gave each person an opportunity to be heard in Task 1 and use this as a way to improve your group's performance in Task 2.

EXPERIMENTAL PROCEDURES (CON'T, CONDITION AB)

You will be given access to **two tools** during the experiment that will help you monitor whether or not the group has been overlooking someone in the discussion.

The first tool monitors how much each person has spoken and presents this information in real-time (while the discussion is taking place) in a display (See Figure 2).

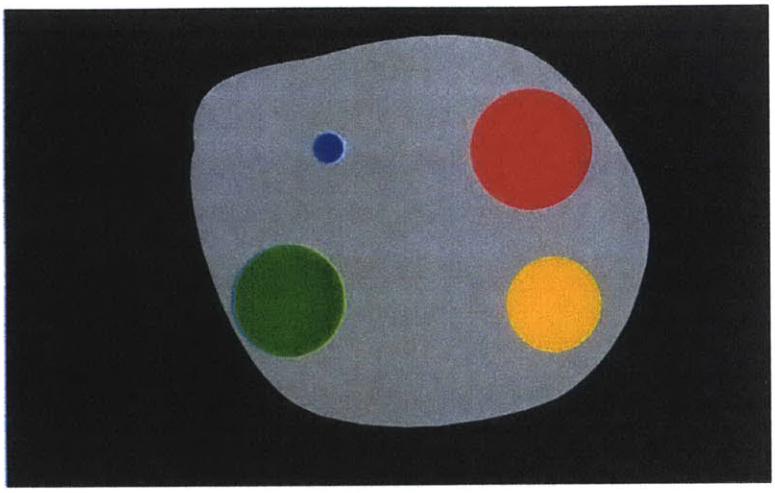


Figure 2: a display of individual participation in a group discussion

Each of you will be assigned a color circle. The circles represent the relative amount each of you has spoken. The larger the circle, the more that person has spoken. The circles grow and shrink in proportion to how much each person has spoken so far in the task.

Before begin the group discussion, we will calibrate the microphones and make sure that each of everyone is clear as to which circle represents each of you on the display.

As stated in the background section above, research has shown that teams perform better when each team member has an opportunity to contribute to the discussion. When one group member is ignored, the team may perform less optimally. For example, the group shown in Figure 3 may have missed out on Person 2's opinion.

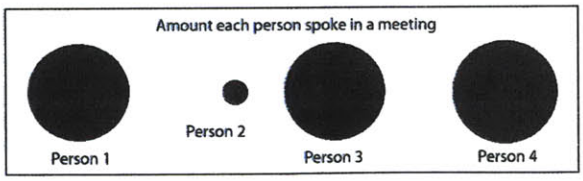


Figure 3: amount each person spoke in a meeting. Note that Person 2 spoke less than the others.

We suggest you use the tool to assist you in making sure each person has had an opportunity to be heard.

The second tool monitors when each person has spoken and then, after the task, you will be given time to replay the turn-taking patterns of the previous conversation (See Figure 4). Before using the display, you will be given some instructions on how to use the interface.

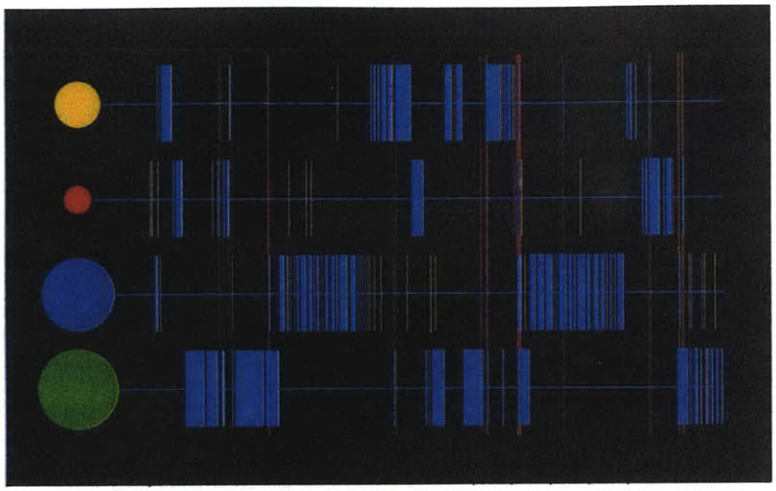


Figure 4: a timeline display of the turn-taking patterns of a group discussion. Each circle represents a person and the vertical blue bars show when that person spoke.

After each task you will be given time to discuss how you believe the group performed and you will be able to use the timeline display to replay the discussion in order to assist you in your review.

As stated in the background section above, research has shown that teams perform better when each team member has an opportunity to contribute to the discussion. When one group member is ignored, the team may perform less optimally. For example, this group shown in Figure 5 may have missed out on Person 2's opinion.

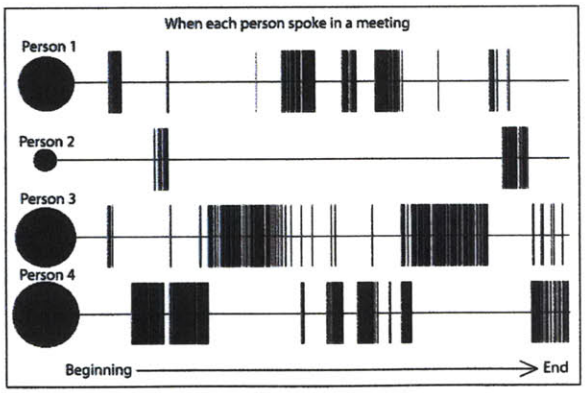


Figure 5: In this timeline, Person 2 has not spoken as much as the other subjects.

