AUTOMATED LANGUAGE-BASED FEEDBACK
FOR TEAMWORK BEHAVIORS

A Dissertation
Presented to the Faculty of the Graduate School
of Cornell University
In Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy

by
Gilly Leshed
August 2009

AUTOMATED LANGUAGE-BASED FEEDBACK
FOR TEAMWORK BEHAVIORS

Gilly Leshed, Ph. D.
Cornell University 2009

While most collaboration technologies are concerned with supporting task accomplishment, members of work teams do not always have the skills necessary for effective teamwork. In this research I propose that providing dynamic feedback generated by automated analysis of language behavior can help team members reflect on and subsequently improve their teamwork behaviors. This prospect is developed based on research in multiple disciplines, including teamwork effectiveness and social behaviors, feedback for training and regulating behaviors, and use of language in group conversations.

To support this research, I directed the design and development of GroupMeter, a web-based chat system that analyzes conversations using a dictionary-based word count technique and visualizes indicators of language. I present a set of requirements for the GroupMeter system and the iterative process in which its design evolved. Findings from experiment 1 included a set of linguistic indicators that may serve as a useful source of automated feedback, such as agreement words and self-references, and that were embedded into the GroupMeter system.

Experiments 2, 3 and 4 used GroupMeter as a research platform to examine the effects of automated linguistic feedback on team members. The experiments identify the conditions under which feedback positively enables reflection on and changes in language use and teamwork behaviors, as well as when it risks distraction and gaming

behaviors. The findings are discussed in light of how feedback visualization shapes interpretations and perceptions of normative teamwork behaviors; ambiguity and benchmarking in representing social behaviors and language use; how to support balance between task-focus and socio-emotional interaction; and, improving teamwork behaviors versus “gaming the system”.

This research contributes on three levels. Theoretically, it develops a three-way relationship between teamwork, feedback, and language, by tying together theories from multiple domains and supporting this relationship with empirical findings. Practically, it demonstrates a novel technique for training people to develop their teamwork skills. And design-wise, my work adds to the accumulating knowledge about groupware technologies that, while keeping the team activity in the center, illuminate peripheral awareness information about social interaction.

BIOGRAPHICAL SKETCH

Gilly Leshed was in the first group of students to enter the Information Science PhD program at Cornell University in August 2004. She received a Bachelor's degree, *Summa Cum Laude*, and a Master's degree, *Cum Laude*, from the Technion, Israel Institute of Technology, in Industrial Engineering at 1996 and in Information Management Engineering in 2000, respectively. Before returning to grad school, Gilly spent six years in industry designing, developing, and evaluating user interfaces for a mobile flight-deck information system for commercial aircraft cockpits crews, and multi-user geographical command and control systems for military use.

During her graduate studies at Cornell, Gilly worked in the Human-Computer Interaction Lab at Cornell with Prof. Geri Gay, collaborating with students and faculty from many disciplines, including information science, computer science, communication, and design. She publishes her research at top ACM conferences such as CHI and CSCW, and received best paper awards. Gilly co-taught both undergraduate and graduate level courses in human-computer interaction, and was constantly involved in mentoring undergraduate, master's, and PhD students in the HCI lab. Gilly initiated and organized the Information Science Breakfast seminar series, now an established lecture series that brings together faculty, researchers, and students.

For her graduate studies, Gilly moved to the United States with her husband, Ofer, their son, Imri, and their daughter, Rotem. The family loves the outdoors, and they enjoy hiking, skiing, and triathlons.

למיכל

ACKNOWLEDGMENTS

I am grateful to Geri Gay for her advice, mentorship, and support all along the way – from the first transatlantic phone call to the moments of sealing this dissertation. I appreciate my remarkable PhD committee members, Jeff Hancock, Poppy McLeod, and Jeremy Birnholtz, each sharing with me their expertise and being available whenever I needed assistance and advice. I would like to thank Dan Cosley for thoughtful discussions, technical help, and advice. I thank Michael Schober and Sven Travis from The New School for sharing with me their early ideas about GroupMeter and letting me embark on this project. Thanks to Françoise Vermeylen at the Cornell Statistical Consulting Unit for her statistical advice.

To all the students who worked with me enthusiastically in designing and building GroupMeter, running experiments, and analyzing data: Jad Atoui, Peter Babinski, Sarah Birman, David Bunker, Emily Etinger, Julia Filimonow, Puneet Gupta, Seth Katzman, Lauren Kehe, Cathryn Lao, Soyoung Lee, Brian Lim, Austin Lin, Cathy Liu, Tiffany Ng, Diego Perez, Ayumi Tokuno, Sada Williams, OJ Zhao. This was a true collaborative effort.

Many thanks to all my friends and companions, co-authors and collaborators, students and faculty, in the Information Science Program and the HCI Lab at Cornell University. Their encouragement, guidance, and just being there for a chat has been invaluable, and our friendships have extended well beyond our shared office and time in the program.

Special thanks to Rosemary Adessa and the Information Science administrative staff, for making things happen, so that I can focus on my academic endeavor.

I thank the Intel PhD Fellowship program for financially supporting my final year of graduate studies.

Finally, it all comes down to my family. To my parents, Daphna and Ehud Naheer, thank you for raising me to excel and be confident in myself to achieve anything I could possibly want. My deepest appreciation goes to my beloved Ofer, for endlessly supporting, encouraging, and believing in me, and for being with me through it all, with unconditional love. I couldn't have done this without you. And lastly, to my dearest and beloved Imri and Rotem, who are always my motivation and inspiration to keep on going.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

PREFACE

Motivation for Study

In the years 1996-1998, I worked as a Human-Computer Interaction designer in an avionics startup company. We developed a flight-deck information system for B747-200 aircraft cockpit crews, the Integrated Crew Information System (ICIS). Each of the three crew members, the captain, co-pilot, and flight engineer, was designated one ICIS unit that would be mounted near their seat. The ICIS was designed to replace the heavy paperwork that exists in the cockpit, including normal and abnormal checklists, operating manuals, flight plans, performance calculations, maps, among others. For instance, it provided access to electronic checklists that keep track of the status of checklists being completed and other features that were not available with paper checklists.

Very soon I realized, as I was observing users in cockpits and simulators, that maintaining coordination within the cockpit crew is an intricate process that consists of social behaviors both directed toward accomplishing the task and those oriented toward maintaining the team well-being. For instance, if the flight engineer pauses the completion of a normal checklist for any reason, he needs to notify the other crew members about it. To support this action we designed a notification that appears in all units when a checklist is left without completion, and a quick link to the unfinished checklist.

As another example, the captain has the highest status in the aircraft and can therefore override any decision or action made by the other crew members. Yet, many decisions made in the cockpit are a matter of personality and social relationships established over time, and while some captains dominate the cockpit, others expect the first officer to share with them the responsibility for flying the aircraft. Further, crew members should be able to fill in for each other in case of unexpected pilot incapacitation. The system was therefore designed to allow all crew members carry out all activities in all units as well as use one unit to complete an activity previously started in another unit. Adding technology to the intricate social mix to support task and relationships is a complex matter. In this case, the design decisions, led by the need to allow for maximum versatility, left outside of the system the social activity of figuring out, and sometimes negotiating, the activities distribution.

Later in my professional career, in the years 2000-2004, I worked as a software engineer in the design of command and control systems for military use at the battalion level. The battalion officers, subordinates, and staff operate in a complex distributed environment in the field, and the command and control system was designed to support their coordination activities in a mission. Again, we took into consideration the complicated social interaction process that exists among officers, subordinates, and staff to accomplish the mission. For instance, back and forth communication with repetitions and further clarifications often took place to understand the exact location of a unit at any certain time, to avoid loss of a unit or friendly fire. To support this task-oriented interaction we designed all commands and reports to be time-stamped and map-based, so that every activity – planned or completed – is anchored to a certain geographical location in the field and point in time to improve situational awareness of the involved parties. Yet, I observed that, despite the availability of high-end communication and information technologies, too often activities were not coordinated and reports were misunderstood and later repeated, causing frustration, annoyance, and disputes among officers.

Later on, during my graduate studies at Cornell, I served as a teaching assistant in group-project-based human-computer interaction classes, and as a research assistant observing distributed teams of civil and material engineering students from Cornell and Syracuse Universities working to design elements for a Space Crew Exploration Vehicle. Both students in the same school, and those distributed across two schools, had several communication and information technologies at their disposal, such as email, instant messaging, online file sharing spaces, video conferencing, and wikis. Students were encouraged to use these tools to improve their communication and information sharing in order to work better together in their teams. Still, I observed many cases in which student teams struggled with teamwork issues, such as communication breakdowns, sub-grouping, time management, and social loafing, all contributing to teams' ineffectiveness.

These experiences led me to the questions that motivated my current research: Why, despite the availability of technologies that are designed to help teams collaborate and do a better job, teams are still facing difficult social interactions? And is there a way in which technology can help teams not only accomplish their tasks better, but also overcome social difficulties?

This chapter serves as a preface for my dissertation. In the following sections, I present the challenges I identify for providing answers to the problems that motivated my research. I then propose a solution and the research questions that arise from it. I continue by defining terms used throughout this dissertation, and conclude with the contributions of this research and my research approach. The following chapters provide a more complete theoretical background, the technology development, and the empirical work of this research.

The Challenge

My experiences and observations during my professional career and as a teaching assistant led me to the challenge I address in this dissertation. A gap exists between acknowledging that social behaviors for maintaining the team well-being are important for the team to accomplish its tasks and achieve its goals, and addressing these specific behaviors in the design of technologies for teams. This gap is exemplified by two chapters in Human Factors in Aviation, edited by Earl Wiener and David Nagel (1988). One chapter deals with issues of crew interaction processes, noting that the complexities of the aircraft operational environment mandate highly coordinated teamwork and that professional individual flight skills of crew members "are often not enough to assure effective performance" (Foushee & Helmreich, 1988, p. 194). Yet, a later chapter in the same volume, reviewing the development of cockpit-crew systems design (Sexton, 1988), fails to treat the crew as a group of interacting individuals, and instead refers to it as one entity interacting with the cockpit systems.

Technologies for groups, also known as groupware technologies, have traditionally been designed to help groups achieve better task performance. The ICIS enabled the B747-200 cockpit crew complete checklists more accurately. The command and control system enabled automatic anchoring of commands and reports to electronic maps to improve situational awareness. Even in less demanding environments such as college, an online meeting space facilitated reaching mutual understanding by offering a shared whiteboard to geographically dispersed students teaming for a class project.

The fundamental premise behind collaboration tools is that introducing them will automatically guarantee successful team outcomes (Guzdial & Turns, 2000; Kreijns, Kirschner, & Jochems, 2003). However, despite – and sometimes because of – the availability of groupware technologies, the social interaction process of the team, beyond task accomplishment, is not always trouble free. People do not necessarily know "how to collaborate effectively and they need to develop these skills to use the tools productively" (Joiner, 2004).

In his analysis of challenges facing the design and development of groupware technologies, Grudin writes: "Groupware may be resisted if it interferes with the subtle and complex social dynamics that are common to groups" (Grudin, 1994b). When evaluating the ICIS in flight simulators, I observed a flight engineer waiting to be told by the captain to perform performance calculations with the ICIS, although such direction was not given when using the paper format. In class groups, I noticed that team members using the shared whiteboard did not know if it is appropriate to overwrite someone else's sketch. Thus, I noticed that teams were expected to work together effectively using collaboration tools without specific guidance on what kinds of social behaviors, apart from task-oriented behaviors, they are expected to demonstrate.

I suggest that groupware technologies are inherently problematic in that their design is focused too much on helping teams getting their tasks done, leaving out elements involving team members' social behaviors and interaction. If teams are expected to work together using groupware, team members should be given guidance on what kinds of social behaviors are beneficial when using this medium, at least while learning to use it. One way to provide such guidance is through functionality designed into the technology itself, ensuring that the social behaviors expected from team members are supported, and not ignored or conflicted, by the design of the system. Such guidance toward the application of certain social behaviors can serve as what Orlikowski describes as "technological frames", shaping members' shared interpretations of the nature and role of technology in the team setting (Orlikowski & Gash, 1994).

But the challenge is not exclusively in the domain of technology design. From a theoretical perspective, there are multiple theories and models from different domains each touching upon certain aspect of this problem. Theories of effective team performance and social behaviors originate in organizational and management studies as well as communication of small groups. Models for coaching and supporting certain behaviors are often derived from self-regulation as well as educational theories. The theoretical challenge is to link together perspectives from multiple domains in order to address the problem described above. The next section, then, presents both a technology proposal to the challenge, as well as a theoretical framework that combines multiple conceptual theories to argue for the practical application.

Proposed Solution

I propose to address this challenge by providing team members with automated dynamic feedback about their social behaviors while they interact with their team. When it comes to communication technologies used by teams, team members' social behaviors can be revealed through their use of language. In an extreme example, one of the reasons for the crash of Northwest Flight 5719 on December 1, 1993, killing 16 people on board, was the captain's overly authoritative communication toward the first officer, and the first officer providing information to the captain in a questioning manner rather than as assertions (Tarnow, 2000).

I therefore suggest supplementing the user interfaces of groupware technologies with dynamic feedback automatically generated from teach conversations about some characteristics of the language used by team members during the team communication activities. This dynamic feedback, presented in an ongoing manner and changing as the communication progresses, can stimulate team members' reflection on and increase their awareness of their use of language while they communicate with their peers. By reflecting on and being aware of their use of language, team members can learn to change the language they use and acquire better teamwork skills. The result is a shift in the role of groupware technology from solely supporting tasks toward paying attention to the social behaviors involved in the team communication.

As I will show in Chapter 2, the theoretical reasoning for providing team members with automated dynamic feedback about their use of language during the team activity to improve the teamwork process comes from multiple domains. The first set of theories explains, from a social psychological perspective, what constitutes effective teamwork, both descriptively and normatively. I also draw from theories and empirical studies in education and psychology, on how teamwork skills can be trained in general, and how feedback in particular can serve as a useful tool for this purpose. I will further illustrate connections between social behaviors and language use, based on communication and psychology studies and theories of language, as well as empirical studies of natural language processing techniques. Finally, I will review previous studies of systems collecting, analyzing, and mirroring social behaviors to teams in the domain of HCI and CSCW to illustrate the efficacy of my proposed solution.

In this dissertation I plan to examine a number of issues. First, besides a theoretical justification required for using feedback from an automated source to improve teamwork behaviors, I need to demonstrate that the proposal outlined above practically works: that team members who receive automated feedback reflect on and change their use of language. I also need to find out what kind of linguistic features can be used as a source of feedback on teamwork behaviors. Such linguistic features need to demonstrate that they are associated with teamwork behaviors, and that users perceive them as a reasonable source of feedback. Finally, augmenting a communication medium with dynamic feedback visualizations will likely cause distraction from the team conversation and task accomplishment. This issue needs to be examined in order to design feedback that allows for an easy shift between the primary team activity and the secondary visualization of about social behaviors.

To address these questions, in this dissertation I present the theoretical underpinning for my proposal, and a system, GroupMeter, I designed to demonstrate automated linguistic feedback in a mediated team environment. I then present a series of empirical studies that examine the efficacy of this proposal, and discuss their results with regards to key research questions in light of both the system design and the theoretical concepts.

Definitions

This research is concerned with providing dynamic peripheral linguistic feedback to team members while they are engaged in a team activity using groupware technology, to stimulate their reflection on, awareness of, and change in their teamwork behaviors. While some of the terms used in my research seem intuitive, a definition for each one will ensure their common and consistent understanding.

The first term to define is groupware. *Groupware* is a piece of technology designed to help people involved in a common task achieve their goals, both at the project-level and when working in small groups (Grudin, 1994a). "Desktop conferencing, videoconferencing, co-authoring features and applications, electronic mail and bulletin boards, meeting support systems, voice applications, workflow systems, and group calendars," (Grudin, 1994b) all fall under the definition of groupware. Since my interest here is on social communicative behaviors in teamwork, I will focus on small-group applications that emphasize communication (Grudin, 1994a).

The next set of terms refers to teamwork. *Teamwork* is the process by which a small number of individuals work together in a team and perform one or more tasks by interacting with each other in a real group setting (as opposed to a nominal group) toward a common goal with specific roles and within an organizational context (Salas, Dickinson, Converse, & Tannenbaum, 1992). The *team* is a social system with boundaries and interdependencies, operating within a larger social system to produce some outcomes for which members have collective responsibility and which can be measured as an indication of the teamwork effectiveness (Hackman, 1990). In order to produce outcomes, the team engages in a social *interaction process*, which includes the ongoing activities that define the team dynamics "as distinct from merely relating initial member and group states and situational conditions to subsequent products or outcomes" (Lebie, Rhoades, & McGrath, 1995). *Teamwork behaviors* are therefore the interaction acts carried out by team members during the team interaction process, and *teamwork skills* represent the competence to perform teamwork behaviors that lead to an effective interaction process and outcomes (Stevens & Campion, 1994).

Another key concept in this research is feedback. *Feedback* is a general systems theory term that describes a situation in which information about the outcome of an activity is provided to the entity performing the activity. The acting entity (person, animal, mechanical system, social system, etc.) can then use the feedback to modify its behavior in order to change the outcomes. The focal entities of my research are individuals within a team receiving feedback about their behaviors in the teamwork interaction process.

A typology of feedback (Geister, Konradt, & Hertel, 2006) distinguishes between different types of feedback, with *outcome feedback* on the one end, and *process feedback*, concerning how one achieves the outcomes, on the other end. Feedback can also be classified using the level at which it is aggregated, at the *individual level* or at the *team level*. Feedback at the individual level can further be classified at the recipient level, provided to each member separately, or being provided to all members (i.e., when all team members see everybody's feedback). Another aspect of feedback is the source providing it, which needs access to the actions or outcomes for generating the feedback information. Feedback can come from a *human source*, either peers evaluating each other within the team, or external observers such as supervisors, instructors, or managers. It can also come from an *automated source*, when a machine traces and analyzes behaviors or outcomes to be used for feedback. Finally, Geister et al. distinguishes between feedback for the purpose of *evaluation*, or for *developmental* reasons.

These classifications are important for framing the kind of feedback I explore in this research. Specifically, I am interested in the use of process feedback at the individual level, provided from an automated source for the purpose of developing and improving one's behaviors. I will also use the term *dynamic feedback* as a specific type of development feedback, when it is provided in an ongoing manner, and changing in real time as the observed behaviors or outcomes change. Finally, I use *linguistic feedback* to characterize process feedback provided on some features of the language used by the recipient.

There are many aspects of *language*, the symbolic system used to communicate meaning in speech or in writing. Here I will refer to *word choice*, the specific words an individual uses in their speech or writing. Pennebaker argues that using words require a certain social skill and as such they can serve as indicators of people's personality and behavior in social situations (Pennebaker, 2002; Pennebaker, Mehl, & Niederhoffer, 2003), making word choice especially suitable to be used as process feedback on teamwork behaviors.

The final set of terms refers to the processes occurring as a result of providing feedback. *Reflection* is a "human activity in which people recapture their experience, think about it, mull it over and evaluate it" (Boud, Keogh, & Walker, 1985). *Awareness*, particularly in the context of my research, is "the state of knowing about the environment in which you exist; about your surroundings, and the presence and activities of others" (Wisneski et al., 1998). With the focus of my research, the concept of awareness can be expanded to *self-awareness*, which refers to knowing about one's own actions, thoughts, emotions, and personality. Although it has been argued that awareness is naturally information that is "being gathered passively, while other workplace activities progress" (Dourish & Bly, 1992), I will use the term *peripheral awareness* to specify information about activities that are not foreground tasks but instead reside at the periphery of one's attention (Gaver, 2002).

Contributions

This research provides the following key contributions:

First, this research shows practically that providing automated linguistic feedback stimulates team members' reflection on and change in their use of language and teamwork behaviors. Triggering reflection and leading to change in the use of language is a challenge, because generating language, and more specifically the production of a certain vocabulary, is spontaneous and difficult to change (Pennebaker & King, 1999). Further, I will show that word choice is associated with teamwork behaviors. As such, by stimulating reflection on and guiding toward certain change in word choice in teamwork situations, people are directed toward changing their teamwork behaviors and can acquire better teamwork skills.

To demonstrate the feasibility of this endeavor, my research also integrates multiple theoretical perspectives from few disciplines. Specifically, theories of self-regulation and reflective practice have been applied in the education and management domains demonstrating how teamwork behaviors can be developed. Further, empirical work in natural language processing has linked automatically generated linguistic features with social behaviors and performance measures in small group conversations. My research contributes on the theoretical level by bringing together three concepts—teamwork behaviors, feedback, and language use—into an integrated model that delineates how automated dynamic feedback on language use, and more specifically, word choice, can lead to reflection on and improvement of teamwork behaviors.

On the design level, this research also addresses the problem of designing technologies with dynamic displays that move back and forth between the center and periphery of the user's attention, requiring minimal cognitive effort for this shift (Shami, Leshed, & Klein, 2005). As such, successful peripheral display design enables users to focus on a primary task while maintaining peripheral awareness of a secondary task (Plaue, Miller, & Stasko, 2004). My research demonstrates the effort in designing a dynamic display in a teamwork situation that is peripheral to the primary team task. The feedback display I designed provides awareness information about language use, as an indication for the team interaction process, because teams need to constantly balance between attending to the task and to the social aspects of their operation (Bales, 1950). I will describe the problems I encountered and my approaches to addressing them in order to achieve clarity and glanceability (Matthews, 2007) as well as some level of ambiguity for rich, contextual interpretation (Gaver, Beaver, & Benford, 2003).

Finally, my research introduces a shift in the role of groupware technology from primarily supporting group tasks to help maintaining the social interaction process in the team. Similar to Wainer & Braga (2001), I argue that technologies that only focus on the team task overlook an inherent part of the teamwork that is crucial to the social well-being of the team. Keeping in mind that team members need to consider both the task and social aspects of teamwork, I present this concept as a complementary approach to, rather than a replacement of, groupware designed to enhance task accomplishment.

Approach to research

As a human-computer interaction (HCI) scholar, the research methodology I apply in this research is an adaptation of the *iterative and incremental model*. According to this model, based on the software iterative and incremental development model (Cockburn, 2001), requirements are gathered based on an initial user study, then translated into a design that is implemented into a prototype and evaluated typically through a user study. The results of the evaluation are then used to revise the requirements and the design, and so on until a satisfied product is being implemented (Zhang, Carey, Te'eni, & Tremaine, 2005).

However, my interest lies not only in developing successful interactive systems, but also in addressing important social-science related issues from a theoretical perspective. Using the system to understand social processes, my approach to research therefore modifies the original HCI model by including a set of theories that are based on empirical findings. These theories then feed into the design of the interactive system I use to address the research questions I am interested in with subsequent user studies that produce further empirical findings (Figure 1).

The iterative model is used as the backbone structure of this dissertation. In Chapter 2, I establish the theoretical foundations of my research, rationalizing my approach of using dynamic automated linguistic feedback for guiding reflection on and change in teamwork behaviors and the primary research questions it raises.

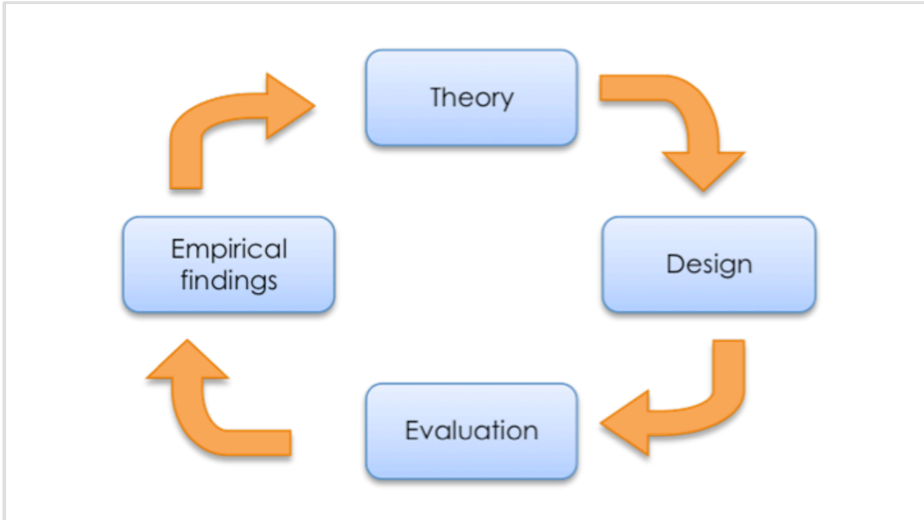


Figure 1. Adapted iterative and incremental model. In this model, social-science related theory takes part in feeding into the design of the interactive system, as well as being updated in light of empirical findings obtained through user studies evaluating the design of the system.

Chapter 3 presents the design of a system, GroupMeter, grounded in theoretical reasoning, which serves as a research platform for the key questions of this research. Chapters 4, 5 and 6 present empirical user studies designed both to address the research questions of my research and to improve the design of GroupMeter. The results of the studies are generally discussed in Chapter 7 in light of both the theoretical perspectives underlying my research as well as the design principles of GroupMeter as a groupware technology prototype.

CHAPTER 2

THEORETICAL BACKGROUND AND RELATED WORK

Theories of Teamwork Effectiveness

Teams can be a powerful tool for learning (Barkley, Cross, & Major, 2005; Roberts, 2005; Stahl, 2006) and for accomplishing tasks (Hackman, 1990; McGrath, 1984). They promote critical thinking, engage their members in the activity, enable the accomplishment of complex tasks, and allow for social support among their members, among other benefits. Barbara Gross Davis notes (1993): "Students learn best when they are actively involved in the process. Researchers report that, regardless of the subject matter, students working in small groups tend to learn more of what is taught and retain it longer than when the same content is presented in other instructional formats. Students who work in collaborative groups also appear more satisfied with their classes" (p. 147). In science and engineering education, small group learning promotes more favorable attitudes as well as greater academic achievement (Springer, Stanne, & Donovan, 1999).

However, the benefits of teamwork in education and work settings do not necessarily guarantee that it is trouble-free. Teams, especially in technology-mediated environments, face a variety of challenges. For instance, they need to engage in coordination activities and this creates a cost that can be a significant barrier to project success (Cramton, 2001; Cummings & Kiesler, 2007; Kraut, 2003; Malone & Crowston, 1994), but can help also control internal team conflicts (Montoya-Weiss, Massey, & Song, 2001). Further, relationship and task conflicts that arise in groups can be associated with decreased satisfaction and intent to keep working with the team (Jehn, 1995) and with reduced team performance (De Dreu, & Weingart, 2003). Distributed teams communicating through leaner media such as email or chat also have to deal with fragile trust by skillfully managing uncertainty and complexity in these environments (Jarvenpaa & Leidner, 1998), and with higher levels of social loafing (Chidambaram & Tung, 2005).

These shortcomings of teamwork has led to a wealth in research on improving teamwork processes under the assumption that successful outcomes are a result of effective teams. Models of group effectiveness have drawn upon systems theory to describe the various inputs and processes contributing to the outputs of teamwork effectiveness. For instance, McGrath first introduced the input-process-output framework (1964) describing a number of input factors that feed into this model of teamwork effectiveness (Figure 2). These include individual level factors, such as the skills brought in by the individuals, their attitudes and traits, group level factors, such as its hierarchical structure, its level of cohesion to start with, and the group size, and environmental factors, such as the characteristics of the task and the reward structure. The input factors are mediated by the group interaction process, which includes the actions group members carry out when interacting with each other, such as

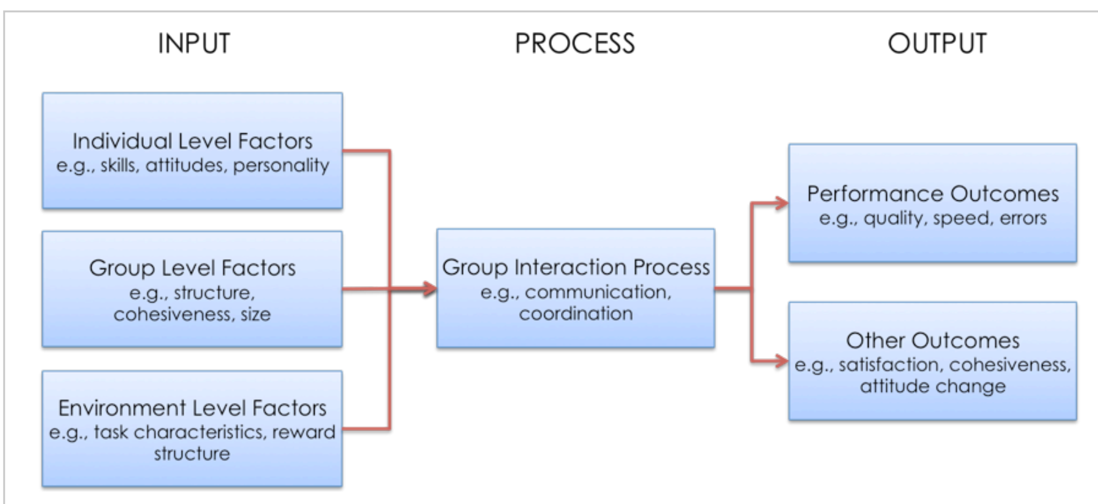


Figure 2. Input-process-output model of team effectiveness. Adapted from McGrath (1964).

communicative actions and coordination activities. Communicative actions can include, for example, expressing ideas to solve a problem and stating agreement or disagreement with others' ideas. The output includes outcomes that can be measurable performance variables as well as other outcomes such as individuals' satisfaction.

In order not only to describe factors inherent in teamwork processes, but also to improve such processes, Hackman expanded the input-process-output framework toward a normative model of group effectiveness (1987). Noting that "the key to understanding the 'group effectiveness problem' is to be found in the ongoing *interaction process* which takes place among group members while they are working on a task" (Hackman & Morris, 1983, p. 331), he identified a broad range of factors that influence team performance, such as the efforts and skills that team members apply in the interaction process. Gladstein also extended McGrath's input-process-output model (1984), considering task characteristics as moderating factors in the interaction process. For example, she suggests that flexible communication patterns during the interaction process are important for high task performance only when the task is highly complex, uncertain, and consists of interdependencies among team members.

Another set of theories applicable to the examination of team effectiveness specifically considers the behaviors team members carry out during the team interaction process, and how they are associated with effective teamwork. One such useful approach is Bales' Interaction Process Analysis (1950), distinguishing between sets of acts that support the completion of the task and those that address the socio-emotional aspects of the group. According to a psychodynamic perspective of group development, these two basic functions are related to the performance and to the maintenance of the team, respectively (Rousseau, Aube, & Savoie, 2006). The two types of acts need to balanced, such that team members need to be aware of the emotional processes in the group that inhibit or contribute to team effectiveness (McLeod & Kettner-Polley, 2004). For instance, working toward completing the task may give rise to tension in the group, and thus attending to socio-emotional aspects is important for releasing the tension, maintaining the team's well being, and reaching outcome goals.

Based on his early work, Bales expanded this applied/theoretical approach into SYMLOG – SYstematic Multi-Level Observation of Groups (Bales & Cohen, 1979). The SYMLOG framework describes three orthogonal and bipolar dimensions of interpersonal behaviors on which people interact: dominance/submissiveness, friendliness/unfriendliness, and task orientation/socio-emotional expressiveness. According to Keyton & Wall (1989), the strength of SYMLOG is both methodological and theoretical: theoretically, SYMLOG defines a conceptual space of behavioral interaction created by the three dimensions; methodologically, it serves to place individuals within that cube based on their observed or perceived communicative behaviors.

The SYMLOG framework is particularly useful for evaluating team interaction process for two primary reasons. First, although it was initially developed as a descriptive approach for the examination of group interaction process, a set of standards have been established on its three dimensions to characterize effective teamwork (see McLeod, Liker, & Lobel, 1992). These standards include evenly divided contribution among team members; group-oriented and friendly behaviors; and a balance between task-oriented and socio-emotional expressive behaviors, leaning toward more task-oriented. Second, the SYMLOG framework was initially developed "with the vision of rapid computer-produced feedback to the group" (Bales

& Cohen, 1979, p. xiv), making it particularly useful for my purposes as I describe in the following sections.

Guiding Teamwork Behaviors

Theories related to teamwork effectiveness further highlight the discrepancy between theoretical views of effective teams, both descriptive and normative, and the difficulties teams often experience, especially in mediated environments. This gap raises the question of what can be done to improve the teamwork interaction process: how can we help teams develop effective teamwork behaviors? It has already been recognized that "learning to use all of the traditional team skills in an environment where most interactions take place through a telecommunications medium is a critical challenge" (Townsend, DeMarie, & Hendrickson, 1998, p. 26). Despite the difficulties in technology-mediated teamwork, and that effective teamwork behaviors are prerequisite to successful team outcomes (Bransford, Brown, & Cocking, 1999), teams are usually assigned or self-organized and then expected to follow their tasks without proper guidance or training on teamwork behaviors and practices per-se (Rummel & Spada, 2005; Swezey & Salas, 1992).

Bosworth (1994) argues that "instructors cannot assume that the students coming into college classrooms have the skills that they need to begin a collaborative process" (p. 31). If people are expected to work in teams, they should be given explicit tools to acquire the appropriate interpersonal skills to overcome the challenges that commonly arise in teamwork situations (Oakley, Felder, Brent, & Elhajj, 2004). Children who worked in groups and received instruction on collaborative skills demonstrated more cooperative and helping behaviors to each other, used more inclusive language, and obtained higher learning outcomes, as compared to untrained groups (Gillies & Ashman, 1996).

There have been several attempts to develop techniques to teach collaborative teamwork skills, mostly in the education discipline. For example, Rummel & Spada (2005) examined two techniques, watching a video that exemplifies successful collaboration, and reading a guiding script. These were administered in a "learning phase", followed by the "collaborative phase" in which the teams applied what they learned and interacted in dyads to complete a task. This technique requires that the learning situation resembles the collaborative phase as closely as possible, and thus raises the issue of how people can learn to transfer the skills they learn and adapt their social behaviors when the circumstances in which they operate change.

Another approach that applies a learning phase includes practice sessions in which team members play roles with slightly different goals, such as peers, teachers, etc. This approach is suggested to help students not only to learn collaboratively, but also to learn how to collaborate, by adopting the cognitive perspectives associated with a role (Burton, Brna, & Treasure-Jones, 1997). Beyond practicing in an initial learning phase, Oakley et al. (2004) suggests adding peer evaluations during the team activity that focus on cooperation and contribution to the team effort. McKinney & Denton (2006) included peer evaluations during the team activity on general teamwork skills, such as "cooperate with a team in an effort to solve problems", provided with examples for good and poor behaviors on each dimension, and reported on improvement in teamwork skills toward the end of the semester in a computer science programming course.

Bosworth (1994) presented a holistic view for collaborative skill development based on Fitts and Posner's approach to skill acquisition (1967). In this approach, learners first follow specific rules used as scaffolding for teamwork behaviors. They then develop deeper understanding of the relationships between the rules and specific situations. Finally, students become able to respond to a variety of conditions. Based on this approach, Bosworth suggested the following five-stage technique to teaching collaborative learning skills: 1) Identify the collaborative skills needed to be trained; 2) Demonstrate successful examples of such skills; 3) Model the skills by breaking them down into practical behaviors; 4) Provide feedback in practice situations; and 5) Provide students with time to process and reflect on their experiences.

My research applies some elements of this approach. Specifically, I suggest providing team members with feedback about specific teamwork behaviors manifested by language use in mediated team conversations, and examining the reflective process in receiving the feedback and responding to it. My research emphasizes the feedback stage of this approach, but it does not ignore the other stages. I use SYMLOG as the framework for teamwork behaviors, and identify the specific linguistic features that are associated with them. I also examine how providing specific guidance for the successful application of these behaviors, and allocating time for reflection, helps team members become aware of and change their behaviors in technology-mediated teamwork situations. This is a difficult problem, as Jarvenpaa & Leidner (1998) suggest that in virtual teams "it is particularly challenging to encourage groups to reflect upon, learn from, and redirect, as appropriate, their communication behaviors."

Automated Feedback for Guiding Teamwork Behaviors

My goal is to help team members develop teamwork skills by becoming reflective about and aware of their teamwork behaviors during the team interaction process. I propose using feedback as the primary tools for raising team members' awareness of their teamwork behaviors, because I believe feedback can trigger reflection on the team interaction process and help team members learn how to recognize, in the moment, how to apply specific behaviors in certain situations. Further, I believe that the main challenge to developing teamwork skills is not learning discursively what the appropriate behaviors *are*, as suggested by theories and models of effective teamwork, but rather it is learning dynamically what *to do* at any given moment.

As reviewed above, models and theories of effective teamwork are abundant. They exist both in the form of theoretical frameworks (e.g., Gladstein, 1984; Hackman, 1987; McGrath, 1964; Rousseau et al., 2006), as well as lists of practical guidelines (e.g., Hoover, 2005; Levi, 2001; Swezey & Salas, 1992). However, more often than not, these models are generic and abstract. For example, it has been argued that open communication of ideas and feelings is beneficial for effective teamwork (Gladstein, 1984). This is a useful general principle. However, in any specific situation, what should team members pragmatically do to make sure everyone communicates their ideas and feelings, and accepts others' thoughts and emotions? As another example, according to the teamwork KSA (Knowledge, Skill, Ability) requirements developed by Stevens & Campion (1994), being able to recognize and encourage desirable, but discourage undesirable, team conflict is important for optimal team performance. Again, different situations and social dynamics would probably give rise to different types and sources of conflict. For instance, a range of views on a problem can be constructive in a brainstorming session or when fleshing out details of a plan, but might be destructive when the group needs to reach a unanimous decision.

Context matters: an academic brainstorming session, a military briefing and an airplane cockpit all likely set different expectations for the pragmatic realization of principles for effective teamwork. In each case, team members should act differently to set the tone and patterns of the conversation. Perhaps the team would need to use different techniques at different stages of the work, or different behaviors might be appropriate at different times. The essential skill of effective teamwork is being able to translate such generic principles to specific behaviors while working together. I suggest that learning to become reflective about one's teamwork behaviors and to adapt them in specific situations is what makes individuals "good teamwork players".

