Influencing Group Participation with a Shared Display

Joan Morris DiMicco, Anna Pandolfo, Walter Bender

MIT Media Lab
E15-320, 20 Ames St.
Cambridge, MA 02139 USA
{joanie, ap, walter}@media.mit.edu

ABSTRACT

During face-to-face interactions, groups frequently overly rely on the dominant viewpoint to lead the group in its decision-making process. We begin with a discussion of this phenomenon and the possibility for technology to assist in addressing it. We then present findings from a behavioral study that examines how a shared display of individual speaker-participation rates can impact the behavior of the group during a collaboration task. The results from the study indicate that the presence of such a display influences the behavior of group participants in the extremes of over and under participation. While influencing the quantity of time someone speaks is not directly equivalent to influencing the topics discussed, we suggest that this approach of providing peripheral displays of social information is promising for improving certain types of group interactions.

Categories and Subject Descriptors

H.5.3 [**Information Interfaces and Presentation**]: Group and Organization Interfaces – C*omputer-supported cooperative work, Evaluation/methodology;*
J.4 [**Computer Applications**]: Social and Behavioral Sciences – *Psychology*

General Terms

Experimentation, Human Factors.

Keywords

CSCW, collocated collaboration, speaker participation, public displays, behavioral experiment.

1. INTRODUCTION

We are interested in building interfaces that assist groups in improving their interaction processes. Our 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. Our interfaces 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.

CSCW'04, November 6–10, 2004, Chicago, Illinois, USA.

The focus of this paper is on the evaluation of one such interface. We hypothesized that shared displays within a group setting would impact the behavior of a group, and we wanted to gain a more complete understanding of how and when some specific behavioral hypotheses applied. For the evaluation, we examined the behavior of groups with and without the presence of a display of social information. In our scenario, the display continuously revealed how much each person in the group had participated in the conversation, in relation to the other group members. We hypothesized that this information would encourage under-participators to speak more and over-participators to speak less, based on the theory that social pressures would encourage a group norm of equal participation [1, 10], yet we did not know what other reactions (emotional or attentive) participants might have to this interface.

We begin this paper with a discussion of group dynamics and how technology may work towards improving group processes. We continue by describing the design of the behavioral experiment and a discussion of the results. We conclude by discussing the larger issues of what shared displays of social information can and should reveal about group interaction for the purpose of improving group processes.

2. GROUP DECISION-MAKING

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 scenario of a 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 [27, 28]. 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 [2, 10, 12, 18, 29]. 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.

We are 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, we hypothesize that technologically-enhanced groups will have the ability to make more effective, higher-quality decisions.

3. PRIOR RESEARCH

Beginning nearly two decades ago, researchers have worked on integrating technology into our group decision-making processes (for some examples see: [6, 19, 25]). The majority of the studies examining the impact of technology on decision-making have compared face-to-face groups to groups that communicate via text through computer interfaces similar to today's most basic email, chat, and instant-messaging applications. Most of the experimental analysis of group behavior has 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 [11, 14, 16, 23–25]

One of the declared benefits of mediated communication over face-to-face is that mediation results in more equal participation across group members [8, 24]. 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 [4] and later expanded by Reid, et al. [21], 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.

Advances in sensing technology and networked applications 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 [17] and affect signals with basic sensors [20]. 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 [22] and conference call applications, such as the Jabber project [13], can capture and present context-sensitive information as the group converses over the network.

Motivated by this increased facility to monitor and simultaneously present context-sensitive information to a group, we are 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.

4. A SHARED DISPLAY

In designing interfaces for group reflection, our goal is to present information that encourages a group to increase the breadth of its discussion so that more viewpoints are considered, with the consequence of increasing the quality of the group decision [12, 29]. Our first work in this area focused on capturing spoken content using voice-recognition software and presenting information to the group within an interactive idea-capturing tool [7].

To understand how a display impacts group decision-making processes, we took a step back and designed a simpler interface. We removed the content-rich speech-recognizer input and replaced it with a simple detection of how much each person had spoken during the meeting. While we see great potential for information-rich interfaces to augment face-to-face meetings, we decided it was valuable to first understand how simpler interfaces in a controlled group setting influenced behavior.