We suggest you use the tool to check to see that you gave each person an opportunity to be heard in Task 1 and use this as a way to improve your group's performance in Task 2.

Appendix C: Experiment Questionnaires

Post-Task Questionnaire

About the Task

1. **What was the group's final decision:**

   _____

2. How **difficult** was the task?

   • • • • • • •

   Very Difficult — **Very Easy**

3. How well do you think the **group performed** on the task you were given?

   • • • • • • •

   Very Poor — **Very Well**

4. How **satisfied** are you with the final decision?

   • • • • • • •

   Very Unsatisfied — **Very Satisfied**

5. How **efficiently** did the group solve the task?

   • • • • • • •

   Very Inefficiently — **Very Efficiently**

6. **How satisfied are you with the** group process **used during the task?**

   • • • • • • •

   Very Unsatisfied — **Very Satisfied**

About You

7. **How satisfied are you with the** amount you contributed **to the task?**

   • • • • • • •

   Very Unsatisfied — **Very Satisfied**

8. **How satisfied are you with the** quality of your contributions **to the task?**

   • • • • • • •

   Very Unsatisfied — **Very Satisfied**

9. **Do you wish you had spoken** more or less **during the task?**

   • • • • • • •

   Much Less — **No change** — **Much More**

10. **How** much did you speak **in relation to the others in the group?**

    • • • • • • •

    Much Less — **Same amt** — **Much More**

11. How much do you think you contributed to the group's final decision?

•	•	•	•	•	•	•
Not much			Avg amount			A lot

About the Group

12. For each other person in the group,

   a. How much **did they contribute to the task (in relation to the others)?**
   b. How valuable **was their contribution to the task (in relation to the others)?**
   c. Do you wish they had spoken more or less **during the task (in relation to the others)?**
   d. Did they exhibit leadership **qualities during the task (in relation to the others)?**

Subject Color:______________________

Quantity of contribution?	• well below average	•	•	• average	•	•	• well above average
Quality of contributions?	• well below average	•	•	• average	•	•	• well above average
Should speak more or less?	• much less	•	•	• just right	•	•	• much more
Exhibit qualities of leadership?	• none	•	•	• some	•	•	• many

Subject Color:______________________

Quantity of contribution?	• well below average	•	•	• average	•	•	• well above average
Quality of contributions?	• well below average	•	•	• average	•	•	• well above average
Should speak more or less?	• much less	•	•	• just right	•	•	• much more
Exhibit qualities of leadership?	• none	•	•	• some	•	•	• many

Subject Color:______________________

Quantity of contribution?	• well below average	•	•	• average	•	•	• well above average
Quality of contributions?	• well below average	•	•	• average	•	•	• well above average
Should speak more or less?	• much less	•	•	• just right	•	•	• much more
Exhibit qualities of leadership?	• none	•	•	• some	•	•	• many

13. For each other person in the group,

a. How **comfortable** were you collaborating with them on this task (compared with the others) ?

b. How well did you **trust** them during the task (compared with the others) ?

c. How well did they **listen** to you during the task (compared with the others) ?

Subject Color:________________________

You were comfortable collaborating:	•	•	•	•	•	•	•
	not at all			**average**			**very much**
You trusted them:	•	•	•	•	•	•	•
	not at all			**average**			**very much**
This person listened to you:	•	•	•	•	•	•	•
	not at all			**average**			**very much**

Subject Color:________________________

You were comfortable collaborating:	•	•	•	•	•	•	•
	not at all			**average**			**very much**
You trusted them:	•	•	•	•	•	•	•
	not at all			**average**			**very much**
This person listened to you:	•	•	•	•	•	•	•
	not at all			**average**			**very much**

Subject Color:________________________

You were comfortable collaborating:	•	•	•	•	•	•	•
	not at all			**average**			**very much**
You trusted them:	•	•	•	•	•	•	•
	not at all			**average**			**very much**
This person listened to you:	•	•	•	•	•	•	•
	not at all			**average**			**very much**

The Real-Time Display

14. Did you look at the display during the task? YES_____ NO_____

15. If yes, what prompted you to look at the display? (check all that apply)

- □ **To look at myself**
- □ **To look at someone else**
- □ **To see if it was working**
- □ **To check my performance relative to others**
- □ **To check the performance of others in relation to myself**
- □ **Other: _____**

16. Did you find the display distracting?

• • • • • • •

Not at all distracting **Very distracting**

17. How useful **did you find the display for completing the task?**

• • • • • • •

Not at all useful **Very useful**

18. How informative **did you find the display about the group's behavior?**

• • • • • • •

Not at all informative **Very informative**

19. Did you discuss the display **during the task?**

YES _____

NO _____

If yes, please explain what was discussed:

20. Any additional comments about the Real-Time Display:

The Replay Display

1. **How** useful **did you find the replay for reflecting on the previous task?**