Why Feedback?

Several theories explain why feedback can be a useful tool to learn and improve behaviors. One set of these theories is related to explaining self-regulation mechanisms, directing purposeful action toward improved performance. For instance, goal-setting theory (Locke & Latham, 1990) suggests that when people have specific and challenging goals, they will work hard to achieve their goals. By receiving feedback, individuals can measure the gap between what they aspired for and what they achieved and as a result adjust the efforts they put in the next time they apply their behavior. According to social cognitive theory (Bandura, 1991), the combination of goals with performance feedback is what heightens motivation toward achieving one's goals. Butler & Winne's developed self-regulated learning model (1995), suggesting that feedback enables individuals to constantly monitor their engagement with the task, providing grounds for "reinterpreting elements of the task and one's engagement with it, thereby directing subsequent engagement" (p. 248). Although these theories were originally developed for performance-oriented goals and behaviors, it has been demonstrated that they can be applied to interpersonal teamwork behaviors as well (Losada, Sanchez, & Noble, 1990; McLeod et al., 1992).

The self-regulation approaches explaining the role of feedback to improve behavior could be seen as mechanistic and positivist, drawing from control theory and cybernetics, as illustrated in Kripperndorff's definition of feedback in his dictionary of cybernetics:

> Feedback: A flow of information back to its origin. A circular causal process in which a system's output is returned to its input, possibly involving other

systems in the loop. Negative feedback or deviation reducing feedback decreases the input and is inherently stabilizing […], e.g., the governor of a steam engine. Positive feedback or deviation amplifying feedback increases the input and is inherently destabilizing, explosive or vicious, e.g., the growth of a city when more people create new opportunities which in turn attract more people to live there. (Krippendorff, 1986)

In this view, the role of feedback is a mechanic factor that operates to increase or decrease behavior in a deterministic manner. We can therefore consider alternative, or complementary approaches to self-regulation mechanisms, such as Schön's concepts of reflection-in-action and reflective practice (1983). Schön describes how practitioners reflect on their actions, using the knowledge they gain from their reflection to solve problems, develop, and improve their practice. Similarly, Gott & Lesgold (2000) report on their experiences in training skills in the workplace, arguing that workers construct increasingly mature mental models by reflecting on their own solutions, ultimately inducing general patterns of problem solving and demonstrating skill transferability to novel situations.

The role of feedback in the reflective practice approach can be seen as allowing individuals to see what the results of their actions are, stimulating deeper reflection and gaining knowledge about their actions: "Through the unintended effects of action, the situation talks back. The practitioner, reflecting on this back-talk, may find new meanings in the situation" (Schön, 1983, p. 135). Each new experience of reflection-in-action enriches the individual's repertoire, enabling him to generalize to other cases "not by giving rise to general principles, but by contributing to the practitioner's repertoire of exemplary themes from which, in the subsequent cases of his practice, he may compose new variations" (p. 140). This description corresponds with the idea proposed earlier that good teamwork skills mean knowing how to adapt one's teamwork behaviors to a variety of situations. As such, one of my goals is to provide the kind of feedback that will initiate reflection on one's teamwork behaviors.

What Kind of Feedback to Provide and How?

Applying feedback as a guiding tool requires considering important factors that would lead to its successful application, such that individuals become reflective about and modify their teamwork behaviors. Feedback Intervention Theory (FIT, Kluger & DeNisi, 1996) provides one theoretical framework that explains the conditions under which feedback should improve or undermine performance. According to FIT, feedback about high-level behaviors, such as coordination and conflict management (see Rousseau et al., 2006) is too abstract and general, draws the individual's attention toward himself or herself, and thus is likely to be less effective. Instead, effective feedback includes low-level action-specific information, drawing individuals' attention to the details of the action needed to achieve a goal, and away from higher levels of attention, such as self-esteem.

FIT also posits that the timing and frequency of feedback is important (Kluger & DeNisi, 1996). Based on the punctuated equilibrium model of group development (Gersick, 1988; Gersick, 1989; Gersick, 1991; Gersick, 1994), there exist naturally occurring transition points in the life spans of a team during which the team members will be maximally receptive to feedback. At such points, team members will be most willing and able to engage in reflection about feedback, facilitating successful navigation of the transition such that team members will make improvements in their teamwork behaviors. Feedback can also be used to induce such transition points, refraining teams from lapsing into ineffective teamwork patterns (Gersick & Hackman, 1990).

Further, Ilgen, Fisher, & Taylor (1979) claim that in most situations more frequent feedback would be beneficial. Feedback is considered dynamic when provided repeatedly during the course of task accomplishment, updating as performance fluctuates. Such dynamic feedback allows individuals to constantly monitor goal-outcome discrepancies, adjusting their goals and efforts over time with the intent to close the gap (Ilies & Judge, 2005). For instance, dynamic feedback about contribution patterns in an asynchronous collaborative environment was shown to have a positive influence on participation, motivation, and problem-solving within learning teams (Reimann & Zumbach, 2003; Zumbach, Mühlenbrock, Jansen, Reimann, & Hoppe, 2002). However, an important issue, investigated in the current research, is how to provide such feedback that would not end with too much distraction from working toward accomplishing the team task.

Other parameters related to the kind of feedback and the manner in which it is provided include the level of unit encapsulated in the feedback information, i.e., whether the feedback information is at the group or individual level, and the degree of publicity of the feedback. Presenting public information about individuals can promote social comparisons (Festinger, 1954) and augment social presence (Short, Williams, & Christie, 1976), and as a result can increase the attention team members dedicate to each other (McLeod, Baron, Marti, & Yoon, 1997), awareness of what others are doing (Birnholtz, Gutwin, & Hawkey, 2007; Tollmar, Sandor, & Schömer, 1996), and social accountability (Erickson & Kellogg, 2000). On the other hand, drawing attention to individuals can promote competition within the group, appropriating the system to achieve goals incongruent with the system's underlying assumptions of cooperative practices (Orlikowski, 1996).

The Feedback Source: Who provides the feedback?

One important factor when providing dynamic feedback on teamwork behaviors is who evaluates the teamwork interaction process and provides the feedback. Traditionally, feedback comes from external members such as expert observers, teachers, and managers. Using the SYMLOG scheme for evaluation and feedback (Bales & Cohen, 1979), expert-based feedback was shown to affect the team interaction process such as balancing contribution in face-to-face settings (McLeod et al., 1992) and increasing socio-emotional interactive sequences in computer-mediated settings (Losada et al., 1990). Despite this promise, to provide teams with accurate dynamic feedback, observers have to be trained and be available during what could be lengthy team meetings, increasing the cost of this approach.

A different approach for providing dynamic feedback on teamwork behaviors could be peer evaluations. Peers observe each other throughout the process and often have more detailed knowledge of others' contributions (Falchikov, 1995). Peer assessment can lead to fair, responsible, accurate, and valid evaluations, in addition to encouraging team members to become more involved in the activity process (Dochy, Segers, & Sluijsmans, 1999; Freeman & McKenzie, 2002; Goldfinch, 1994; Jarzabkowski & Bone, 1998; Keaten & Richardson, 1992; Laybourn, Goldfinch, Graham, MacLeod, & Stewart, 2001; Lejk & Wyvill, 2001; Saavedra & Kwun, 1993), and achieving better learning results than expert-based feedback (Cho, Chung, King, & Schunn, 2008). Timely peer assessment also assists in improving factors important for teamwork, such as workload balance and dealing with skills shortages (Harkins & Jackson, 1985; Scott, van der Merwe, & Smith, 2005), and inter-group relationships and task focus (Druskat & Wolff, 1999; Smith, Cooper, & Lancaster, 2002; Turner & Schober, 2007).

The potential benefit of peer-based feedback is that the act of engaging in peer evaluation may itself trigger reflection and result in improved outcomes (Dominick, Reilly, & McGourty, 1997). However, peer feedback has drawbacks, such as variation in team members' motivation and evaluation skills (Cho & Schunn, 2007), and lack of willingness to receive feedback from peers when they are considered novices (Zhang, 1995).

An automated system for assessing teamwork behaviors and providing teams with feedback may be able to overcome the shortcomings of human-originated feedback. Computing systems designed to change people's behaviors falls under the definition of *persuasive technology* (Fogg, 2003). *Self-monitoring technology* is a type of persuasive tools that give users information about their behaviors in real time, allowing users "to monitor themselves to modify their attitudes or behaviors to achieve a predetermined goal or outcome" (Fogg, 2003, p. 44). Fogg suggested that such technology can be useful to change language behavior, and presented a conceptual design of a mobile phone application that helps reduce the use of the common word "like" in everyday speech:

> A word recognition system would listen to them as they talked on the mobile phone. Whenever they used the word "like," the phone would give a signal, making them aware of it. The signal could be a vibration or a faint audio signal that only the speaker could hear. In this way, the speaker could be trained to use the word "like" less frequently. (Fogg, 2003, p. 46)

Hence, a system that provides users with dynamic automated feedback about their teamwork behaviors with the purpose of helping them change their behaviors can be defined as a self-monitoring persuasive technology.

When teams work together in mediated environments, technology can capture some elements of the team interaction and carry out an analysis on them that can then be fed back to the team. For instance, when writing together a report using a shared document such as wiki or another online service, an automated process can examine who in the team is contributing which parts to the document and when they are doing so, if team members are deleting others' contributions, and what kinds of contributions are being made such as global additions, minor edits, or rearrangements of the different parts of the document. Of course, only part of the team interaction is captured in this way. Teams often have a variety of collaboration tools at their disposal, including project management tools, shared calendars, email, videoconferencing tools, and more (Dubé & Paré, 2001), and each of these technologies can instantiate different social interactions given its affordances and the context of its use (DeSanctis & Poole, 1994).

Communication tools are one type of collaboration technologies, facilitating the team's conversational interactions. Since in many workplace and educational settings teamwork is essentially conversational, analyzing the contents of the team communication data can reveal much about the teamwork process (Donnellon, 1996; Paulus, 2005). A number of frameworks for evaluating group dynamics from team conversations exist (Cannon-Bowers, Tannenbaum, Salas, & Volpe, 1995; McIntyre & Salas, 1995; Poole, 1981; Poole, 1983a; Poole, 1983b; Poole & Roth, 1989a; Poole & Roth, 1989b; Rousseau et al., 2006), mostly designed for manual coding of team conversations. However, an alternative evaluation approach is required to automate the analysis of the team conversation when captured by technology.

Such technique does not necessarily need to adhere to the "media equation" (Reeves & Nass, 1996); we should not expect that recipients of automated feedback would treat and respond to it in the same way that they treat and respond to human-originated feedback. Still, an automated approach will need to demonstrate that it provides relevant feedback that can be effectively used by team members to reflect on and change their behaviors, according to factors such as those defined by Feedback Intervention Theory (Kluger & DeNisi, 1996). The results need to be understandable, leading to the recipients' insight of what they can do to change their teamwork behaviors. Acting as an information source, an automated system also needs to be perceived as credible, in terms of both its trustworthiness and expertise (Fogg & Tseng, 1999). These issues pertain to how team members interpret and experience feedback they receive from an automated source, and are investigated in the current research.

What can an automated system provide feedback about?

Automating the assessment of communication data collected in technology-mediated environment is made available by capturing both nonverbal aspects of conversations, such as audio volume (Bergstrom & Karahalios, 2007a), eye-gaze (Kulyk, Wang, & Terken, 2005; Vertegaal, Slagter, van der Veer, & Nijholt, 2001), and body movements (Sundström, Ståhl, & Höök, 2007), as well as the content and language in speech (Cassell, 2000) and, most widely spread, in writing, such as email (Viégas & Golder, 2006), online communities (Burke, Joyce, Kim, Anand, & Kraut, 2007; Burke & Kraut, 2008), and instant messaging (Avrahami & Hudson, 2006).

Considerable work has been dedicated to the creation of visualizations based on nonverbal dimensions of team conversations, taken both from audio and from lingual input sources. Most of these works base their visualizations on some aspects of amount of participation in the conversation. Assuming that evenly divided participation among team members is advantageous (Kelly & Duran, 1985; McLeod et al., 1992), these systems were designed to encourage participants to equally "share the floor" during team conversations.

For instance, the Babble system represents individuals in an electronic discussion room as colored dots around a circle that represents the group, placing individuals closer to or farther from the center based on their level of activity in the discussion (Erickson et al., 1999). Similarly, Chat Circles presents chatroom participants as circles that grow and become brighter for more active participants (Viégas & Donath, 1999). The Participation Tool visualizes team members' level of participation based on the number and length of messages they post in a chatroom (Janssen, Erkens, Kanselaar, & Jaspers, 2007). Other systems derive their visualizations from audio input to visualize amount of participation and turn taking patterns, such as Conversation Clock (Bergstrom & Karahalios, 2007a), Second Messenger (DiMicco, Pandolfo, & Bender, 2004), and Meeting Mediator (Kim, Chang, Holland, & Pentland, 2008).

When evaluated through user studies, these systems were shown to provide useful feedback to teams, helping users reflect on and become aware of their participation behaviors, and in some cases changing them with the attempt to equalize participation levels. For example, when seeing visualizations of their participation patterns based on audio input in Second Messenger, talkative members were able to decrease the amount of their participation and quiet participants spoke more (DiMicco, Hollenbach, Pandolfo, & Bender, 2007). In a study of Conversation Clock, participants found interest in monitoring their interaction behaviors by looking at the dynamic visualization, and altered either the length or number of their turns, trying to balance participation in the group (Bergstrom & Karahalios, 2007b). The Meeting Mediator was also shown to help equalize participation levels and reduce overlapping speech among team members (Kim et al., 2008). Finally, use of the Participation Tool, visualizing number and length of chatroom messages, was shown to be associated with more equal distribution of long messages among team members and with identifying free riders in the team (Janssen et al., 2007).

Automated Linguistic Indicators of Teamwork Behaviors

The tools and studies discussed above are promising in that they demonstrate the potential of dynamic feedback visualizations to stimulate reflection on and encourage behavioral changes toward equal participation in a group. However, looking only at nonverbal behaviors, and in particular at participation patterns, misses important aspects of the team interaction process. For instance, Bales' IPA framework (1950) distinguishes between interaction acts that support the completion of the task and that address socio-emotional team dynamics. Higher or lower levels of participation of team members do not necessarily indicate whether they are contributing toward accomplishing the task or whether they are helping (or hindering) to improve the team's social well being. There is more in the team interaction process than mere participation patterns, which can be revealed when the content and the language are captured and available for analysis.

Previous research has shown that linguistic analysis can uncover features that correspond to social behaviors and team performance. For example, Martin & Foltz (2004) used Latent Semantic Analysis (LSA), a computational linguistic technique that measures the semantic similarity between bodies of text, to predict team performance in a simulated UAV flight based on similarity of the entire team discourse with other performance-rated discourses, and by categorizing members' statements into classes that are correlated with team performance. Using the same data, Gorman et al. (2003) used LSA to develop indicators such as *communication density* and *lag coherence* that they found to be correlated with team performance. These indicators evaluate, respectively, the degree to which meaningful information is communicated concisely and the degree to which utterances shift from one topic to the other between and within communicators. Gorman et al. suggested feeding back these indicators to teams and their supervisors in real time in order to "quickly pinpoint shortcomings in team information processing." However, these works used manually transcribed team conversations, since automated speech recognition tools for unstructured group conversations typically achieve lower performance scores (Eide, Gish, Jeanrenaud, & Mielke, 1995).

Further, machine-learning approaches have shown promise in identifying dialogue contributions in which learners construct arguments that build on others' contributions in collaborative chat learning settings (Joshi & Rosé, 2007). The alternative rule-based approach for automated text analysis has been used to identify, in real-time, when a team expresses more agreement or more critical discussion in synchronous online conversations (Janssen, Erkens, & Kanselaar, 2007). The latter approach has shown promising results in visualizing this information to users, leading to more positive perceptions of the team collaboration process, a different set of interaction acts, and improved performance in certain parts of their tasks (Janssen et al., 2007).

The category-based word count approach is a widely used technique to automatically identify linguistic features. In this technique, developed by Pennebaker and colleagues into the Linguistic Inquiry and Word Count (LIWC) program (Pennebaker, Francis, & Booth, 2001), dictionary words are mapped into content and style categories such as emotionally-charged words, self-references and other pronouns, and assents. For example, the words "office", "class", and "product" are mapped into the content category "work", and the words "and", "also", and "plus" are mapped into the style category "inclusion". LIWC counts what percentage of words in a block of text fall into these categories, and unlike advanced natural language processing techniques such as LSA and machine learning, can produce linguistic markers in near-real time.

Pennebaker argues that linguistic features based on word choice, as captured by some of the LIWC categories, can serve as measures of conversation style and social behavior (Pennebaker, 2002; Pennebaker et al., 2003). For example, he suggests that:

> The use of 1st person singular (I, me, my) versus 1st person plural (we, us, our) provides insight into people's social identity and "ownership" of their speaking or writing topic. By the same token, references to other people suggest an awareness and, often, integration with others. (Pennebaker, 2002, p. 8)

This premise was previously examined. It has been shown that the use of collective versus individual first person pronouns ("we" vs. "I") can indicate level of involvement in a dyadic conversation (Cegala, 1989), a sense of group belonging (Cassell & Tversky, 2005), and the degree of perceived inter- or independency in marital relationships (Sillars, Shellen, McIntosh, & Pomegranate, 1997). Further, it has been shown that increased use of first person singular pronouns is associated with increased attention individuals pay to themselves rather than to the external environment (Davis & Brock, 1975) and a higher probability that a message in an online community will receive a reply (Burke et al., 2007), and that decreased use of such pronouns is associated with psychological distancing surrounding September 11, 2001 in blog writings (Mehl & Pennebaker, 2003a) and with deceptive communication in instant messaging (Hancock, Curry, Goorha, & Woodworth, 2005).

Other linguistic features based on word-level analysis were examined in the context of group and dyadic situations. For example, design team members who used more emotionally-charged words in chat conversations were perceived by their peers as poor collaborators, and teams reduced the use of such words after participating in peer feedback procedures (Turner & Schober, 2007). Also, dating couples used more emotion words in their instant messaging communication when they were keeping a journal of their feelings and thoughts, and their use of positive emotion words was correlated with relationship stability (Slatcher & Pennebaker, 2006). Finally, lying in an instant messaging conversation was shown to be associated with more sense-related words, e.g., seeing and touching (Hancock et al., 2005).

Similarity measures of use of certain categories within groups were also examined. Pairs exhibiting high trust between them were similar in their use of words representing optimism (a sub-category of positive emotion words), whereas pairs exhibiting low trust were similar in their use of negative emotion words in instant messaging conversations (Scissors, Gill, Geraghty, & Gergle, 2009). Examining style categories, such as prepositions and articles, it was shown that dyads tend to match each other's linguistic style in a computer-mediated conversation (Niederhoffer & Pennebaker, 2002). This measure of linguistic style matching (LSM) was also shown to be positively associated with cohesiveness within 4-6 member groups communicating face-to-face and via chat (Gonzales, Hancock, & Pennebaker, in press).

These works, demonstrating relationships between linguistic indicators, specifically at the word level, and aspects of interpersonal and social behaviors, suggest that they can be serve as a valuable source of feedback during the team interaction process. Such feedback, based on automatically distilled linguistic features that are proxy representations of teamwork behaviors, follows the criteria of effective feedback outlined above: dynamic, credible, and detail-level, action-specific. Which specific linguistic features to use is a matter of investigation applied in the current research, which will have to demonstrate connections between the features and relevant teamwork behaviors, as well as the degree to which team members can adapt their use of language on the dimensions indicated by the features. Further, changing linguistic behavior upon these features needs to occur in a direction that improves, rather than impedes, teamwork behaviors, another issue examined in the current research.

Research Questions

The literature reviewed here brings together three concepts, including teamwork, feedback, and language, with theoretical and empirical foundations that connect between them only to a partial extent. The first connection is between feedback and teamwork: theories of effective teamwork and of feedback and empirical evidence demonstrate that appropriate feedback can guide teams toward improving

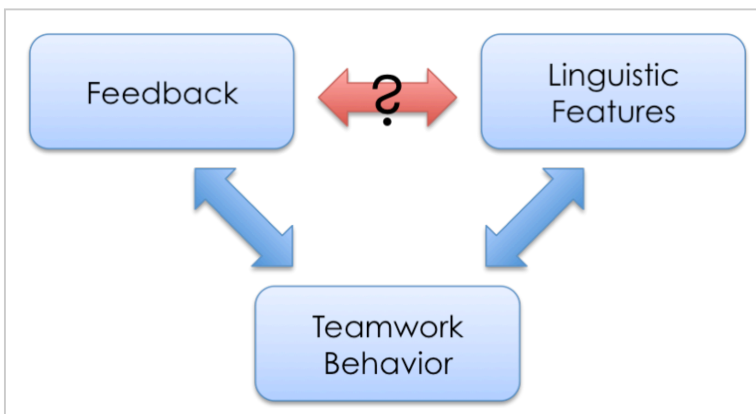


Figure 3. The three elements explored in this research—feedback, linguistic features, and teamwork behaviors—and studied vs. unexplored connections between them. Previous work shows the efficacy of feedback based on observations or automated analysis to improve teamwork behaviors, as well as associations between linguistic features and teamwork behaviors. In this research I add the third, missing, link: using linguistic analysis that produces indicators of teamwork behaviors as a source of feedback for improving these behaviors.

their interaction process and performance. The second connection is between the products of linguistic analysis and teamwork, highlighting relationships between word choice and social behaviors in group interactions.

The missing link is, therefore, bringing together these two sets of theoretical and empirical efforts (Figure 3). This consists of using automated linguistic analyses such as word level approaches to reveal indicators that would serve as a source of feedback to help team members attend to, reflect on, and improve their communicative and teamwork behaviors. Feedback based on the word count approach can provide dynamic and detailed information about teamwork behaviors (Pennebaker et al., 2003), and as such is in line with principles of Feedback Information Theory (Kluger & DeNisi, 1996) and reflection-in-action (Schön, 1983). Provided with appropriate benchmarks, for example, that offering reasoned criticism is important for enhancing teamwork (Wegerif, Mercer, & Dawes, 1999), this kind of feedback also follows concepts of self-regulation such as in goal-setting theory (Locke & Latham, 1990), social cognitive theory (Bandura, 1991), and self-regulated learning model (Butler & Winne, 1995).

My research is proposed to bridge the gap that exists in the previous literature in connecting the three concepts of teamwork, feedback, and language, and it addresses several questions that emerge from this intersection. The overall two key questions I explore in this research are:

RQ1. Does providing automated linguistic feedback based on word choice stimulate team members' reflection on their use of language in a teamwork setting?

RQ2. Does this feedback motivate team members to change their use of language in ways that may lead to improved teamwork behaviors?

Based on previous theories and empirical results, I designed GroupMeter, a tool that presents dynamic feedback visualizations of linguistic indicators during a computer-mediated teamwork conversation. Using GroupMeter as a research platform, the experiments that follow were designed to address the two key research questions formulated above, as well as additional questions necessary for achieving the purpose of this research:

1) Which linguistic features correspond to teamwork behaviors and can be a valuable source of feedback to teams? This question was examined in the first experiment, by correlating linguistic analyses of team conversation transcripts with peer-rated teamwork behaviors, and examining how language changed as a response to peer feedback (Chapter 4).

2) How do users experience and interpret automated linguistic feedback? This question was examined in experiment 2, in response to findings in which participants reflected on but did not change their use of language when receiving feedback, and in experiment 3, which examined user responses to two feedback designs (Chapter 5).

3) Does visualizing dynamic linguistic feedback distract team members from working toward their task? How can we design feedback that would help balance between task focus and attention to social team interaction? These questions were addressed in experiment 3, comparing two different visualizations of the same feedback information (Chapter 5).

4) How can automated linguistic feedback help guiding team members to change their behaviors in a particular direction? This question was examined in experiment 4, in which guidance toward favorable teamwork behaviors was provided and examined in interaction with automated linguistic feedback (Chapter 6).

Taken together, the results of all four experiments suggest factors important for understanding the potential of automated linguistic feedback to support reflection on and change of language use and teamwork behaviors, shedding light on theoretical, practical, and design aspects of the relationship between teamwork, feedback, and language.

CHAPTER 3

THE DESIGN RATIONALE AND EVOLUTION OF THE GROUPMETER SYSTEM

In the preceding chapter I described how automated feedback on word choice could potentially help team members reflect on their use of language and improve their teamwork behaviors. Following these theoretical foundations, I designed GroupMeter, a research platform that instantiates these concepts and enables the empirical investigation of the relationship between teamwork behaviors, language use, and feedback.

The basic design of GroupMeter consists of a web-based system in which groups communicate through a chatroom to perform their tasks and receive feedback through a visualization that appears near the chatroom window. The feedback visualization represents linguistic features that are automatically extracted from the chatroom conversation text, and changes dynamically as team members change their use of language during the conversation.

This chapter is composed of two sections. The first describes and discusses the requirements that inform the design of GroupMeter. In the second section, I report on the iterative nature in which the design and implementation of the GroupMeter system evolved as a response to the experiments I carried out in this research. A full description of the experiments and their results appears in the following chapters.

Design Requirements

In this section, I present a set of requirements I identified as important for the design and implementation of GroupMeter. These requirements are: 1) Computer-mediated text-based communication, 2) Dynamic display of feedback, 3) Peripheral display of feedback, 4) Public display of individual-level feedback, 5) Context-aware not, and 6) Modular technology architecture. The requirements were developed based on theories and prior empirical work from multiple disciplines, including human-computer interaction, social psychology, cognitive science, and software engineering, among others.

1. Computer-Mediated Text-Based Communication

The first requirement for the design of GroupMeter is that it should incorporate computer-mediated text-based team communication. This requirement follows both theoretical and practical considerations. From a theory perspective, I argued that distributed teams that use groupware technologies to collaborate experience a variety of challenges, and need to develop teamwork skills for working together in mediated environments. Therefore, in order to build the necessary skills for communicating in computer-mediated environments, teams need to practice working in such situations. This is congruent with situated learning theory, which stresses that effective learning occurs in the same context in which it is applied (Lave & Wenger, 1991).

From a practical point of view, a system needs to be able to capture the team conversation in order to process its content and provide feedback about the language used by team members in real time. DiMicco et al. (2004), Bergstrom & Karahalios (2007a), and Kim et al. (2008) all used sets of microphones to capture the conversation in a face-to-face meetings. However, they did not process the content or language of the conversations, but instead provided feedback based on audio cues. On the other hand, Martin & Foltz (2004) and Gorman et al. (2003) processed manually transcribed conversations, ruling out the option of providing real-time feedback to teams.

As long as speech recognition tools do not yield satisfactory outcomes for conversational speech (Eide et al., 1995), synchronized text communication is necessary for analyzing conversational language in real time. Further, mediated text communication is not necessarily a limitation in terms of the interpersonal relations that develop as compared to face-to-face (Walther, 1996), and is widely used these days in forms such as email, text-messaging, and instant messaging. Based on these reasons, I chose to employ chat, a synchronized text-communication medium that allows several participants to take part in a conversation, as the primary communication medium for GroupMeter.

2. Dynamic Display of Feedback

The second requirement for the design of the GroupMeter interface is that the feedback provided to team members on their language use should be dynamic. By that I mean that feedback is presented continually and updates when the data underlying the feedback information changes: as team members intentionally or unintentionally modify their use of language throughout the conversation, the feedback should seamlessly reflect these changes. Consistent with Feedback Intervention Theory (Kluger & DeNisi, 1996), dynamic feedback enables individuals to see more clearly how their behaviors are linked to the feedback they receive. Further, it can support a metacognitive process, in which, by monitoring and reflecting on their behaviors, individuals construct knowledge about their own learning process and behaviors, and as a result can consciously control behavioral changes they would like to make (Brown, 1987; Flavell, 1979).

According to Gersick's punctuated equilibrium model of group development, teams show little change in their behavioral patterns during the process of working together (Gersick, 1988; Gersick, 1989; Gersick, 1991). Instead, naturally occurring transition points in the life spans of teams provide them with opportunities to revise their methods and alter their behaviors to the next phase of their collaborative work. The dynamic nature of feedback does not assume that any specific times are appropriate for reflection and transition, but leaves it to the team members to recognize these points based on their monitoring of the feedback display and the team process.

3. Peripheral Display of Feedback

The next design requirement for the GroupMeter interface is that feedback should be displayed at the periphery of the primary team communication medium. This requirement is based on notions of what constitutes effective teamwork, cognitive processing in dual-task settings, and previous works in HCI on peripheral displays.

According to Bales' Interaction Process Analysis (1950), effective teamwork involves balancing between carrying out task-related behaviors and socio-emotional behaviors, aimed at completing the task and maintaining the team well being, respectively. While it is easy to focus on task-related behaviors, nonconscious processes shape our socio-emotional or maintenance behaviors (McLeod & Kettner-Polley, 2004). Maintaining awareness of the socio-emotional behaviors is therefore important for the effective operation of a team, and can be achieved by displaying information about such behaviors at the periphery of the task-oriented communication medium. Dynamic feedback about social interaction behaviors presented at the periphery of the task-oriented communication medium can therefore serve as a mirroring tool (Soller, Martinez, Jermann, & Muehlenbrock, 2005), keeping awareness of individuals' and teammates' behaviors in the periphery while working on a primary task (Gaver, 2002; Plaue et al., 2004).

Presenting awareness information in a peripheral display requires individuals to divide their attention between the primary and the secondary information. This involves two cognitive processes that need to be taken into consideration in the design of the peripheral display: resource allocation and shifting (Wickens & Hollands, 2000). First, our ability to attend to several resources of information simultaneously is restricted (Broadbent, 1971), requiring to allocate the limited cognitive resources to the primary and secondary information. Efficient allocation depends both on the individual's skills and on intrinsic properties of the information, demanding more or less cognitive resources (Navon & Gopher, 1979). The second process refers to the cost of switching between tasks (Rogers & Monsell, 1995), resulting in a tendency to continue lower-priority tasks longer than necessary in order to avoid switching (Wickens & Hollands, 2000). Therefore, successful peripheral display design allows efficient resource allocation as well as smooth attention shifting back and forth between the central and peripheral sources of information.

One way to present peripheral information is using the principle of glanceability, defined by Matthews as "enabling quick and easy visual information uptake" (Matthews, Rattenbury, & Carter, 2007). Glanceability refers to the interpretation of information after the user has paid attention to the display (Matthews, 2007). A glanceable display helps users monitor their secondary tasks while they multitask, and can be achieved, for example, by using simple renditions with clear colors (Matthews, 2007).

A different model of peripheral systems uses a classification framework over three orthogonal dimensions: Interruption, Reaction, and Comprehension (IRC, McCrickard, Chewar, Somervell, & Ndiwalana, 2003). Interruption is defined as the event promoting transition and reallocation of attention focus from the primary task to the secondary task. Reaction is the rapid and accurate response to the stimuli provided by the peripheral display. Comprehension refers to making sense and remembering the peripheral information at a later time. When designing a peripheral display, one should take into account where on each of the dimensions the display needs to be, and design the display accordingly. For example, Consolvo et al. designed a peripheral display of one's own physical activities on their mobile phone screen using the principle of unobtrusiveness – that the information should be available to the user whenever they need it, without interrupting or calling attention to it (Consolvo, McDonald, & Landay, 2009). Using the IRC framework, GroupMeter needs to be high on comprehension in that it should promote reflection and understanding of the information it conveys, low on reaction in that no immediate response is mandatory, and intermediate on interruption, balancing attention between task and feedback.

A profusion of peripheral displays has emerged in the past two decades, supporting awareness of one own's background activities as well as social awareness of others' activities. Systems such as Sideshow (Cadiz, Venolia, Jancke, & Gupta, 2001) and digital handwritten notes (Hsieh, Wood, & Sellen, 2006) present information for personal use on the edge of one's desktop or on a second monitor. Social awareness through peripheral displays was first supported by video streaming (Dourish & Bly, 1992), and later by iconic representations (Greenberg, 1996). Awareness of group behaviors during team meetings was supported by peripheral displays projected on a screen (DiMicco et al., 2004), a table (Bergstrom & Karahalios, 2007a), or in group members' cell phone screens (Kim et al., 2008).

A question may arise of why not present information using different modalities, such as audio or tactile. In a seminal work, Allport, Antonis, & Reynolds measured response time to multiple signals, showing that people can attend to signals from two distinct modalities, i.e., visual and auditory (1972). Sweller transferred these ideas from cognitive psychology to the education discipline, arguing that because working memory is limited in the number of elements it can contain simultaneously, reducing the working memory load can lead to effective learning (1988). One way to reduce cognitive load is by combining input modalities, such as visual and auditory information (Mousavi, Low, & Sweller, 1995).

Previous works in HCI demonstrated successful use of different modalities to convey peripheral awareness information, using haptics (Weiser & Brown, 1996), audio (Alexanderson, 2004; de Guzman, Yau, Gagliano, Park, & Dey, 2004), and olfactory (Bodnar, Corbett, & Nekrasovski, 2004), as well as modality combinations (Dahley, Wisneski, & Ishii, 1998; Ishii & Ullmer, 1997; Ishii et al., 1998; Pedersen & Sokoler, 1997). But mixing input modalities in a dual-task setting does not necessarily lead to superior performance (Wickens & Liu, 1988; Wickens, 1991). For example, there is evidence that non-linear presentation of information can be advantageous for learning (Stanton, 1994), but auditory information is provided in a linear dimension. Employing principles such as glanceability and unobtrusiveness as well as concepts of visual design (Tufte, 1983; Tufte, 1990) can help create a peripheral display that conveys complex information in a simple, visually appealing way.

4. Public Display of Individual-Level Feedback

In order for the feedback information to have more influence on team members, both in terms of their reflection about and change of their behaviors, I followed a decision to present feedback at the individual-level to all team members. That is, every member should see the feedback information of every other team member, rather than just seeing their own feedback or seeing an aggregate of the team. This requirement is based on the assumption, derived from the established social comparison theory (Festinger, 1954), that by presenting public information about individuals, members can interpret the feedback about their own behaviors in comparison to the behaviors of others.

When everyone sees the feedback all others received, it draws attention to the individuals in the team. Increased attention to members who deviate from the majority was shown to be associated with positive influence of the minority members on the majority opinions and on the group decision (McLeod et al., 1997). Further, based on functional leadership theory (Hackman & Wageman, 2005), a stronger team member, perhaps functioning as a team leader, can take advantage of seeing how others are doing and intervene actively to motivate and direct others to change their behaviors (Klein, Ziegert, Knight, & Xiao, 2006). Public display of feedback can also motivate individuals with lower feedback than others, stimulating them to change their behavior and increase their productivity (McLeod & Kravec, 2008).

The presence of public individual-level feedback is also aligned with the notion of social translucence for designing systems that support social interactions (Erickson & Kellogg, 2000). Social translucence emphasizes making social information visible within a system, supporting the adherence to social norms through awareness of others and accountability of the individual's own behaviors.

5. Context-Aware Not

An important requirement applied in the design of the GroupMeter interface is that it should avoid modeling contextual factors as much as possible. GroupMeter is comprised of two basic interface elements: the chatroom through which team members communicate, and the feedback visualization presented near the chat window. The feedback visualizes data on language use, directly analyzed from the chat communication text. To this end, there are many contextual factors that could have, but have not been modeled into the system: factors at the individual level, such as members' gender, personality, or skills; at the group level, such as its size, hierarchical structure, roles, and development phase; and exogenous factors, such as the task assigned to the group and its characteristics, the reward structure, and cultural norms of using language. Just as the chatroom through which teams communicate does not know if the team is brainstorming ideas or trying to reach consensus, so should the feedback visualizations.

By leaving out contextual factors, one can create a system that is applicable to a wider range of settings and that allows for richer interpretation and reflection on the experiences it bears:

> Although a designer is not able to control the general context in which a person will use his product, this context can influence the experiences of the users when interacting with this product. (Overbeeke, Djajadiningrat, Hummels, Wensveen, & Frens, 2003, p. 10)

Ranganathan et al. argue that modeling contextual information and adding them to the interface may enrich chat conversations (Ranganathan, Campbell, Ravi, & Mahajan, 2002). Modeling and designing these factors into the system can also reduce the costs of grounding (Clark & Brennan, 1991), establishing mutual understanding about such factors. For instance, the presence of an expert in the team may suggest a social norm that this person should talk more than others when discussing an issue related to his or her expertise. In another setting, for example when trying to reach mutual consensus, equal participation is considered to improve the group process and task performance (McLeod et al., 1992).

The decision not to model such contextual factors was deliberate and follows two primary ideas. First, introducing such factors may restrict the range of settings in which the system can be applied. For instance, if the system encapsulates a model in which team members enact certain roles such as leader, expert, and peer, it might not be applicable for teams in which some of these roles are absent or other roles are present. As another example, Group Decision Support Systems (GDSS) may be useful in helping teams reach high-quality decisions (DeSanctis & Gallupe, 1987; McGrath & Hollingshead, 1994), but if the team is to carry out a task with different characteristics such as brainstorming or planning (McGrath, 1984), the tool may be less useful. As such, leaving contextual factors out of the system allows for greater control by teams to appropriate the technology for their own purposes (Poole & DeSanctis, 1989). For instance, if GroupMeter provides feedback about levels of agreement expression, team members can construct the interpretation of high or low levels of agreement based on the task and the phase of the group development. They can perceive high agreement as beneficial when attempting to reach consensus, and as detrimental when trying to critically discuss and negotiate solutions.

Second, the decision not to represent contextual factors in the system corresponds with the idea from critical design in HCI that technologies should not necessarily be designed to “convey a single, specific, clear interpretation of what they are for and how they should be used and experienced” (Sengers & Gaver, 2006, p. 99). Instead, ambiguity enables the user to develop multiple understandings of what the system is for and how they should use and experience it in different contexts (Gaver et al., 2003). This is also in line with the idea of interpretative flexibility: “not only that there is flexibility in how people think of or interpret artifacts but also that there is flexibility in how artifacts are *designed*. There is not just one possible way or one best way of designing an artifact” (Bijker, Hughes, & Pinch, 1987, p. 40). For instance, when providing feedback about team members’ use of the word “I”, users can interpret it as self-focus or as high involvement in the conversation. This can enrich users’ experiences when interacting with their teammates through GroupMeter, although, as I discuss later, it also entails a challenge in guiding teams to change their language use and improve their teamwork behaviors in a certain direction.