4.1 A Display of Speaker Participation

Our simpler application consists of a shared display that shows how much each person has spoken during a meeting in relation to the others. Shown in Figure 1, the display has a dynamically adjusting histogram and, along the top of the screen, a visual representation of who has spoken over the previous 30 seconds. 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. Each bar is numbered to assist users in identifying their bars.

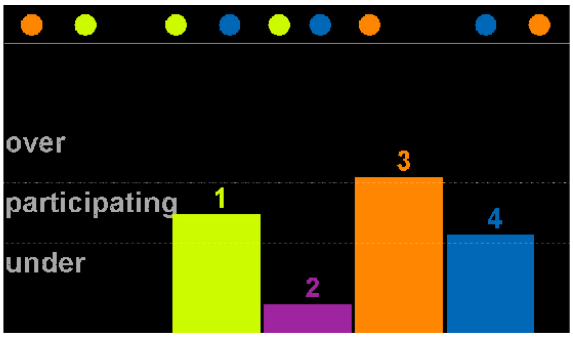


Figure 1. Speaker participation interface.

To detect speaking time, each user of the system wears a microphone and a client application performs sound level detection. As an approximation of word counting, the application 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 quick "oh's" and "um's" and other isolated non-lexical vocalizations from being interpreted as speech. We have found that this interpretation of speech

provides an accurate reflection of how much time each person spends speaking in a meeting in relation to the others. An additional benefit of having this speech detection done across client machines is that we are able to accurately interpret simultaneous speech as such.

This interface is meant to be a tool for individual reflection and also to imply a standard of acceptable 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, our choice of words was deliberate to clarify our intended message.

4.2 Information Sharing & Speaker Participation

Equal participation is not necessarily ideal for all types of meetings, but when meetings are focused on collaborative decision-making, particularly when information must be exchanged between group members in order to make the decision, 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. For the evaluation of our display of speaker participation, we are explicitly interested in this type of meeting.

Behavioral experiments on information sharing study how different ways of presenting a task to a group can impact the way information is shared. Many of the studies in this area reveal that the framing of the decision impacts the amount of information shared. They also demonstrate that one indication of poor information sharing is when there is an imbalance between the amount of time a person speaks and the amount of information that the person has to share [11, 27].

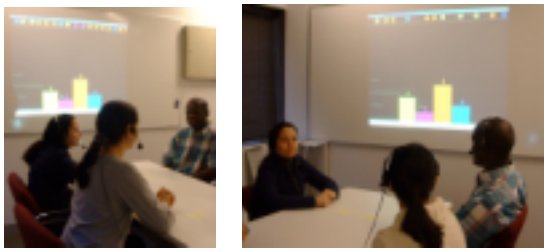
In a study run by Stasser and Taylor [26], they directly examined how speaker participation rates related to the issue of information sharing. As one would expect, subjects that spoke the most in the experimental task shared the most information. Somewhat less expected was that the researchers were able 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 had the most influence on the decision. While this finding does not universally apply to all group decisions, we find it suggestive that speaker participation in a group setting is a valid metric for quantifying an aspect of information sharing.

5. BEHAVIORAL STUDY

To understand how a display of speaker participation impacts group interaction, we designed and ran a behavioral study with an information-sharing task. 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 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. We allowed subjects 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, we 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.

A detail we did not explain to the subjects was that one of them held information that would be critical in determining the best answer. During each task this became clear to subjects, such that they could identify who held the critical information afterwards, but we did not explicitly tell them this was part of the experiment's design because we wanted to establish that each subject's contribution to the discussion was of equal importance. Telling them one subject held the answer to the task would have encouraged the group to abandon their assignment to discuss each person's viewpoint.

The experiment had two conditions: control and experimental. In the control condition, the groups discussed the two decisions 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.



Figures 2A & 2B. Subjects in the experimental condition.

Figures 2A & 2B 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.

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.

5.1 Study Hypotheses

Based on theories on the formation of group norms [1, 10] and on the self-regulation of behavior [3], we predict that any display of information about group performance will 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, we 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 our 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, we 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, we expected it 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 [9, 23].