   • • • • • • •

   Not at all useful — Very useful

2. **How** informative **did you find the replay about the group's behavior?**

   • • • • • • •

   Not at all informative — Very informative

3. **Are there ways you wish the group would change its interaction that were demonstrated in the replay of the meeting?**

   YES ____

   NO ____

   If yes, please explain:

4. **Any additional comments about the Replay Display:**

Post-Experiment Questionnaire

Final Demographic Questions

1. **Sex: Male _____ Female: _____**

2. **Age: _____**

3. **Ethnicity: _____**

4. **Citizenship: _____**

5. **Employment: Student: _____ Non-Student: _____**

6. **Hours online everyday (on average): _____hours**

Final Display Questions

7. **How** accurate **do you think the system was in detecting speaking frequencies and patterns?**

•	•	•	•	•	•	•
Not at all accurate						**Very accurate**

8. **How** comfortable **were you with having this information shown to the group during the task?**

•	•	•	•	•	•	•
Very uncomfortable						**Very comfortable**

9. **How** comfortable **do you think the other group members were having this information shown?**

•	•	•	•	•	•	•
Very uncomfortable						**Very comfortable**

10. **Would you want to have this type of information available in** other types of meetings **you attend?**

•	•	•	•	•	•	•
Never			**In some types of meetings**			**In all my meetings**

11. Do you have **any final comments** about the technology, regarding its appearance, its features, or your interaction with it?

Appendix D: Experiment Tasks

STUDENT SELECTION TASK

INSTRUCTIONS

- Read the following Decision Task and the information on the following pages.
- Write down your initial decision on the last page.
- Discuss the decision with the group until you reach consensus.

DECISION TASK:

MIT's incoming class for the Fall of 2005 has been chosen, with one additional spot still open. A candidate off of the waiting list must be chosen to fill this spot. Your task is to select which one of the following three waitlist candidates should be admitted into MIT's incoming class. While a student's future success as an MIT student can not be known, MIT has outlined a set of criteria for admitting people off of the waitlist.

A successful MIT student, at a minimum, graduates, doesn't fail a class, doesn't transfer to another school or leave school indefinitely, and doesn't get suspended or kicked out of MIT for academic or disciplinary reasons. An ideal MIT student participates in non-academic activities, ideally in a leadership role, receives grades of B's or above, graduates in four years, and successfully gets a job or is admitted to graduate school upon graduation

The three students you must select between are:

- **Sam Smith**
- **Adam Adams**
- **Jane Jones.**

NAME	**Sam Smith**
HOMETOWN	Dayton, Ohio
HIGH SCHOOL	Private high school
SAT VERBAL	790
SAT MATH	770
GPA	3.5 out of 4.0
COURSES	Sam is taking 4 Advanced Placement classes and 1 non-AP class.
EXTRACURRICULAR	He plays varsity football and is a member of the debate team
AWARDS	Sam won his school's science prize.
PARENTS	His father works in sales and his mother owns a bakery.
SIBLINGS	Sam's sister attends Ohio State University
PLANS TO ATTEND	Ohio State University
NOTES FROM RECOMMENDATIONS	
.	Previous students from his school who attended MIT did very well academically and socially
	(HIDDEN FACT) He is undecided in his major.
	(HIDDEN FACT) Sam's football coach says he is a very talented player and could compete at the college level.
	(HIDDEN FACT) A note on his transcript indicates that Sam was suspended from school for 3 days for consuming alcohol

	during a school field trip.
	(HIDDEN FACT) His drama teacher says in his recommendation that Sam has shown a natural talent for acting.
NAME	**Adam Adams**
HOMETOWN	Los Angeles, CA
HIGH SCHOOL	Large public high school, middle-to-low income neighborhood
SAT VERBAL	600
SAT MATH	680
GPA	3.2 out of 4.0
COURSES	He is taking five regular (non advanced-placement) courses this term.
EXTRACURRICULAR	He is a member of the engineering club.
AWARDS	**(HIDDEN FACT)** He won first place in the annual state-wide science fair
PARENTS	Divorced. Mother works in retail. Father works in construction.
SIBLINGS	His brother attends UCLA
PLANS TO ATTEND	Carnegie Mellon University
NOTES FROM RECOMMENDATIONS	
	No one from his high school has ever applied to MIT.
	His father holds a vocational certification and his mother completed high school.
	(HIDDEN FACT) He intends to major in electrical engineering.
	(HIDDEN FACT) He worked 2 summers at a electrical engineering firm.
	(HIDDEN FACT) His summer boss writes that Adam is extremely dedicated and talented employee.
NAME	**Jane Jones**
HOMETOWN	Newton, Massachusetts
HIGH SCHOOL	Large high school, high-to-middle income neighborhood
SAT VERBAL	680
SAT MATH	750
GPA	3.8 out of 4.0
COURSES	She is taking 3 AP classes and 1 non-AP class.
EXTRACURRICULAR	
	Jane took classes at her local community college during the summers.
	She is a member of the yearbook staff and the honor society.
	(HIDDEN FACT) She volunteers at her church's soup kitchen.
PARENTS	Her father is an engineer and her mother is a housewife.
SIBLINGS	Jane is the eldest of four children
PLANS TO ATTEND	Tufts University
NOTES FROM RECOMMENDATIONS	
	Previous students from her school who attended MIT had some academic difficulties.
	(HIDDEN FACT) Plans to major in math.
	(HIDDEN FACT) Jane's recommendations indicate that her teachers do not know her very well.
	(HIDDEN FACT) She has only taken 4 courses each term of high school, instead of the average 5.