6. Modular Software Architecture

Finally, GroupMeter was designed and developed in response to the requirement that its underlying infrastructure should be modular to the extent that each of its modules could be replaced with minimal changes to the interface and the backend code. This requirement, drawn from software engineering, enables "loose coupling", such that different software components are independent of their surrounding entities (Jacobson, 2004), and as a result changes in the design have a local effect (Sommerville, 1992).

The three main software components in GroupMeter are the communication medium, the feedback visualization, and the linguistic analysis. A modular architecture of the system allows flexibility in implementing and refining the system, such that any one component can be changed with minimal modifications required to the other components. For example, switching from word-based linguistic analysis to semantic LSA similarity measures, would still be based on using the conversation text from the communication medium, and could feed into the same feedback visualization.

Iterative Design and Implementation

In this section, I describe how the design and implementation of GroupMeter evolved over several iterations during three years. Some changes were motivated by technical and design issues the design and development teams faced. Other changes were informed by findings from the experiments, reported in the following chapters, to make the system fulfill its promise of stimulating reflection on and changes of teamwork behaviors using dynamic automated linguistic feedback.

Early design

The first year was dedicated to the architecture and implementation of the backend infrastructure of the GroupMeter software. This included a web client-server architecture (Figure 4), in which the server, running under Apache Tomcat, includes modules for managing a team meeting session, the chat communication, and the linguistic processing of the chat conversation text. It also included the design and construction of the database and the administrator interface, allowing an authorized user to add user accounts, assign them to teams, create sessions for teams to

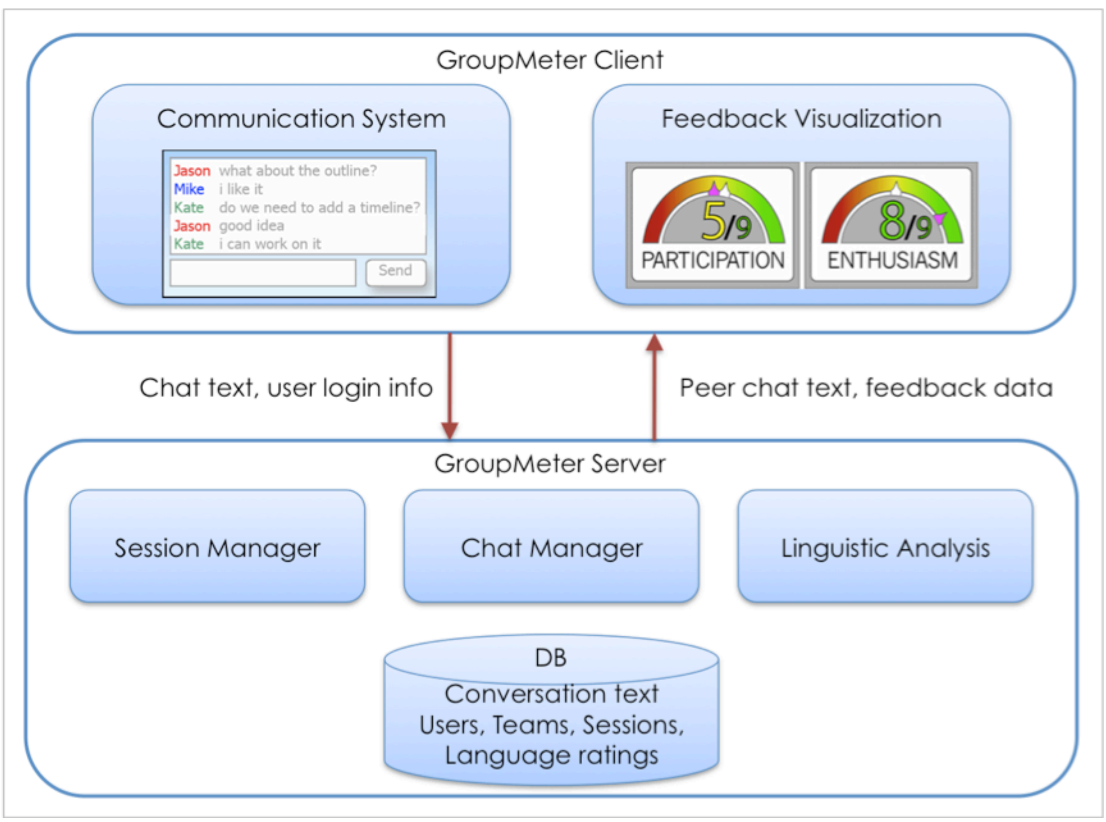


Figure 4. The GroupMeter software architecture.

communicate, and change parameters of the sessions and the feedback.

The design of the client side user interface was roughly sketched by students from The Parsons School of Design (Figure 5), and was presented in a workshop at CSCW 2006 (Leshed et al., 2006). At this point the idea was to design the GroupMeter feedback features in a browser window external to an existing chat client, such as AIM or iChat. This was based on the thinking that many users are familiar with such applications, but it also imposed two challenges. First, the designers needed to create a feedback visualization that is independent of the chat interface. This left them with less control over where the visualization would be located on the screen relative to the chat interface and whether it would be hidden by the chat window or

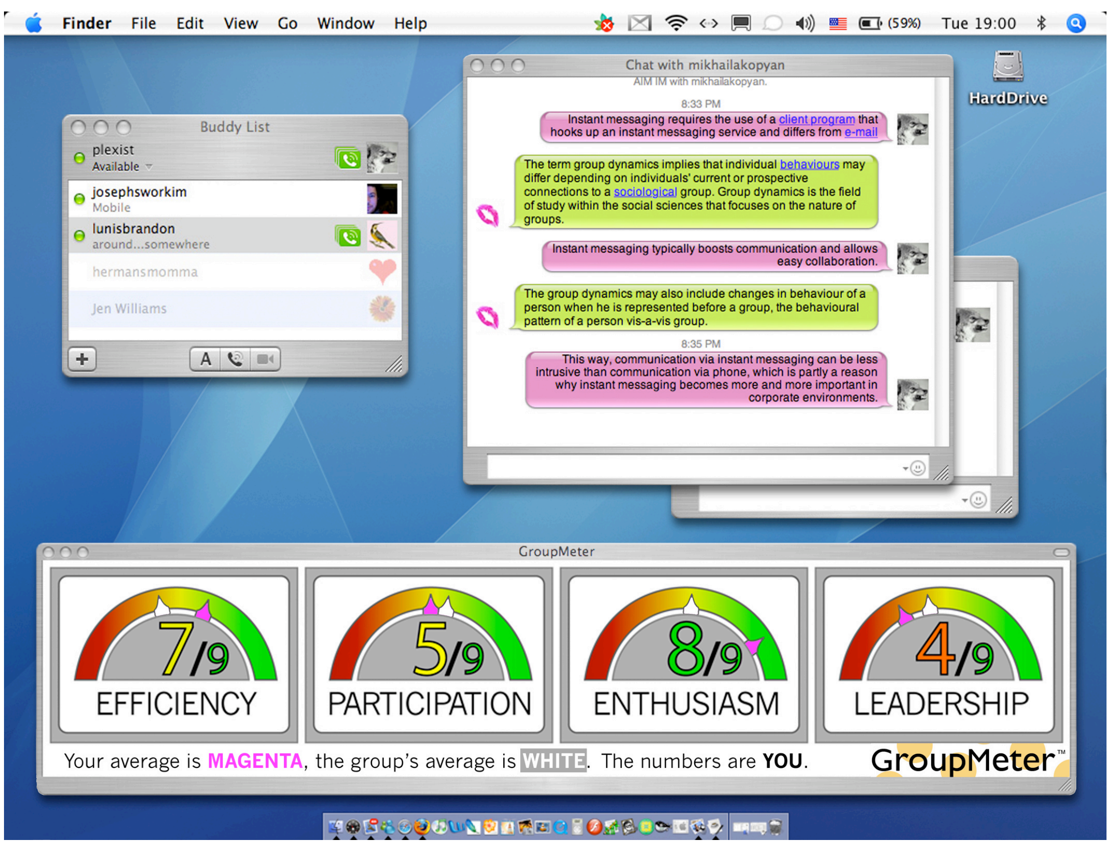


Figure 5. Early design of GroupMeter: external, off-the-shelf chat interface with feedback dials in a separate window.

other windows. Second, on the technical side, a bot had to be created as an interface between GroupMeter and the off-the-shelf chat client. This enabled the GroupMeter backend to capture the conversation text within the chat client, but was found to be technically risky since changes in the chat API would not controlled by the GroupMeter team but by the provider of the chat application.

The design sketch followed the requirement of visualizing the feedback in a peripheral display, and was based on the metaphor of a car dashboard, one dial for each dimension. At that point, prior to experiment 1 (see Chapter 4), I thought to translate the linguistic features from the automated analysis directly into teamwork behaviors such as those defined by Freeman & McKenzie (2002), and to combine them with peer evaluations. However, this visualization remained in the sketching phase and did not reach the implementation phase.

Version 1: Web-Based Chat with Bar-Graph Feedback Visualizations

Few issues in the early design were addressed in the first functioning version of GroupMeter. First, to create a more coherent design and to overcome technical issues in connecting with external proprietary chat protocols, the development team made a decision to implement a chatroom embedded in a web browser. Together with a designer, we created a new design for GroupMeter that included the fish logo, the color scheme, the chat window, and the feedback bar meters, shown in Figure 6. In this design, presented at GROUP 2007 (Leshed et al., 2007), every team member is associated with a color that appears in their name, as a colored star in front of their chat entries, and in the feedback bars.

Further, the green-red dials shown in Figure 5 imply certain norms to be achieved by team members – toward the green and away from the red. This design was based on the assumption that providing benchmarks for interpreting the feedback would create a perception of an outcome-goal discrepancy, guiding team members to change their behaviors in an attempt to close the gap (Locke, Shaw, Saari, & Latham, 1981; McLeod et al., 1992).

However, different situations call for different behavioral norms. For instance, one dial presents level of participation, assuming that high level of participation is always desirable, but in some cases keeping silent or listening to others can be more of a contribution than talking. Instead of tailoring the standards to each set of contextual factors such as task type, progress through a task, hierarchy, and cultural norms, I decided not to provide normative benchmarks. This follows the requirement that contextual factors should not be captured, assuming that this ambiguity will motivate people to create interpretations for their own needs (Gaver et al., 2003), and allow teams to appropriate the system for their own purposes (Poole & DeSanctis, 1989).

Second, Figure 5 shows the individual's score on each meter compared to an

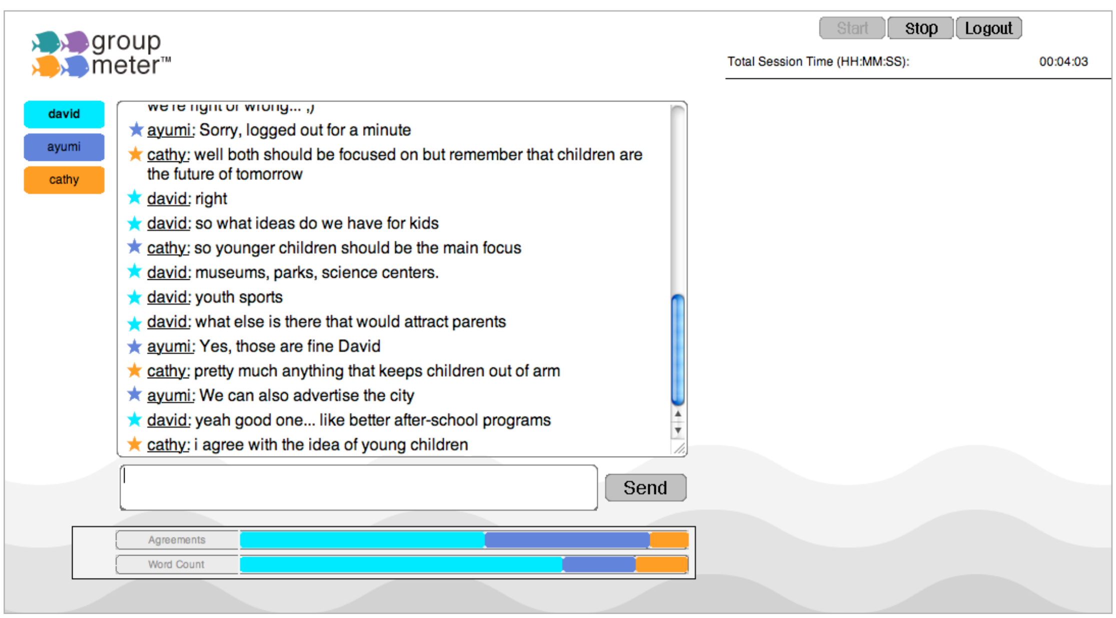


Figure 6. Version 1 of GroupMeter: web-based chat with bar-graph visualizations of the feedback at the bottom.

aggregate average of the group. However, as discussed above, I theorized that public feedback at the individual level would allow team members to compare themselves to specific others in contrast to an abstract aggregate (Festinger, 1954), and reflect on both their own and their teammates' behavior. At the same time, I wanted to ensure the design did not encourage too much competition arising from team members comparing themselves to others, which might result in negative interpersonal processes such as low trust, low coordination of effort, and attempts to mislead others (Johnson & Johnson, 1975).

The feedback visualization in version 1 consists of horizontal bar graphs, one for each feedback dimension. The bars change in length based on each team member's behavior on the dimensions chosen by the administrator. The linguistic analysis that feeds into the bars operates every one minute on the text entered by each member in the past five minutes, creating a moving average. We chose to present each feedback dimension as an aggregate stacked bar to reduce the possibility that people would process the visualization in a competitive way. An alternate display using a standard bar chart might cause people to meticulously compare the length of their bars to others' instead of referring to the general length as compared to others. Further, showing individuals' behavior as an aggregate bar emphasizes the idea of being part of a group. Locating the bars at the bottom of the chat window allowed us to place them on-screen without the need for scrolling.

The front end of version 1 was fully implemented in the second year using Ajax and became operational for carrying out experiment 2 and experiment 3, reported in Chapter 5.

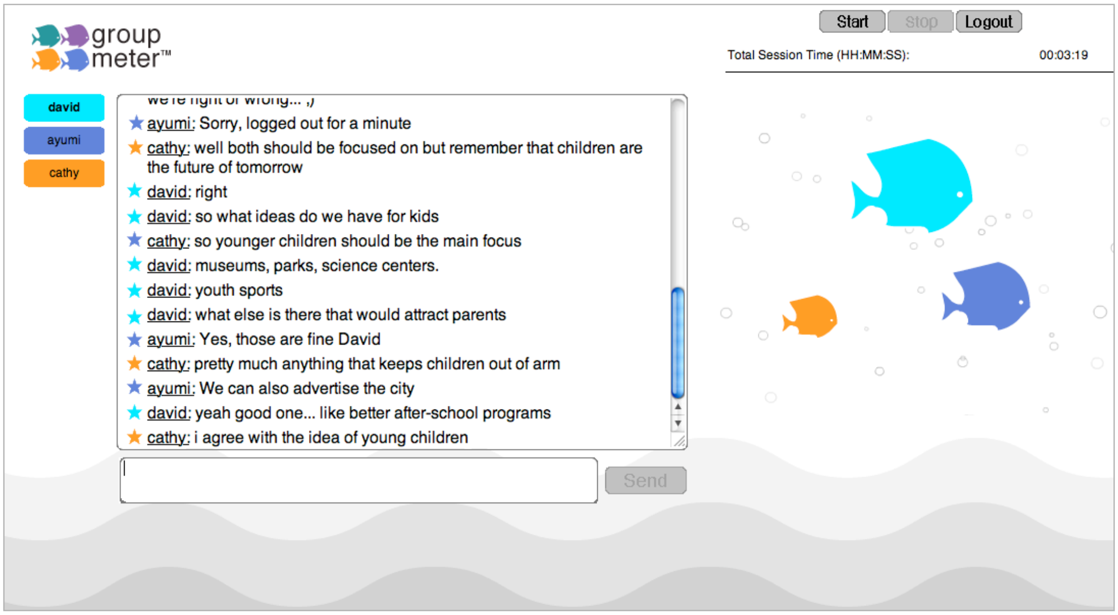


Figure 7. Version 2 of GroupMeter: web-based chat with fish visualizations of the feedback in a circular position.

Version 2: The Fish Metaphor

The results of experiment 2 showed that the feedback about language use stimulated reflection on teamwork behaviors, but they did not demonstrate that people changed their communication patterns when receiving linguistic feedback through the bars. One reason could be that people may have made subtle or inconsistent changes that the experiment's measures did not pick up. Another possibility is that the bar graphs, located unobtrusively at the bottom of the chatroom, were too easy to ignore.

To address the second possibility, I worked in the third year with a new development team on how to make the visualizations of the feedback more engaging and leading to change in communication behaviors. In an attempt to design a more playful and interesting feedback visualization, I used the metaphor of a school of fish swimming together. This metaphor, in which every person is represented by a fish, was used before in persuasive technology to promote physical activity (Lin, Mamykina, Lindtner, Delajoux, & Strub, 2006) and in a shared display to represent workplace activity (Farrell, 2001). I found the school of fish metaphor inspiring because it symbolizes “togetherness”, breaks away from conventional charts, and has a natural and serene connotation. I hoped that the school of fish would instill the feeling of being part of a group, using aesthetically pleasing visualizations that “tell a story about the data” (Tufte, 1983, p. 177).

In this visualization, colored fish represent individual team members, matching members’ colors in the chat window, as shown in Figure 7. The fish start in a circular formation, all at the same size and equidistant to the center. The fish visualization is animated, dynamically changing the size of the fish and their distance from the center depending on the feedback dimensions. This way, the visualization can represent two feedback dimensions. The team considered different movement axes, such as moving in an X-Y space as well as changing size. I chose the circular form instead of a horizontal/vertical movement, since I felt it conveys a better sense of unity and community, and as such can promote the feeling of “teamness”.

One purpose of the lively nature of the fish was to draw more attention to the feedback information represented by their position and size. The shape of the visualization again dictated its placement. Locating it below the chat window would have required users to scroll to see it; placing the visualization to the right of the chat window allowed it, like the bars in version 1, to be continually visible. As in the bars visualization, the fish update their size and position based on linguistic analysis that occurs every one minute, based on the past five minutes of conversation.

In the third year, the programming team fully implemented this design using Adobe Flash, making it operational for experiment 3, reported in Chapter 5 and presented at CHI 2009 (Leshed et al., 2009).

Version 3: Bubble Trails and Directive Guidance

The circular fish visualization have shown promising results in making users reflect on and change their communication behaviors, as reported in the results of experiment 3, in Chapter 5. However, to notice changes in the feedback visualizations users had to constantly monitor them, taking their attention away from the task.

To improve the glanceability of the display (Matthews et al., 2007), I came up with idea to enrich the feedback visualization with a history view (Hill & Hollan, 1994) showing how the feedback has been changing during the conversation. It has been argued that historical information is critical for understanding changes in social behaviors within a collaborative space (Smith, 1999; Viégas, Wattenberg, & Dave, 2004; Viégas & Golder, 2006). Further, this enables users to consider their own behavior not in isolation, but in relation to the trend they see in the historical

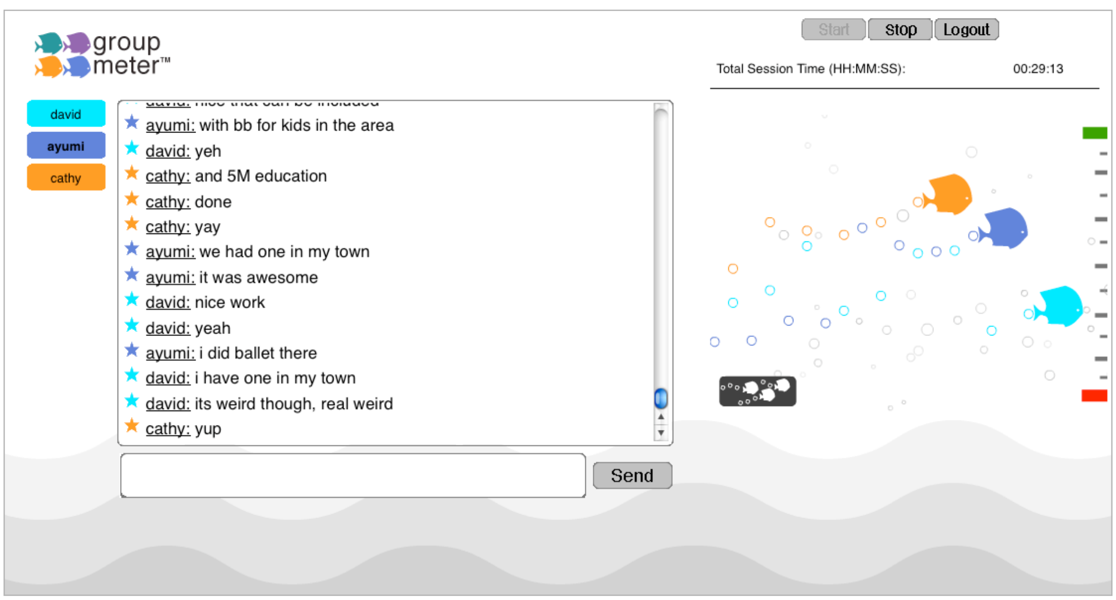


Figure 8. Version 3 of GroupMeter: fish move up and down in response to one feedback dimensions and leave bubble trails behind them as they update their position. A ruler on the right with green-red marks serves as directive guidelines.

visualization (Consolvo et al., 2009). In this way, persistency exists not only the text that remains available in the chatroom as the conversation proceeds, but also in the feedback visualization. Similar to a technique suggested by Bezerianos et al. (2006), if a user has been occupied by the primary team task and did not look at the display for a while, he or she could glance at the display and catch up with not only what is going on right now, but also the trend of the feedback over the conversation.

I chose to design the history view as trails of bubbles, shown in Figure 8. Every time a fish updates its position it leaves a bubble behind it in its previous location. To simplify the presentation of the history view and the number of feedback dimensions represented by the visualization, I designed the fish moving up and down on the vertical axis in response to one feedback dimension. As a result, the bubble trails appear on the horizontal axis, giving an impression of the fish swimming from left to right. Also implemented into this design is a button that appears at the bottom-left of the fish visualization. Clicking on the trail button opens a window with the full history view, since the space for the visualization only allows presenting the trend of the past nine minutes.

Another change in this version refers to the earlier decision to avoid designing for a single interpretation as to what counts as good or poor behavior, with the purpose of allowing teams to develop multiple meanings of the feedback in different contexts. However, the mix of multiple meanings participants came up with might have been one reason for no significant changes in communication patterns in experiment 2. This time I wanted to demonstrate how the feedback can guide team members to interpret the feedback and change their behavior toward a certain direction. Therefore, I added to the visualization a ruler that appears on the right hand side of the fish visualization. It presents green-red markers on the top and bottom of the ruler, which can be reversed by an administrator. The markers serve the purpose of experiment 4 (explained in Chapter 6), guiding users to change their behaviors to move their fish toward the green marker and away from the red marker.

I now turn to describe the experiments I carried out to address the questions underlying this research and to further inform the design of GroupMeter. Experiment 1 examines which linguistic features to present using the GroupMeter feedback visualizations. Experiments 2 and 3 are exploratory in nature, examining the general efficacy of the feedback visualizations of GroupMeter to elicit reflection and change in behavior. In experiment 4, I examine specific hypotheses about the promise of GroupMeter's automated linguistic feedback to help team members improve their teamwork behaviors.

CHAPTER 4
EXPERIMENT 1:
LINGUISTIC FEATURES FOR TEAMWORK BEHAVIORS[1]

The previous chapter describes the evolution of the GroupMeter research platform into a web-based chat system augmented with a dynamic, peripheral visualization of linguistic features automatically extracted from the chat conversation. It also reported on the key principles for designing the chat communication medium and the feedback visualizations. However, one key factor was left out: the linguistic features represented by the feedback visualization. One reason for this can be found in the sixth design principle: the software should be modular as much as possible. In other words, the automated analysis of the conversation text carried out in the backend of the system, producing linguistic features, should be independent of the visualization. The linguistic analysis in the backend provides the front-end with numbers on a set of one or more linguistic features, and the front-end translates the numbers into the visualization. To that end, the visualization is disconnected from the actual linguistic features, be it 1st person singular pronouns, agreement words, reasoning words, or other features.

The missing piece for operationalizing GroupMeter is therefore the linguistic features to be mapped into the feedback visualizations. This chapter presents an experiment that fills this gap. The purpose of the experiment was to find linguistic features extracted from team conversations that can serve as useful sources of automated feedback on teamwork behaviors.

---

[1] This chapter is based on a paper published in GROUP 2007 (Leshed, Hancock, Cosley, McLeod, & Gay, 2007).

I asked teams to perform a decision-making task using a chat application, and to provide peer feedback on each other's teamwork behaviors. I used the SYMLOG framework to collect peer evaluations on teamwork behaviors, asking group members to rate each other along the three dimensions it defines: dominance/submissiveness (also labeled *participation level*), individual/group orientation (*friendliness*), and task focus/socioemotional expressiveness (*task focus*) (Bales & Cohen, 1979). I then used Linguistic Inquiry and Word Count (Pennebaker et al., 2001), a dictionary-based tool for analyzing language features. LIWC counts what percentage of words in a block of text occurs in various content categories such as emotion words, self-references, and assents, that can be seen as measures of conversation style (see Chapter 2).

This experiment addresses two key research questions that a system that provides linguistic feedback must answer:

RQ1. What linguistic features, as indicators of conversation style, predict ratings of teamwork behaviors?

RQ2. Does feedback on teamwork behaviors affect subsequent conversation style, as measured by linguistic features?

By collecting the chat conversation and processing them with LIWC and recording explicit peer ratings of team members' behaviors, I was able to address these questions. First, associations between peer evaluations on the SYMLOG dimensions and the linguistic features provided by LIWC indicators provide an answer to RQ1. Second, changes in the LIWC indicators following peer feedback on teamwork behaviors addresses RQ2. Altogether, the results of this study enabled me to choose the linguistic features that satisfy these questions and can be rationalized to be embedded into the GroupMeter feedback visualizations.

Method

Participants. One-hundred and four undergraduate students (62 females, 42 males) at Cornell University volunteered to participate in the experiment for course credit and for a chance to win $40. Participants were randomly assigned to mixed-gender 4-person groups in one of two conditions: *Feedback* (*FB*, 13 groups) or *No Feedback* (*no-FB*, 13 groups).

Procedure. An overview of the experiment procedure appears in Figure 9. Upon arrival, participants were seated at isolated computer workstations. Participants were informed that they would be working as a team on a decision making task and that their part is very important for the team's success. The instructions included a description of the three SYMLOG dimensions and behavioral propositions associated with effective teamwork on these dimensions.

Groups used iChat to communicate over the task. In the task, *Lost on the Moon*, teams need to reach a decision with respect to the ranking of 15 items necessary for their survival as a team of astronauts on the moon. The correct solution

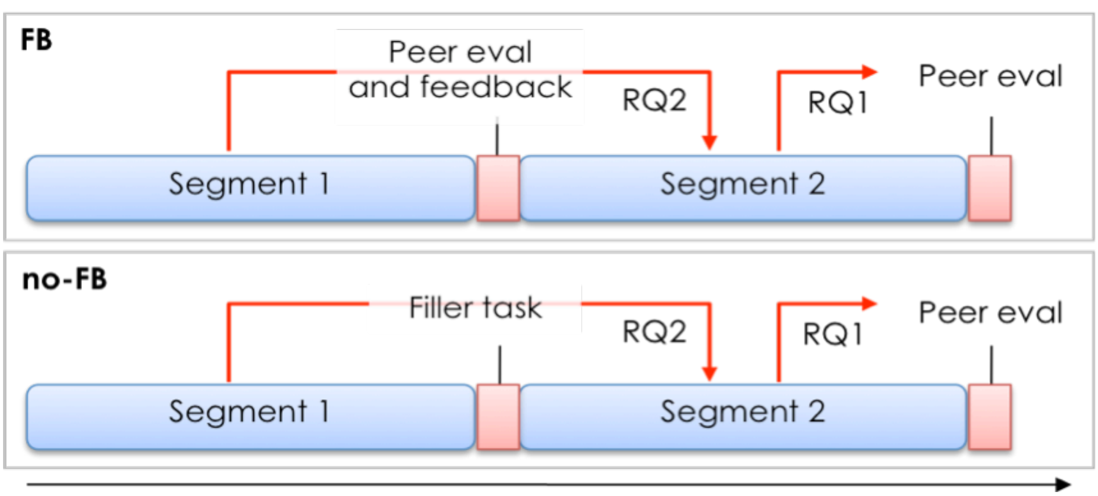


Figure 9. An overview of the procedure of experiment 1.

of this task, provided by NASA's experts, includes items such as oxygen and water at the top of the list, and compass and matches at the bottom. I chose not to reveal group members' actual names (to reduce potential gender biases); members were identified by a color: Blue, Red, Green, or Yellow.

An experimenter monitored the group's progress. When the group completed ranking seven items out of 15, they were prompted to pause the conversation and fill out an evaluation questionnaire. *FB* groups completed peer evaluations that consisted of three 7-point scales corresponding to the three SYMLOG dimensions along with open-ended responses to explain each rating they provided. *no-FB* groups evaluated the chat user interface as a filler task, based on the user interface satisfaction questionnaire by Chin, Diehl, and Norman (1988). Once all group members completed the questionnaire, they continued the group task. The experimenter tallied the ratings and sent *FB* groups a summary. The summary was an image with three bar graphs showing the group members' average ratings and reiterated the behavioral propositions from the initial instructions.

Upon completion of the task, all participants filled out post-session peer evaluations identical to those completed by the *FB* groups earlier. Participants were then debriefed and thanked for their participation. The whole session lasted about 50 minutes. The experiment instruments appear in Appendix A.

Results

Each transcript was divided into two segments, as shown in Figure 9. Segment 1 and segment 2 correspond to the conversation before and after the intervention, respectively. I used LIWC to analyze language use in each segment both at the individual and at the group level. For individuals, I extracted their contribution to the conversation and used that as input to LIWC, generating a list of linguistic features and the percentage of the text that corresponds to each feature. For groups, I submitted the entire transcript of the conversation to LIWC, less communication to and from the experimenter. The group measurements are effectively an average of each individual's conversational style, weighted by the amount each individual contributed overall to the group.

Communication Style Predicts Feedback

My first research question was "What elements of communication style predict peer evaluations on the SYMLOG dimensions?" To address this question, I examined the relationship between individuals' LIWC results in segment 2 and peer evaluations provided at the end of the session (RQ1 in Figure 9). I created a number of

Table 1. Estimates of parameter coefficients in hierarchical linear models using linguistic measures as covariates and peer ratings on SYMLOG dimensions as predicted variables. Each value represents the coefficient parameter estimate on a single predictor model. (Significance: **p*<0.1, ***p*<0.05, ****p*<0.01).

		SYMLOG peer evaluation ratings		
		Participation	**Friendliness**	**Task focus**
Linguistic measures	**Mean (SD)**	5.29 (0.89)	5.80 (0.89)	5.49 (0.85)
Word count	137.6 (91.7)	0.005***	0.0002	0.002*
Achievement (*best, solve, win*)	1.32 (1.16)	0.151**	0.111*	0.148**
Inclusive (*also, and, plus*)	2.90 (1.71)	0.157***	0.061	0.086**
Agreement *(ok, yes, agree)*	6.64 (5.91)	-0.063***	-0.028**	-0.036***
Affect *(funny, hate, good)*	7.12 (4.85)	-0.061***	-0.019	-0.036**
Positive emotion *(love, lucky, neat)*	6.54 (4.91)	-0.061***	-0.020	-0.036**

hierarchical linear models using data from all participants, with individual nested within group, and group nested within feedback condition, to control for non-independence within groups. Each model used one LIWC indicator to predict the rating of one SYMLOG dimension—that is, does this indicator of language use correspond to higher or lower ratings? Because this is an exploratory study of the effect of language use, I examined 25 potential LIWC indicators. Table 1 presents a summary of the linguistic factors that were significant predictors of peer ratings.

Participants' word production was positively related to peer evaluations of group participation and task focus. Achievement words (e.g., *best, goal*) and inclusive words (e.g., *also, plus*) were also positively related to peer evaluations of participation and task focus. Conversely, using affect-laden words, particularly related to positive emotion, was negatively related to participation and task focus evaluations. The use of agreement words (e.g., *yes, agree*) was also negatively related to all three evaluation dimensions.

The open-ended responses in which group members explained their ratings provide insight into why team members with various language styles were rated as better or worse collaborators. Those who received high ratings on participation were evaluated as "active", "leader", and "keeps us on track", whereas those with lower ratings were evaluated as "didn't talk much", "passive", and "only responds when necessary". Those with high ratings on task focus were evaluated as "organizes things" and "helped the group come to decisions", whereas those with low ratings were evaluated as "humorous" and "makes jokes". Finally, low evaluations were also followed by comments such as "seemed just to go with the flow and agree with what everybody else was saying" and "didn't seem to have many insights".

The open-ended responses help explain the positive relationship between peer evaluations and word count, achievement words, and inclusive words: 1) people may interpret verbosity as contribution, 2) being inclusive may be interpreted as suggestive and detailed, and 3) discussing achievement may be interpreted as being interested in the team's success. Alternatively, frequently using agreement terms may have been perceived as passivity (i.e. passively agreeing without active contribution). Using affect terms, especially positive emotion, may have been interpreted as straying off the task at hand.

Feedback Affects Communication Style

My second research question was "How does feedback affect subsequent collaboration practices, as indicated by communication style?" To answer this question, I examined whether the feedback intervention resulted in linguistic style change from segment 1 to segment 2 (RQ2 in Figure 9), comparing changes in LIWC results at the group level between the two conditions.

Figure 10 shows three LIWC indicators on which the *FB* and *no-FB* conditions

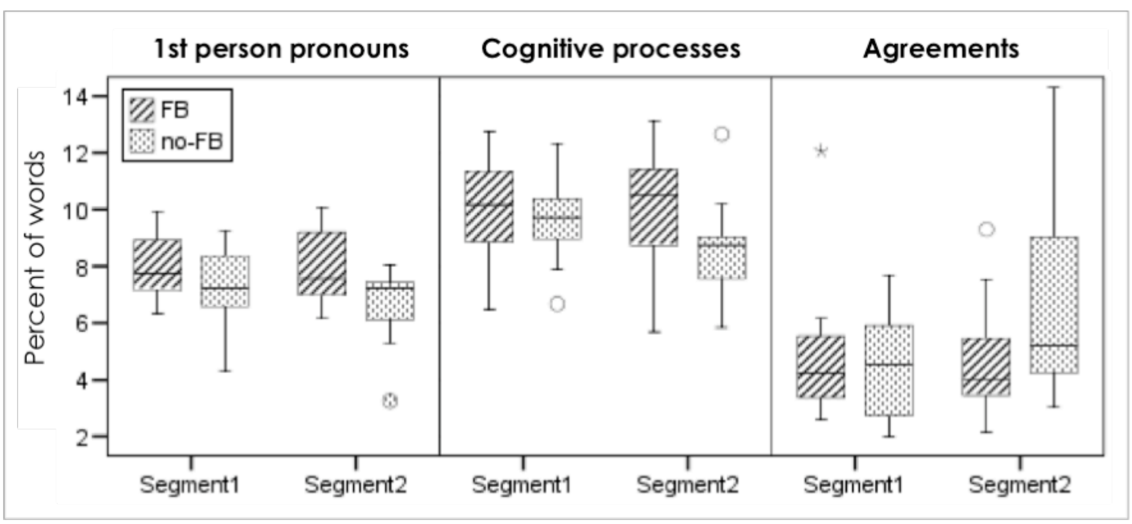


Figure 10. Use of first person pronouns, cognitive process terms, and agreement words in segments 1 and 2.

differed. While *FB* groups did not differ from *no-FB* groups in segment 1 on the use of total first person pronouns (*I*, *me*, *we*, *us*), cognitive process terms (*cause*, *know*, *think*, *should*), and agreement words (*ok*, *agree*, *yes*), in segment 2 *FB* groups used significantly more first person pronouns, more cognitive process terms, and less agreement words. As demonstrated in Figure 10, these differences between *FB* and *no-FB* groups resulted from the *FB* groups not changing their communicative behavior from segment 1 to segment 2, whereas the *no-FB* groups decreased their self pronoun use and cognitive process terms, and increased the use of agreements.

I argue that the pattern of change observed in the *no-FB* condition reflects the natural communication course for this task without a feedback intervention. Toward the end of the session team members sought consensus, leading to fewer self pronouns and cognitive process terms and more assents (less "I think" and more "yeah, yeah, yeah"). It appears that the feedback intervention helped groups stay on track and keep the teamwork going with the same cognitive effort and involvement as in the first part of the conversation. This finding is also aligned with Losada, Sanchez, & Noble (1990), suggesting that feedback increases self-focus compared to no feedback in a technology-mediated environment. Note that the *no-FB* groups talked just as much as the *FB* groups in segment 2 (*FB*: M=506.9, SD=385.4; *no-FB*: M=558.1, SD=269.1; $t(24)$=-0.39, p=0.70). This shows that conversation style can change while word count remains the same and supports the idea of using more sophisticated language analysis tools to understand group dynamics.

Discussion

In this experiment I demonstrated the potential of automated linguistic features based on the word-level analysis technique to measure teamwork behaviors and to be influenced by feedback. The first set of results, responding to RQ1, resonates with previous findings that link language indicators with social behaviors. For example, the findings on affect-laden words replicates Turner & Schober's finding that team members using more emotionally-charged words were perceived as poor collaborators by their peers (2007). Further, the association between overly expressing agreement and low peer ratings is aligned with findings that "exploratory talk", when team members disagree with and challenge each others' ideas (Mercer, 1996), improves group reasoning (Wegerif et al., 1999).