We summarize these predictions in 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.*

H3: *The display would have minimal-to-no impact on the group's perception of interpersonal dynamics.*

Beyond these hypotheses, we planned 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.

5.2 Method

5.2.1 Experimental Task

The specific decision tasks given to the subjects were to select a student to admit into the university's undergraduate program and to select a location for a new 24-hour convenience store in the local metropolitan area. The topics were counter-balanced within each condition.

These topics were based on previous tasks used by Stasser, Hollingshead, and others [5, 11, 27] that examine information sharing between individuals in a group setting. In an information-sharing task, the challenge is to effectively share the private information in order to discover the best decision choice. We additionally chose to manipulate the role each participant played in the conversation: for 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.

The task topics 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. Also with this design, we could manipulate who was the most informative person in each given task. Each subject was required to be from the university community, so that they would have a similar perspective on the topic. Because we aimed to create a scenario where equal group participation was ideal for the decision, our goal was to instill an assumption of equality over the group.

5.2.2 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 (a protocol violation) and as a result the subject elected to only actively participate in the task topic unrelated to university life, rendering measurements of change in group interaction irrelevant. As a result, the remaining analysis contains 11 control groups (44 subjects) and 12 experimental groups (48 subjects).

5.3 Results

Our experiment results indicate that the shared display of speaker participation impacted 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 our findings, we also discovered 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 our three hypotheses.

5.3.1 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 4:40 minutes to 28:39 minutes, with an average time of 13:45 minutes. With this high rate of variance (stdev = 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. We believe 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 future experiments we hope to have more difficult tasks, so that task accuracy and the amount of information shared can be measures of success.)

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, we asked subjects to rate 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. We attribute the perception of increased task difficulty to an increased effort applied by the 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 our 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 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%).

5.3.2 Evaluation of Hypothesis 1

Our first experimental hypothesis was that we predicted 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 our hypothesis, we divided subjects into three categories: under-participators, middle-participators, and over-participators. We wanted 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, we pooled the subjects within each condition and categorized each person 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 stdev and over-participators were defined as those participating at a rate higher than the mean plus one stdev. 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 we chose to make 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, we then calculated the change in participation rate for each subject (participation in Task 2 – participation in Task 1) and examined these new datasets 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 significance 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 improved changes, we decided to conduct the analysis of under, middle, and over-participators without the two problematic groups. We do not have 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 this 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, we performed a paired t-test to first determine if there were significant changes within each group. Then we performed a t-test of independent samples comparing the changes in participation across the two

conditions. These results are shown in Table 1 and graphically in Figure 3.

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 ($t(4)= -.507$, $p=.635$, 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).

Table 1. Participation rates for each participation category and condition, with statistics for significance of change and difference.

Participation & Condition		N	Task 1	Task 2	Avg Chg	Std Error	Paired t-test	Ind 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**	

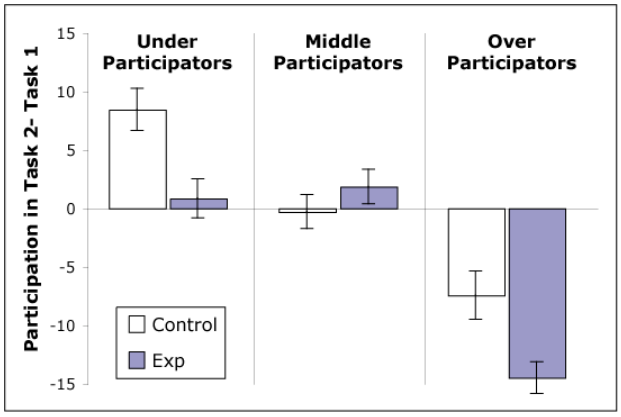


Figure 3: Average change in participation rates, by participation category and condition.

Our 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 our hypothesis, we found that 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.

As stated with our hypotheses, we expected the display to 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 our expected way, while under-participators did not. One explanation of these apparently opposing results is that under-participators may have responded to the information about themselves differently than the over-participators did. This conjecture will be revisited in Section 6.3.5.