STORE SELECTION TASK

INSTRUCTIONS

1. Read the following Decision Task and the information on the following pages.
2. Write down your initial decision on the last page.
3. Discuss the decision with the group until you reach consensus.

DECISION TASK:

A 24-hour convenient store is opening a new location and needs to decide between three locations. The possible locations are three different squares in Boston. Your task is to evaluate the different aspects of these three locations and choose the best one for the business. While the financial success of the convenient store cannot be predicted for certain, there are several criteria upon which you should make your choice.

The success of convenient store depends entirely on its annual revenue. Many factors impact a store's revenue: the type of goods sold, the number of customers, the volume of sales, rent and taxes.

The three store locations you must select between are

- **Dewey Square**
- **Elliot Square**
- **Fredrick Square.**

NAME	**Dewey Square**
SUBWAY	3 blocks from the Dewey T stop.
BUS LINES	2 bus lines pass by this location
DESCRIPTION	This is a small business district with a small number of sandwich shops, 1 restaurant, and 1 drug store.
SALES & EXPENSES	At this location, the store would offer the usual things (snack food, cold beverages, and magazines) and a few items not offered in the area at night (deli sandwiches, soups, hot beverages). **(HIDDEN FACT)** Businesses within university buildings qualify for tax exemptions (estimated to be equivalent to a 2% increase in revenue). **(HIDDEN FACT)** According to market research, similar businesses in the area take in sales of approximately 4 times their rental costs. **(HIDDEN FACT)** Rent is $3,000/month.
COMPETITION	There are 3 liquor stores nearby that sell similar products (tobacco, cold beverages, lottery tickets). There are 2 existing 24-hour convenience stores within walking distance.
POPULATION	The area has 3,200 residents.
DEMOGRAPHICS	Half of the residents in the area are students. The other half is middle-to-lower income. There is a homeless shelter 2 blocks away with 130 beds. There is 1 major hotel in the area.
CRIME	There have been 2 armed robberies in the area in the past 5 years. **(HIDDEN FACT)** The nearby university police patrol the area frequently because the store is within a university-owned building.
NAME	**Elliot Square**
SUBWAY	Directly next to the T stop.
BUS LINES	Elliot Square is a transfer point for 6 different bus routes.
DESCRIPTION	The neighborhood is an active business area with many restaurants, bars, and night clubs.
SALES & EXPENSES	At this location, the store will sell snack food, cold beverages, and magazines.

	(HIDDEN FACT) There are be no foreseeable increases in taxes in this district in the next 5 years. **(HIDDEN FACT)** According to market research, similar businesses in the area take in sales of approximately 4 times their rental costs. **(HIDDEN FACT)** Rent is $1,900/month
COMPETITION	All competing stores in the area close by 10pm. There is 1 nearby gas station, open 24-hours, selling some of the same products.
POPULATION	There are 5,500 residents in the area.
DEMOGRAPHICS	Most of the residents in the area are young professionals. There is a baseball stadium 4 blocks away that brings spectators to the area during baseball season. There are 2 major business hotels nearby.
CRIME	There have been 3 armed robberies in the area within the past 2 years. **(HIDDEN FACT)** In the surrounding blocks, street-level disturbances require police calls.
NAME	**Fredrick Square**
SUBWAY	1 block from the Fredrick Square T stop
BUS LINES	There are 4 bus routes passing by this location. There is a taxi stand within 1000 feet of the store location.
DESCRIPTION	This neighborhood is a business district with small restaurants, neighborhood bars, small coffee shops.
SALES & EXPENSES	This store would sell an extensive selection of items typical of other stores in the area (diary products, newspapers, frozen food, tobacco, lottery tickets, snack food, and small souvenirs). **(HIDDEN FACT)** The taxes on business owners in this location will be increasing in the next year, effectively reducing the revenue of the store by 3.1% **(HIDDEN FACT)** According to market research, similar businesses in the area take in sales of approximately 3 times their rental costs. **(HIDDEN FACT)** Rent is $2,400/month.
COMPETITION	There are 3 drug stores in the area selling similar products that close at 9pm.
POPULATION	There are 7,800 residents in the area.
DEMOGRAPHICS	Most of the neighboring residents are students. There are no major hotels in the area. During the day there is a shuttle from this location to a local mall and movie theatre.
CRIME	The neighborhood police patrol this area. **(HIDDEN FACT)** There have only been minor incidents of crime in the area in the past 5 years.

Bibliography

Abowd, G.D. "Classroom 2000: An Experiment with the Instrumentation of a Living Educational Environment." *IBM Systems Journal, Special issue on Pervasive Computing*, vol. *38*, no. 4, pp. 508-530, 1999.

Anderson, A., O'Malley, C., Doherty-Sneddon, G., Langton, S., Newlands, A., Mullin, J., Fleming, A. and Velden, V. "The Impact of VMC on Collaborative Problem Solving," in Finn, K., Sellen, A. and Wilbur, S. eds. *Video-Mediated Communication.* Lawrence Erlbaum Associates, Hillsdale, NJ, 133–155, 1997.

Argyle, M. *Bodily Communication.* International Universities Press, Inc, Madison, CT, 1988.