The second set of results, addressing RQ2, show the potential of providing feedback to teams for changing their conversation style as marked by linguistic features such as first person pronouns, thinking words, and agreement words. The peer feedback intervention might have encouraged team members to become more involved in the activity process (Dochy et al., 1999), and this was revealed in their use of language. Further, as part of an impression management process, the feedback may have helped team members monitor the impressions others form of them and could motivate them to communicate in a certain way in order to control how others see them (Goffman, 1959; Leary & Kowalski, 1990).

One reason for changes observed in conversation style could be the act of engaging in peer evaluation, raising individuals' awareness of their impression management, and not necessarily the feedback itself (Dominick et al., 1997). Thus, if individuals change their use of language when receiving peer feedback, then the next question will be if they would change their use of language when receiving automated feedback, without being involved in the evaluation process. This question is examined in the following chapters.

Based on these findings and on previous work, I have chosen the linguistic features that will be embedded in the GroupMeter feedback visualizations. These features are word count, representing amount of talk (see also Janssen et al., 2007), percentage of emotion words (see also Turner & Schober, 2007), percentage of self-references (1st person singular pronouns) (see also Cegala, 1989; Pennebaker, 2002), and percentage of agreement words (see also Janssen et al., 2007; Kelly & Duran, 1985). Not all linguistic features were visualized in all versions of GroupMeter. Following the results of experiment 2, reported in Chapter 5, I chose to drop features that individuals had a difficulty to interpret, such as emotion words. Further, linguistic feedback had to demonstrate that it helps individuals change their use of language, resulting in the selection of agreement words as the sole linguistic feature for experiment 4, reported in Chapter 6.

My next steps were to run experiments to evaluate the extent to which automated linguistic feedback fulfills its promise to trigger reflection on individuals' use of language, and to help them improve their teamwork behaviors. These experiments also contribute to the development of a theoretical framework that elucidates the linkages between teamwork behaviors, feedback and language use. I present these experiments in the next two chapters.

CHAPTER 5

EXPERIMENTS 2 & 3:

EXPLORATORY STUDIES OF AUTOMATED LINGUISTIC FEEDBACK

With a functioning GroupMeter system and the linguistic features chosen, I was now ready to start an empirical examination of the primary questions of my research. The first objective of the two studies described in this chapter was to serve as a preliminary investigation into the two research questions: whether providing automated linguistic feedback stimulates team members' reflection on their use of language (RQ1), and results in them changing their use of language (RQ2). Second, these studies also look qualitatively into the meanings that individuals make of the linguistic features and of different visualizations, in order to make informed decisions about the selection and refinement of various parameters, features, and design elements. Third, with one of the key design principles of GroupMeter to display feedback at the periphery of the main team activity, the studies address the distraction that might be created by dynamic feedback visualizations, and how to design such feedback that would help balance between task focus and attention to social behaviors. Finally, experiments 2 and 3 helped fine-tuning valuable procedure details for experiment 4.

Experiment 2: Pilot Study

Purpose

To assess the initial version of GroupMeter, and in preparation for the primary experiment, I ran an experiment in which teams used GroupMeter's chat to communicate over a decision-making task. The purpose of the study, in congruence with the primary questions of the current research, was to evaluate the potential of

GroupMeter to: 1) encourage team members to reflect on their teamwork behaviors, and 2) ultimately change these behaviors. Thus I was interested both in finding out how users interpret the feedback visualizations and the linguistic features they represent, and whether receiving feedback leads them to change or reflect more carefully on their language use and teamwork behaviors. While the quantitative data did not yield statistically significant results to shed light on differences between providing linguistic feedback or not, valuable insights emerged from the qualitative data collected in this experiment.

This study used version 1 of GroupMeter with the bar-graph visualizations (Figure 6). Based on the findings from experiment 1, I chose to represent three linguistic features with the feedback visualizations, one for each bar graph: word count, self-references (1st person singular pronouns), and emotion words. The self-references and emotion words categories used the LIWC dictionary (Pennebaker et al., 2001).

Method

Participants. Eighty-eight undergraduate students (53 females) at Cornell University participated for course credit and a chance to win $40. Participants were randomly assigned to mixed-gender 3- or 4-person teams in one of two conditions: *Feedback* (*FB*, 12 groups) or *No Feedback* (*no-FB*, 11 groups).

Procedure. Participants were seated at a table facing each other, with a laptop computer in front of each participant. They logged into GroupMeter using their first name as an identifier. Participants were told that they would work as a team on a decision making task and were given a few minutes to get to know each other face-to-face. The experimenter then explained the ranking task, *Lost on the Moon*, also used in experiment 1 (see Chapter 4 and Appendix A).

The team was informed that their goal was to reach the most accurate ranking, and that the best team would receive $40 for each of its members. The experimenter then explained that the team would be using a chatroom to communicate, and told *FB* teams that they would be receiving feedback about their communication behavior during the task. No additional information was given to *FB* groups about the feedback. *no-FB* groups had a version of GroupMeter with only the chatroom and no feedback visualizations.

Dividers prevented eye contact between participants during the chat conversation. Upon task completion the dividers were removed and, in order to get a deeper understanding of participants' thoughts, interpretations, and experiences, the experimenter carried out a semi-structured group interview. Participants were prompted to openly share their thoughts about their awareness of their own and others' communication styles, the GroupMeter user interface, and the feedback meters (the latter for *FB* participants only). All interviews but three were audio-recorded with participants' permission, and later transcribed and analyzed. Sessions lasted about 50 minutes total. The experiment instruments appear in Appendix B.

Results

Data from this experiment consisted of transcripts of the chat conversations, transcripts of the end-of-session interviews, and the rankings each team reached for the task. In summary, the interview data and task performance results suggest that the feedback stimulated participants to reflect on their communicative behavior, although linguistic analyses of the chat conversations revealed that individuals did not significantly alter their use of language.

Reflection on communication style. One goal was to determine if feedback provokes people to reflect on their behaviors. All participants from the *FB* condition reported that they noticed the feedback, although they attended to it in varying degrees: some only glanced at it once or twice, while others looked more closely at the visualizations.

Participants in 10 of the 12 *FB* groups reported that the visualizations made them aware of their language use. One participant noted that she was looking at the meters and "I noticed I was talking more than others, and if maybe I was kind of slacking off I would notice at the bottom." Another participant said: "I was looking at the meter at the bottom about halfway through. Nobody used any language about themselves, and then that switched when we had our list and people had opinions and were trying to change things." Similar to previous findings (DiMicco et al., 2007), these results indicate that the feedback visualizations induced some team members to pay attention to the group's conversation style.

(No) change in linguistic behavior. My second goal was to find out whether providing feedback about communication behaviors had an effect on team members' linguistic style, specifically on the dimensions on which they were given feedback. To address this question, I processed the chat transcripts using LIWC's dictionary-based word count analysis (Pennebaker et al., 2001) and compared the conditions on the measures of word count, emotion words, and self-references at both the individual and group levels. Because I did not give normative guidelines for "appropriate" use of these categories, I did not predict specific changes in linguistic behaviors, though I did expect them to change in the *FB* groups compared to *no-FB* group. I also looked at the within-group standard deviation of these variables to see whether *FB* groups were

more or less similar in how they used each type of language than the control groups. In all cases, there were no significant differences in language use between participants or groups in the two conditions, suggesting that although *FB* participants reflected on their language use, their reflection did not lead to systematic linguistic differences from the control group. An examination of the interview data surfaced three themes that may explain why there was no systematic change in linguistic behavior across conditions: tension between task and process, difficulty in changing linguistic behavior, and need for guidance.

Tension between attending to task and process. Unlike experiment 1, in which the task was interrupted for the purpose of providing and receiving feedback, here I wanted to explore the effect of dynamically providing ambient feedback without interrupting the task. The interview data suggest this might have caused tension between attending to feedback and accomplishing the task. Participants in 6 *FB* groups noted that they felt tension between accomplishing the task and thinking about the feedback: "I saw that the values were changing, but because we had limited time I guess I just focused on getting the task done." Another participant reported that she was thinking about how to change the way that she talked, but "just dismissed it because I didn't want to get behind all the conversation."

The experimental setup might also have encouraged people to focus on the task. Because I did not want to affect how people naturally used the feedback indicators, the experimenter did not explicitly instruct participants to attend to them. This might have led participants to focus more on the task. Finally, the reward structure may have also contributed to the focus on task: teams had no incentive to form long-term effective collaborations, but did have incentive to perform well on the task, with a promised award for the best performing team.

To a certain degree, the task performance results suggest that processing the feedback might have distracted teams from focusing on the task. Compared to the correct solution, *FB* groups performed worse than *no-FB* groups (*FB*: N=12, M=32.5, SD=10.9; *no-FB*: N=11, M=27.3, SD=6.8; lower values indicate better performance), although by a marginally significant amount ($t(21)$=1.36, p=0.09).

Linguistic style may be hard to change. One of the interview questions asked participants to what degree they were conscious of their individual and their teammates' word choice. *FB* participants and *no-FB* participants were similar in their responses to this question. Most said that they were focusing more on the content of their communication rather than on choosing their words: "I didn't pay too much attention to exact words as long as I get the idea I wanted to convey", and "we were just worried about the information, not like how we were getting it across." Indeed, research on psycholinguistics suggests that much of our language production and planning is automatic, that is, without attention or conscious awareness (Levelt, 1989). As such, more explicit guidance on how to interpret feedback and adapt one's behavior may be needed to effect change in subsequent communicative behavior.

Ambiguity may not cultivate guidance. Participants responded in different ways to the same feedback information. For instance, one participant interpreted the feedback on self-references in a way that she sought to push down this measure:

> "I was looking at the meter that was talking about how much you talk about yourself, and I was hoping it would be lower, because I didn't want to be that person that's just talking about themselves all the time."

However, another participant interpreted this measure in the opposite direction:

> "I don't think it's bad to say "I", 'cause sometimes it's better to convince people by saying, well this is how I feel, but I might be wrong. You know you've said "I" twice there, but you're just trying to be nice by saying, you don't have to think what I think."

The missing guidance is also highlighted by one participant pondering: "I was like, is it a bad thing if I say "me" or "I" more often or less often?" If effective teamwork in certain situations involves certain communicative behaviors, then without guidance as to what effective collaborative behaviors look like, interpretations can vary drastically, which may impede systematic behavioral changes.

Discussion

The results of this pilot study shows that providing dynamic feedback about linguistic behavior succeeded in giving participants an opportunity to experience, reflect, and richly interpret visualizations of their own behaviors – but it failed to change their behaviors in a systematic fashion. I believe that one way to achieve behavioral changes is to make the feedback more salient, thus drawing more attention to the reflective process on the team activities and behaviors.

Intensifying the salience of the feedback can be done in several ways. First, the bar-graph visualization, purposely designed for unobtrusive display of feedback and located below the chatroom window, might have been too subtle and easy to disregard. This was addressed by designing the engaging, playful fish visualization for version 2 of GroupMeter (Figure 7), and examined in experiment 3 below.

Second, as a result of this preliminary study, I reconsidered my selection of linguistic features represented by the feedback visualization. In particular, whereas the interview data revealed that participants reflected on the amount of their talk and on their use of self-references, little reflection emerged with respect to emotion words.

It was not obvious to participants what words accumulate into this feedback dimension, as one participant noted: “I was wandering how it was measuring emotional words, like, I was wandering what words are emotional.” The LIWC dictionary maps into this category any word that has a positive or a negative valence, which might be too broad. It was also hard for them to figure out what this dimension means for their interaction with the group, as another participant said: “I thought the more “I” versus “we should do this”, that made more sense, but the emotional words, I guess that was a little more vague.” As a result, participants reported that they ignored the bar of the emotion words measure, and spent more time thinking about the word count and self-references feedback. For these reasons, I decided to drop this feature off from future experiments of automated linguistic feedback.

Third, the results suggested that the system, or the setting in which it is being applied, should provide more guidance on how to interpret and respond to feedback. Although ambiguity is recognized as one of the design principles in this research based on Gaver et al. (2003) and Sengers and Gaver (2006), providing more information about the linguistic indicators and how they are mapped into the visualization could facilitate interpretation in task-oriented settings. This was applied in experiment 3, reported below. Further, it might be that when specific behaviors are known to be more or less desirable, explicit guidance toward specific use of language could assist team members in directing them toward improving their teamwork behaviors. This idea was implemented in the design of version 3 of GroupMeter (Figure 8) and applied in experiment 4 (Chapter 6).

Experiment 3: Feedback Visualizations of Linguistic Features[2]

Purpose

The purpose of this experiment was to evaluate the potential of two feedback visualizations, version 1 and version 2 of GroupMeter (Figure 6 and Figure 7, respectively), to trigger individuals' reflection on and changes in their communicative behaviors. This experiment also examined the peripheral characteristic of the feedback visualizations, attempting to balance between focusing on the team task and attending to the social interaction process. Finally, this experiment also served to refine the procedural details for the implementation of experiment 4.

This study addressed four research questions. The first two are the questions that motivated the entire research, and that were examined in this study as a precursor to experiment 4:

RQ1. Does automated language-based feedback cause people to reflect on their language use, especially about the specific dimensions of language the feedback represents?

RQ2. Do people change their communication patterns when provided with feedback compared to when not?

Another question is driven by one of the key principles of the GroupMeter design. While I am interested in the value that automated feedback has on team members' reflections and behaviors, feedback cannot be at the center of the team activity. A long-recognized principle of teamwork effectiveness is that group members must attend both to their task and to social matters (Bales, 1950; Hackman, 1990; McLeod & Kettner-Polley, 2004; Rousseau et al., 2006). To ensure that team members are able to focus on their task, the feedback visualizations in GroupMeter

[2] This section is based on a paper published at CHI 2009 (Leshed et al., 2009).

were designed as peripheral displays to the team activity chatroom, as described in Chapter 3. It is important that the peripheral visualization effectively conveys awareness information (Cadiz, Venolia, Jancke, & Gupta, 2002; Plaue et al., 2004) in ways that minimize cognitive load and do not significantly impact users' performance on the primary task (Maglio & Campbell, 2000). Still, my aim is to stimulate reflection on and awareness of language use and teamwork behaviors, leading to a possible tension between attending to peripheral feedback visualizations and accomplishing a task. The third research question, therefore, is how to create designs that are cognizant of this balance between task and social process:

RQ3. Does automated feedback about language use distract from the team task?

In this experiment I chose to present feedback on word count and percentage of agreement words, the latter using the LIWC assent category modified for chat communication style (e.g., using "k" for "okay", "def" for "definitely"). This decision was based on experiment 1's findings showing a positive association between amount of contribution to discussion and peer-rated teamwork behaviors, and a negative association between the use of agreement words and teamwork behaviors. The same feedback information was presented by two different visualizations. In version 1 of GroupMeter, each linguistic feature is presented in one horizontal bar graph. In version 2, fish move closer to the center when participants use more agreement words, and they increase in size when participants speak more.

As such, I wanted to understand the differences between feedback visualizations not only in their efficacy to stimulate reflection and behavioral changes, but also in the experiences they bear. This marks a distinction between improving usability, productivity and effectiveness, encapsulated by examining the distraction the feedback visualization might engender, and considering enjoyment of the experience.

As Overbeeke et al. write: "A user may choose to work with a product despite it being difficult to use, because it is challenging, seductive, playful, surprising, memorable or rewarding, resulting in enjoyment of the experience" (2003, p. 11). I therefore compared version 1 of GroupMeter (bar graphs, see Figure 6), with version 2 (fish, see Figure 7), asking:

RQ4. How do people feel about their experience of both designs? What aspects of each do they like and dislike?

Method

Participants. Twenty-five undergraduate students (9 females) from a Human-Computer Interaction course at Cornell University received course credit for their participation in the study. As HCI students, these participants are likely sensitive to interface design issues and I sought their opinions about GroupMeter's designs.

Procedure. Figure 11 shows an overview of the structure of a session. Using a within-subjects design, participants were randomly assigned to seven mixed-gender groups of 2 to 5 members. Each group went through three conditions: Fish visualization, Bars visualization, and None (i.e., no visualization, or control). The order of the conditions was randomized for each group.

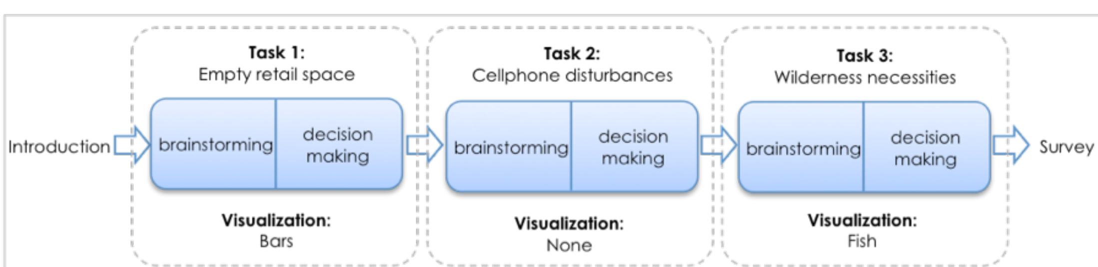


Figure 11. An example session of experiment 3. Task order remained constant while visualization order was randomized across groups.

To engender a feeling of distributed chat collaboration, I did not require participants to come to the lab. Instead, they were scheduled to log into the GroupMeter website at particular times from wherever they were, using either Internet Explorer or Firefox on a PC or a Mac.

Once logged in, participants were greeted by the experimenter (via the GroupMeter chat) and informed that they would be working as a team on three tasks. They then had a few minutes to talk, to get used to the interface and to break the ice. The experimenter then explained the nature of the study and that in each of the tasks they would see a different visualization of the language they used. The linguistic features word count and agreements were then described in terms of how they are measured in the chat and presented in the visualizations. This explanation was provided in an attempt to increase the salience of the feedback as a lesson learned from experiment 2.

For each task, the team was given five minutes to brainstorm ideas for solving a problem, and then five more minutes to choose the top three solutions. The tasks were to 1) decide how to fill an empty retail space near campus; 2) develop strategies to reduce cell phone disturbances on public transportation; and 3) choose items necessary for wilderness survival in Alaska. The experimenter sent a message at the end of each brainstorming period and each task. Due to a small number of groups, it was not possible to counterbalance both the visualization and task orders. Since the tasks were all of the same general type, I felt that the risk of carryover effects was smaller than it would be for visualizations. I therefore decided to keep the task order constant for all groups and randomize the order in which visualizations were presented to groups.

Upon completion of all tasks, participants filled out an online survey that included scale items to assess their perceived awareness of their language use during each task, distraction by the visualizations, and open-ended questions to capture their experience. I kept the task times short and waited to the end of all three tasks to fill out the survey to avoid possible contamination between the tasks and to reduce the chance of participant fatigue. Sessions lasted about 60 minutes total. The experiment instruments appear in Appendix C.

Data Analysis

The data from this experiment consisted of numerical and open-ended responses to the survey and the transcripts of the chat conversations. To examine communication patterns within teams, each chat entry was coded with one of six codes representing a behavioral move that is related to the nature of the tasks and the feedback provided through the visualizations. The coding scheme is explained in Table 2.

The coding excluded entries about study administration, troubleshooting, digressive remarks, and fillers. Two coders coded the transcripts independently, 1269

Table 2. Coding scheme for the chat transcripts in Experiment 3.

Idea	Suggesting a solution for the problem-solving task. This included both new ideas as well as repeating ideas already mentioned (22% of all Idea entries). An example for the wilderness task was “a water purifier thingy.”
Agreement	Expressing agreement with someone else’s statement, e.g.: “oh, that’s a good idea.”
Disagreement	Expressing disagreement with or opposition to an entry, e.g.: “it won't need purification.”
Discussion	Discussing an idea, typically elaborating on an idea or reasoning one’s thoughts, e.g.: “maybe they are in the center of alaska where there is no water...”
Ranking	Proposing or stating a ranking of the ideas for reaching consensus and completing the task, e.g.: “water purifier is important, i think.”
Feedback	Mentioning the visualization, e.g.: “wow my fish just grew a lot.”

entries in total, assigning each entry the most prominent code. Inter-coder agreement was 80%; disagreements were resolved by discussion.

The data were hierarchical in nature, with observations nested within participants and participants nested within teams. I used hierarchical linear models to account for non-independence of the statistical data, calculating variations within and between participants and teams, and used post-hoc analyses for the statistical tests.

Results

Awareness of Language Use

Since *RQ1* was framed in terms of participants' perceptions, I relied on their responses to survey questions in which they evaluated the degree to which the visualizations made them think about and change the words they used. Three five-point Likert scale items addressing these issues were aggregated into a single scale

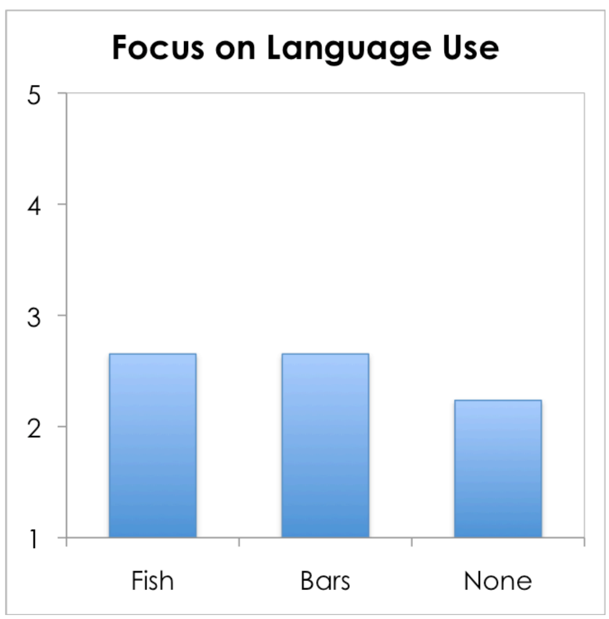


Figure 12. Aggregate scale of participants' responses to survey questions asking about focus on language use.

(Cronbach's α = 0.77): "I focused on the words I used rather than on the content of the task," "The visualization caused me to choose words differently as the task progressed," and "My behavior changed as a result of the visualization" (Figure 12).

Participants reported that both visualizations made them more aware of their language use compared to no visualization (Fish: M=2.65, SD=1.00; Bars: M=2.65, SD=0.99; None: M=2.24, SD=0.72; $F(2,46)=7.87$, $p=0.001$). Pairwise comparisons between the three conditions show that both the Fish and Bars visualizations were significantly different from None (p=0.001), and that Fish and Bars did not differ from each other. There was no significant effect of the order in which visualizations were presented ($F(2,46)=0.62$, *ns*), meaning that participants were more aware of their language use in the feedback conditions regardless of the visualization order.

This provided an initial answer to RQ1: feedback caused a reported increase in reflection on language use.

Communication Patterns and Group Dynamics

RQ2 asked whether the feedback visualizations affected communication patterns and group dynamics. I therefore used the coded chat transcripts as behavioral data in answering this question. Based on McGrath's classification of tasks (1984), I assumed that the dynamics and processes used by groups to accomplish the brainstorming and decision-making subtasks would also differ. For example, while brainstorming requires divergent group thinking, decision-making calls for convergence. I therefore divided the transcripts into two segments, according to the 5-minute subtask periods during which they were typed.

Figure 13 shows the relative proportions of each coded entry type by condition and segment. The data for the bar charts appears as percentages in Table 3. For each entry type, I used post-hoc testing for main effects of *Condition* and *Segment* (the latter is also accounted for as a repeated factor in the hierarchical models), as well as interaction effects of these factors. As with the questionnaire data above, I found no main effects of *Visualization Order* for any of the entry types, ruling out the possibility that the order in which the visualizations were administered affected the results.

Idea entries (Figure 13a). I was interested in whether the experimental condition had any effect on the number of ideas generated. Differences between the conditions might indicate, for instance, that feedback distracted people from the brainstorming task, or motivated them to produce more ideas. However, there was no main effect of *Condition* ($F(2,48)=0.78$, *ns*) or *Condition×Segment* interaction effect ($F(2,72)=0.79$, *ns*). There were also no statistically significant differences between conditions in the number of unique ideas generated (Fish: $M=7.6$, $SD=4.5$; Bars: $M=7.7$, $SD=3.0$; None: $M=8.4$, $SD=2.0$; $F(2,12)=0.19$, *ns*).

Table 3. Percentages of entry types by condition and segment (BS=Brainstorming, DM=Decision Making).

	Fish		**Bars**		**None**	
	BS	**DM**	**BS**	**DM**	**BS**	**DM**
Idea	27.0	9.8	26.5	2.9	29.9	5.3
Agreement	31.3	44.3	24.6	45.4	23.9	32.2
Disagreement	6.7	5.2	16.2	2.8	9.6	5.5
Discussion	29.7	15.3	31.0	20.6	33.7	27.0
Ranking	1.7	24.9	1.7	28.3	2.9	29.9
Feedback	3.7	0.6	0.0	0.0	0.0	0.0
Total	100	100	100	100	100	100

Figure 13. Proportions of each entry type by condition and segment. For each condition/segment pair, its bars sum to 100% across the six entry types, as shown in Table 3.

A statistically significant main effect of *Segment* ($F(1,72)=64.94$, $p<0.001$) indicated, as expected, that more ideas were presented during the brainstorming segment.

Agreement entries (Figure 13b). Because participants received feedback on the frequency of their use of agreement words, differences between conditions on their proportion of agreement entries would strongly suggest that the feedback affected communication behavior.

A main effect of *Condition* ($F(2,48)=3.45$, $p=0.04$) was observed, supporting our expectation that the feedback triggered changes in the expression of agreement. There was a marginally higher proportion of agreement entries in the Fish condition compared to None ($t(120)=1.84$, $p=0.07$), and a significantly higher proportion in the Bars condition compared to None ($t(120)=2.35$, $p=0.02$). Although the experimenter did not tell participants that frequent agreement were desirable, they expressed more agreement when receiving feedback than when they did not, suggesting that they inferred an implicit norm from the feedback.

Once again, I found a main effect of *Segment*, with participants using a higher proportion of agreement entries during the decision-making segment as compared to the brainstorming segment ($F(1,72)=16.12$, $p<0.001$). Again, this was expected given that participants were instructed to reach consensus toward the end. There was no *Condition*×*Segment* interaction effect ($F(2,72)=0.95$, *ns*).

Disagreement entries (Figure 13c). I did not find a main effect of *Condition* on disagreement entries ($F(2,48)=2.00$, *ns*), but, interestingly, there was a *Condition*×*Segment* interaction effect ($F(2,72)=6.29$, $p=0.003$). Participants in the Bars condition used a higher proportion of disagreements in the brainstorming segment compared to the other two conditions and showed a larger decrease in disagreement in the decision-making segment. This finding was unexpected. One possible explanation is that the bars visualization provided participants with the clearest feedback, which in turn, may have prompted the most behavioral change.

Why this occurred for disagreements and not for other statement types remains unclear. This suggests directions for future research; perhaps the bars created a more "business-like" feel to the session, and this may have led to greater sensitivity to disagreement than to agreement feedback.

Again, I observed the expected changes across segments ($F(1,72)=17.98$, $p<0.001$), with participants reducing the proportion of disagreement entries from the brainstorming to the decision-making segment. This complements findings on the increase of agreement entries from the brainstorming to the decision-making segment; both changes suggest that groups attempted to reach consensus toward the end of the task.

Discussion entries (Figure 13d). As with idea generation, differences in the amount of discussion between conditions would indicate that the feedback was affecting how teams performed the task. There was a marginal main effect of *Condition* ($F(2,48)=2.92$, $p=0.06$) that resulted from a lower rate of discussion entries in the Fish condition compared to None ($t(118)=-2.51$, $p=0.01$). In other words, when viewing the Fish visualization, teams tended to discuss ideas less compared to when not receiving feedback. I speculate that the visualization of agreements might have induced more focus on agreement rather than on discussing ideas. This finding also suggests, regarding RQ3, that the fish drew more attention to the specific behaviors that influence the visualization as compared to completing the task at hand.

Across all conditions, a main effect of *Segment* showed that groups discussed ideas more in the brainstorming segment than in the decision making segment ($F(1,72)=11.16$, $p=0.001$), as expected. The interaction effect between *Condition* and *Segment* was not significant ($F(2,72)=0.70$, *ns*).

Ranking entries (Figure 13e). I examined proportions of ranking entries to see if the feedback had an effect on the extent to which participants expressed their preferences for ideas. As expected, a main effect of *Segment* shows that more ranking entries occurred during decision-making than brainstorming ($F(1,72)=87.72$, $p<0.001$). There was no main effect of *Condition* ($F(2,48)=1.18$, *ns*), nor an interaction effect ($F(2,72)=0.32$, *ns*).

Feedback entries (Figure 13f). As one indicator of the extent to which different visualizations changed team interaction, I looked at the number of entries that explicitly referenced the feedback. These could indicate that the visualizations captured the attention of participants and became a conversation topic. Only 11 (< 1%) of the 1269 entries where coded as Feedback entries, however. They only appeared in the Fish condition, and only for 3 groups, as in the following chat excerpt:

m: wait, weren't our fishes supposed to swim?

b: they get closer with more agreement

m: ah i just saw one move

Moreover, these entries provide behavioral indicators relative to the distraction effect and the experience of the feedback visualizations (RQ3 and RQ4). Comments such as the example above suggest that the fish visualization did not remain in the periphery to the same extent that the bars visualization did. The fish may thus have been somewhat more distracting than the bars visualization, eliciting comments that deviated from the task. I present further evidence relative to RQ3 and RQ4 below.

Total conversational activity. Finally, I looked at total conversational activity to see whether the visualizations had any effect on the duration or substance of group conversation. Neither the total number of entries nor the total word counts posted by team members had statistically significant differences by *Condition, Segment,* or *Condition×Segment*.

Considered together, these analyses allow addressing RQ2. Feedback impacted the communication pattern of groups in several key ways: 1) feedback about the frequency of agreement words tended to increase agreement statements, 2) in the Bars feedback condition participants drastically reduced their disagreement statements as they transitioned from brainstorming to decision-making, and 3) in the Fish condition participants discussed ideas less compared to None. While it is possible that some of the differences observed in the proportion of agreement statements stem from participants attempting to manipulate or "game" the feedback display, they nonetheless changed behavior in ways that made sense for the specific activity they were doing (e.g., brainstorming vs. decision-making). Further, I found little evidence of gaming the system in reading the transcripts.

Automated analysis captures communication patterns. An important assumption in my approach is that automated linguistic analysis can effectively capture communication patterns. To assess this assumption, I compared the results of an automated LIWC analysis with the relevant codes produced by human coders. For example, the LIWC *assent* category should correlate with the human coding of agreement entries. Likewise, the *negate* category should correlate with disagreement statements. This was the case.

To address this question, I could not simply examine correlations given the non-independence in the group data. Instead, I created hierarchical linear models nesting participants within groups. One model tested whether the rate of *assent* terms identified by the automated LIWC analysis could predict the proportion of agreement entries per participant identified by human coders. The second model predicted the proportion of disagreement entries identified by human coders using the LIWC category *negate*.

The results of both models suggest that automated linguistic analysis, such as that produced by LIWC in the GroupMeter system, is a powerful method for capturing the communication pattern and tone of entries as interpreted by human coders: the *assent* category significantly predicted the proportion of agreement entries ($F(1,69)=56.18$, $p<0.001$), and the *negate* category significantly predicted the proportion of disagreement entries ($F(1,69)=24.72$, $p<0.001$). This finding is interesting, since although the human coders were looking for linguistic cues when coding for agreement and disagreement statements, they were instructed to consider the tone of messages beyond the words they contained. For instance, the entry "me too" was coded as agreement but does not contain agreement words.

Distraction of Feedback Visualizations

In addition to the behavioral data discussed earlier, I addressed RQ3 by asking participants to evaluate their task focus in each of the tasks, using the 5-point item "I

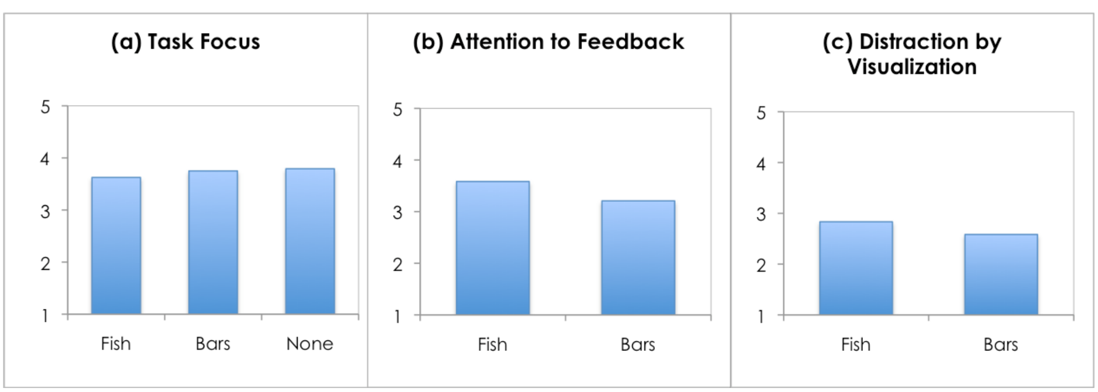


Figure 14. User responses to survey questions: (a) Focus on task, (b) Attention paid to feedback visualizations, (c) Distraction by feedback visualization. Responses were on a 5-point scale.

remained focused on the task throughout the exercise" (Figure 14a). Responses were slightly, but not significantly, lower in the Fish condition (Fish: M=3.66, SD=1.10; Bars: M=3.75, SD=1.03; None: M=3.79, SD=0.98, $F(2,46)$=0.68, *ns*). This suggests preliminarily that the visualizations did not increase participants' cognitive load.

To specifically address the difference between the two visualizations, the survey asked participants to rate the item "I was paying attention to the feedback" (Figure 14b). Participants indicated paying attention more to the fish than to the bars (Fish: M=3.58, SD=1.32; Bars: M=3.21, SD=1.28; $t(23)$=1.81, p=0.04), suggesting that, as designed, the fish attracted more attention than the bars. Participants also found them marginally more distracting (Figure 14c), based on an aggregate score of two 5-point items (Cronbach's α = 0.77): "The feedback I received was interruptive" and "The visualization distracted me from the task." (Fish: M=2.81, SD=1.11; Bars: M=2.54, SD=1.07, $t(23)$=1.64, p=0.06).

Thus, with regard to RQ3, the fish visualization seemed to draw more attention and was perceived as more distracting than the bars visualization. This is also supported by two behavioral effects reported above: people in the Fish condition were the only ones to talk about the visualization, and they had a smaller number of discussion entries compared to None.

Experience of Feedback Designs

Finally, I was interested in participants' qualitative experiences with and opinions about the visualizations (RQ4). The survey asked participants in open-ended questions to express their reactions to and what they liked and disliked about each of the visualizations. They reported that they generally liked the visualizations and thought they were "cool". They liked that the fish were "fun to look at", "cute", and "dynamic". However, the fish were also considered by users to be harder to understand than the bars, and more distracting with their lively animations and movements on the screen. In contrast, whereas users liked the bars' ease of interpretation, they were referred to as "boring" and "just there". The behavioral findings presented earlier also support these self-reports. The fewer discussion entries in the Fish condition compared to None and the three groups talked about the fish but not about the bars suggest that the fish drew attention and were worth discussing during a timed exercise.

Together with the results for RQ3, these responses suggest that users experienced the fish visualization as engaging and even enchanting (as defined by McCarthy, Wright, Wallace, & Dearden, 2006), as they offered an experience of being "caught up and carried away". However, users did feel that the playfulness designed into the fish sacrificed ease of use, glanceability, and peripherality, which are important in the task-related setting for which GroupMeter is designed. They liked the unobtrusiveness of the bars, but also criticized their lack of excitement.

Discussion

The findings suggest that the feedback visualizations, both the fish and the bars, affected participants' awareness of their language use and their communication patterns. The findings also address issues of peripheral feedback displays and the experience of the feedback designs.

Addressing RQ1, the findings suggest that receiving visual feedback made users more aware of their use of language as compared to not seeing feedback. While it is possible that the within-subjects design led participants to pay special attention to the feedback when it was present, research on psycholinguistics suggests that much of our language production and planning is unconscious (Levelt, 1989). Thus, I believe that stimulating reflection on one's own word choice is important in making people change their communication patterns, particularly in a conversational setting. This is a critical first step in understanding how to improve teamwork behavior via visual feedback, and is further examined in a between-subject experimental design in experiment 4.

With respect to RQ2, the results show that people changed their behavior when receiving feedback. They spent more time agreeing with each other, less time discussing the brainstormed ideas when seeing the fish, and drastically decreased their disagreement when seeing the bars. The fact that GroupMeter was able to elicit changes in communication behavior is especially encouraging given the findings in experiment 2 and previous research suggesting that people's choice of words in conversation is largely spontaneous, unintentional, and uncontrolled (Chung & Pennebaker, 2007; Pennebaker, 2002). In this experiment, participants were nonetheless able to make changes in their communication patterns (e.g., agreeing) in both visualization conditions.

These results thus demonstrate the power of automated linguistic analysis for teamwork feedback in stimulating reflection on and change in behavior. This leaves open the question, however, of which specific behavioral changes are desirable. The instructions of the experiment did not explicitly pose normative guidelines under the premise that the visualizations will serve as a mirror to teams, helping them construct their own interpretation of what the appropriate behavioral response to the feedback is. However, length of bar and distance of fish from the center possibly implied that more agreements were preferred. This, in addition to the mere presence of the feedback guiding participants toward self-focused awareness, might have led them to conform more to the group (Duval & Wicklund, 1972) and to a groupthink process (Janis, 1972). If fewer agreements and more discussion are favorable behaviors, as found in experiment 1, integrating appropriate guidelines for effective teamwork into the design might be important for task-oriented groups. Both the results of experiment 2 and experiment 3 suggest that the interplay of ambiguity and directedness designed into the feedback is an important open question worthy of future research.

Regarding RQ3, the findings demonstrate the complexity in designing feedback visualizations that raise awareness of and reflection about social processes without distracting the team from its task. Both feedback designs presented here triggered awareness of language use, and their impact on behavior was similar in some elements (e.g. agreements) but different in others (e.g. discussion and disagreements). However, the fish were perceived as drawing more attention and distracting than the bars, suggesting that given the desired reflection and behavioral change, the salience or subtleness of the visualizations need to be considered. One limitation of these results is the use of self-reports to estimate how distracted participants were. Future work could involve more objective measures of distraction, such as via eye tracking or evaluation of participant recall of specific conversational elements.