5.3.3 Evaluation 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.

Our second behavioral hypothesis was that we predicted 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, we pooled subjects 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. We then looked for systematic changes in participation between tasks and across conditions.

As detailed in Table 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).

While we did not find a significant difference between control and experimental subjects, we did find that 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, 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). While

this is weak support for Hypothesis 2, we believe it is a potential indicator that the display encouraged modulation in participation based on the information held by the subject. We 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.

Table 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	Avg Chg	Std Error	Paired t-test	Ind 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	

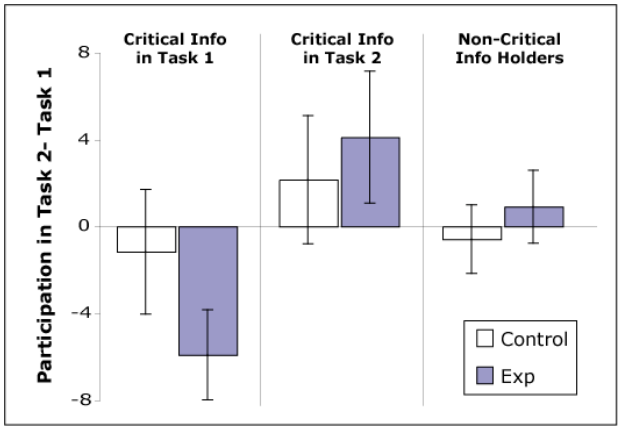


Figure 4. Average change in participation rates, by information held and condition.

5.3.4 Testing 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 we did not explicitly manipulate these variables, we will evaluate this hypothesis 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, we see 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 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 3.

Table 3. 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	Avg Response	Std 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. 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

To measure the impact of the display on group dynamics, we asked subjects to rate 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 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).

We conclude this examination of Hypothesis 3 by stating that 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, we have no other indications that the display impacted task performance, the group's ability to focus on the task, or the development of healthy group interaction.

5.3.5 Perception vs. Reality

To quantify subjects' understanding of their participation during the tasks, we asked subjects in the experimental condition to rate 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 our 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. We also found 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, we 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 5 contains these results.

Table 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	Std Error	Ind t-test
under	5	3.40	.600	t(8)=–2.324
over	5	5.20	.490	**p<0.05**
middle	27	4.41	.359	—

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.

Perception of Participation

To understand this discrepancy between perception of participation and actual participation, in the control condition, we asked an additional question we did not ask 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, we transformed the subjects' actual rate of participation for the two tasks into seven equally-divided bins that correlated with the bins of the 7-point Likert scale on the questionnaire.

Table 6. Self-rated vs. actual participation of subjects in control condition, on a 7-point Likert scale.

Control (N=44)	Participation	Avg Response	Std 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**

Table 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.

We also believe that this finding explains 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 [3] explain how individuals self-regulate their behavior and this is what we believe we have observed with the over-participators in our 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 our 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. 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) [3]. What we 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.

5.4 Study 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 we found that the introduction of the display disrupted the trend of increasing group trust that was found in the control condition. Lastly, we found that subjects over-estimated their level of participation in a conversation and that under-participators rated the display as a less accurate reflection of their behavior than over-participators did.

This last finding assists us 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 [15]. The authors also concluded that a discrepancy between one's mental state and the feedback standard explained these observations.

Taken together, our 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 therefore 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 our informal observations of subjects and combine well with our 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.

6. CONCLUSION

In our behavioral study, we found that over-participators responded to the display by restricting their comments, while under-participators did not increase their participation levels. We also found that 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 impacted trust development between group members. Our discovery that subjects had inflated self-ratings of their participation helped illuminate why under-participators did not respond as expected to the display.