Bales, R.F. *Interaction process analysis: a method for the study of small groups.* Addison-Wesley, Reading, MA, 1950.

Bales, R.F. *Social Interaction Systems: Theory and Measurement.* Transaction Publishers, New Brunswick, NJ, 1999.

Bales, R.F. and Cohen, S.P. *SYMLOG: A System for the Multiple Level Observation of Groups.* Free Press, New York, 1979.

Basu, S., Choudhury, T., Clarkson, B. and Pentland, A. "Towards Measuring Human Interactions in Conversational Settings," In *Proceedings of the IEEE Int'l Workshop on Cues in Communication (CUES 2001) at CVPR 2001* (Kauai, Hawaii). 2001.

Beenen, G., Ling, K., Wang, X., Chang, K., Frankowski, D., Resnick, P. and Kraut, R.E. "Using social psychology to motivate contributions to online communities," In *Proceedings of the ACM Conference on Computer Supported Cooperative Work* (Chicago, IL). pp. 212-221, 2004.

Begeman, M., Cook, P., Ellis, C., Graf, M., Rein, G. and Smith, T. "Project Nick: meetings augmentation and analysis," In *Proceedings of the ACM Conference on Computer Supported Cooperative Work.* 1986.

Bettenhausen, K. and Murnighan, J.K. "The emergence of norms in competitive decision-making groups." *Administrative Science Quarterly*, vol. *30*, no., 350–372, 1985.

Bradner, E., Kellogg, W.A. and Erickson, T. "Babble: Supporting Conversation in the Workplace." *SIGGROUP Bulletin*, vol. *19*, no. 3, pp. 8-9, 1998.

Brave, S. and Dahley, D. "inTouch: A Medium for Haptic Interpersonal Communication," In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI 97)* (Atlanta, GA). 363–364, 1997.

Bray, R.M., Johnson, D. and Chilstrom, J.T.J. "Social Influence by Group Members with Minority Opinions: A Comparison of Hollander and Moscovici." *J. of Personality and Social Psychology*, vol. *43*, no. 78-88, 1982.

Brooks, R.A. "The Intelligent Room Project," In *Proceedings of the Second International Cognitive Technology Conference (CT'97)* (Aizu, Japan). 1997.

Brown, R. "Group Polarization," in Brown, R. ed. *Social Psychology*. Free Press, New York, 200–248, 1986.

Burger, S., MacLaren, V. and Yu, H. "The ISL Meeting Corpus: The Impact of Meeting Type on Speech Style," In *Proceedings of the ICSLP-2002*. 2002.

Carver, C.S. and Scheier, M.F. *On the Self-Regulation of Behavior*. Cambridge University Press, Cambridge, UK, 1998.

Chen, M. "Visualizing the Pulse of a Classroom," In *Proceedings of the ACM International Conference on Multimedia (MM'03)* (Berkeley, CA). 2003.

Choudhury, T. and Pentland, A. "Modeling Face-to-Face Communication using the Sociometer," In *Proceedings of the International Conference on Ubiquitous Computing* (Seattle, WA). 2003.

Cook, P., Ellis, C., Graf, M., Rein, G. and Smith, T. "Project Nick: meetings augmentation and analysis." *ACM Transactions on Information Systems (TOIS)*, vol. *5*, no. 2, 1987.

Daft, R. and Lengel, R. "Organizational Information Requirements, Media Richness and Structural Design." *Management Science*, vol. *32*, no. 5, 554 - 571, 1986.

Davis, J., Zaner, M., Farnham, S., Macjan, C. and McCarthy, B. "Wireless Brainstorming: Overcoming Status Effects in Small Group Decisions," In *Proceedings of the Hawaii International Conference on System Sciences (HICSS'03)* (Hawaii). 2003.

Delbecq, A.L., Ven, A.H.V.d. and Gustafson, D.H. *Group Techniques for Program Planning: a guide to nominal group and delphi processes*. Scott, Foresman and Company, Glenview, IL, 1975.

Dennis, A.R., Hilmer, K.M. and Taylor, N.J. "Information Exchange and Use in GSS and Verbal Group Decision Making: Effects of Minority Influence." *J. of Management Information Systems*, vol. *14*, no. 3, 61-88, 1998.

DeSanctis, G. and Gallupe, R.B. "Group Decision Support Systems: A New Frontier." *DATA BASE*, vol. *16*, no. 2, 3-10, 1985.

DiMicco, J.M. and Bender, W. "Second Messenger: Increasing the Visibility of Minority Viewpoints with a Face-to-face Collaboration Tool," In *Proceedings of the ACM Conference on Intelligent User Interfaces (IUI'04)* (Funchal, Madeira, Portugal). 2004a.

DiMicco, J.M., Lakshmipathy, V. and Fiore, A.T. "Conductive Chat: Instant Messaging With a Skin Conductivity Channel," In *Proceedings of the ACM Conference on Computer Supported Cooperative Work (CSCW02)* (New Orleans, LA). 2002.

DiMicco, J.M., Pandolfo, A. and Bender, W. "Influencing Group Participation with a Shared Display," In *Proceedings of the ACM Conference on Computer Supported Cooperative Work (CSCW04)* (Chicago, IL). 2004b.