Finally, regarding RQ4, the qualitative findings suggest that the two designs elicited quite distinct experiences from users. They enjoyed the playfulness and liveliness of the fish, and valued the efficiency of information conveyed through the bars. This further highlights the complexity of the design task. On the one hand, it is useful to have a design that engages and entertains users, which may in itself trigger reflection on behavior. On the other hand, ease of interpretation is also important so that users can derive meaningful lessons and alter their behavior in useful ways.

In conclusion, the findings from this experiment suggest that dynamic feedback visualizations of teamwork process information have the potential to affect team interaction in a mediated environment. Results from manual coding of chat transcripts, as well as user responses to scale items and open-ended questions, demonstrate that automated feedback on linguistic behavior alters how people think about, communicate in, and experience their teamwork practices. I found evidence that automated feedback can make people aware of and reflect on their language use, a first step in training them to acquire better teamwork skills. The findings also show that automated detailed feedback about word choice can be powerful in more than raising awareness of one's own language use – it can also cause people to alter their communication patterns. Finally, the results demonstrate that the design of peripheral feedback systems should carefully consider the balance between focus on the primary team activity and maintaining awareness of peripheral information, and the power of different visualizations to bring about diverse user experiences. These issues are further refined and examined with the redesign of version 3 of GroupMeter and the procedural setup of experiment 4, described next.

CHAPTER 6

EXPERIMENT 4: GUIDING TEAMWORK BEHAVIORS

The purpose of experiment 4 was to further address the role of automated linguistic feedback in teamwork settings to motivate team members to reflect on and change their use of language. Experiments 2 and 3 provided insights about the feasibility of automated linguistic feedback to bring about such effects. This experiment applies lessons learned from these experiments and tests specific hypotheses about the effects of automated linguistic feedback on reflection and change in team members' language use and on their teamwork behaviors.

Hypotheses

Reflecting on Language

The GroupMeter feedback visualizations were designed using principles of peripheral display design, such that they are glanced at when necessary, but also easy to ignore when focusing on the task at hand. However, this issue exists not only at the surface of attending to the display or understanding the information it conveys. At a more profound level, this issue represents the difficulty in balancing between thinking about and working toward completing the team task, and reflecting on one's social behaviors that accrue to the team's well-being (McLeod & Kettner-Polley, 2004; Rousseau et al., 2006).

To help team members reflect on their behaviors as revealed through the feedback display, this experiment adds an intermission for reflection between tasks as an inherent element of the feedback. This explicit time for team members to review the feedback they receive and reflect on it can lessen the competition over cognitive resources during task completion. This way, individuals can get real-time feedback and interpret it at a glance, and dedicate time outside of working on the task to consider ways in which they would like to change their behavior based on the feedback they receive. The feedback remains at the periphery during task performance, but becomes salient at a certain point in time, not interrupting the task.

This idea is aligned with Gersick's model of team development, that certain points in a team's life, such as intermissions between two tasks, serve as natural transition points in which team members are willing to engage in reflective actions in order to revise their behaviors toward improving the teamwork in the next phase (Gersick, 1988; Gersick, 1989; Gersick, 1991). In his five-stage approach to teaching collaborative learning skills, Bosworth also suggested providing students with time to reflect on their experiences (1994).

Previous studies of feedback on teamwork behaviors, including expert-based (Losada et al., 1990; McLeod et al., 1992), and peer-based (Dominick et al., 1997), applied this principle of providing the feedback at an intermission between tasks, allowing people to take it in and apply changes in their behavior in a subsequent team session. Even with the availability of dynamic automated feedback, it has been found to be more useful when explicit time is given to review and reflect on the feedback (DiMicco et al., 2007).

Deriving from Gersick's theory on team development and previous research, I therefore propose that such automated linguistic feedback, applied with a pause for reflection, will cause individuals to be more reflective of their behaviors in general, and of their use of language and word choice in particular:

H1a. Individuals receiving automated linguistic feedback will reflect more on their behavior.

H1b. Individuals receiving automated linguistic feedback will reflect more on their language use and word choice.

Experiments 2 and 3 have already shown that dynamic automated linguistic feedback led to participants reporting they were thinking about their language. However, these experiments were exploratory: in experiment 2 the findings were qualitative in nature, and in experiment 3 it is possible that the within-subjects design contributed to this finding. This experiment is designed to test these hypotheses in more ideal conditions.

Changing Use of Language

Having individuals reflect on how they talk with their teammates is important, but it does not fulfill the entire promise of feedback. For the feedback to be beneficial, people receiving it need not only to think about it, they also need to apply what they learned by changing their behavior. Especially when it comes to applying changes in use of language, this is a difficult problem because of the automaticity in language production and word choice during conversations (Chung & Pennebaker, 2007; Levelt, 1989; Pennebaker, 2002). Indeed, experiment 2 showed that participants did not change their communication patterns when receiving feedback, and experiment 3 only showed changes in manually coded communication patterns. The question whether feedback can also affect individuals' word choice remains yet open.

One possibility for these missing, or partial findings in experiments 2 and 3 could be the ambiguity with respect to the normative behaviors expected from individuals receiving the feedback. Although ambiguity enables rich interpretations and reflections (Gaver et al., 2003; Sengers & Gaver, 2006) and as such it is the basis for one of GroupMeter's design requirements, it might have also made it difficult for participants to adapt their behaviors toward a specific direction. Especially in a time-constrained experimental setting, it could be that the team did not have the time to construct norms for what counts as good or poor behaviors given the feedback they receive and other contextual factors such as the task they were performing.

One way to alleviate this problem is to accompany the feedback with behavioral standards established in advance. Based on goal-setting theory (Locke & Latham, 1990), the benchmark for interpreting the feedback can then help create a perception of an outcome-goal discrepancy, guiding team members to change their interpersonal behaviors in an attempt to close this gap (Locke et al., 1981; McLeod et al., 1992). Feedback about a certain category of words allows team members to monitor how often they use such words. When accompanied with a benchmark, they can try to use more or fewer of them in the conversation in order to accord with the standard. Based on this reasoning, I propose that when there is a behavioral standard for using certain kinds of words, individuals receiving feedback about their word choice will increase or decrease their use of these words to match the standard, as compared to when not receiving feedback.

In this experiment I decided to provide feedback on one linguistic feature only—the LIWC category 'assent'. This feature includes words that represent agreement, such as 'yes', 'okay', and 'agree'. The reason for choosing only one linguistic dimension was to simplify the level of interpretation required by participants. Furthermore, using this category, both high and low standard of use can be justified: by frequently accepting others' ideas, high level of agreement can represent collective behavior, leading to cohesion and bonding of the group, and to improved task performance (Driskell & Salas, 1992). On the other hand, low level of agreement can be reasoned as motivating critical and analytical discussion (Janssen et al., 2007). I therefore propose the following hypotheses:

H2a. When provided with a high standard for using agreement words, individuals receiving feedback about percentage of agreement words will express more agreement as compared to individuals not receiving feedback.

H2b. When provided with a low standard for using agreement words, individuals receiving feedback about percentage of agreement words will express less agreement as compared to individuals not receiving feedback.

Improving Teamwork Behaviors

A fundamental premise of this research is that automated linguistic feedback can help individuals acquire better teamwork skills. It is therefore important to demonstrate that individuals receiving automated linguistic feedback not only modify their use of language in response to the feedback, but also develop their interpersonal teamwork behaviors. This issue is essential because linguistic feedback, especially when provided with a benchmark, can guide team members to use certain language in the conversation in order to manipulate the visualization, but this does not necessarily improve the team interaction process. For example, just saying 'yes' all the time to see how the fish move up in the display might be detrimental to the performance of the team, resulting in groupthink (Janis, 1972). Such behavior, when individuals systematically attempt to reach a certain superficial goal without engaging in meaningful thought about their actions, can be identified as "gaming the system", and has been shown to reduce learning when using a system that offers feedback (Baker, Corbett, Koedinger, & Wagner, 2004).

This issue is examined here using Bales and Cohen's SYMLOG framework, introduced in Chapter 2. The framework describes three dimensions of interpersonal behavior: dominance/submissiveness, friendliness/unfriendliness (also group-/individual-orientation), and task focus/socio-emotional expressiveness. Effective teamwork behaviors are characterized on these dimensions respectively as evenly divided participation among team members; group-oriented and friendly behaviors; and a balance between task-oriented and socio-emotional expressive behaviors (Losada et al., 1990; McLeod et al., 1992). If automated linguistic feedback helps team members not only alter their word choice but also develop better teamwork behaviors, then moving closer to the ideal behaviors on the three SYMLOG dimensions should reflect this change. The following hypotheses are proposed:

H3a. Participation will be more evenly distributed among members of teams receiving automated linguistic feedback.

H3b. Individuals will demonstrate a more friendly behavior in teams receiving automated linguistic feedback.

H3c. Individuals will demonstrate a better balance between task-orientation and socio-emotional expressiveness in teams receiving automated linguistic feedback.

Method

Participants. One-hundred and twenty three undergraduate students (66 females) at Cornell University participated for course credit. Participants were randomly assigned to one of 41 mixed-gender 3-member teams.

Design. Two variables were manipulated in the experiment. The first manipulated variable was *feedback*: each team either received feedback (F1) about their use of language or did not receive feedback (F0). The second variable was agreement instructions: team members received instructions to agree more (A1) or to agree less (A0) with each other. This created four conditions each team was randomly assigned to: A0F0 (*N*=30, 10 groups), A0F1 (*N*=33, 11 groups), A1F0 (*N*=30, 10 groups), A1F1 (*N*=30, 10 groups).

Procedure. Participants assigned to the same team were scheduled to arrive at the laboratory at the same time. Upon arrival, they were seated at a round table facing each other, with a laptop computer in front of each participant. Participants were told that they would be working as a team on two tasks and were asked to introduce themselves to each other. The experimenter then explained that they will be using the GroupMeter chatroom to communicate for completing the tasks, and gave feedback groups only (A0F1 and A1F1) an explanation about the feedback visualization. These groups used version 3 of the GroupMeter interface (Figure 8). They were told that a fish for each team member would move up and down as a response to their frequency of agreement words usage (with examples for such words provided by the experimenter), updating approximately every one minute and leaving behind a trail of bubbles. All groups were then given a few minutes to chat in the GroupMeter chatroom, to get familiar with each other and with the GroupMeter interface. Groups in the no-feedback condition (A0F0 and A1F0) used a version of GroupMeter with the chat window only. The experimenter pointed out the feedback visualization to A0F1 and A1F1 participants, and encouraged them to try typing in different agreement words at the trial session and see how the fish move up and down. They were also shown the trail button and how they can open the full bubble trail. After 3-4 minutes the experimenter stopped the GroupMeter trial session and asked participants to logout.

Before receiving the first task, teams were instructed to agree with each other more (A1F0 and A1F1) or less (A0F0 and A0F1), with the following excerpts from the experimenter's instructions (see Appendix D for the full instructions):

A1 (more): "Research has shown that to effectively work together it is better to express more agreements among team members. This leads to a more positive environment, a more cohesive team process, and as a result working more efficiently toward a team decision. As such, it is recommended that you show more agreements and fewer objections to your team members' arguments."

A0 (less): "Research has shown that to effectively work together it is better to express fewer agreements among team members. This leads to a more analytical discussion, thinking about different ways to solve the problem, and as a result working more efficiently toward a team decision. As such, it is recommended that you challenge your team members' arguments and not necessarily agree with them too often."

In addition to these instructions, groups receiving feedback saw a ruler with green and red ends at the right hand side of the fish visualization of the GroupMeter feedback display (see Figure 8 in Chapter 3): A1F1 participants saw the green end at the top and the red end at the bottom, whereas A0F1 participants saw the red end at the top and the green end at the bottom. This served as a subtle visual reminder of the normative standards provided by the instructions.

For each of the two tasks, the team was given ten minutes to brainstorm ideas for solving a problem, and then five more minutes to choose the best alternative. The tasks were to brainstorm and decide: 1) how the imaginary city of Brooksfield should spend a 10 million dollar Urban Renewal Grant (based on Wickham & Walther, 2007), and 2) what Cornell University's president can do to improve the connections with the Ithaca community. The tasks were given to participants one at a time in paper format. Once they received the task, participants were instructed to login back into GroupMeter and start discussing the task. To simulate a distributed chat session, the experimenter put up dividers between participants to prevent eye contact. While dividers were up the experimenter sent instructions at the end of each brainstorming period and at the end of each task via the GroupMeter chat.

Upon completing the first task, participants were asked to fill a short survey. Groups receiving feedback (A0F1, A1F1) completed a “reflection survey” (see Appendix D), while groups not receiving feedback (A0F0, A1F0) completed a user interface evaluation survey as a filler task as in experiment 1. The purpose of the reflection survey was to encourage the team to become aware of the visualization and reflect on their use of language as a response to the feedback they received. Once all group members completed the survey, they continued to the second task. Because the tasks were similar in nature, their order was kept constant for all groups.

At the end of the second task all participants filled out a survey, after which dividers were removed and participants were debriefed and thanked. The whole session lasted about 60 minutes.

Measurements and Data Analysis

The data from this experiment consists of the group communication transcripts collected by the GroupMeter chat server and participants’ responses to the post-session survey that included several scales.

Chat Transcripts

Automated analysis. Due to a technical problem the communication data of one group in the A1F0 condition were incorrectly captured and so they were pulled out. The chat transcripts were first processed using a Perl script to separate between entries of the three group participants, after which each participant’s set of entries were processed using LIWC (Pennebaker et al., 2001). The categories of interest were word count and assent, the latter corrected by adding words such as ‘def’ for ‘definitely’ and ‘kk’ for ‘okay’, commonly used in chat communication. For exploratory reasons, I also looked at other categories such as ‘cognitive mechanisms’, i.e., words that represent thinking such as ‘should’ and ‘because’, and 1st person singular pronouns, i.e., words in which the speaker refers to themselves, such as ‘I’ and ‘me’.

Manual coding. In addition to automated processing with LIWC, the chat transcripts were manually coded, such that each chat entry was examined by a coder who gave it one of eight codes representing behavioral moves that are related to the nature of the tasks and the feedback provided by the visualizations. For example, because of the instructions to express more or less agreement, and that the feedback

Table 4. Coding scheme for the chat transcripts in experiment 4.

Idea	Suggesting a solution for the brainstorming task. This only included new ideas that were not mentioned before. E.g., “maybe a renovation of the downtown area?”
Agreement	Expressing agreement with someone else’s statement, e.g., “sure”.
Disagreement	Expressing disagreement with or opposition to someone else’s statement. E.g., “but that’s not necessarily bringing the community together”.
Discussion	Offering an explanation for an idea, discussing a certain idea, expressing more thoughts about an idea, or repeating an idea. E.g., “that would attract new families who are just starting out and looking for places to raise their children”.
Choice	Stating preference or suggesting choosing alternative(s) for the group’s final decision. E.g., “I like the community events”
GroupMeter	Talking about the GroupMeter application, including referring to the chat interface or to the fish visualization, or problems with the application. E.g., “everyone’s fish moved up”.
Management	Talking about the organization, management, or coordination of how to perform the task. E.g., “let’s narrow down our list”.
Other	None of the above. These entries included mostly fillers and digressive remarks, were about 14% of the total entries, and were excluded from further analysis.

visualized percentage of agreement words, it was important to indicate whenever participants' expressed agreement and disagreement in their entries. The coding scheme is explained in Table 4.

Three coders independently coded the transcripts. After training over three transcripts and discussing disagreements in the coding, the coders reached inter-coder reliability of 87%, calculated using Krippendorff's formula (1980), and standardized for three coders using the Spearman-Brown formula (Schmidt & Hunter, 1996). The remaining transcripts were then split among the three coders who coded them separately. In total, 8100 chat entries were coded, and 6935 remained after removing entries coded as 'Other'.

Survey Instruments

The full set of instruments used in this experiment appears in Appendix D. This section describes the sources for the instruments and the rationales for using them.

Consensual and critical group norms. As a manipulation check, to find out if

Table 5. Consensual and critical group norms scales from Postmes et al. (2001).

Consensual group norms scale
Cons1. In this group you should contribute to the group's goals Cons2. In this group you should conform to the others Cons3. In this group you ought to align yourself with the opinions of other members
Critical group norms scale
Crit1. In this group you are expected to make an independent contribution Crit2. In this group you should think critically Crit3. In this group you ought to act independently

participants assimilated the instructions priming them to agree more or less, they completed two scales of group norms (based on Postmes, Spears, Sakhel, & de Groot, 2001). Each scale consisted of three items, measuring whether participants perceived their group as having consensual group norms or critical group norms, respectively (see Table 5). I assumed that when instructed to agree more with each other, participants would perceive the group norms to be more consensual and less critical, and the opposite when instructed to agree less with each other. The order of the six questions was randomized in the survey.

An aggregate consensual norms measure was created based on Cons2 and Cons3 (Cronbach's α=0.818), since Cons1 did not yield good correlation with the other two questions (r=-0.14, p=0.1 with Cons2; r=-0.09, p=0.3 with Cons3). A critical norms item was aggregated based on the critical norms scale items (α=0.612). Higher scores on these variables represent more consensual or critical group norm perception, respectively.

Perception of reflection on and change of language use. This scale was developed specifically for the experiment to measure if participants reflected on their use of language and their perceptions of changing their language use. It included eight 7-item questions, four asking participants about their reflection and four asking about their perception of changing their language use. The questions within each scale were constructed so that two asked about general language use, and the other two specifically asked about agreements. Also, two of the questions asked specifically about the role of GroupMeter in their reflection or perception of change. The questions appear in Table 6, and their order was randomized in the survey. An aggregate measure of reflection on use of language was computed from the first four items (Cronbach's $\alpha = 0.810$), and an aggregate measure of perception of language change was computed from the latter four items (Cronbach's $\alpha = 0.761$).

As another measurement of reflection, I used the Self-Reflection and Insight Scale (SRIS, Grant, Franklin, & Langford, 2002). According to Carver & Scheier (1998), reflection, the inspection and evaluation of one's thoughts, feelings and behavior, and insight, the clarity of understanding of one's thoughts, feelings, and behavior, are metacognitive factors central to the process of purposeful, directed change. I modified the tense of the 6-point items in the instrument from present to past, so that participants could relate to their specific experiences during the

Table 6. Scales used for perception of reflection about and change of language use. Each scale asked two questions about general language, and the other two where specific about agreements. Two questions in each scale referred specifically to the GroupMeter system.

Reflection		
General Language		I spent time thinking about the words I was using while communicating with my team members.
General Language	GroupMeter	The GroupMeter system helped me think about how I talk with my teammates.
Agreement		I spent time thinking about how often I agree with my team members.
Agreement	GroupMeter	The GroupMeter system helped me think about how often I agreed with my team members.
Change		
General Language		During the conversation I tried to change the words I was using when communicating with my team members.
General Language	GroupMeter	The GroupMeter system helped me adjust the words I was using during the conversation.
Agreement		During the conversation I tried to change how often I agreed with my team members.
Agreement	GroupMeter	The GroupMeter system helped me adjust how often I agree with my team members during the conversation.

experiment. The order of the questions was randomized in the survey. Following Grant et al., the instrument yielded two aggregate measures: engagement in self-reflection (12 items, Cronbach's α=0.887), and insight into self (8 items, α=0.826). An example for a self-reflection item is: "I frequently took time to reflect on my thoughts". An example for an insight item is: "I usually had a very clear idea about why I've behaved in a certain way".

Peer assessments of teamwork behaviors. As in experiment 1, I used ratings on the three SYMLOG dimensions (Bales & Cohen, 1979) as a measure of the teamwork behaviors each participant showed. Each participant was asked to rate all three group members, including themselves, on the three dimensions using 7-item scales. The score for each participant on each dimension was calculated as an average of the rating the other two group members gave him or her.

Perception of distraction. To examine the degree to which participants perceived that they were distracted during the task, they responded to three 7-item questions: "Some features in GroupMeter distracted me from focusing on the task", "Thinking about the words I used in the chatroom distracted me from focusing on the task", and "I remained focused on the task throughout the conversation". The wording of the questions was such that participants in groups who received feedback and those who did not could both answer them (they did not specifically ask about the feedback visualizations). After reversing the score of the third item to account for distraction rather than focus, the responses for the items were averaged to a single measure of distraction (Cronbach's α=0.486), with higher scores representing more distraction from the team task.

Table 7. Ten-Item Personality Inventory (Gosling et al., 2003) and the personality dimensions each item aggregates to.

Item		**Personality Dimension**
1	Extraverted, enthusiastic	Extraversion
2	Critical, quarrelsome	Agreeableness (Reversed)
3	Dependable, self-disciplined	Conscientiousness
4	Anxious, easily upset	Emotional stability (Reversed)
5	Open to new experiences, complex	Openness to experiences
6	Reserved, quiet	Extraversion (Reversed)
7	Sympathetic, warm	Agreeableness
8	Disorganized, careless	Conscientiousness (Reversed)
9	Calm, emotionally stable	Emotional stability
10	Conventional, uncreative	Openness to experiences (Reversed)

Personality measurements and gender. It is likely that variables such as gender and personality affect how participants talk with each other and how they think about their communication and interaction with others. To control for such extraneous variables, I recorded participants' gender and used the Ten-Item Personality Inventory (TIPI, Gosling, Rentfrow, & Swann, 2003, see Table 7). The short instrument for a personality measure was necessary given time constraints. The scores for each of the "big five" personality dimensions was combined from two 7-item questions, of which one was reversed. The five dimensions with the mean scores and standard deviations are: extraversion (M=5.1, SD=1.34), agreeableness (M=4.9, SD=1.17), conscientiousness (M=5.5, SD=1.15), emotional stability (M=5.1, SD=1.29), and openness to experiences (M=5.4, SD=0.88).

Data Analysis

The statistical analysis included the SPSS MIXED syntax to calculate linear mixed models that predict the response variables—behavioral or self-reports—using the two manipulated variables feedback and agreement instructions. Unlike linear regression models that assume independence among observations, mixed models were necessary to account for non-independence between participants grouped together in the same experimental session. In each model the intraclass correlation coefficient (ICC) is calculated, producing the degrees of freedom for that model. ICC represents the proportion of variance that can be explained by differences between the groups, rather than between participants (Koch, 1982). When ICC is low, the response to the observed variable is independent of the grouping of participants into teams, and the number of degrees of freedom is closer to the number of participants. When ICC is

```
MIXED
RESPONSE_VARIABLE  BY AG FB gender
                   WITH PerExtr PerAgre PerCons PerStable PerOpen
  /PRINT = SOLUTION
  /RANDOM INTERCEPT | SUBJECT(GroupID) COVTYPE(VC)
  /FIXED = AG FB AG*FB gender PerExtr PerAgre PerCons PerStable PerOpen| SSTYPE(3)
  /EMMEANS = TABLES(AG*FB) COMPARE(AG) ADJ(BONFERRONI).
```

```
Where:
RESPONSE_VARIABLE = the variable being predicted

AG = the Agreement instructions manipulated variable; categorical with two levels:
More (A1) / Less (A0)

FB = the Feedback manipulated variable; categorical with two levels: With Feedback
(F1) / No Feedback (F0)

gender = gender of participant

PerExtr, PerAgre, PerCons, PerStable, PerOpen = five personality measures of
extravertion, agreeableness, conscientiousness, stability, and openness; continuous
variables.
```

Figure 15. The general syntax of the mixed model in SPSS used for analyzing the data of experiment 4.

high, a greater part of the variance in the response variable can be explain by the grouping, and the number of degrees of freedom is closer to the number of groups in the experiment.

In each of the statistical analyses, I also controlled for gender and for the five personality measures computed from the TIPI instrument. Controlling variables that were found not to have statistically significant effect on the response variable were removed from the model one-by-one to increase statistical power.

The generic syntax of the statistical models appears in Figure 15. The manipulating variables, feedback and agreement instructions, and the controlling variables, gender and personality measures, are all considered fixed factors. Participants were randomly grouped into three-member teams, hence the grouping in the RANDOM line. The line that starts with EMMEANS represents the calculation of the Estimated Marginal Means to examine specific contrasts of how feedback interacts with agreement instructions: it tests at each of the feedback levels (F1/F0) the difference between the two agreement instructions levels (A1/A0). Significant differences between A1 and A0 at the F1 level but not at the F0 level suggest that the feedback manipulation was powerful to elicit these changes. Bonferroni adjustment is used for multiple contrast tests, raising the standard for proof needed when testing few contrasts simultaneously (e.g., whether A0F0 differs from A1F0 and whether A0F1 differs from A1F1).

For the most part, I was not interested in the interaction effects between feedback and agreement instructions, since it involves six comparisons between the variables A0F0, A1F0, A0F1, A1F1, and in particular the comparisons between A0F0 and A1F1 and between A1F0 and A0F1 are not of interest in this work. However, adding this element to the model helps identify an interesting specific contrast between levels of one variable within each level of the other variable, and is required for carrying out this contrast at the syntax level.

Results

Manipulation Check

Consensual group norms. Participants' perceptions of consensual norms appear in Table 8a and Figure 16a. A main effect of the agreement instructions on the perception of consensual norms ($F(1,37)=15.42$, $p<0.001$) suggests that participants in groups that were instructed to agree more perceived the group norms as more consensual as compared to those instructed to agree less with each other. Running additional specific contrast comparing the two agreement instructions levels within each of the feedback levels reveals that this effect exists both when feedback is present ($F(1, 37)=9.63$, $p=0.004$) and when not ($F(1, 37)=6.04$, $p=0.02$). This suggests that regardless of seeing feedback, participants perceived higher consensual norms when they were in groups that were instructed to agree more with each other. Interestingly, feedback had a marginal main effect on the perception of consensual norms ($F(1,37)=3.05$, $p=0.09$), suggesting that those receiving feedback perceived higher consensual norms than those not receiving feedback, but specific contrast between the feedback levels within each of the two agreement instructions levels did not indicate significant differences between the groups on each of the agreement levels (A0: $F(1,37)=0.91$, *ns*; A1: $F(1,37)=2.27$, *ns*). The controlling variables, gender and the five personality measures, were not found as associated with the perception of consensual norms and were thus removed from the model.

Critical group norms. The results of participants' perceptions of critical norms are shown in Table 8b and Figure 16b. A main effect of agreement instructions on the

Table 8. Consensual and critical group norms means and standard deviations.

(a) Consensual Norms		Feedback		
		FB	noFB	Total
Agree instructions	More	4.18 (1.68)	3.55 (1.52)	3.87 (1.62)
	Less	2.91 (1.39)	2.52 (1.05)	2.72 (1.25)
	Total	3.52 (1.66)	3.03 (1.40)	3.28 (1.55)

(b) Critical Norms		Feedback		
		FB	noFB	Total
Agree instructions	More	4.76 (1.17)	4.97 (.77)	4.86 (.99)
	Less	5.49 (1.02)	5.43 (.86)	5.47 (.94)
	Total	5.14 (1.15)	5.20 (.84)	5.17 (1.01)

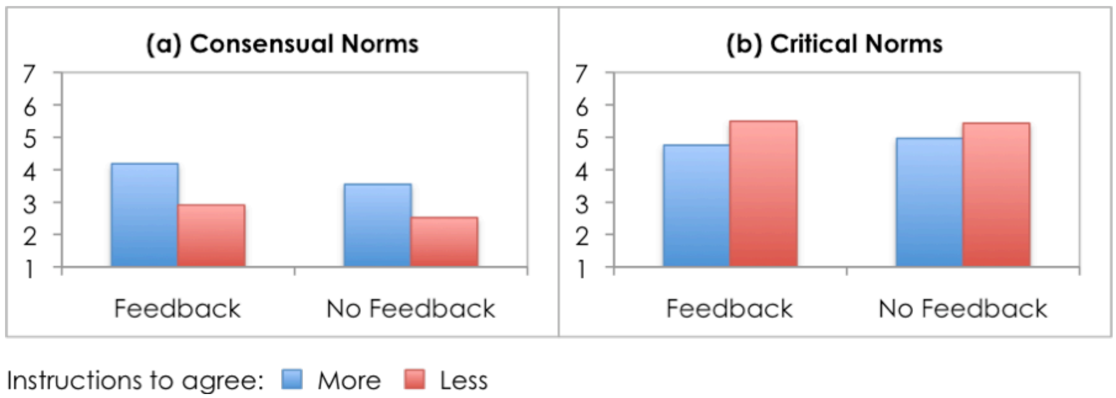


Figure 16. Consensual and critical group norms.

perception of critical norms ($F(1,118)=15.78$, $p<0.001$) suggests that participants in groups that were instructed to agree more perceived the group norms as less critical than those instructed to agree less with each other. Additional specific contrast comparing the agreement instructions within each level of feedback reveal that this effect was stronger when feedback was provided ($F(1,118)=13.67$, $p<0.001$) as compared to when not ($F(1,118)=3.66$, $p=0.06$). No main effect of feedback was found ($F(1,118)=0.10$, *ns*). Of the controlling variables, extraversion was found to have a significant association with critical group norms, with high extraverts perceiving higher critical group norms ($t(118)=2.78$, $p=0.006$), perhaps because they are more assertive than introverts (Costa & McCrae, 1992).

An interesting difference was found in the degrees of freedom between the statistical analyses for consensual norms and for critical norms. This difference results from a high intraclass correlation coefficient (ICC) to start with in the measurement of consensual norms (high portion of the variability is explained by the grouping), as compared to a low ICC to start with for critical norms (response to this measure is independent of the team grouping). In other words, the degree to which participants perceived consensual norms was dependent on the group in which they operated, whereas their perception of critical norms was independent of their group.

In summary, the manipulation of instructing teams to agree more or less with each other deemed successful. Encouraging higher levels of agreement by justifying it as supporting effective consensus reaching led individuals to perceive higher consensual norms and lower critical norms. On the other hand, encouraging lower levels of agreement by justifying it as enabling analytical discussion led to perceptions of lower consensual norms and higher critical norms. Furthermore, the data suggest that the feedback visualizations reinforced the effect of the agreement instructions: visualizing agreement words had a marginal effect on perceiving higher consensual norms, and strengthened the effect of the perception of critical norms.

Reflection

It was expected that individuals receiving feedback would engage more in reflecting on their behavior in general (H1a) and on their linguistic behavior in particular (H1b).

Reflection on behavior. To examine H1a, Grant et al's Self-Reflection and Insight Scale (2002) produced two measures, engagement in self-reflection and insight into self, presented in Table 9 and in Figure 17. A main effect of feedback on engagement in self-reflection (Table 9a and Figure 17a) suggests that participants who received feedback perceived that they were engaging in self reflection more than when

Table 9. Self-reflection and insight means and standard deviations.

(a) Self-Reflection		Feedback		
		FB	noFB	Total
Agree instructions	More	3.73 (0.88)	3.53 (0.75)	3.63 (0.82)
	Less	3.90 (0.82)	3.28 (0.89)	3.61 (0.90)
	Total	3.82 (0.85)	3.40 (0.82)	3.62 (0.86)

(b) Insight into self		Feedback		
		FB	noFB	Total
Agree instructions	More	5.34 (0.72)	5.19 (0.70)	5.26 (0.71)
	Less	4.98 (0.78)	5.33 (0.72)	5.15 (0.77)
	Total	5.15 (0.77)	5.26 (0.71)	5.21 (0.74)

Figure 17. Engagement in self-reflection and insight into self.

not receiving feedback ($F(1,118)=6.17$, $p=0.01$). Running additional specific contrast between feedback and no-feedback groups within each of the agreement instruction levels reveal that this effect was apparent when groups were instructed to agree less with each other ($F(1,118)=6.11$, $p=0.02$), but not when they were instructed to agree more ($F(1,118)=1.05$, *ns*). No main effect of the agreement instructions was found ($F(1,118)=0.03$, *ns*).

Of the controlling variables, agreeableness was found to have an effect on the engagement in self-reflection ($t(118)=2.03$, $p=0.04$): participants with higher agreeableness measures perceived themselves as more engaged in self-reflection. Agreeable individuals consider getting along with others as important (Costa & McCrae, 1992), and it is therefore plausible that these individuals were concerned about the team getting along together and as such their self-reflection measures where higher.

When examining the second measure, insight into self (Table 9b and Figure 17b), I did not find a main effect of feedback on this variable ($F(1,36.7)=0.82$, *ns*). However, although the agreement instructions also did not have a main effect ($F(1,38.1)=0.43$, *ns*), a significant agreement×feedback interaction effect was revealed ($F(1,39.4)=4.41$, $p=0.04$). To examine this effect, specific contrast between the two agreement instructions levels at each of the feedback levels was carried out, revealing that when provided with feedback, participants in groups instructed to agree more perceived having higher levels of insight into self, as compared to groups instructed to agree less ($F(1,40.9)=3.78$, $p=0.06$). This difference was not significant in groups not receiving feedback ($F(1, 36.8)=1.08$, *ns*). With feedback, perhaps those guided to agree more had a better understanding of their behavior, seeing the connection between their use of agreement words and the instructions they received. On the other hand, individuals instructed to agree less with each other could see less connection between the visualization and how they were guided to act. This is further discussed in the summary and discussion section below.

The personality measure of conscientiousness was found to have a significant effect on the measure of insight: individuals with high conscientiousness measures perceived that they had more insight into self ($t(117.0)=3.87$, $p<0.001$). These individuals are more self-controlled and directed (Costa & McCrae, 1992), and thus it is easier for them to understand their own behaviors and feelings.

The findings support H1a: participants receiving feedback were more engaged in reflecting on their behavior as compared to those not receiving feedback, especially when they were instructed to agree less with each other. Further, congruent with Grant et al.'s findings (2002), being involved in self-reflection did not necessarily lead to better understanding of one's feelings and behavior: although participants receiving feedback were more reflective, there was no difference between receiving and not receiving feedback on participants' self insight. Still, among those who received feedback, instructions to agree more led to higher perception of insight as compared to instructions to agree less.

Reflection on linguistic behavior. To test H1b, that linguistic feedback will stimulate individuals' reflection on their use of language, I looked at the aggregate measure of participants' perception of reflecting on their word choice (Table 10a and Figure 18a). A main effect of feedback (F(1,38.5)=20.77, p<0.001) suggests that

Table 10. Individuals' reflection on and perception of changing their word choice means and standard deviations.

(a) Reflection on word choice		Feedback		
		FB	noFB	Total
Agree instructions	More	4.48 (1.70)	3.80 (1.12)	4.14 (1.47)
	Less	5.09 (1.01)	3.52 (1.41)	4.34 (1.44)
	Total	4.80 (1.41)	3.66 (1.27)	4.24 (1.45)

(b) Perception of language change		Feedback		
		FB	noFB	Total
Agree instructions	More	4.22 (1.45)	3.57 (0.97)	3.89 (1.27)
	Less	4.12 (1.07)	3.18 (1.30)	3.67 (1.27)
	Total	4.17 (1.25)	3.38 (1.16)	3.78 (1.27)

Figure 18. Individuals' reflection on and perception of changing their word choice.

participants in groups receiving feedback reflected more on their use of language compared to groups not receiving feedback. Running additional specific contrast comparing the two levels of feedback reveals that this difference was apparent in both agreement instruction levels, but stronger for participants in groups instructed to agree less ($F(1,36.5)=19.36$, $p<0.001$) than for participants in groups instructed to agree more ($F(1,38.7)=4.62$, $p=0.04$). This finding is congruent with the findings for H1a, suggesting that when instructed to agree less, the effect of feedback on reflection is stronger. There was no main effect of agreement instructions ($F(1,36.2)=0.35$, *ns*). These findings replicate the conclusions of experiments 2 and 3 and support H1b: participants who received feedback reflected more on their use of language than those who did not receive feedback.

Of the controlling variables, emotional stability and openness to experience were found to have a marginally significant effect on reflection on language use: participants with higher emotional stability had lower measures of reflection ($t(117.0)=-1.81$, $p=0.07$), perhaps because they are less likely to feel tense (Costa & McCrae, 1992) and so they less feel the need to think about the words they choose in a conversation; those with higher openness measures had higher measures of reflection ($t(117.0)=1.72$, $p=0.09$), perhaps because they tend to be more attentive to their inner feelings (Costa & McCrae, 1992).

As suggested by the conclusions of experiment 3, the strength of dynamic feedback to elicit reflection might come with the cost of being distracted. On the aggregate measure of distraction, participants reported being more distracted when they received feedback as compared to when not (FB: $M=4.0$, $SD=1.34$; no-FB: $M=3.6$, $SD=1.16$), although this difference failed to reach statistical significance ($F(1,37)=2.713$, $p=0.11$).

Modifying Use of Language

Perception of changing language use. An aggregate measure from the post-session survey assessed whether participants perceived that they were changing their use of language during the experiment (Table 10b and Figure 18b). A significant main effect of feedback suggests that participants receiving linguistic feedback believed they were trying to change their word choice more than those who did not receive feedback ($F(1,37)=10.09$, $p=0.003$). Additional specific contrast comparing the two feedback levels within each level of agreement instructions reveals that this difference was significant in groups that were instructed to agree less ($F(1,37)=7.21$, $p=0.01$) and marginally significant in groups instructed to agree more ($F(1,37)=3.30$, $p=0.08$). There was no significant effect of agreement instructions ($F(1,37)=0.92$, *ns*). Gender and personality measures were not found to have a statistical effect on the measure of perception of language change.

According to the second set of hypotheses, I expected that individuals receiving feedback would change their use of language in the direction of the instructions they were given. In other words, when instructed to agree more with each other, individuals receiving feedback on their percentage of agreement words were expected to express more agreement compared to those not receiving feedback (H2a), and when instructed to agree less, individuals receiving feedback were expected to express less agreement compared to those not receiving feedback (H2b). These hypotheses call for behavioral measures beyond participants' self-reports on whether they modified their word choice or not.