In conclusion, we are satisfied to find that the display encouraged more vocal members to restrain themselves from dominating further, while not restricting the contributions of critical information holders. Yet our findings that under-participators spoke more and that trust increased more in the *control condition* indicate that there can be drawbacks to a public display of social information. In our future work, we are going to explore these issues in several ways. First, our subjects used these interfaces for a very short period of time with a very specific type of task. We intend to run future studies with established, real-world groups over longer periods of time to gauge longer-term behavior trends and more general feedback on when this type of display is useful. Second, we are interested in further exploring the issue of subjects over-estimating their own participation and how different displays might assist users in correcting this misconception. In particular, if information about behavior were sent as private messages to subjects, instead of as public displays, perhaps less vocal members would have responded differently. Lastly, we will continue working towards our original goal of discovering types of information displays that encourage a group to expand the breadth of its discussion during a decision-making meeting for the purpose of increasing the quality of the group interaction.

This paper discusses the possibility that flaws in our group decision-making processes can be corrected by interfaces that make us aware of the imbalances in our discussion and consideration of alternative choices. To explore this possibility, we built a simple display of speaker participation to examine how such a display would impact individual participation in a group task. While the quantity of time someone speaks is not directly equivalent to the influence they have over a group decision, our findings still suggest that displays of social information influence individual behavior enough that this approach is promising for improving certain types of group interactions.

7. ACKNOWLEDGMENTS

We would like to thank all of our study participants for their contribution to this work. We also thank the Electronic Publishing research group, Larissa Welti-Santos, Judy Olson, and the anonymous reviewers who provided valuable feedback and guidance in the data analysis. This work was funded by the MIT Media Lab research consortia.

8. REFERENCES

[1] Bettenhausen, K. and Murnighan, J.K. The emergence of norms in competitive decision-making groups. *Administrative Science Quarterly*, vol *30* (1985), 350–372.

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

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

[4] Daft, R. and Lengel, R. Organizational Information Requirements, Media Richness and Structural Design. *Management Science*, vol *32*, 5 (1986), 554–571.

[5] 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*, 3 (1998), 61–88.

[6] DeSanctis, G. and Gallupe, R.B. Group Decision Support Systems: A New Frontier. *DATA BASE*, vol *16*, 2 (1985), 3–10.

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

[8] 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* (1991), 119–146.

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

[10] 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.

[11] Hollingshead, A.B. Information Suppression and Status Persistence in Group Decision Making – the Effects of Communication Media. *Human Computer Research*, vol *23*, 2 (1996), 193-219.

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

[13] Kazman, R., Hung, W. and Mantei, M. Dynamic Meeting Annotation and Indexing, In *Proc. of the 1995 Pacific Workshop on Distributed Multimedia Systems* (Honolulu, HI). 1995, 11–18.

[14] Kiesler, S. and Sproull, L. Group Decision Making and Communication Technology. *Organizational Behavior and Human Decision Processes*, vol *52* (1992), 96–123.

[15] Losada, M., Sanchez, P. and Noble, E.E. Collaborative technology and group process feedback: their impact on interactive sequences in meetings, In *Proc. of the Conference on Computer-Supported Cooperative Work* (Los Angeles, California). ACM Press, 1990, 53–64.

[16] 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*, 5 (1997), 706–718.

[17] Morgan, N., Baron, D., Bhagat, S., Carvey, H., Dhillon, R., et al. Meetings About Meetings: Research at ICSI on Speech in Multiparty Conversations, In *Proc. of the ICASSP–2003* (Hong Kong). 2003.

[18] Myers, D.G. and Bishop, G.D. The enhancement of dominant attitudes in group discussion. *J. of Personality and Social Psychology*, vol *20* (1971), 385–391.

[19] 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*, 7 (1991), 40–61.

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

[21] 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*, 3 (1997), 241–262.

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

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

[24] Siegel, J., Dubrovsky, V., Kiesler, S. and McGuire, T.W. Group processes in computer-mediated communication. *Organizational Behavior and Human Decision Processes*, vol *37* (1986).

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

[26] Stasser, G. and Taylor, L.A. Speaking Turns in Face-To-Face Discussions. *J. of Personality and Social Psychology*, vol *60*, 5 (1991), 675–684.

[27] 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* (1987), 81–93.

[28] 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 *28* (1996), 95–113.

[29] Whyte, G. Decision failures: Why they occur and how to prevent them. *Academy of Management Executive*, vol *5* (1991), 23–31.