Donath, J., Karahalios, K. and Viégas, F. "Visiphone: Connecting Domestic Spaces with Audio," In *Proceedings of the International Conference on Auditory Display (ICAD)* (Atlanta, GA). 2000.

Dubrovsky, V.J., Kiesler, S. and Sethna, B.N. "The Equalization Phenomenon: Status Effects in Computer-Mediated and Face-to-Face Decision-Making Groups." *Human-Computer Interaction*, vol. *6*, no., 119-146, 1991.

Eagle, N., Singh, P. and Pentland, A. "Common Sense Conversations: Understanding Casual Conversation using a Common Sense Database," In *Proceedings of the Artificial Intelligence, Information Access, and Mobile Computing Workshop at IJCAI* (Acapulco, Mexico). 2003.

Erickson, T. and Kellogg, W.A. "Social translucence: an approach to designing systems that support social processes." *ACM Transactions on Computer-Human Interaction (TOCHI)*, vol. *7*, no. 1, 59-83, 2000.

Erol, B., Lee, D.-S. and Hull, J.J. "Multimodal Summarization of Meeting Recordings," In *Proceedings of the IEEE International Conference on Multimedia & Expo*. 2003.

Fay, N., Garrod, S. and Carletta, J. "Group Discussion as Interactive Dialogue or as Serial Monologue: The Influence of Group Size." *American Psychological Society*, vol. *11*, no. 6, pp. 481-461, 2000.

Fogg, B.J. *Persuasive Technology: Using Computers to Change What We Think and Do*. Morgan Kaufmann Publishers, Boston, 2003.

Fry, B. and Reas, C. "Processing Programming Language, Massachusetts Institute of Technology and Interaction Design Institute Ivrea, www.processing.org, 2004 - 2001.

Galley, M., McKeown, K., Fosler-Lussier, E. and Jing, H. "Discourse Segmentation of Multi-Party Conversation," In *Proceedings of the ACL*. 2003.

Gersick, C. and Hackman, R. "Habitual routines in task-performing groups." *Organizational Behavior and Human Decision Processes*, vol. *47*, no., 67-97, 1990.

Greenspan, S., Goldberg, D., Weimer, D., and Basso, A. "Interpersonal Trust and Common Ground in Electronically Mediated Communication," In *Proceedings of the Conference on Computer-Supported Cooperative Work* (Philadelphia, PA). 2000.

Griffin, E. "Groupthink of Irving Janis," in *A First Look at Communication*. McGraw-Hill, Inc, New York, 1997.

Grosz, B. and Sidner, C. "Attention, Intentions, and the Structure of Discourse." *Computational Linguistics*, vol. *12*, no. 3, 175-204, 1986.

Hackman, J.R. "The Design of Work Teams," in Lorsch, J.W. ed. *Handbook of Organization Behavior*. Prentice Hall, Englewood Cliffs, NJ, 1987.

Hackman, J.R. "Group influences on individuals in organizations," in Dunnette, M.D. and Hough, L.M. eds. *Handbook of Industrial and Organizational Psychology*, 1992.

Hillard, D., Ostendorf, M. and Shriberg, E. "Detection of Agreement vs. Disagreement in Meetings: Training with Unlabeled Data," In *Proceedings of the HLT-NAACL Conference* (Edmonton, Canada). 2003.

Hollan, J. and Stornetta, S. "Beyond Being There," In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI 92)* (Monterey, California). 119-125, 1992.

Hollingshead, A.B. "Information Suppression and Status Persistence in Group Decision Making - the Effects of Communication Media." *Human Computer Research*, vol. *23*, no. 2, 193-219, 1996.

Huang, A., Russell, D. and Sue, A. "IMHere: Public Instant Messaging on Large, Shared Displays for Workgroup Interactions," In *Proceedings of the Conference on Human Factors in Computing Systems (CHI2004)* (Vienna, Austria). ACM Press, 2004.

Hudson, J.M. and Bruckman, A. "The Bystander Effect: A Lens for Understanding Patterns of Participation." *Journal of the Learning Sciences*, vol. *13*, no. 2, pp. 169-199, 2004.

Isaacs, E. and Tang, J.C. "What Video Can and Cannot Do for Collaboration." *Multimedia Systems*, vol. 2, no., 63-73, 1994.

Ishii, H., Kobayashi, M. and Grudin, J. "Integration of interpersonal space and shared workspace: ClearBoard design and experiments." *ACM Transactions on Information Systems*, vol. *11*, no. 4, 349-375, 1993.

Janis, I.L. *Groupthink: Psychological studies of policy decisions and fiascos.* Houghton Mifflin, Boston, 1982.

Jebara, T., Ivanov, Y., Rahimi, A. and Pentland, A. "Tracking Conversational Context for Machine Mediation of Human Discourse," In *Proceedings of the AAAI Fall 2000 Symposium.* 2000.

Jermann, P.R. *Computer Support for Interaction Regulation in Collaborative Problem-Solving.* PhD Thesis, *Faculté de Psychologie et des Science de l'Education,* l'Université de Genève, Geneva, 2004.

Johanson, B., Fox, A. and Winograd, T. "The Interactive Workspaces Project: Experiences with Ubiquitous Computing Rooms *IEEE Pervasive Computing Magazine*, 2002.