Automated analysis of language use. To examine hypotheses H2a and H2b, I analyzed participants' entries in the chat transcripts using LIWC (Table 11 and Figure 19). The LIWC category of primary interest is 'assent', or 'agreement' (Table 11a and

Table 11. Individuals' use of language, based on LIWC analysis, means and standard deviations.

(a) Agreement words		Feedback		
		FB	noFB	Total
Agree instructions	More	6.88 (2.51)	4.09 (1.74)	5.56 (2.58)
	Less	4.52 (1.67)	3.81 (1.87)	4.18 (1.79)
	Total	5.64 (2.41)	3.94 (1.80)	4.84 (2.29)

(b) Word count		Feedback		
		FB	noFB	Total
Agree instructions	More	566.2 (214.2)	501.3 (174.3)	535.5 (197.3)
	Less	500.6 (207.3)	469.5 (192.5)	485.8 (199.4)
	Total	531.8 (211.5)	484.5 (183.1)	509.4 (199.1)

(a) Thinking words		Feedback		
		FB	noFB	Total
Agree instructions	More	18.8 (2.57)	20.0 (2.3)	19.4 (2.5)
	Less	20.7 (2.5)	19.5 (2.6)	20.1 (2.6)
	Total	19.8 (2.7)	19.7 (2.4)	19.7 (2.5)

(b) Self references		Feedback		
		FB	noFB	Total
Agree instructions	More	3.78 (1.37)	2.73 (1.06)	3.28 (1.33)
	Less	2.42 (1.24)	2.87 (1.18)	2.63 (1.23)
	Total	3.07 (1.47)	2.8 (1.12)	2.94 (1.31)

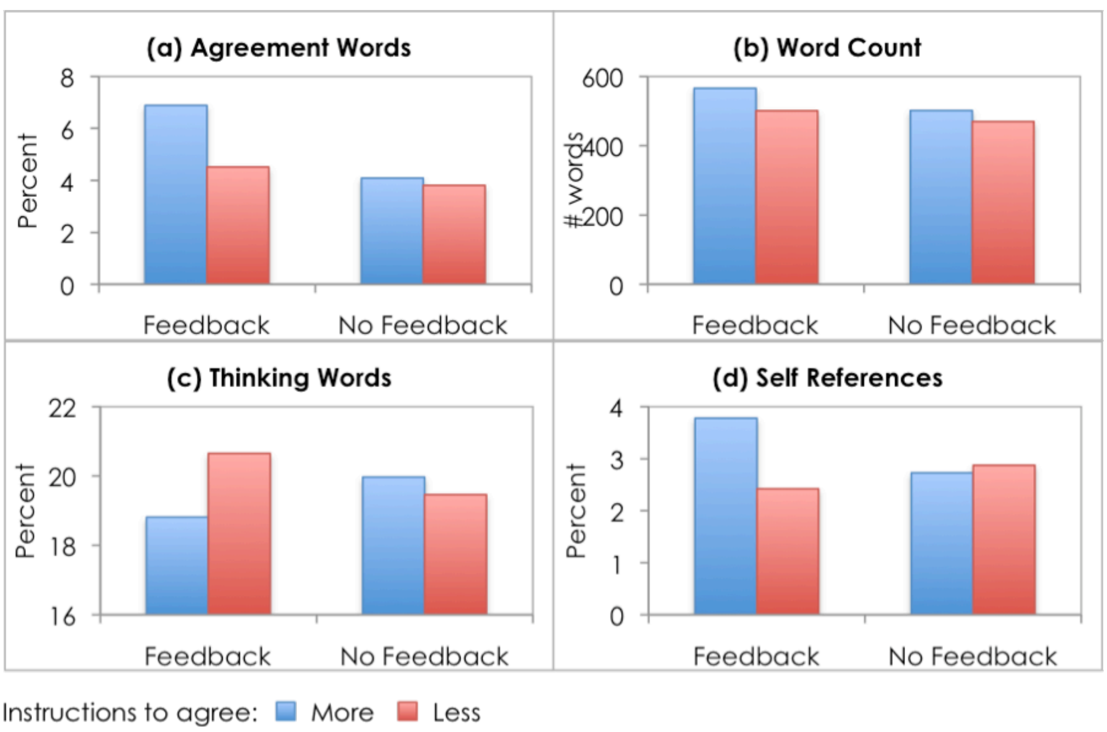


Figure 19. Results of LIWC analysis on four categories.

Figure 19a), which includes words that represent agreement. A main effect of the agreement instructions was found, further supporting the success of the manipulation: when told to agree more, participants had higher percentages of agreement words as compared to those instructed to agree less ($F(1,37.0)=9.62$, $p=0.004$). Also, a main effect of feedback ($F(1,36.0)=14.30$, $p=0.001$) indicates that participants who received feedback used more agreement words than those who did not receive feedback. A significant interaction effect ($F(1,36.8)=5.96$, $p=0.02$) was followed by additional specific contrast between the two feedback levels at each of the agreement instruction levels, showing that when instructed to agree more, the difference between the two feedback levels was significant ($F(1,36.1)=18.50$, $p<0.001$), whereas when instructed to agree less the difference was not significant ($F(1,36.7)=0.91$, *ns*).

The latter findings suggest that H2a was supported, such that when instructed to agree more with each other, receiving feedback motivated participants to type in more agreement words as compared to those who did not receive feedback. However, the effect was not reversed when instructed to agree less, in which case participants did not significantly differ in their production of agreement words, and so H2b was not supported.

Among the controlling variables, gender was found to have a significant effect: females used significantly more agreement words than males ($t(102.1)=2.48$, $p=0.02$). This is similar to other studies finding that women can be characterized by expressing agreements in online discussions (Schler, Koppel, Argamon, & Pennebaker, 2006) and in spoken conversation (Jurafsky, Ranganath, & McFarland, 2009). Extraversion was also found to be associated with percentage of agreement words, with high extraverts expressing less agreement ($t(94.8)=-2.04$, $p=0.04$), perhaps because they tend to be more assertive and dominant than introverts (Costa & McCrae, 1992).

Table 11 and Figure 19 also show the results of the LIWC categories word count (b), words that represent thinking such as 'consider' and 'because' (c), and references to self (1st person singular pronouns) such as 'I' and 'me' (d). While not directly related to the hypotheses, these categories shed more light on the potential of automated linguistic feedback to affect how team members communicate with each other.

Regarding word count, from Table 11b and Figure 19b it seems that those receiving feedback talked somewhat more than those not receiving feedback. The effect of feedback was marginally significant ($F(1,37.5)=3.46$, $p=0.07$), suggesting that seeing the feedback visualization might have motivated participants to type in more words to the chatroom. The effect of the agreement instructions was not found to be statistically significant ($F(1,36.6)=0.21$, *ns*). Controlling for gender and personality measures, extraversion was found to have a significant effect: high extraverts talked more than introverts ($t(99.4)=5.55$, $p<0.001$), being more socially active (Costa & McCrae, 1992).

This effect was strengthened when counting the total number of entries participants sent to the chatroom (A1F1: $M=94.5$, $SD=51.2$; A0F1: $M=59.3$, $SD=24.3$; A1F0: $M=57.7$, $SD=31.4$; A0F0: $M=58.4$, $SD=28.7$). Participants receiving feedback sent significantly more chat entries ($F(35.9)=5.84$, $p=0.02$). A marginally significant agreement×feedback interaction effect ($F(1,36.5)=2.95$, $p=0.09$) was followed by specific contrast between the two feedback levels at each of the agreement instruction levels, finding that when instructed to agree more, participants typed in significantly more entries when receiving feedback as compared to when not ($F(1,36.0)=8.16$, $p=0.007$), whereas the difference was not significant when instructed to agree less ($F(1,36.5)=0.25$, *ns*). In other words, the when instructing participants to agree more, showing them feedback about their agreements induced them to send more messages to the chatroom.

When examining words that represent thinking (Table 11c, Figure 19c), no significant main effects of feedback or of agreement instructions were found (Feedback: $F(1,36.0)=0.002$, *ns*; Agreement instructions: $F(1,36.0)=1.57$, *ns*). However, the two variables significantly interact ($F(1,37.2)=4.11$, $p=0.05$), and therefore specific contrast was carried out between the two levels of agreement instructions at each of the feedback levels. This revealed that when feedback was provided, participants instructed to agree more had lower percentages of thinking words than those instructed to agree less ($F(1,37.1)=5.63$, $p=0.02$). When feedback was not present the direction of change was opposite although not statistically significant ($F(1,36.2)=0.30$, *ns*). This suggests that when providing feedback about agreement expressions, guiding groups toward less agreement (justified by a more methodical teamwork process) can lead to higher levels of thinking expressions in the group conversations. Of the controlling variables, it was found that women typed in higher percentages of thinking words as compared to men, in line with previous findings showing that women tend to use more words representing discrepancy, such as ‘should’, and ‘would’ (Mehl & Pennebaker, 2003b; Pennebaker & King, 1999).

Examining self-references (Table 11d, Figure 19d), no effect was found of feedback ($F(1,36)=1.32$, *ns*), but there was a significant main effect of agreement instructions ($F(1,36)=5.35$, $p=0.03$), suggesting that participants instructed to agree more expressed higher percentages of self-references as compared to those instructed to agree less. Further, a significant agreement×feedback interaction effect ($F(1,36)=8.24$, $p=0.007$) was followed by specific contrast between the agreement instructions levels at each of the feedback levels, revealing that when feedback was present, the difference between the two agreement instruction levels was significant

($F(1,36)=14.14$, $p=0.001$), but not when feedback was not provided ($F(1,36)=0.15$, *ns*). These results suggest that when providing feedback, guiding groups to agree more could lead to higher levels of involvement in the group interaction process, indicated by higher levels of self-references (Cegala, 1989). None of the controlling variables were associated with this measure.

Manual coding of communication patterns. To further examine hypotheses H2a and H2b, I looked at the results of the manual coding of the chat transcripts. The distribution of the seven types of chat entries across the four experimental conditions appears in Table 12 and Figure 20. As can be seen, the bulk of the conversation across all conditions was spent on discussing the ideas that were raised, followed by expression of agreement. Figure 21 shows the average percentage of each entry type out of the total number of entries in every condition.

The key chat entry type for addressing the hypotheses is agreement entries (Figure 21b). A significant main effect of feedback indicates that regardless of the instructions to agree more or less, receiving feedback was associated with higher percentages of agreement entries ($F(1, 36.1)=10.76$, $p=0.002$). Participants instructed to agree more had somewhat a higher percentage of agreement entries than those instructed to agree less, although this effect failed to reach statistical significance ($F(1,37.4)=2.08$, *ns*). Of the controlling variables, extraversion had a marginally significant effect on percentage of agreement entries: extravert participants had lower percentages of agreement entries ($t(110.2)=-1.84$, $p=0.07$), similar to the LIWC results on agreement words.

These results provide only a partial answer to H2a and H2b. Feedback had an effect on the expression of agreement as indicated by manual coding of the chat transcripts, but not necessarily in the direction of the agreement instructions provided

Table 12. Percentages of chat entry types means and standard deviations.

Feedback	FB		noFB	
Agreement Instructions	**More**	**Less**	**More**	**Less**
Idea	7.3 (5.2)	7.8 (4.9)	11.0 (6.3)	10.1 (7.0)
Agreement	23.5 (7.8)	23.3 (8.7)	19.3 (8.8)	15.1 (9.0)
Disagreement	3.8 (5.3)	2.9 (3.0)	3.0 (3.0)	4.5 (4.0)
Discussion	48.2 (15.6)	51.3 (12.3)	51.1 (12.1)	54.7 (14.7)
Choice	3.5 (2.6)	3.8 (2.8)	3.5 (3.1)	5.1 (4.1)
GroupMeter	7.6 (9.2)	1.1 (2.0)	2.6 (2.8)	1.0 (1.7)
Management	6.1 (5.1)	9.9 (6.2)	9.6 (8.1)	9.5 (6.5)

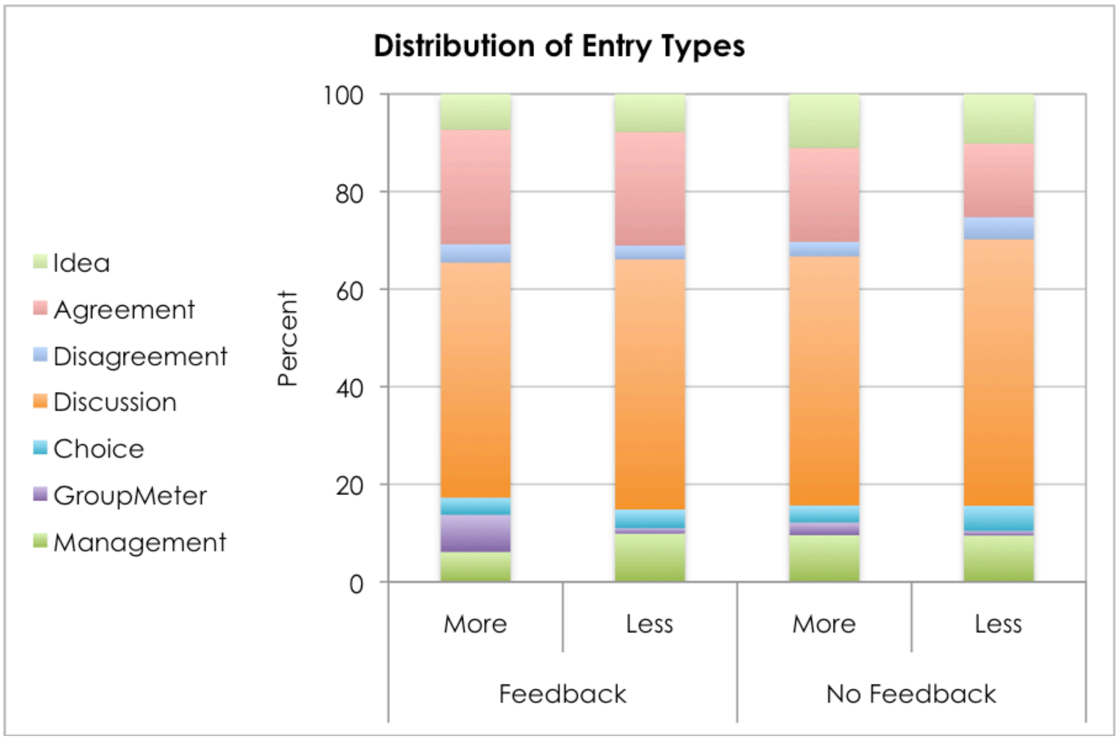


Figure 20. Distribution of manually coded entry types across the conditions.

to participants. This suggests that the visualization had a strong effect on motivating participants to express more agreement despite of what they were told to do, and is further discussed in the discussion section.

Although the other chat entry types are not directly related to the hypotheses, their examination can reveal important aspects of how automated linguistic feedback

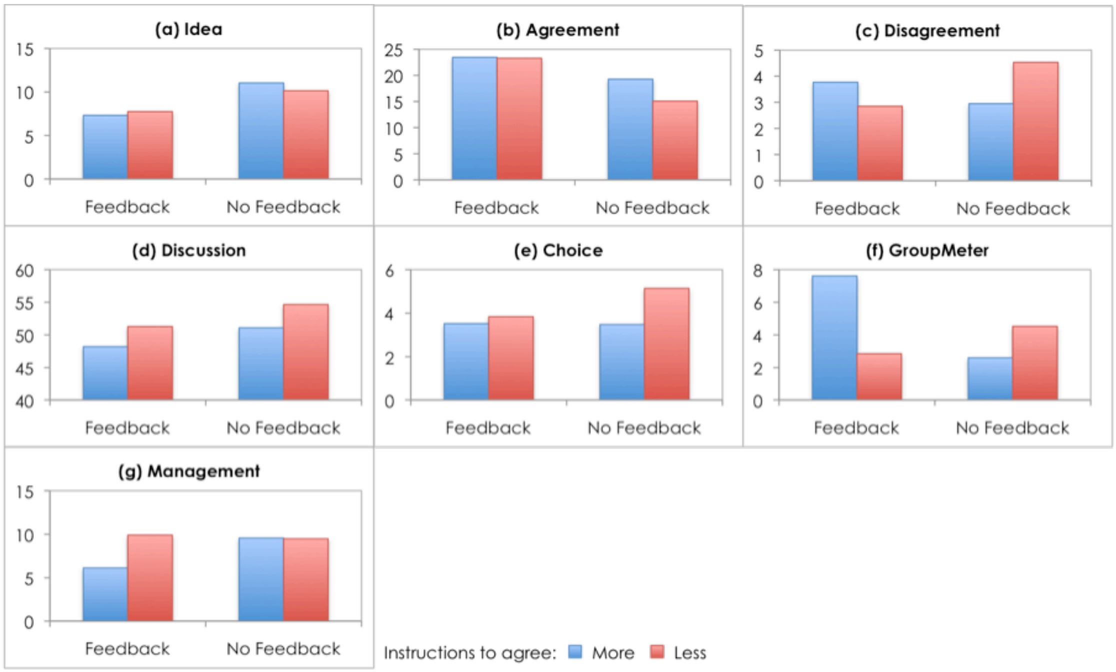


Figure 21. Proportions of each entry type by condition. For each condition, its bars sum to 100% across the seven entry types.

affects communication patterns in teams.

A significant effect of feedback on idea entries (Figure 21a) suggests that regardless of the agreement instructions, when receiving feedback participants had lower percentages of idea entries as compared to those not receiving feedback ($F(1, 35.8)=4.98$, $p=0.03$). Note that this does not necessarily mean that participants receiving feedback produced fewer ideas, but that a lower fraction of their entries were idea entries. To examine this, I counted the total number of ideas generated by groups, which can also be viewed as a measure of performance for the brainstorming task. The number of ideas in groups receiving feedback was indeed lower compared to groups not receiving feedback (FB: $M=12.5$, $SD=5.2$; no-FB: $M=13.5$, $SD=5.9$), although the difference was not significant ($t(38)=0.57$, *ns*). No effect of the agreement instructions was found on percentage of idea entries ($F(1,36.6)=0.20$, *ns*).

Of the controlling variables, extraversion and conscientiousness had a significant effect on percentage of idea entries: high extraversion was associated with lower idea entry percentages ($t(93.0)=-2.87$, $p=0.005$), and high conscientiousness was associated with lower idea entry percentages ($t(105.0)=-2.02$, $p=0.05$). Again, this does not mean that extravert participants and conscientious participants had fewer ideas. Rather, their distribution of entry types was such that a lower fraction of their entries were coded as ideas.

Another significant effect was found in the chat entries in which participants talked about the GroupMeter system and user interface (Figure 21f). In total, 222 entries were coded as GroupMeter entry type (3% of all coded entries). Still, an interesting pattern emerged. Specific contrast comparing the two agreement instructions levels at each of the feedback levels revealed that when feedback was provided, participants who were instructed to agree more had a higher percentage of GroupMeter entry types as compared to participants instructed to agree less ($F(1,36)=9.21$, $p=0.004$). However, there was no such significant difference when feedback was not provided ($F(1,36)=0.53$, *ns*). Perhaps the combination of the feedback and the instructions to agree more inspired participants to comment about the visualization, as in the following example from one of the groups in the A1F1 condition:

A: why is my fish low

B: i think its upset

A: am i being mean?

C: yes

C: yes you are

B: you just want your fish to move up

This part of conversation exemplifies how the feedback enabled participant A to reflect on her social behavior, worrying about conforming to the appointed standard for behavior. It also exemplifies that some individuals started to say positive remarks regardless of discussion context, as pointed out by B's final remark responding to C's two entries, a behavior that can be inferred as trying to "game the system".

The entry types disagreement (Figure 21c), discussion (Figure 21d), choice (Figure 21e), and management (Figure 21g) did not reveal significant main effects of the feedback or the agreement instruction.

In total, the findings provide partial support for H2a and H2b. Participants reported that they were trying to change their use of language when exposed to linguistic feedback. The behavioral findings also support that receiving feedback stimulated participants to type in higher percentage of words representing agreement and higher percentage of entries expressing agreement. H2a was supported by the word-based analysis results, in that participants who were instructed to agree more typed in higher percentages of agreement words when exposed to feedback. However, participants instructed to agree less did not type in lower percentages of agreement words when receiving feedback, and as such H2b was not supported. Further, this interaction was not replicated with the manual coding data. To sum, receiving changed participants' perceptions of and actual use of language, but not always in the direction toward which they were guided.

Change in Teamwork Behavior

The final set of results examines if feedback affects dimensions of interpersonal behavior characterizing the team interaction process. It was hypothesized that when receiving automated feedback, teams will move closer to what are considered to be effective teamwork behaviors, as defined by positioning on the space created by the three SYMLOG dimensions (Table 13 and Figure 22): even distribution of participation within the team (H3a), friendliness and group-orientation (H3b), and a balance between task-focus and socio-emotional expressiveness (H3c).

Participation. To measure participation levels, I looked both at participants' peer ratings on the SYMLOG participation dimension, ranging from submissive to dominant, and at their actual input to the group chat conversation. Although participants sent significantly more entries to the chatroom when receiving feedback and especially when instructed to agree more with each other (see word counts and entry counts reported above), their scores on the participation dimension (Table 12a and Figure 22a) were not significantly affected by the feedback ($F(1,38.9)=1.35$, *ns*) or by the agreement instructions ($F(1,38.2)=0.12$, *ns*). Similar to the behavioral

Table 13. Peer ratings on the three SYMLOG dimensions means and standard deviations.

		(a) Participation			(b) Friendliness			(c) Social/Task Focus		
		Feedback			Feedback			Feedback		
		FB	noFB	Total	FB	noFB	Total	FB	noFB	Total
Agree instructions	More	5.72 (1.22)	5.22 (0.94)	5.47 (1.11)	6.28 (0.97)	6.27 (0.61)	6.28 (0.80)	4.23 (1.26)	4.97 (1.10)	4.60 (1.23)
	Less	5.21 (0.84)	5.32 (1.00)	5.26 (0.91)	5.85 (0.77)	5.98 (0.74)	5.91 (0.75)	5.32 (0.91)	5.30 (0.91)	5.30 (1.02)
	Total	5.45 (1.06)	5.27 (0.96)	5.36 (1.01)	6.06 (0.89)	6.13 (0.69)	6.09 (0.79)	4.80 (1.21)	5.13 (1.13)	4.96 (1.18)

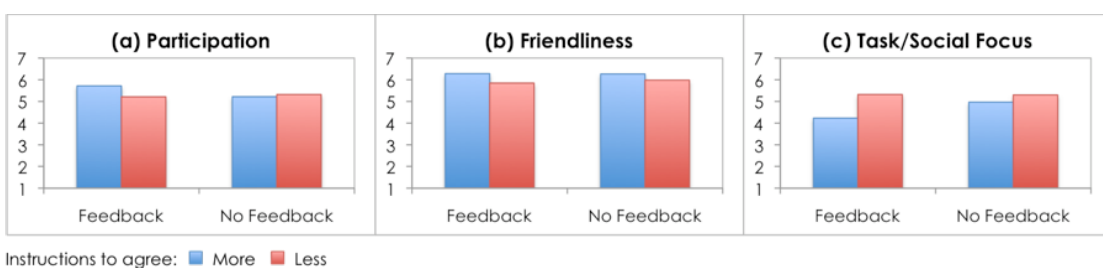


Figure 22. Individuals' scores on the SYMLOG dimensions, based on peer evaluations.

findings, participants with high extraversion measures received higher participation scores ($t(104.9)=4.59$, $p<0.001$), perhaps because they are more socially active (Costa & McCrae, 1992).

To test H3a, I calculated the range of participation scores within each team, between the member who received the highest average score and the member who received the lowest average score (A1F1: $M=1.8$, $SD=1.3$; A0F1: $M=1.5$, $SD=0.7$; A1F0: $M=1.4$, $SD=0.7$; A0F0: $M=1.6$, $SD=1.0$). Smaller ranges would characterize more equal distribution of participation. The results show no significant effect of the feedback ($F(1,37)=0.12$, *ns*) or the agreement instructions ($F(1,36)=0.03$, *ns*). This was replicated when looking at the actual range of word count within each team, between the team member who spoke the most and the one who spoke the least (A1F1: $M=296.3$, $SD=182.1$; A0F1: $M=330.9$, $SD=201.2$; A1F0: $M=279.7$, $SD=79.8$; A0F0: $M=347.7$, $SD=193.3$). There was no significant effect of the feedback ($F(1,36)=0.0$, *ns*) or the agreement instructions ($F(1,36)=0.865$, *ns*). These findings suggest that H3a was not supported – I did not find evidence that participation was more evenly distributed in teams receiving automated linguistic feedback.

Friendliness. Participants' scores on the SYMLOG friendliness dimension, ranging from individual-oriented to group-oriented, as rated by their peers, appear in Table 13b and Figure 22b. No significant effect of feedback was revealed ($F(1,35.5)=0.103$, *ns*). Participants instructed to agree more with each other seemed to receive more group-oriented scores, although this tendency failed to reach statistical significance ($F(1,35.1)=2.84$, $p=0.10$). The scores were generally high, averaging around 6 on a seven-point scale, and the standard deviations small – less than 1 point in all conditions. This can be explained by the setup of the experiment in which participants were trying to work together, had no incentive not to be friendly, and there was no reason for tension to arise in the group. This might be a possible explanation for the missing results; examining the effect of feedback in real groups outside the lab or in settings where individuals have conflicting individual goals could help address the absence of the current results.

Among the controlling variables, conscientiousness was significantly associated with peer ratings on the friendliness dimension: participants with highly conscientious measures received more individual-oriented (i.e., lower) scores ($t(101.3)=-2.83$, $p=0.006$). This might be attributed to the idea that conscientious individuals strictly adhere to principles and are achievement-oriented (Costa & McCrae, 1992), which could be interpreted as being unfriendly.

In sum, H3b was not supported – I did not find evidence that participants were more friendly or group-oriented when receiving automated linguistic feedback.

Task/social orientation. Table 13c and Figure 22c show participants' average peer ratings on the SYMLOG task/social orientation dimension, ranging from task-focused to socio-emotionally expressive. In this experimental setting that encouraged task-oriented behaviors (no time for chit chat, no history or anticipated future interactions), I expected to see high ratings on this dimension, indicating that participants were perceived as highly task-focused. As such, lower ratings that are closer to the midpoint of the 7-point scale would indicate a balance between task- and social-orientation.

Participants who received feedback had scores closer to the center of the task/social-orientation dimension, as compared to participants not receiving feedback who had more task-focused scores, although this failed to reach statistical significance ($F(1,37)=1.704$, *ns*). Interestingly, a main effect of agreement instructions on the scores on this dimension was found ($F(1,37)=6.70$, $p=0.01$), such that when instructed to agree more with each other, participants were perceived as less task-focused and more socially-oriented. Specific contrast within each level of feedback suggest that the difference between the two agreement instruction levels exists when feedback was provided ($F(1,37)=8.03$, $p=0.007$), but not when there was no feedback ($F(1,37)=0.72$, *ns*). This suggests that when feedback was provided, instructing teams to agree more with each other led members to be perceived as more socially-oriented, as compared to instructing them to agree less in which they were rated as more task-focused. None of the controlling variables were found to have significant effects on this variable.

It could be that because participants were instructed to agree more and saw the feedback about their agreement word percentage, they typed in more agreements (see response to H2a above), and because they expressed more agreements, they were viewed by their peers as more socio-emotionally expressive. To test if the expression of agreement mediated the relationship between agreement instructions, feedback, and task/social orientation ratings, I added the LIWC 'assent' category to the model as a controlling variable. A main effect of percentage of agreement words on task/social orientation scores was found, such that those expressing more agreement were perceived as less task focused and more socially-oriented ($t(111.6)=-2.53$, $p=0.01$). However, when receiving feedback, individuals instructed to agree more still received more socially-oriented ratings compared to those instructed to agree less ($F(1,40.5)=3.80$, $p=0.06$).

These findings provide partial support for H3c: individuals were perceived as more task/socially-balanced when receiving automated linguistic feedback, but only when they were guided to express agreement with each other to a greater extent.

Summary of Findings and Discussion

The three sets of hypotheses for this experiment proposed that visualizing automated linguistic feedback about percentage of agreement words would: stimulate team members to reflect on their behavior in general (H1a) and on their word choice in particular (H1b); guide team members to express more agreement (H2a) or less agreement (H2b) with each other according to the instructions they were given; and lead to improved teamwork behaviors with more equal distribution of participation (H3a), friendlier behaviors (H3b), and a better balance between task-focus and socio-emotional expressiveness (H3c).

First, the findings of this experiment support H1a and H1b: participants who received dynamic feedback about their language use engaged more in reflection on their behavior and on their word choice. The decision to allocate explicit time for reflection on the feedback deemed valuable, as it allowed team members to engage in thinking about what they can do to change their behaviors given the feedback they received, in line with Gersick's model of team development (1988; 1989; 1991). The next step would be to examine more specifically the interaction between dynamic feedback and reflection time: is it the dynamic feedback, the explicit time allocated for reflection, or are both factors important for stimulating reflection and further changes in behavior? Further research is needed to clarify this question.

The effect found was particularly stronger for participants instructed to agree less with each other. In line with an embodied view of language representations (Lakoff & Johnson, 1981; Richardson, Spivey, Barsalou, & McRae, 2003), the incompatibility between the language instructions (less agreement is better) and their spatial representation (more agreement moves fish higher; higher is better) might have required more thought when translating the instructions into action that would be represented in the visualization.

Second, the findings support H2a, but do not support H2b. Automated analysis of the chat transcripts revealed that when instructed to agree more with each other, feedback about percentage of agreement words led to higher rates of agreement word production, supporting H2a. This can be explained, again, by an embodied cognition view – the movement of the fish up with higher percentages of agreements was compatible with the instructions to agree more for effective teamwork, thus leading to the expression of more agreement words when seeing the visualization.

However, when instructed to agree less, participants did not exhibit lower rates of agreement word production when provided with feedback, thus not supporting H2b. Further, manual coding of the chat transcripts revealed that when provided with feedback on agreement word percentage, participants typed in higher rates of entries that represent agreement, regardless of the instructions they received to agree more or less. Perhaps interpreting the fish moving up with a more desirable behavior, based on the embodied cognition perspective described above, created nonconscious goal priming that was stronger than the explicit instructions provided at the beginning of the experimental session (Chartrand & Bargh, 1996). Reversing the visualization, such that the fish move down with more agreement words, or alternatively move up with the expression of disagreement words, could have settled this asymmetry in the findings, and deserves future investigation.

Finally, the results do not show support for H3a and for H3b, and they partially support H3c. Teams receiving feedback did not exhibit a more equal distribution of participation, either perceived or actual (H3a), and their members were not rated as friendlier than in teams that did receive feedback (H3b). When instructed to agree more with each other, individuals receiving feedback were rated as more task/socially balanced, but this difference did not occur when instructed to agree less, thus partially supporting H3c. It could be that these hypotheses were too ambitious, since participants did not receive guidance that would connect their linguistic behavior represented by the feedback visualizations to effective teamwork behaviors over the SYMLOG framework, as suggested by Feedback Intervention Theory (Kluger & DeNisi, 1996).

Still, the findings supporting H3c are encouraging in light of previous findings in which participants found it hard to be more socio-emotionally expressive in a task-oriented study (McLeod et al., 1992). One possible explanation is that the combination of guiding teams to express more agreement and visualizing these expressions encouraged a more relaxed and informal style of interaction within the team (Whittaker, Frohlich, & Daly-Jones, 1994). Another possibility is that because feedback made individuals more reflective about their behavior (H1a), they became aware of the emotional processes shaping their interpersonal behaviors that are important for effective teamwork (McLeod & Kettner-Polley, 2004). A less optimistic explanation is that participants were simply exhibiting gaming behavior, their excessive use of agreement words resulting in them being perceived as less task-focused. More research is needed to fully understand the distinction between users modifying their language in order to manipulate the real-time visualization, and the actual adoption of effective teamwork skills outside of the laboratory.

CHAPTER 7

GENERAL DISCUSSION

Previous research on teamwork and feedback has demonstrated that providing process feedback to teams, i.e., feedback that includes information about the social interaction process rather than task performance, can guide team members to reflect on their social behaviors and strive to change them toward improving the team interaction process. This has been demonstrated using both human-generated feedback (Cho et al., 2008; Losada et al., 1990; McLeod et al., 1992; Turner & Schober, 2007) and automated feedback (Bergstrom & Karahalios, 2007b; DiMicco et al., 2007; Janssen et al., 2007; Janssen et al., 2007).

At the same time, there has been a line of research for tying together language use and social behaviors in group settings. A long tradition of small group research, spanning many decades, used the analysis of communication acts in team conversations to achieve a descriptive understanding of how teams operate (e.g., Bales, 1950; Poole, 1981; Poole, 1983a; Poole, 1983b; Poole & Roth, 1989a; Poole & Roth, 1989b), as well as normative models of effective behaviors for team performance (Futoran, Kelly, & McGrath, 1989; Morris, 1966). More recent work applied state-of-the-art automated language processing techniques for identifying elements of team conversation that can be linked to team performance and learning (Gorman et al., 2003; Janssen et al., 2007; Joshi & Rosé, 2007; Martin & Foltz, 2004) and to social behaviors (Hancock et al., 2005; Mehl & Pennebaker, 2003a; Scissors et al., 2009; Slatcher & Pennebaker, 2006; Turner & Schober, 2007).

The goal of the present research has been to connect these two lines of research together, examining the efficacy of automated language analysis to reveal linguistic indicators that would serve as a source of feedback to help team members attend to, reflect on, and improve their communicative and teamwork behaviors. Therefore, the two questions that underlie this research, and that were threaded through the theory development, technology design, and empirical research were:

RQ1. Does providing automated linguistic feedback stimulate team members' reflection on their use of language in a teamwork setting?

RQ2. Does this feedback motivate team members to change their use of language toward improved teamwork behaviors?

The research platform designed to address these questions, GroupMeter, was developed through an iterative process over three years. The design requirements leading to the design of the system were drawn from multiple disciplines, such as human-computer interaction, social psychology, cognitive science, and software engineering. The requirements were to allow groups interact via computer-mediated text-based communication (1), display feedback visualization that is dynamic (2), peripheral (3), and public at the individual level (4), avoid modeling contextual factors (5), and implement a modular system that enables iterative future changes (6). The experiments that followed where used to improve and refine the GroupMeter system, and more importantly, to answer the research questions and explain the nuanced factors that underlie their theoretical foundations.

Summary of Research Findings

The first experiment in this research supported the idea that visualizing feedback (in this case, peer-based) can result in changes in teamwork behaviors, and identified a set of candidate linguistic features to be used as a source of automated feedback in computer-mediated team conversations. Associating feedback based on peer evaluations with automated analysis of conversation transcripts using LIWC (Pennebaker et al., 2001), led to the discovery of linguistic features that could predict peer evaluations, and that changed as a result of providing peer feedback. In line with previous research, I found evidence that particularly promising linguistic features based on word-level analysis are word count, representing amount of contribution (see also Bergstrom & Karahalios, 2007a; DiMicco et al., 2004; Janssen et al., 2007; Kim et al., 2008; Viégas & Donath, 1999), percentage of emotionally-charged words (see also Scissors et al., 2009; Turner & Schober, 2007), percentage of agreement words (see also Janssen et al., 2007), and percentage of 1st person pronouns (see also Burke et al., 2007; Cassell & Tversky, 2005; Cegala, 1989; Hancock et al., 2005; Mehl & Pennebaker, 2003a). These linguistic features were embedded into the GroupMeter feedback visualizations, making it ready for experiments 2, 3 and 4.

One possible explanation for the findings of experiment 1 was that the act of engaging in the peer evaluation procedure, and not the feedback per-se, increased participants' awareness of their behaviors and how others see them, and as a result participants changed their communicative behavior (Dominick et al., 1997) in order to manage the impressions others make of them (Goffman, 1959; Leary & Kowalski, 1990). Thus, experiments 2 and 3 were exploratory in nature, seeking to understand under what circumstances automated language-based feedback can fulfill its promise to trigger reflection and behavioral changes toward improved teamwork behaviors.

Experiments 2 and 3 provided initial evidence for RQ1 and RQ2, with implications spanning a theoretical perspective for understanding the connections between of language use, automated feedback, and teamwork process, methodological issues applied in the design of experiment 4, and design considerations that were implemented into the following versions of the GroupMeter system.

Experiment 2 served primarily as a pilot study. Participants worked in teams via the GroupMeter chatroom and either saw feedback visualizations in the form of bar graphs on their word count, percentage of emotion words, and percentage of 1st person singular pronouns, or not. The results demonstrated that dynamic visualizations of automated linguistic behavior helped participants experience, reflect on, and richly interpret their own communication behaviors. However, I did not find evidence that participants receiving feedback changed their use of language compared to participants who did not see the feedback visualization.

One methodological understanding from the results of experiment 2 was that the linguistic features chosen, especially emotion words and self-references, were not clearly explained and linked to teamwork behaviors. This might have led to a range of competing interpretations of what effective responses to the feedback should be, leading to contradicting uses and balanced out behavioral effects (Hayes & Walsham, 2000). Design-wise, the unobtrusive bar graphs at the bottom of the chat window were too subtle and easy to disregard. And theoretically, if the feedback visualization is ignorable and it is hard to decipher how to respond upon perceiving it, then it is only natural to avoid changing one's language production that is considered spontaneous, unintentional, and uncontrolled in the first place (Chung & Pennebaker, 2007; Levelt, 1989; Pennebaker, 2002).

Experiment 3 examined RQ1 and RQ2 again, after addressing methodological and design issues from experiment 2. Using a smaller scale study and a within-subject design, and in preparation for experiment 4, the experiment tested two different designs of the feedback visualizations, the bar graphs and the newly designed fish, comparing them to each other and to a version that had no feedback visualizations. Further, I chose to use the linguistic features word count and percentage of agreement words, and designed into the experimental procedure an explanation of how they are computed and translated into the feedback.

The findings of experiment 3 were encouraging in the sense that they provided evidence for both research questions: participants reported that they were more reflective on their use of language when receiving feedback, and analysis of team conversation transcripts revealed distinct communication patterns when receiving feedback.

The results of experiment 3 also raised some issues. Participants' self-reports as well as behavioral findings suggested that the fish visualization was engaging, interesting, and thought provoking, but attending to it came with a cost of distracting teams from the task. The need to constantly monitor changes in the display was addressed by designing the bubble trails, applied in the next version of GroupMeter. However, beyond perceptual distraction and on a more theoretical level, the question raised is whether this shift of focus represents tension between attending to the task and to socio-emotional behaviors.

Another issue in the results of experiment 3 relates to the direction of observed behavioral changes. Although experiment 1's findings demonstrated a negative association between the production of agreements and peer ratings, participants became more agreeable when seeing feedback about their expression of agreement. The target behaviors were intentionally left ambiguous, to see how teams develop their own interpretations of what desirable behaviors are. Theoretically, this suggests that linguistic feedback could lead to detrimental team process, in this case, groupthink (Janis, 1972), and raises a question of how we can steer behavioral changes toward more productive directions. As a methodological move, I decided for experiment 4 to provide explicit guidance as to how to respond communicatively upon receiving linguistic feedback.

After applying methodological adjustments based on lessons from experiments 2 and 3, experiment 4 again showed evidence that participants receiving feedback were thinking about their behaviors in general and their use of language in particular. It also tested the hypothesis that providing linguistic feedback (agreement word percentage) together with behavioral standards (more agreement is better or less agreement is better) would guide participants to change their behavior in the direction of the given standard. The findings provided partial evidence for this hypothesis, such that participants receiving feedback increased their expression of agreements when directed to do so, but did not decrease it when directed the opposite. The design of the fish, moving up with the expression of more agreement, might have been in conflict with the directed goal to decrease agreement, according to an embodied cognition perspective (Lakoff & Johnson, 1981; Richardson et al., 2003). Further, the increase in agreement expression when instructed to might have been in an attempt to manipulate the visualization and game the system, and not necessarily to improve the teamwork process. Reading through the chat transcripts revealed that some participants used positive remarks excessively, regardless of discussion context, supporting this possibility.

The latter inference leads to the final hypothesis tested in experiment 4, that teams receiving feedback will exhibit improved teamwork behaviors. Using Bales & Cohen's three SYMLOG dimensions (1979), an improvement would be indicated by evenly distributed participation among team members; group-oriented and friendly behaviors; and a balance between task-oriented and socio-emotional expressive behaviors (Losada et al., 1990; McLeod et al., 1992). Of the three dimensions, the findings suggested that with the goal to express more agreements, individuals receiving feedback were rated as more task/socially-balanced. This is encouraging, suggesting a theoretical link between automated linguistic analysis of team conversations, feeding the features it generates back to team members, and as a result improving the team interaction process. Becoming more social in a highly task-oriented setting such as a laboratory experiment could have resulted from a more informal interaction style the instructions and visualization encouraged (Whittaker et al., 1994), or from increased awareness of the team interaction process and the socio-emotional behaviors that facilitate effective teamwork (McLeod & Kettner-Polley, 2004). On the negative side, it might be that excessive use of agreements was viewed by peers as gaming behaviors, not engaging in meaningful thought about their actions, and unnecessarily being lightweight, jokey, and social.

Discussion

The fundamental argument that motivated this research is that the design of groupware technologies and theories of teamwork are too often focused on helping teams get their tasks done effectively. Instead, I argue that we also need to consider ways for guiding teams toward developing better social behaviors toward an improved interaction process. The findings of this research suggest few themes for discussion, in light of theoretical, practical, and technology design aspects to achieve this goal.

Interpreting Social Behaviors and Language Use

In this research I deliberated on how effective teamwork is translated into social behaviors, which can then be represented by linguistic behavior, and how change in linguistic behavior translates back to social team interaction. The constant strive to create a connection between theoretical frameworks of teamwork, such as McGrath's (1964), Hackman's (1987), Gladstein's (1984), and Bales and Cohen's (1979), with communicative behaviors and linguistic markers revealed in team conversations, has been of primary interest in this research.

However, it is clear that the researcher is not the only one to consider this connection. As initially assumed, and indicated by the research findings, members of teams engage in reflection of what kinds of communicative and linguistic behaviors are acceptable and favorable in the setting in which they interact with each other. For this reason, one of the initial underlying principles of GroupMeter was to leave out of the design contextual factors and normative benchmarks as for how to respond to the feedback. Based on the perspective that ambiguity can give rise to the development of multiple interpretations beyond what the researcher meant (Gaver et al., 2003), the system was designed to provide objective signals for the emergence of subjective social experiences and interpretations, as has been recently done in the field of affective computing (Leahu, Schwenk, & Sengers, 2008). This principle also gives the team more control to appropriate the system for its own purposes given the situation in which they are operating (Poole & DeSanctis, 1989).

However, another design principle, based on social comparison theory (Festinger, 1954), was to provide public feedback at the individual level. With this kind of feedback, team members were expected to interpret the feedback they receive about their own behaviors in comparison to others. This principle worked together with the graphic design of the feedback visualizations, changing length of bars, size of fish, and location of fish. According to an embodied cognition view (Lakoff & Johnson, 1981), long bars, and fish that are large, central, and up, are interpreted as better than short bars and fish that are small, peripheral, and down. It is no surprise then, that participants expressed more agreement when seeing feedback about their percentage of agreement words visually designed this way.

That benchmarks for interpreting desirable behavior can be conveyed with visual design is compelling especially when users are provided with separate instructions as for how to act. If the directive guidance is congruent with what the design implies, for example, the explicit goal is to express many agreements, and the fish move up with higher percentage of agreement words, the effect might be intensified to the point that users exhibit gaming behaviors. On the other hand, the externally set goal might be in conflict with what the design suggests, for example, telling users to express less agreements and moving the fish up with the expression of more agreements. In this case, the findings of this research suggest that the ongoing display of the feedback takes over. In both case, users reflecting on their social and linguistic behaviors in the particular situation in which they are operating can question whether the suggested goals, provided externally or through the design, are appropriate.

Further, the question remains open as for the extent to which the system (broadly defined, including the visualization, the linguistic features being measured, and the instructions given) should fall between the two extremes of being ambiguous or prescriptive with respect to the norms of behavior: On the one hand, providing valence-free information leaves it for teams to construct their own interpretations given the context in which they operate and may initiate deeper reflection. On the other hand, offering explicit goals for linguistic behavior and unequivocal interpretation for how it translates into social interaction can guide task groups toward more effective behaviors.

Balancing Task/Social Attention

Another issue that is woven through this research is the level of attention team members should be paying to their social behaviors, including choosing their words during team conversations, as compared to focusing on performing the team task. The findings of this research suggest that automated linguistic feedback can result in shifting attention away from the task toward more social aspects of the team process.

This idea is not new, that team members carry out different types of acts or assume certain roles to support the completion of the task and the social maintenance of the team (Bales, 1950; Benne & Sheats, 1948). It has been argued that effectively functioning teams find the equilibrium between channeling their efforts toward socio-emotional and task-related needs of the team (Bales, 1953), with more weight given to task behaviors in task-oriented groups (Levi, 2001).

The findings are therefore encouraging, since it has long been known that laboratory experiments are characterized by overly task-oriented behaviors (O'Rourke, 1963), especially with ad-hoc groups who do not exhibit the same socio-emotional behaviors as do naturally occurring teams (Chang, Bordia, & Duck, 2003). Further, based on the equilibrium model of group development (Bales, 1953), it is assumed that directing efforts to socio-emotional needs is done at the expense of task-related needs (Chidambaram & Bostrom, 1997). The impression that an ad-hoc group in a lab setting needs to focus on accomplishing its task and avoid socio-emotional expressions could therefore be one reason that the feedback visualization, not being part of the task, would be considered as a possible distraction.

To this end, there could be two sources of task distraction. One source is perceptual, caused by the visual elements, their color, shape, and movement. This possibility was addressed by designing the feedback visualization as a peripheral display (Cadiz et al., 2002; Maglio & Campbell, 2000; Matthews, Dey, Mankoff, Carter, & Rattenbury, 2004; McCrickard et al., 2003), and by evaluating it in the context of the primary team activity (Shami et al., 2005). Indeed, the switch from unobtrusive bar graphs to playful fish design increased perceptual distraction, as expected. However, on a more significant level, a second source of task distraction could be the reflective process, including understanding what the information conveyed by the visualization is, what this information means for the team process, and whether to react by changing one’s word choice, and if so, how.

Further research is needed to distinguish between perceptual distraction and distraction caused by reflection on social behavior. This distinction can help researchers and designers think of ways to reduce the visual/perceptual distraction, which stands in the way of transparent interaction (Abowd, 1999), and embrace the process of reflection-in-action, leading to new knowledge and improved practices (Schön, 1983).

Improving Behavior or Gaming the System

By providing automated linguistic feedback to members interacting together, I hoped that improving the team interaction process would occur through three steps: first, the feedback would engage team members in reflection on their use of language. Second, they will apply changes in their communicative behaviors by choosing their words differently in the conversation. And third, seeing how the feedback changes dynamically given the changes they apply in their in word choice, will lead them to understand how their language use is related to an improved team interaction process, and by this they will acquire improved teamwork skills.

However, one possible explanation for individuals changing their linguistic behavior in response to the visualization is that they were trying to game the system. This kind of behavior is indicated by systematic attempts to achieve high feedback scores by exploiting properties of the system instead of learning the knowledge or behaviors aimed for by the training system (Baker et al., 2004). That is, team members were excessively typing certain words to manipulate the visualization, to the extent that the conversation became artificial and off-task.

On the surface, such gaming behavior, especially when characterized by unnecessary off-task talk, could lead to ineffective team interaction process and result in poor task performance. As such, progressing through the three steps of improving teamwork behaviors through linguistic feedback would discontinue after the second step – changing one's language use. But perhaps an exaggerated behavior, deviating from how people normally talk, could in fact stimulate team members' reflection. The disproportionate use of language could serve as a magnifying glass, promoting deeper reflection and understanding of what this kind of behavior means for the team interaction process.

Further, a team operating in the long run is likely to develop social norms (Tuckman, 1965), and to adjust the norms at certain transition points in time when they stop and evaluate their progress (Gersick, 1988; Gersick, 1989). These norms represent member's approval and disapproval of what behaviors, including use of language, are appropriate, resulting in regulating members' behaviors (Hackman, 1992). A long-term study could reveal if and how real teams in real organizations develop over time certain norms with respect to language use in response to linguistic feedback visualizations, and if these norms carry on when the feedback is no longer available.

Contribution

The contribution of this research extends three levels: practical, theoretical, and design.

On the practical level, this work shows a novel technique for training people to develop their teamwork skills. Other examples of teamwork skills training techniques include observing examples of successful team interactions during a learning phase (Rummel & Spada, 2005), role-playing in practice sessions (Burton et al., 1997), and peer evaluations during collaborative team activities (McKinney & Denton, 2006; Oakley et al., 2004). I see the technique developed in this research, in which team members receive automated feedback about their language use while they communicate with each other, as complementary to the ones previously suggested. Practitioners in education and business can benefit from applying this technique to guide individuals to be sensitive to their use of language in team conversations.

On the theoretical level, my research demonstrates a three-way relationship between teamwork, language, and feedback. It ties together theories from multiple domains, including theories of effective teamwork (e.g., McGrath, 1964) and interpersonal behaviors (e.g., Bales & Cohen, 1979), self-regulation (e.g., Locke & Latham, 1990) and feedback (e.g., Kluger & DeNisi, 1996), and communication and psychological perspectives on language use in social settings (e.g., Pennebaker, 2002). My research supports this three-way relationship using empirical findings demonstrating how teamwork behaviors are linked to indicators of language use, and under what circumstances dynamic feedback based on linguistic indicators can stimulate changes in language use and teamwork behaviors.

On a design level, my work adds to the accumulating knowledge in human-computer interaction and computer-supported cooperative work about designing groupware technologies that, while keeping the team activity in the center, illuminate peripheral awareness information about social interaction. To achieve this, I combined principles from human-computer interaction (Erickson et al., 1999) and visualization (e.g., Tufte, 1983), human factors models of cognitive processing (Wickens & Hollands, 2000), social psychology theories of team interaction (Bales, 1953), and computer-mediated communication (Walther, 1996). The result was the GroupMeter research platform, a mediated team communication system augmented with peripheral visualizations reflecting social aspects of the team interaction process. The design of the GroupMeter system evolved over time with managing tradeoffs and addressing design lessons I learned from the experiments in which the system was tested. I hope that my experience in designing, re-designing, implementing and evaluating GroupMeter can help other designers think about factors that add up to successful design of technologies for teams.

General Limitations and Future Work

As with all research, there are limitations to the methodology, the interpretation of the results, and other issues that need to be considered when generalizing this research to broader issues of interest. While I have already pointed out to some of the limitations, the following is a discussion of some of the more general issues, and propositions for future work that can help overcome them.

This research used a natural language analysis technique based on Pennebaker's word-level LIWC (Pennebaker et al., 2001). While this technique has demonstrated convincing findings in analyzing people's writings (Newman, Groom, Handelman, & Pennebaker, 2008), spoken language (Pennebaker & Lay, 2002), and conversations (Hancock, Curry, Goorha, & Woodworth, 2008), it might still suffer from flaws rooted in the bag-of-words model. In this model, the text or speech being analyzed is represented as a collection of words, assuming independence between words and therefore disregarding grammar and order of words.

However, we cannot always assume independence between the words used in text, speech, or conversation (Nanas & Vavalis, 2008). For example, when providing feedback on agreement word percentage, the word "yes" is always counted into the agreement category since it generally represents agreement. However, when "yes" is followed by "but", as in the phrase "yes, but…", the word "yes" represents positive politeness in an expression of disagreement (Brown & Levinson, 1978; Holtgraves, 1997), and in this case it should not be counted into the agreement category.

This research demonstrated significant findings on the associations between certain linguistic features LIWC provides and teamwork behaviors, and their potential to serve as a source of feedback for individuals to reflect on and improve their teamwork behaviors. However, advanced language processing techniques can be useful to overcome the bag-of-words approach and to add context to the results of the automated language analysis. For example, compositional semantics has been shown to successfully extract positive or negative sense from a compound expression as a function of the meaning of its different words and the syntactic rules combining them (Choi & Cardie, 2008). Further language processing approaches can therefore yield information that not only is accurate in the eyes of the researcher, but also makes sense to the users of the system, who are constantly judging its trustworthiness, expertise, and credibility (Fogg & Tseng, 1999).

Another area for future research is to expand the experiments carried out in this research. One such extension is toward cross-cultural studies, already underway (Diamant, Lim, Echenique, Leshed, & Fussell, 2009), examining how members from different cultures working together in teams respond differently to linguistic feedback. For example, given Hofstede's dimensions for cultural differences (Hofstede, 1983), Chinese members are expected to pay more attention to the feedback than their American peers. Other studies could examine more into depth the impact of automated linguistic feedback in helping team members become aware of their socio-emotional behaviors, which can in turn facilitate effective teamwork (McLeod & Kettner-Polley, 2004).

Another effort worth undertaking is to keep developing and improving the GroupMeter system. In its current form, GroupMeter enables teams to communicate via web-based chat and receive automated feedback about discrete linguistic features. One way to expand the GroupMeter program is to work with other widely used communication media. For example, email is considered a critical tool for team collaboration (Tang, Lin, Pierce, Whittaker, & Drews, 2007), as is instant messaging (Isaacs, Walendowski, Whittaker, Schiano, & Kamm, 2002). These two mediums are both based on text communication, enabling linguistic analysis without the need for transcribing or speech recognition. Adding language-based feedback visualizations to email clients or instant messaging applications can help people become aware of and change their use of language in their everyday communication activities.

Finally, a general argument against controlled laboratory experiments is that they lack external validity (Shaw, 1981). One could argue that there is a limitation in generalizing the results of this research beyond the lab to natural teams operating in real organizations. The teams in the experiments were ad-hoc, without a shared history or anticipated future interaction, although I did make an effort to assign them with tasks that exemplify tasks that real teams do. The strength of my research is in demonstrating causality between the different variables of interest – feedback, language use, and teamwork behaviors. The next step would therefore be to take the theoretical, practical, and technology design concepts developed in this research outside of the lab. This could include long-term studies of teams in real organizations, educational or business, using different communication media such as instant messaging, video conferencing, email, and face-to-face, to accomplish a variety of tasks. Such studies could shed light on how members of teams develop effective teamwork skills, and the potential of technology, providing dynamic feedback based on automated language analysis, to support this process.

APPENDIX A

Experiment 1 Instruments

Instructions

Team Collaboration
This is a training session that will help you develop your collaboration skills. You will be collaborating with a team through AIM chat to complete a task. This is a decision making task that has been developed by NASA experts and used in many settings to demonstrate effective team work.

Your part is very important for the success of the team in the task. Participation in this exercise will assist you in building your collaborative skills. In addition, members of the best performing group, in terms of performance on the task and the collaborative behaviors it exhibits, will each receive a $40 gift certificate to the Cornell Store.

Peer Evaluation and Feedback
At two points during the session you will be prompted to evaluate your peers about their collaborative skills. We would like to examine whether peers are able to accurately evaluate their peers' collaborative skills, and use these evaluations as peer feedback. Therefore, you are encouraged to provide the most accurate assessments you can. You will be instructed to click the NEXT button when it is time to provide evaluations. Evaluations will be provided through a webpage. After all evaluations have been processed, you will receive feedback about your collaborative performance. The feedback will assist you in enhancing the group task as well as better develop your personal collaboration skills.

The first peer feedback will occur about halfway through the group task. The second peer feedback will occur at the end of the group task. At the end of the session you will be requested to complete a questionnaire through a webpage.

Collaborative Skills
The collaborative skills upon which you will be evaluated are:
- PARTICIPATION: The level of participation should be roughly equal among all group members. Dominance of few members and submissiveness of others impedes the contribution of all members to the group task.

- FRIENDLINESS: Group members should behave in a friendly manner to each other, instead of being unfriendly or individually-oriented.

- TASK/SOCIAL-ORIENTATION: There should be a balance between focusing on the task at hand and focusing on social and emotional expression, with a tendency to be a little more focused on the task.

Group Task

The Challenge - Lost on the Moon

Your spaceship has just crashed on the lighted surface of the moon. You were scheduled to rendezvous with a mother ship 200 miles away, also on the lighted side, but the rough landing has ruined your ship and destroyed all the equipment on board except for the 15 items listed in the next page.

Your crew's survival depends on reaching the mother ship, so you must choose the most critical items available for the 200-mile trip.

Your group task is to rank the 15 items in terms of their importance for your team's survival. Place a number 1 by the most important item, number 2 by the second most important, and so on, through number 15, the least important. Your objective is to work toward a team solution that all members of your group are willing to support. Record your team ranks in the column labeled "Rank." You will have 30 minutes to discuss the team ranking with your group through the chat.

You may assume:

- The number of people is the same as the number on your team.
- You are the actual people in the situation.
- All items are in good condition.

Items	**Rank**
Box of matches	
Food concentrate	
50 feet of nylon rope	
Parachute silk	
Solar-powered portable heating unit	
Two .45 caliber pistols	
One case of dehydrated milk	
Two 100-pound tanks of oxygen	
Stellar map (of the moon's constellations)	
Self-inflating life raft	
Magnetic compass	
5 gallons of water	
Signal flares	
First-aid kit containing injection needles	
Solar-powered FM receiver-transmitter	

Peer Evaluation Survey

Please rate your team members on the following scales. For each rating, provide an explanation of why you rated your peer this way. The explanation can be in the form of an example of this member's behavior in the group. You are encouraged to provide the most accurate assessments you can.

1. Participation

	Passive, introverted, contributes little						Active, dominant, contributes a lot	
Red	1	2	3	4	5	6	7	Explain:
Blue	1	2	3	4	5	6	7	Explain:
Yellow	1	2	3	4	5	6	7	Explain:
Green	1	2	3	4	5	6	7	Explain:

2. Friendliness

	Unfriendly, individually -oriented						Friendly, group-oriented	
Red	1	2	3	4	5	6	7	Explain:
Blue	1	2	3	4	5	6	7	Explain:
Yellow	1	2	3	4	5	6	7	Explain:
Green	1	2	3	4	5	6	7	Explain:

3. Task/Social-Orientation

	Shows feelings, focused on socializing						Task-oriented, analytical	
Red	1	2	3	4	5	6	7	Explain:
Blue	1	2	3	4	5	6	7	Explain:
Yellow	1	2	3	4	5	6	7	Explain:
Green	1	2	3	4	5	6	7	Explain:

Filler Survey

Please rate the user-friendliness of the chat program you are using to communicate with your peers on the following scales:

1. Overall reaction to the program							
1=terrible, 7=wonderful	1	2	3	4	5	6	7
1=difficult, 7=easy	1	2	3	4	5	6	7
1=frustrating, 7=satisfying	1	2	3	4	5	6	7
1=dull, 7=stimulating	1	2	3	4	5	6	7
1=rigid, 7=flexible	1	2	3	4	5	6	7

2. Learning to use the program	Totally Disagree						Totally Agree
Learning to operate the program is easy	1	2	3	4	5	6	7
It is easy to remember the use of commands	1	2	3	4	5	6	7
Performing tasks is straightforward	1	2	3	4	5	6	7

3. System capabilities	Totally Disagree						Totally Agree
Program response time to user actions is adequate	1	2	3	4	5	6	7
The program appears to be reliable	1	2	3	4	5	6	7
The program is designed for all levels of users	1	2	3	4	5	6	7

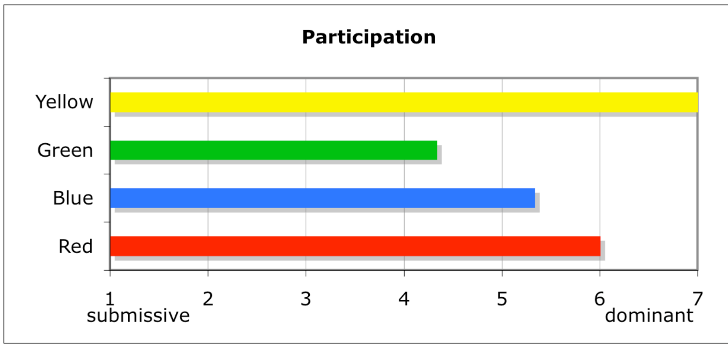
4. List the most positive aspects of the program:

5. List the most negative aspects of the program:

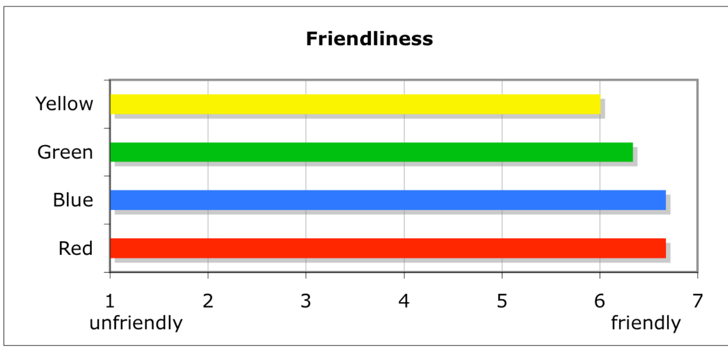
Peer Feedback Summary Form Example

Peer Feedback

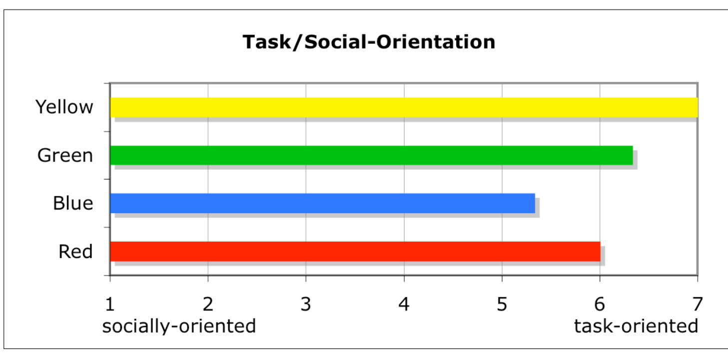
Please take a moment to consider the following feedback you received from your peers on the three collaborative behavior dimensions. Think about ways that you might want to change your behavior to move your group toward the effective group profile.



In effective groups there are no distinct dominant or submissive members. Participation should be roughly equal among all group members.



Group members should behave in a friendly manner to each other, instead of being unfriendly or individually-oriented.



There should be a balance between focusing on the task at hand and focusing on social and emotional expression, with a tendency to be a little more focused on the task.

APPENDIX B

Experiment 2 Instruments

Instructions

These instructions were read out by the experimenter.

Thanks for coming to this experiment. We will start by everyone telling us your name and major, and your best vacation ever:

You are going to carry out a task as a team. Imagine that you are a team of astronauts that are lost on the moon and need to make a 200-mile trip to your mother spaceship. You have 15 items that you will need to rank based on their importance to the success of reaching the spaceship and for your survival as a team.

You are competing with other teams in this experiment for the most accurate solution given the ranking provided by NASA's experts to this task. The best team will get $40 for each of its group members as a gift certificate to the Cornell Store.

One of our goals is to find out how people communicate through a chatroom to make group decisions. Therefore, you will be working on the task through a chatroom; please only use the chatroom, and do not speak to each other or look at each other as if you were in different places and not at the same room.

(Feedback groups only): During the chatroom conversation you will receive feedback about some features of your language use from the system. It will appear at the bottom of the page, below the chatroom.

After you complete the task I will ask you to come back to the central table for a short interview.

Interview questions

1. How well do you think your team did in the task?
2. To what extent were you conscious of the words you were using while talking with your teammates?
   What do you think are the reasons for it?
3. To what extent were you conscious of the words your teammates were using while talking?
   What do you think are the reasons for it?
4. To what extent do you think choosing your words changed over the course of the session?
   What are the reasons for it?
5. What do you think about GroupMeter?
   Did you like the user interface?
   What did you like about it?
   What didn't you like about it?
   What would you change about GroupMeter?
6. (Feedback groups only)
   What do you think about the feedback meters?
   Did you notice them? Why or why not?
   Was attending to them interruptive? Why or why not?
   Did you believe the feedback you received? Why?
   Did they make you think about the way you talked with your team?
   Did they make you think about how your teammates talked with you?
   What do you think the dimensions of feedback (word count, 1st person pronouns, emotion words) meant?

APPENDIX C

Experiment 3 Instruments

Instructions

These instructions were sent by the experimenter via GroupMeter chat. Each paragraph was sent separately as one chat entry.

Please tell everyone your name and major.

You are going to carry out three brainstorming and decision making tasks as a team. First, you will come up with ideas for businesses that can fill empty retail space in Collegetown. Second, you will think of ways that the local public transportation system can deal with current reports from passengers regarding the disruptive use of cell phones by other passengers. Third, you will help Billy and Katie think of the three most important things they should each bring on their "Survivor: Couples" excursion in chilly Alaska.

One of our goals is to find out how people communicate through a chatroom to make group decisions. Therefore, you will be working on the task through a chatroom.

For two of the three sections of the experiment, you will receive feedback about some features of your language use from the system. During one section, it will appear at the bottom of the page, below the chatroom. During another section, it will appear at the right of the chatroom. During a third section, you will not receive any feedback.

After you complete the three tasks, I will ask you to fill out a questionnaire that addresses each section of the experiment.

Feedback Explanations

Each of these explanations was sent by the experimenter via GroupMeter chat at the beginning of the session, according to the condition to which this session was assigned. Each paragraph was sent separately as one chat entry.

Fish visualization

You will receive feedback from the system regarding your behavior in the form of fish swimming in a tank to the right of the chatroom window. Each group member is represented by a fish. The system calculates your participation based on the amount of words you type into the chatroom.

The size of your fish is determined by your individual word count; so, the more you contribute, the larger your fish will be.

The system also measures how often you agree with your group members, counting agreement words that you use, such as "okay," "agree," "yes," etc. The closeness of the school of fish is based upon agreement between you and the other members of the group; so, the more an individual agrees, the closer their fish will be to the center.

Bars visualization

You will receive feedback from the system about your behavior in the form of horizontal bar graphs at the bottom of the chatroom window. Each group member is represented by a colored bar. The system calculates your participation based on the amount of words you type into the chatroom.

The size of your bar in the Word Count group is determined by your individual word count; so, the more you contribute, the longer your bar will be. The system also measures how often you agree with your group members, counting agreement words you are using, such as "okay," "agree," "yes," etc.

The size of your bar in the Agreements graph is based upon agreement between you and the other members of the group; so, the more an individual agrees, the longer their bar will be.

No visualization

You will not be experiencing any feedback during this task.

Group Tasks

Each task was sent by the experimenter via GroupMeter chat at the beginning of the session, after receiving the appropriate explanation about the feedback condition. Each paragraph was sent separately as one chat entry.

Task #1 (Empty retail space in Collegetown): A restaurant has been closed in Collegetown. What sort of businesses do you think could successfully fill this retail space? Come up with as many ideas as you can. At the end of the session, choose the group's top three ideas.

You have 10 minutes to complete this task.

Task #2 (Cell phones and buses): Recently, passengers of the local public transportation system have been reporting disturbances created on the buses by other passengers talking on their cell phones. The transportation company has been thinking of ways to alleviate this problem. Come up with as many ideas as you can to amend this issue. At the end of the session, choose the group's top three choices.

You have 10 minutes to complete this task.

Task #3 (Wilderness necessities): Your best friends, Billy and Katie, have just been invited to be contestants in the next season of "Survivor: Couples," which will take place in a cold climate of Alaska. They are only allowed to bring 3 things each, beyond the clothes on their backs. They are having a tough time deciding what to bring, and ask you for your help. What 3 things should Billy and Katie each bring with them? Come up with as many ideas as you can. At the end of the session, choose the group's top three ideas for what Billy and Katie should bring.

You have 10 minutes to complete this task.

End of session survey

Page 1: Identification and demographics

Participant Group #:______________
Your name:_____________________
Gender: Male/Female
Age:___________________________

Page 2: Self Evaluation

Please rate your performance on the task. You are encouraged to provide the most accurate assessments you can.
1=strongly disagree, 2=disagree, 3=neutral, 4=agree, 5=strongly agree

	No Visualization	Bars Visualization	Fish Visualization
I actively participated in the group task.	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
I contributed many of my own ideas to the discussion.	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
I agreed with many of the ideas that were generated by the other group members.	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
I remained focused on the task throughout the exercise.	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
I focused on the words I used rather than on the content of the task.	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5

Page 3: Awareness, Distraction and Performance

Please complete the following questions. You are encouraged to provide the most accurate assessments you can.
1=strongly disagree, 2=disagree, 3=neutral, 4=agree, 5=strongly agree

	No Visualization	Bars Visualization	Fish Visualization	Please explain why you gave these ratings
The feedback I received was interruptive.		1 2 3 4 5	1 2 3 4 5	
The visualization helped me be a better team member during the task.	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	
My behavior changed as a result of the visualization.	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	
The visualization improved my performance on the task.	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	
I was paying attention to the feedback.		1 2 3 4 5	1 2 3 4 5	
The visualization distracted me from the task.	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	
I was able to adequately focus on the group conversation.	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	
The visualization caused me to choose words differently as the task progressed.	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	

Page 4: Preferences

Please answer the following questions. You are encouraged to provide the most accurate assessments you can.

Did you find that the visualizations were helpful and useful to complete your task? Why or why not?

Did you find any one of the visualizations more helpful than the other? Why or why not?

What did you like about the fish visualization? Did you notice them?

What didn't you like about the fish visualization? What would you change?

What did you like about the bars visualization? Did you notice them?

What didn't you like about the bars visualization? What would you change?

Was attending to the fish visualizations interruptive? Why or why not?

Did you believe the feedback you received? Why?

What do you think the dimensions of the fish feedback mean: Size of the fish? Distance between the fish?

APPENDIX D

Experiment 4 Instruments

Instructions

These instructions were read out by the experimenter.

Thanks for coming to this experiment. My name is _____, and I will be the experimenter today. Today you will be working as a team to carry out two tasks.

Before we start, let's tell everybody your name and major to get to know each other a little bit.

(everybody says their name and major)

In each one of the tasks you will be working on you will need to come up with as many ideas as possible for some problem. You will have 10 minutes to brainstorm and produce the ideas, and then 5 more minutes to discuss the ideas and choose the best alternative.

One of our goals is to find out how people communicate through a chatroom to complete group tasks. For this reason, you will be working through a chatroom called GroupMeter; Please only use the chatroom, and do not speak to each other or look at each other. Behave as if you were in different places and not at the same room. I will later put up dividers between you to avoid eye contact and help simulate that you are in different places.

(F1 only) In addition to the chatroom, GroupMeter has a feature that measures how often you agree with your group members, by counting how frequently you type into the chatroom words that represent agreement, such as okay, agree, yes, and so forth. The system will provide you feedback about your level of agreement in the form of colored fish swimming near the chatroom window. Each group member is represented by a fish. The more you agree with your group members relative to the total amount of your talk in the chatroom, the higher up your fish will swim. The fish will start moving after about a minute into the chat session, and then will refresh approximately every one minute. This feedback is provided for you so that you can think about how you are communicating with your team members and make it more effective.

We will start with a short trial of GroupMeter. Please log into GroupMeter using your first name-underscore-last name as your username and the word "password" for the password. Then take a few minutes to get to know each other a bit more through the chatroom. You can talk about your hometown, your favorite vacation, what you did this summer, or anything else interesting you would like to share with the group.

(F1 only) Please notice that the fish started moving, and that they leave a bubble behind them. If you said a word that is considered an agreement word, such as "okay" or "yes", then your fish moved up higher. Otherwise, it is down at the bottom.

Before you start, I wanted to give you some information that can be useful for your teamwork.

(A0 only) Research has shown that to effectively work together it is better to express fewer agreements among team members. This leads to a more analytical discussion, thinking about different ways to solve the problem, and as a result working more efficiently toward a team decision. As such, it is recommended that you challenge your team members' arguments and not necessarily agree with them too often.

(A1 only) Research has shown that to effectively work together it is better to express more agreements among team members. This leads to a more positive environment, a more cohesive team process, and as a result working more efficiently toward a team decision. As such, it is recommended that you show more agreements and fewer objections to your team members' arguments.

I will now put up the dividers and give you the first task to read and complete through GroupMeter. When you are done, I will ask you to complete a short questionnaire and then give you the second task. When you are done with the second task I will ask you to fill out a second questionnaire.

Task #1

****BACKGROUND****
The City of Brooksfield

Brooksfield, a city of 150,000 inhabitants, is similar to many cities of its size in that its traditional industrial-based economy has been undergoing a transition to a more diversified economy. The leading growth sectors in the city have been professional and technical services, construction and retail. The current unemployment rate is 5 percent. The demographic trends show an overall increase in the average age, with a decline in the number of school-age children. This region does not have much tourism, as Brooksfield is not considered a "destination" spot. Its downtown nevertheless offers tourists a safe and relatively pleasant visiting experience. The public school system ranks slightly above average on most measures. The Community College is located within the city limits, and the State University is 90 miles away.

Brooksfield has just received a $10 milling Urban Renewal Grant, and your group has been brought together to brainstorm ideas about how the city can use these funds.

**** THE BRAINSTORMING QUESTION****

How many ideas can your group produce in ten minutes about how Brooksfield can use the grant?
After 10 minutes of brainstorming, you will have 5 minutes to discuss the ideas and choose the best alternative.

Task #2

BACKGROUND*
Cornell and the community

Cornell President David Skorton was quoted in the Cornell Daily Sun as saying: "…we need to be good neighbors…The University is many places... Each one of those places automatically becomes my community." In a community leaders brunch he also said: "Cornell will continue to look for, and consider opportunities to be supportive of long-term community needs."

****THE BRAINSTORMING QUESTION*****

How many ideas can your produce in ten minutes about what President Skorton could do to improve the relationship between Cornell and its local community in the Ithaca, NY area?
After 10 minutes of brainstorming, you will have 5 minutes to discuss the ideas and choose the best alternative.

Reflection on Feedback Survey (F1 participants)

Please take a look at the fish visualization with the bubble trails. Each fish represents one team member, and the bubbles represent the frequency of agreement words used by that member over time.

(A1 participants) Remember, research as shown that to work together effectively it is better to express more agreements among team members. This leads to a more positive environment, a more cohesive team process, and as a result working more efficiently toward a team decision.
(A0 participants) Remember, research as shown that to work together effectively it is better to express fewer agreements among team members. This leads to a more analytical discussion, thinking about different ways to solve the problem, and as a result working more efficiently toward a team decision.

1. Please choose the rating that best describes your communication with your team in task #1, Investments in Brooksfield:

1=strongly disagree, 2=moderately disagree, 3=disagree a little, 4=neither agree nor disagree, 5=agree a little, 6=moderately agree, 7=strongly agree

I frequently agreed with my group members	1 2 3 4 5 6 7
I agreed with my group members too little over the course of the entire session	1 2 3 4 5 6 7
I am satisfied with the changes I observed in my use of agreement words	1 2 3 4 5 6 7
I should have agreed more than what I did at the beginning of the session	1 2 3 4 5 6 7
I should have agreed more than what I did at the end of the session	1 2 3 4 5 6 7

2. Think about how the visualization representing the team conversation is linked to the team process. The words I chose to type in the conversation in task #1:

Contributed to a cohesive and positive team process	1 2 3 4 5 6 7
Encouraged an analytical and thoughtful discussion	1 2 3 4 5 6 7
Contributed to an effective team conversation	1 2 3 4 5 6 7

3. In the coming conversation for task #2:

I will try to agree with my group members often	1 2 3 4 5 6 7
I will try to agree with my group members infrequently	1 2 3 4 5 6 7

4. Given the fish visualization, reflect on your choice of words in task #1, and describe how choosing your words (specifically agreement words) during the conversation affected the team process.

5. How do you plan to change the words you use (in particular agreement words) in the coming task #2? How would you choose words to make the group conversation more effective?