Kazman, R., Hung, W. and Mantei, M. "Dynamic Meeting Annotation and Indexing," In *Proceedings of the 1995 Pacific Workshop on Distributed Multimedia Systems* (Honolulu, HI). 11-18, 1995.

Kiesler, S. and Sproull, L. "Group Decision Making and Communication Technology." *Organizational Behavior and Human Decision Processes*, vol. *52*, no., 96-123, 1992.

Kristjansson, T., Ramesh, P., Huang, T.S. and Juang, B.H. "A Unified Structure-Based Framework for Indexing and Gisting of Meetings," In *Proceedings of the IEEE International Conference on Multimedia Computing and Systems (ICMCS)* (Florence, Italy). 572-577, 1999.

Kurlander, D., Skelly, T. and Salensin, D. "Comic Chat," In *Proceedings of the ACM SIGGRAPH Conference.* 1996.

Lochbaum, K. "A Collaborative Planning Model of Intentional Structure." *Computational Linguistics*, vol. *24*, no. 4, 525-572, 1998.

Losada, M., Sanchez, P. and Noble, E.E. "Collaborative technology and group process feedback: their impact on interactive sequences in meetings," In *Proceedings of the Conference on Computer Supported Cooperative Work* (Los Angeles). 1990.

Maier, N.R. "Assets and liabilities in group problem solving: The need for an integrative function." *Psychological Review*, vol. *74*, no., 239-249, 1967.

Mandryk, R.L. and Inkpen, K.M. "Physiological Indicators for the Evaluation of Co-located Collaborative Play," In *Proceedings of the Conference on Computer Supported Cooperative Work* (Chicago, IL). 2004.

Mantei, M. "Capturing the Capture Lab concepts: A case study in the design of computer supported meeting environments," In *Proceedings of the Computer Supported Cooperative Work (CSCW'88).* 1988.

McCowan, I., Gatica-Perez, D., Bengio, S., Lathoud, G., Barnard, M. and Zhang, D. "Automatic Analysis of Multimodal Group Actions in Meetings." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. *27*, no. 3 (March 2005), p. 305-317, 2005.

McGrath, J.E. *Groups: Interaction and Performance.* Prentice-Hall, Inglewood, NJ, 1984.

McLeod, P.L., Baron, R.S., Marti, M.W. and Yoon, K. "The Eyes Have It: Minority Influence in Face-To-Face and Computer-Mediated Group Discussion." *J. of Applied Psychology*, vol. *82*, no. 5, 706-718, 1997.

McNeill, D. *Hand and Mind: What Gestures Reveal about Thought.* The University of Chicago Press, Chicago, 1992.

Morgan, N., Baron, D., Bhagat, S., Carvey, H., Dhillon, R., Edwards, J., Gelbart, D., Janin, A., Krupski, A., Peskin, B., Pfau, T., Shriberg, E., Stolcke, A. and Wooters, C. "Meetings About Meetings:

Research at ICSI on Speech in Multiparty Conversations," In *Proceedings of the ICASSP-2003* (Hong Kong). 2003.

Moscovici, S. and Zavalloni, M. "The Group as a Polarizer of Attitudes." *J. of Personality and Social Psychology*, vol. *12*, no., 125-135, 1969.

Mota, S. and Picard, R.W. "Automated Posture Analysis for Detecting Learner's Interest Level," In *Proceedings of the 2003 Conference on Computer Vision and Pattern Recognition Workshop* (Madison, Wisconsin). 2003.

Myers, D.G. and Bishop, G.D. "The Enhancement of Dominant Attitudes in Group Discussion." *J. of Personality and Social Psychology*, vol. *20*, no., 385-391, 1971.

Nemeth, C.J. and Staw, B.M. "The tradeoffs of social control and innovation in groups and organizations." *Advances in Experimental Social Psychology*, vol. *22*, no., 1989.

Nunamaker, J.F., Dennis, A.R., Valacich, J.S., Vogel, D.R. and George, J.F. "Electronic Meeting Systems To Support Group Work." *Communications of the ACM*, vol. *34*, no. 7, 40-61, 1991.

O'Grady, T. and Greenberg, S. "A Groupware Environment for Complete Meetings," In *Proceedings of the Conference on Human Factors in Computing Systems (CHI'94)* (Boston, MA). 1994.

O'Malley, C., Langton, S., Anderson, A., Doherty-Sneddon, G. and Bruce, V. "Comparison of Face-to-Face and Video Mediated Interaction." *Interacting with Computers*, vol. *8*, no. 2, 177-192, 1996.

Ochsman, R.B. and Chapanis, A. "The Effects of 10 Communication Modes on the Behavior of Teams During Co-operative Problem-Solving." *International J. of Man-Machine Studies*, vol. *6*, no., 579-619, 1974.

Olson, G.M. and Olson, J.S. "Distance Matters." *Human-Computer Interaction*, vol. *15*, no. 2/3, 139-178, 2000.

Olson, M. *Technological Support for Work Group Collaboration.* Lawrence Erlbaum Associates, Publishers, Hillsdale, NJ, 1989.

Paese, P.W., Beiser, M. and Tubbs, M.E. "Framing Effects and Choice Shifts in Group Decision Making." *Organizational Behavior and Human Decision Processes*, vol. *56*, no., 149-165, 1993.

Parker, K. "Speaking Turns in Small Group Interaction: A Context-Sensitive Event Sequence Model." *Journal of Personality and Social Psychology*, vol. *54*, no. 6, pp. 965-971, 1988.

Picard, R.W. *Affective Computing.* MIT Press, Cambridge, MA, 1997.

Picard, R.W. and Scheirer, J. "The Galvactivator: A Glove that Senses and Communicates Skin Conductivity," In *Proceedings of the 9th International Conference on Human-Computer Interaction* (New Orleans). 2001.

Prante, T., Magerkurth, C. and Streitz, N. "Developing CSCW tools for Idea Finding - Empirical Results and Implications for Design," In *Proceedings of the Conference on Computer-Supported Cooperative Work* (New Orleans, Louisiana, USA). ACM Press, 106-115, 2002.

Raven, B.H. "Groupthink, Bay of Pigs, and Watergate Reconsidered." *Organizational Behavior and Human Decision Processes*, vol. *73*, no. 2/3, 1998.

Reid, F., Ball, L., Morley, A. and Evans, J. "Styles of Group Discussion in Computer-Mediated Decision Making." *British J. of Social Psychology*, vol. *36*, no. 3, 241-262, 1997.

Rekimoto, J., Ayatsuka, Y., Uoi, H. and Arai, T. "Adding another communication channel to reality: an experience with a chat-augmented conference," In *Proceedings of the Conference on Human Factors in Computing Systems (CHI 98)* (Los Angeles, CA). 271 - 272, 1998.

Rice, R. "Task Analyzability, Use of New Media, and Effectiveness: A multi-site exploration of media richness." *Organization Science*, vol. *3*, no., 475-500, 1992.

Rocco, E. "Trust Breaks Down in Electronic Contexts but Can be Repaired by Some Initial Face-to-Face Contact," In *Proceedings of the Conference on Human Factors in Computing Systems (CHI98)* (Los Angeles). 496-502, 1998.

Rosenberger, T.M. and Smith, B.K. "Fugue: A Conversational Interface that Supports Turn-Taking Coordination," In *Proceedings of the 32nd Hawaii International Conference on Systems* (Hawaii). 1998.

Russell, D.M., Streitz, N.A. and Winograd, T. "Building disappearing computers." *Communications of the ACM*, vol. *48*, no. 3, pp. 42-48, 2005.

Sack, W. "What Does a Very Large-Scale Conversation Look Like?," In *Proceedings of the ACM SIGGRAPH 2001* (Los Angeles). ACM Press, 2001.

Selker, T., Lockerd, A., Martinez, J. and Burleson, W. "Eye-aRe, a Glasses-Mounted Eye Motion Detection Interface," In *Proceedings of the Conference on Human factors in Computing Systems (CHI01).* 2001.

Siegel, J., Dubrovsky, V., Kiesler, S. and McGuire, T.W. "Group processes in computer-mediated communication." *Organizational Behavior and Human Decision Processes*, vol. *37*, no., 1986.

Singh, P. "Open Mind Common Sense, MIT http://commonsense.media.mit.edu, 2003.

Smith, M. "Netscan: A tool for measuring and mapping social cyberspaces *http://netscan.research.microsoft.com*, 2001.

Sproull, L. and Kiesler, S. *Connections: New Ways of Working in the Networked Organization.* MIT Press, Cambridge, MA, 1991.

Stasser, G. and Stewart, D. "Discovery of Hidden Profiles by Decision-making Groups: Solving a problem versus making a Judgment." *J. of Personality and Social Psychology*, vol. *63*, no., 426-434, 1992.

Stasser, G. and Taylor, L.A. "Speaking Turns in Face-To-Face Discussions." *J. of Personality and Social Psychology*, vol. *60*, no. 5, 675-684, 1991.

Stasser, G. and Titus, W. "Effects of Information Load and Percentage of Shared Information on the Dissemination of Unshared Information during Group Discussion." *J. of Personality and Social Psychology*, vol. *53*, no., 81-93, 1987.

Stefik, M., Foster, G., Bobrow, D., Kahn, K., Lanning, S. and Suchman, L. "Beyond the chalkboard: computer support for collaboration and problem solving in meetings." *Communications of the ACM*, vol. *30*, no. 1, 32-47, 1987.

Stewart, D.D. and Stasser, G. "The Sampling of Critical Unshared Information in Decision-Making Groups: The Role of an Informed Minority." *European J. of Social Psychology*, vol. *23*, no., 95-113, 1998.

Tversky, A. and Kahneman, D. "The Framing of Decisions and the Psychology of Choice." *Science*, vol. *211*, no., 453-458, 1981.

Vemuri, S., DeCamp, P., Bender, W. and Schmandt, C. "Improving Speech Playback Using Time-Compression and Speech Recognition," In *Proceedings of the Conference on Human Factors in Computing Systems (CHI 2004)* (Vienna, Austria). 2004.

Viégas, F. and Donath, J. "Chat Circles," In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI 99)* (Pittsburgh, PA). 1999.

Whyte, G. "Decision failures: Why they occur and how to prevent them." *Academy of Management Executive*, vol. *5*, no., 23-31, 1991.

Wittenbaum, G.M., Hollingshead, A.B. and Botero, I.C. "From Cooperative to Motivated Information Sharing in Groups: Moving Beyond the Hidden Profile Paradigm." *Communication Monographs*, vol. *17*, no. 3, 2004.