Filler Survey (F0 participants)

Please rate the user-friendliness of the chat program you are using to communicate with your peers on the following scales:

1. Overall reaction to the program							
1=terrible, 7=wonderful	1	2	3	4	5	6	7
1=difficult, 7=easy	1	2	3	4	5	6	7
1=frustrating, 7=satisfying	1	2	3	4	5	6	7
1=dull, 7=stimulating	1	2	3	4	5	6	7
1=rigid, 7=flexible	1	2	3	4	5	6	7

2. Learning to use the program

	Totally Disagree						Totally Agree
Learning to operate the program is easy	1	2	3	4	5	6	7
It is easy to remember the use of commands	1	2	3	4	5	6	7
Performing tasks is straightforward	1	2	3	4	5	6	7

3. System capabilities

	Totally Disagree						Totally Agree
Program response time to user actions is adequate	1	2	3	4	5	6	7
The program appears to be reliable	1	2	3	4	5	6	7
The program is designed for all levels of users	1	2	3	4	5	6	7

4. List the most positive aspects of the program:

5. List the most negative aspects of the program:

End of session survey

Page 1. Demographics

1. Your gender: Male/Female
2. Your age: ____________
3. Major: _______________
4. Level of study: Undergrad/Grad/Other (please specify): ____________

Page 2. Awareness of Language Use

For each of the following statements, please select the rating that describes you best:

1=strongly disagree, 2=moderately disagree, 3=disagree a little, 4=neither agree nor disagree, 5=agree a little, 6=moderately agree, 7=strongly agree

I was looking at the visual feedback display during the conversation. *	1 2 3 4 5 6 7
Thinking about the words I used in the chatroom distracted me from focusing on the task.	1 2 3 4 5 6 7
During the conversation I tried to change the words I was using when communicating with my team members.	1 2 3 4 5 6 7
During the conversation I tried to change how often I agreed with my team members.	1 2 3 4 5 6 7
The feedback visualization had an effect on my choice of words in the chatroom. *	1 2 3 4 5 6 7
I spent time thinking about how often I agree with my team members.	1 2 3 4 5 6 7
Some features in GroupMeter distracted me from focusing on the task.	1 2 3 4 5 6 7
The GroupMeter system helped me think about how often I agreed with my team members.	1 2 3 4 5 6 7
The feedback I received accurately showed how much I agreed with my teammates. *	1 2 3 4 5 6 7
I spent time thinking about the words I was using while communicating with my team members.	1 2 3 4 5 6 7
The feedback display distracted me from focusing on the task. *	1 2 3 4 5 6 7
The GroupMeter system helped me adjust how often I agree with my team members during the conversation.	1 2 3 4 5 6 7
The GroupMeter system helped me think about how I talk with my teammates.	1 2 3 4 5 6 7
I remained focused on the task throughout the conversation.	1 2 3 4 5 6 7
The GroupMeter system helped me adjust the words I was using during the conversation.	1 2 3 4 5 6 7
The feedback I received made me think about the words I used during the conversation. *	1 2 3 4 5 6 7

*F1 participants only saw this question

Page 3. Peer Assessment

Please rate your team members on the following scales, including yourself. For each rating, provide an explanation of why you rated your peer this way. The explanation can be in the form of an example of this member's behavior in the group. You are encouraged to provide the most accurate assessments you can.

1. Participation

	Passive, introverted, contributes little						Active, dominant, contributes a lot
Member 1	1	2	3	4	5	6	7
Member 2	1	2	3	4	5	6	7
Member 3	1	2	3	4	5	6	7

Please explain why you gave these ratings:

2. Friendliness

	Unfriendly, individually -oriented						Friendly, group-oriented
Member 1	1	2	3	4	5	6	7
Member 2	1	2	3	4	5	6	7
Member 3	1	2	3	4	5	6	7

Please explain why you gave these ratings:

3. Task/Social-Orientation

	Shows feelings, focused on socializing						Task-oriented, analytical
Member 1	1	2	3	4	5	6	7
Member 2	1	2	3	4	5	6	7
Member 3	1	2	3	4	5	6	7

Please explain why you gave these ratings:

Page 4. Group Norms

For each of the following items, please choose the rating that best describes your group.
1=strongly disagree, 2=moderately disagree, 3=disagree a little, 4=neutral, 5=agree a little, 6=moderately agree, 7=strongly agree

In this group you should contribute to the group's goals	1 2 3 4 5 6 7
In this group you should think critically	1 2 3 4 5 6 7
In this group you should conform to the others	1 2 3 4 5 6 7
In this group you are expected to make an independent contribution	1 2 3 4 5 6 7
In this group you ought to align yourself with the opinions of other members	1 2 3 4 5 6 7
In this group you ought to act independently	1 2 3 4 5 6 7

Page 5. Self-Reflection and Insight

For each of the following statements, please select the choice that best describes your thoughts and behaviors during the session.
1=strongly disagree, 2=moderately disagree, 3=disagree a little, 4=agree a little, 5=moderately agree, 6=strongly agree

I was not really interested in analyzing my behavior	1 2 3 4 5 6
I was often confused about the way that I really felt about things	1 2 3 4 5 6
I usually had a very clear idea about why I've behaved in a certain way	1 2 3 4 5 6
I frequently examined my feelings	1 2 3 4 5 6
I often thought about the way I felt about things	1 2 3 4 5 6
I frequently took time to reflect on my thoughts	1 2 3 4 5 6
It was important to me to try to understand what my feelings meant	1 2 3 4 5 6
I didn't really think about why I behaved in the way that I did	1 2 3 4 5 6
Thinking about my thoughts made me more confused	1 2 3 4 5 6
It was important for me to evaluate the things that I did	1 2 3 4 5 6
My behavior often puzzled me	1 2 3 4 5 6
I usually knew why I felt the way I did	1 2 3 4 5 6
It was important to me to be able to understand how my thoughts arose	1 2 3 4 5 6
I rarely spent time in self-reflection	1 2 3 4 5 6
I was usually aware of my thoughts	1 2 3 4 5 6
I didn't often think about my thoughts	1 2 3 4 5 6
I was very interested in examining what I thought about	1 2 3 4 5 6
I had a definite need to understand the way that my mind worked	1 2 3 4 5 6
I was often aware that I was having a feeling, but I often didn't quite know what it was	1 2 3 4 5 6
I found it often difficult to make sense of the way I felt about things	1 2 3 4 5 6

Page 6. Evaluation of GroupMeter

Please answer the following questions:

1. What did you like about the GroupMeter system? How was it helpful for you?

2. What did you dislike about GroupMeter?

3. What would you change in GroupMeter?

4. Please let us know of any other thoughts you have about GroupMeter.

Page 7. Personality Traits

Here are a number of personality traits that may or may not apply to you. For each statement, please select the choice that best indicates the extent to which you agree or disagree with that statement. You should rate the extent to which the pair of traits applies to you, even if one characteristic applies more strongly than the other.

1=strongly disagree, 2=moderately disagree, 3=disagree a little, 4=neither agree nor disagree, 5=agree a little, 6=moderately agree, 7=strongly agree

Extroverted, enthusiastic	1 2 3 4 5 6 7
Critical, quarrelsome	1 2 3 4 5 6 7
Dependable, self-disciplined	1 2 3 4 5 6 7
Anxious, easily upset	1 2 3 4 5 6 7
Open to new experiences, complex	1 2 3 4 5 6 7
Reserved, quiet	1 2 3 4 5 6 7
Sympathetic, warm	1 2 3 4 5 6 7
Disorganized, careless	1 2 3 4 5 6 7
Calm, emotionally stable	1 2 3 4 5 6 7
Conventional, uncreative	1 2 3 4 5 6 7

REFERENCES

Abowd, G. D. (1999). Software engineering issues for ubiquitous computing. *ICSE '99: Proceedings of the 21st International Conference on Software Engineering,* Los Angeles, CA. 75-84.

Alexanderson, P. (2004). Peripheral awareness and smooth notification: The use of natural sounds in process control work. *NordiCHI '04: Proceedings of the Third Nordic Conference on Human-Computer Interaction,* Tampere, Finland. 281-284.

Allport, A., D., Antonis, B., & Reynolds, P. (1972). On the division of attention: A disproof of the single channel hypothesis. *Quarterly Journal of Experimental Psychology, 24*, 225-235.

Avrahami, D., & Hudson, S. E. (2006). Communication characteristics of instant messaging: Effects and predictions of interpersonal relationships. *CSCW '06: Proceedings of the 2006 20th Anniversary Conference on Computer Supported Cooperative Work,* Banff, Alberta, Canada. 505-514.

Baker, R. S., Corbett, A. T., Koedinger, K. R., & Wagner, A. Z. (2004). Off-task behavior in the cognitive tutor classroom: When students "game the system". *CHI '04: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems,* Vienna, Austria. 383-390.

Bales, R. F. (1950). *Interaction process analysis: A method for the study of small groups*. MA: Cambridge.

Bales, R. F. (1953). The equilibrium problem in small groups. In T. Parsons, R. F. Bales & E. A. Shils (Eds.), *Working papers in the theory of action* (pp. 111-161). Glencoe, IL: Free Press.

Bales, R. F., & Cohen, S. P. (1979). *SYMLOG: A system for the multiple level observation of groups*. New York: Free Press.

Bandura, A. (1991). Social cognitive theory of self-regulation. *Organizational Behavior and Human Decision Processes, 50*(2), 248-287.

Barkley, E. F., Cross, K. P., & Major, C. H. (2005). *Collaborative learning techniques*. San Francisco, CA: Jossey-Bass.

Benne, K. D., & Sheats, P. (1948). Functional roles of group members. *Journal of Social Issues, 4*(2), 19-49.

Bergstrom, T., & Karahalios, K. (2007a). Conversation clock: Visualizing audio patterns in co-located groups. *HICSS'07: 40th Annual Hawaii International Conference on System Sciences,* Waikoloa, Big Island, HI. 78-86.

Bergstrom, T., & Karahalios, K. (2007b). Seeing more: Visualizing audio cues. *INTERACT 2007: 11th IFIP TC13 International Conference on Human-Computer Interaction,* Rio de Janeiro, Brazil. 29-42.

Bezerianos, A., Dragicevic, P., & Balakrishnan, R. (2006). Mnemonic rendering: An image-based approach for exposing hidden changes in dynamic displays. *UIST '06: Proceedings of the 19th Annual ACM Symposium on User Interface Software and Technology,* Montreux, Switzerland. 159-168.

Bijker, W. E., Hughes, T. P., & Pinch, T. J. (Eds.). (1987). *The social construction of technological systems: New directions in the sociology and history of technology*. Cambridge, MA: MIT Press.

Birnholtz, J. P., Gutwin, C., & Hawkey, K. (2007). Privacy in the open: How attention mediates awareness and privacy in open-plan offices. *GROUP '07: Proceedings of the 2007 International ACM Conference on Supporting Group Work,* Sanibel Island, FL. 51-60.

Bodnar, A., Corbett, R., & Nekrasovski, D. (2004). AROMA: Ambient awareness through olfaction in a messaging application. *ICMI '04: Proceedings of the 6th International Conference on Multimodal Interfaces,* State College, PA. 183-190.

Bosworth, K. (1994). Developing collaborative skills in college students. *New Directions for Teaching and Learning, 59*, 25-31.

Boud, D., Keogh, R., & Walker, D. (Eds.). (1985). *Reflection: Turning experience into learning*. London: Kogan Page.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (Eds.). (1999). *How people learn: Brain, mind, experience, and school*. Washington, DC: National Academy Press.

Broadbent, D. E. (1971). *Decision and stress*. New York: Academic Press.

Brown, A. L. (1987). Metacognition, executive control, self-regulation, and other more mysterious mechanisms. In F. E. Weinert, & R. H. Kluwe (Eds.), *Metacognition, motivation, and understanding* (pp. 65-116). Hillsdale, NJ: Lawrence Erlbaum.

Brown, P., & Levinson, S. C. (1978). *Politeness: Some universals in language usage*. Cambridge: Cambridge University Press.

Burke, M., Joyce, E., Kim, T., Anand, V., & Kraut, R. (2007). Introductions and requests: Rhetorical strategies that elicit response in online communities. *C&T '07: Third International Conference on Communities & Technologies 2007,* East Lansing, MI. 21-40.

Burke, M., & Kraut, R. (2008). Mind your Ps and Qs: The impact of politeness and rudeness in online communities. *CSCW '08: Proceedings of the ACM 2008 Conference on Computer Supported Cooperative Work,* San Diego, CA. 281-284.

Burton, M., Brna, P., & Treasure-Jones, T. (1997). Splitting the collaborative atom: How to support learning about collaboration. In B. du Boulay, & R. Mizoguchi (Eds.), *Artificial intelligence in education: Knowledge and media in learning systems* (pp. 135-142). Amsterdam: IOS Press.

Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research, 65*(3), 245-281.

Cadiz, J. J., Venolia, G. D., Jancke, G., & Gupta, A. (2001). *Sideshow: Providing peripheral awareness of important information.* Technical Report No. MSR-TR-2001-83. Redmond, WA: Microsoft Research.

Cadiz, J. J., Venolia, G., Jancke, G., & Gupta, A. (2002). Designing and deploying an information awareness interface. *CSCW '02: Proceedings of the 2002 ACM Conference on Computer Supported Cooperative Work,* New Orleans, LA. 314-323.

Cannon-Bowers, J. A., Tannenbaum, S. I., Salas, E., & Volpe, C. E. (1995). Defining competencies and establishing team training requirements. In R. A. Guzzo, & E. Salas (Eds.), *Team effectiveness and decision making in organizations* (pp. 333-380). San Francisco, CA: Jossey-Bass.

Carver, C. S., & Scheier, M. F. (1998). *On the self-regulation of behavior*. New York: Cambridge University Press.

Cassell, J. (2000). Embodied conversational interface agents. *Communications of the ACM, 43*(4), 70-78.

Cassell, J., & Tversky, D. (2005). The language of online intercultural community formation. *Journal of Computer-Mediated Communication, 10*(2).

Cegala, D. J. (1989). A study of selected linguistic components of involvement in interaction. *Western Journal of Speech Communication, 53*, 311-326.

Chang, A., Bordia, P., & Duck, J. (2003). Punctuated equilibrium and linear progression: Toward a new understanding of group development. *The Academy of Management Journal, 46*(1), 106-117.

Chartrand, T. L., & Bargh, J. A. (1996). Automatic activation of impression formation and memorization goals: Nonconscious goal priming reproduces effects of explicit task instructions. *Journal of Personality and Social Psychology, 71*(3), 464-478.

Chidambaram, L., & Bostrom, R. (1997). Group development (I): A review and synthesis of development models. *Group Decision and Negotiation, 6*(2), 159-187.

Chidambaram, L., & Tung, L. L. (2005). Is out of sight, out of mind? An empirical study of social loafing in technology-supported groups. *Information Systems Research, 16*(2), 149-168.

Chin, J.P., Diehl, V.A., & Norman, K.L. (1988). Development of an instrument measuring user satisfaction of the human-computer interface. *CHI '88: Proceedings of the SIGCHI conference on Human factors in computing systems,* Washington, DC, 213-218.

Cho, K., Chung, T. R., King, W. R., & Schunn, C. D. (2008). Peer-based computer-supported knowledge refinement: An empirical investigation. *Communications of the ACM, 51*(3), 83-88.

Cho, K., & Schunn, C. D. (2007). Scaffolded writing and rewriting in the discipline: A web-based reciprocal peer review system. *Computers & Education, 48*(3), 409-426.

Choi, Y., & Cardie, C. (2008). Learning with compositional semantics as structural inference for subsentential sentiment analysis. *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing,* Honolulu, HI. 793-801.

Chung, C. K., & Pennebaker, J. W. (2007). The psychological function of function words. In K. Fiedler (Ed.), *Social communication: Frontiers of social psychology* (pp. 343-359). New York: Psychology Press.

Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In L. B. Resnick, R. M. Levine & S. D. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 127-149). Washington, DC: APA.

Cockburn, A. (2001). *Agile software development*. Boston, MA: Addison-Wesley.

Consolvo, S., McDonald, D. W., & Landay, J. A. (2009). Theory-driven design strategies for technologies that support behavior change in everyday life. *CHI '09: Proceedings of the 27th International Conference on Human Factors in Computing Systems,* Boston, MA. 405-414.

Costa, P. T., & McCrae, R. R. (1992). *NEO personality inventory professional manual*. Odessa, FL: Psychological Assessment Resources.

Cramton, C. D. (2001). The mutual knowledge problem and its consequences for dispersed collaboration. *Organization Science, 12*(3), 346-371.

Cummings, J., & Kiesler, S. (2007). Coordination costs and project outcomes in multi-university collaborations. *Research Policy, 36*(10), 138-152.

Dahley, A., Wisneski, C., & Ishii, H. (1998). Water lamp and pinwheels: Ambient projection of digital information into architectural space. *CHI '98: CHI 98 Conference Summary on Human Factors in Computing Systems,* Los Angeles, CA. 269-270.

Davis, B. G. (1993). *Tools for teaching*. San Francisco, CA: Jossey-Bass.

Davis, D., & Brock, T. C. (1975). Use of first person pronouns as a function of increased objective self-awareness and performance feedback. *Journal of Experimental Social Psychology, 11*(4), 381-388.

De Dreu, C. K. W., & Weingart, L. R. (2003). Task versus relationship conflict, team performance, and team member satisfaction: A meta-analysis. *Journal of Applied Psychology, 88*(4), 741-749.

de Guzman, E. S., Yau, M., Gagliano, A., Park, A., & Dey, A. K. (2004). Exploring the design and use of peripheral displays of awareness information. *CHI '04: CHI '04 Extended Abstracts on Human Factors in Computing Systems,* Vienna, Austria. 1247-1250.

DeSanctis, G., & Gallupe, R. B. (1987). A foundation for the study of group decision support systems. *Management Science, 33*(5), 589-609.

DeSanctis, G., & Poole, M. S. (1994). Capturing the complexity in advanced technology use: Adaptive structuration theory. *Organization Science, 5*(2), 121-147.

Diamant, E. I., Lim, B. Y., Echenique, A., Leshed, G., & Fussell, S. R. (2009). Supporting intercultural collaboration with dynamic feedback systems: Preliminary evidence from a creative design task. *CHI EA '09: Proceedings of the 27th International Conference Extended Abstracts on Human Factors in Computing Systems,* Boston, MA. 3997-4002.

DiMicco, J. M., Hollenbach, K. J., Pandolfo, A., & Bender, W. (2007). The impact of increased awareness while face-to-face. *Human-Computer Interaction, 22*(1), 47-96.

DiMicco, J. M., Pandolfo, A., & Bender, W. (2004). Influencing group participation with a shared display. *CSCW '04: Proceedings of the 2004 ACM Conference on Computer Supported Cooperative Work,* Chicago, IL. 614-623.

Dochy, F., Segers, M., & Sluijsmans, D. (1999). The use of self-, peer and co-assessment in higher education: A review. *Studies in Higher Education, 24*(3), 331-350.

Dominick, P. G., Reilly, R. R., & McGourty, J. W. (1997). The effects of peer feedback on team member behavior. *Group & Organization Management, 22*(4), 508-520.

Donnellon, A. (1996). *Team talk: The power of language in team dynamics.* Boston, MA: Harvard Business School Press.

Dourish, P., & Bly, S. (1992). Portholes: Supporting awareness in a distributed work group. *CHI '92: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems,* Monterey, CA. 541-547.

Driskell, J. E., & Salas, E. (1992). Collective behavior and team performance. *Human Factors, 34*(3), 277-288.

Druskat, V. U., & Wolff, S. B. (1999). Effects of timing of developmental peer appraisals in self-managing work groups. *Journal of Applied Psychology, 84*, 58-74.

Dubé, L., & Paré, G. (2001). Global virtual teams. *Communications of the ACM, 44*(12), 71-73.

Duval, S., & Wicklund, R. A. (1972). *A theory of objective self-awareness*. New York: Academic Press.

Eide, E., Gish, H., Jeanrenaud, P., & Mielke, A. (1995). Understanding and improving speech recognition performance through the use of diagnostic tool. *ICASSP-95: International Conference on Acoustics, Speech, and Signal Processing.* Detroit, MI. 221-224.

Erickson, T., & Kellogg, W. A. (2000). Social translucence: An approach to designing systems that support social processes. *ACM Transactions of Computer Human Interaction, 7*(1), 59-83.

Erickson, T., Smith, D. N., Kellogg, W. A., Laff, M. R., Richards, J. T., & Bradner, E. (1999). Socially translucent systems: Social proxies, persistent conversation, and the design of 'babble'. *CHI '99: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Pittsburgh, PA. 72-79.

Falchikov, N. (1995). Peer feedback marking: Developing peer assessment. *Innovations in Education and Training International, 32*(2), 175-187.

Farrell, S. (2001). Social and informational proxies in a fishtank. *CHI '01: CHI '01 Extended Abstracts on Human Factors in Computing Systems,* Seattle, WA. 365-366.

Festinger, L. (1954). A theory of social comparison processes. *Human Relations, 7*(2), 117-140.

Fitts, P. M., & Michael, I. P. (1967). *Human performance*. Belmont, CA: Brooks/Cole.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist, 34*(10), 906-911.

Fogg, B. J. (2003). *Persuasive technology: Using computers to change what we think and do.* San Francisco, CA: Morgan Kaufmann.

Fogg, B. J., & Tseng, H. (1999). The elements of computer credibility. *CHI '99: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems,* Pittsburgh, PA. 80-87.

Foushee, H. C., & Helmreich, R. L. (1988). Group interaction and flight crew performance. In E. L. Wiener, & D. C. Nagel (Eds.), *Human factors in aviation* (pp. 189-227). San Diego, CA: Academic Press.

Freeman, M., & McKenzie, J. (2002). SPARK, a confidential web-based template for self and peer assessment of student teamwork: Benefits of evaluating across different subjects. *British Journal of Educational Technology, 33*(5), 551-569.

Futoran, G. C., Kelly, J. R., & McGrath, J. E. (1989). TEMPO: A time-based system for analysis of group interaction process. *Basic & Applied Social Psychology, 10*(3), 211-232.

Gaver, W. W. (2002). Provocative awareness. *Computer Supported Cooperative Work, 11*(3-4), 475-493.

Gaver, W. W., Beaver, J., & Benford, S. (2003). Ambiguity as a resource for design. *CHI '03: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems,* Ft. Lauderdale, FL. 233-240.

Geister, S., Konradt, U., & Hertel, G. (2006). Effects of process feedback on motivation, satisfaction, and performance in virtual teams. *Small Group Research, 37*(5), 459-489.

Gersick, C. J. G. (1988). Time and transition in work teams: Toward a new model of group development. *The Academy of Management Journal, 31*(1), 9-41.

Gersick, C. J. G. (1989). Marking time: Predictable transitions in task groups. *The Academy of Management Journal, 32*(2), 274-309.

Gersick, C. J. G. (1991). Revolutionary change theories: A multilevel exploration of the punctuated equilibrium paradigm. *Academy of Management Review, 16*(1), 10-36.

Gersick, C. J. G. (1994). Pacing strategic change: The case of a new venture. *The Academy of Management Journal, 37*(1), 9-45.

Gersick, C. J. G., & Hackman, J. R. (1990). Habitual routines in task-performing groups. *Organizational Behavior and Human Decision Processes, 47*(1), 65-97.

Gillies, R. M., & Ashman, A. F. (1996). Teaching collaborative skills to primary school children in classroom-based work groups. *Learning and Instruction, 6*(3), 187-200.

Gladstein, D. L. (1984). Groups in context: A model of task group effectiveness. *Administrative Science Quarterly, 29*(4), 499-517.

Goffman, E. (1959). *The presentation of self in everyday life*. Garden City, NY: Doubleday Anchor.

Goldfinch, J. (1994). Further developments in peer assessment of group projects. *Assesment & Evaluation in Higher Education, 19*, 29-35.

Gonzales, A. L., Hancock, J. T., & Pennebaker, J. W. (in press). Language indicators of social dynamics in small groups. *Communications Research,*

Gorman, J. C., Foltz, P. W., Kiekel, P. A., Martin, M. J., & Cooke, N. J. (2003). Evaluation of latent semantic analysis-based measures of team communications content. *Proceedings of the Human Factors and Ergonomics Society 47th Annual Meeting.* Santa Monica, CA. 424-428.

Gosling, S. D., Rentfrow, P. J., & Swann, W. B. (2003). A very brief measure of the big-five personality domains. *Journal of Research in Personality, 37*, 504-528.

Gott, S. P., & Lesgold, A. M. (2000). Competence in the workplace: How cognitive performance models and situated instruction can accelerate skill acquisition. In R. Glaser (Ed.), *Advances in instructional psychology: Vol. 5. Educational design and cognitive science* (pp. 239-327). Hillsdale, NJ: Lawrence Erlbaum.

Grant, A. M., Franklin, J., & Langford, P. (2002). The self-reflection and insight scale: A new measure of private self-consciousness. *Social Behavior and Personality, 30*(8), 821-836.

Greenberg, S. (1996). Peepholes: Low cost awareness of one's community. *CHI '96: Conference Companion on Human Factors in Computing Systems,* Vancouver, British Columbia, Canada. 206-207.

Grudin, J. (1994a). Computer-supported cooperative work: History and focus. *IEEE Computer, 27*(5), 19-26.

Grudin, J. (1994b). Groupware and social dynamics: Eight challenges for developers. *Communications of the ACM, 37*(1), 92-105.

Guzdial, M., & Turns, J. (2000). Effective discussion through a computer-mediated anchored forum. *Journal of the Learning Sciences, 9*(4), 437-469.

Hackman, J. R. (1992). Group influences on individuals in organizations. In M. D. Dunnette, & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* (pp. 199-267). Palo Alto, CA: Consulting Psychologists Press.

Hackman, J. R. (1987). The design of work teams. In J. Lorch (Ed.), *Handbook of organizational behavior* (pp. 315-342). Englewood Cliffs, NJ: Prentice-Hall.

Hackman, J. R., & Wageman, R. (2005). A theory of team coaching. *Academy of Management Review, 30*(2), 269-287.

Hackman, J. R., & Morris, C. G. (1983). Group tasks, group interaction, and group performance effectiveness. In H. H. Blumberg, A. P. Hare, V. Kent & M. Davies (Eds.), *Small groups and social interaction*. Chichester, UK: Wiley & Sons.

Hackman, J. R. (1990). *Groups that work (and those that don't): Creating conditions for effective teamwork*. San Francisco, CA: Jossey-Bass Inc., Publishers.

Hancock, J. T., Curry, L. E., Goorha, S., & Woodworth, M. (2008). On lying and being lied to: A linguistic analysis of deception in computer-mediated communication. *Discourse Processes, 45,* 1-23.

Hancock, J. T., Curry, L., Goorha, S., & Woodworth, M. (2005). Automated linguistic analysis of deceptive and truthful synchronous computer-mediated communication. *HICSS'05: 38th Annual Hawaii International Conference on System Sciences,* Waikoloa, Big Island, HI.

Harkins, S. G., & Jackson, J. J. (1985). The role of evaluation in eliminating social loafing. *Personality and Social Psychology Bulletin, 11*, 457-465.

Hayes, N., & Walsham, G. (2000). Competing interpretations of computer-supported cooperative work in organizational contexts. *Organization, 7*(1), 49-67.

Hill, W. C., & Hollan, J. D. (1994). History-enriched digital objects: Prototypes and policy issues. *The Information Society, 10*(2), 139-145.

Hofstede, G. (1983). Dimensions of national cultures in fifty countries and three regions. In J. B. Deregowski, S. Dziurawiec & R. C. Annis (Eds.), *Explications in cross-cultural psychology* (pp. 335-355). Lisse NL: Swets and Zeitlinger.

Holtgraves, T. (1997). Yes, but... *Journal of Language and Social Psychology, 16*(2), 222-239.

Hoover, J. D. (2005). *Effective small group and team communication* (2nd ed.). Belmont, CA: Thomson/Wadsworth.

Hsieh, G., Wood, K., & Sellen, A. (2006). Peripheral display of digital handwritten notes. *CHI '06: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems,* Montréal, Québec, Canada. 285-288.

Ilgen, D. R., Fisher, C. D., & Taylor, M. S. (1979). Consequences of individual feedback on behavior in organizations. *Journal of Applied Psychology, 64*(4), 349-371.

Ilies, R., & Judge, T. A. (2005). Goal regulation across time: The effects of feedback and affect. *Journal of Applied Psychology, 90*(3), 453-467.

Isaacs, E., Walendowski, A., Whittaker, S., Schiano, D. J., & Kamm, C. (2002). The character, functions, and styles of instant messaging in the workplace. *CSCW '02: Proceedings of the 2002 ACM Conference on Computer Supported Cooperative Work,* New Orleans, LA. 11-20.

Ishii, H., & Ullmer, B. (1997). Tangible bits: Towards seamless interfaces between people, bits and atoms. *CHI '97: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems,* Atlanta, GA. 234-241.

Ishii, H., Wisneski, C., Brave, S., Dahley, A., Gorbet, M., Ullmer, B., et al. (1998). ambientROOM: Integrating ambient media with architectural space. *CHI '98: CHI 98 Conference Summary on Human Factors in Computing Systems,* Los Angeles, CA. 173-174.

Jacobson, I. (2004). *Object-oriented software engineering: A use case driven approach*. Redwood City, CA: Addison-Wesley.

Janis, I. (1972). *Victims of groupthink*. Boston, MA: Houghton-Mifflin.

Janssen, J., Erkens, G., Kanselaar, G., & Jaspers, J. (2007). Visualization of participation: Does it contribute to successful computer-supported collaborative learning? *Computers & Education, 49*(4), 1037-1065.

Janssen, J., Erkens, G., & Kanselaar, G. (2007). Visualization of agreement and discussion processes during computer-supported collaborative learning. *Computers in Human Behavior, 23*(3), 1105-1125.

Jarvenpaa, S. L., & Leidner, D. E. (1998). Communication and trust in global virtual teams. *Journal of Computer-Mediated Communication, 3*(4).

Jarzabkowski, P., & Bone, Z. (1998). A 'how-to' guide and checklist for peer appraisal of teaching. *Innovations in Teaching and Training International, 35*, 177-182.

Jehn, K. A. (1995). A multimethod examination of the benefits and detriments of intragroup conflict. *Administrative Science Quarterly, 40*(2), 256-282.

Johnson, D. W., & Johnson, R. T. (1975). *Learning together and alone: Cooperation, competition, and individualization*. Englewood Cliffs, NJ: Prentice-Hall.

Joiner, R. (2004). Supporting collaboration in virtual learning environments. *CyberPsychology & Behavior, 7*(2), 197-200.

Joshi, M., & Rosé, C. P. (2007). Using transactivity in conversation for summarization of educational dialog. *SLaTE Workshop on Speech and Language Technology in Education,* Farmington, PA.

Jurafsky, D., Ranganath, R., & McFarland, D. A. (2009). Extracting social meaning: Identifying interactional style in spoken conversation. *North American Chapter of the Association for Computational Linguistics - Human Language Technologies (NAACL-HLT) 2009,* Boulder, CO.

Keaten, J. A., & Richardson, M. E. (1992). A field investigation of peer assessment as part of the student group grading process. *Western Speech Communication Association Convention,* Albuquerque, NM.

Kelly, L., & Duran, R. (1985). Interaction and performance in small groups: A descriptive report. *International Journal of Small Group Research, 1*, 182-192.

Keyton, J., & Wall, V. D., Jr. (1989). SYMLOG theory and method for measuring group and organizational communication. *Management Communication Quarterly, 2*(4), 544-567.

Kim, T., Chang, A., Holland, L., & Pentland, A. (2008). Meeting mediator: Enhancing group collaboration using sociometric feedback. *CSCW '08: Proceedings of the ACM 2008 Conference on Computer Supported Cooperative Work,* San Diego, CA. 457-466.

Klein, K. J., Ziegert, J. C., Knight, A. P., & Xiao, Y. (2006). Dynamic delegation: Shared, hierarchical, and deindividualized leadership in extreme action teams. *Administrative Science Quarterly, 51*(4), 590-621.

Kluger, A. N., & DeNisi, A. (1996). Effects of feedback intervention on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin, 119*(2), 254-284.

Koch, G. G. (1982). Intraclass correlation coefficient. In S. Kotz, & N. L. Johnson (Eds.), *Encyclopedia of statistical sciences* (pp. 213-217). New York: John Wiley & Sons.

Kraut, R. E. (2003). Applying social psychological theory to the problems of group work. In J. Carroll (Ed.), *HCI models, theories and frameworks: Toward A multidisciplinary science* (pp. 325-356). New York: Morgan Kaufmann.

Kreijns, K., Kirschner, P. A., & Jochems, W. (2003). Identifying the pitfalls for social interaction in computer-supported collaborative learning environments: A review of the research. *Computers in Human Behavior, 19*(3), 335-353.

Krippendorff, K. (1980). *Content analysis: An introduction to its methodology*. Beverly Hills, CA: Sage Publications.

Krippendorff, K. (1986). *A dictionary of cybernetics*. Norfolk, VA: The American Society for Cybernetics.

Kulyk, O., Wang, J., & Terken, J. (2005). Real-time feed-back on nonverbal behaviour to enhance social dynamics in small group meetings. *MLMI'05: Proceedings of the International Workshop on Machine Learning for Multimodal Interaction,* Edinburgh. 150-161.

Lakoff, G., & Johnson, M. (1981). *Metaphors we live by*. Chicago: University of Chicago Press.

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge: Cambridge University Press.

Laybourn, P., Goldfinch, J., Graham, J., MacLeod, L., & Stewart, S. (2001). Measuring changes in group-working skills in undergraduate students after employer involvement in group skill development. *Assessment & Evaluation in Higher Education, 26*(4), 367-380.

Leahu, L., Schwenk, S., & Sengers, P. (2008). Subjective objectivity: Negotiating emotional meaning. *DIS '08: Proceedings of the 7th ACM Conference on Designing Interactive Systems,* Cape Town, South Africa. 425-434.

Leary, M. R., & Kowalski, R. M. (1990). Impression management: A literature review and two-component model. *Psychological Bulletin, 107*(1), 34-47.

Lebie, L., Rhoades, J. A., & McGrath, J. E. (1995). Interaction process in computer-mediated and face-to-face groups. *Computer Supported Cooperative Work, 4*(2/3), 127-152.

Lejk, M., & Wyvill, M. (2001). Peer assessment of contributions to a group project: A comparison of holistic and category-based approaches. *Assessment & Evaluation in Higher Education, 26*(1), 61-72.

Leshed, G., Hancock, J. T., Cosley, D., McLeod, P. L., & Gay, G. (2007). Feedback for guiding reflection on teamwork practices. *GROUP '07: Proceedings of the 2007 International ACM Conference on Supporting Group Work,* Sanibel Island, FL. 217-220.

Leshed, G., Perez, D., Hancock, J. T., Cosley, D., Birnholtz, J., Lee, S., McLeod, P. L., & Gay, G. (2009). Visualizing real-time language-based feedback on teamwork behavior in computer-mediated groups. *CHI '09: Proceedings of the*

27th International Conference on Human Factors in Computing Systems, Boston, MA. 537-546.

Leshed, G., Schober, M., Travis, S., Hancock, J. T., Gay, G., & McLeod, P. L. (2006). GroupMeter: A collaborative design effort of CSCW software. *Paper Presented at the Design and CSCW Workshop at CSCW'06: Computer-Supported Cooperative Work,* Banff, Alberta, Canada.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

Levi, D. (2001). *Group dynamics for teams*. Thousand Oaks, CA: Sage Publications.

Lin, J. J., Mamykina, L., Lindtner, S., Delajoux, G., & Strub, H. B. (2006). Fish'n'Steps: Encouraging physical activity with an interactive computer game. *UbiComp 2006: 8th International Conference on Ubiquitous Computing,* Orange County, CA. 261-278.

Locke, E. A., & Latham, G. P. (1990). *A theory of goal setting and task performance*. Englewood Cliffs, NJ: Prentice Hall.

Locke, E. A., Shaw, K. N., Saari, L. M., & Latham, G. P. (1981). Goal setting and task performance: 1969-1980. *Psychological Bulletin, 90*(1), 125-152.

Losada, M., Sanchez, P., & Noble, E. E. (1990). Collaborative technology and group process feedback: Their impact on interactive sequences in meetings. *CSCW '90: Proceedings of the 1990 ACM Conference on Computer-Supported Cooperative Work,* Los Angeles, CA. 53-64.

Maglio, P. P., & Campbell, C. S. (2000). Tradeoffs in displaying peripheral information. *CHI '00: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems,* The Hague, The Netherlands. 241-248.

Malone, T. W., & Crowston, K. (1994). The interdisciplinary study of coordination. *ACM Computing Surveys, 26*, 87-119.

Martin, M. J., & Foltz, P. W. (2004). Automated team discourse annotation and performance prediction using LSA. *Proceedings of the Human Language*

Technology and North American Association for Computational Linguistics Conference (HLT/NAACL), Boston, MA. 97-100.

Matthews, T. (2007). *Designing and evaluating glanceable peripheral visualizations.* Unpublished PhD, EECS Department, University of California, Berkeley.

Matthews, T., Rattenbury, T., & Carter, S. (2007). Defining, designing, and evaluating peripheral displays: An analysis using activity theory. *Human-Computer Interaction Journal, 22*(1), 221-261.

Matthews, T., Dey, A. K., Mankoff, J., Carter, S., & Rattenbury, T. (2004). A toolkit for managing user attention in peripheral displays. *UIST '04: Proceedings of the 17th Annual ACM Symposium on User Interface Software and Technology.* Santa Fe, NM. 247-256.

McCarthy, J., Wright, P., Wallace, J., & Dearden, A. (2006). The experience of enchantment in human–computer interaction. *Personal and Ubiquitous Computing, 10*(6), 369-378.

McCrickard, D. S., Chewar, C. M., Somervell, J. P., & Ndiwalana, A. (2003). A model for notification systems evaluation - assessing user goals for multitasking activity. *ACM Transactions on Computer-Human Interaction, 10*(4), 312-338.

McGrath, J. E. (1984). *Groups: Interaction and performance*. Englewood Cliffs, NJ: Prentice-Hall.

McGrath, J. E., & Hollingshead, A. B. (1994). *Groups interacting with technology: Ideas, evidence, issues, and an agenda*. Thousand Oaks, CA: Sage Publications.

McGrath, J. E. (1964). *Social psychology: A brief introduction*. New York: Holt, Rinehart and Winston.

McIntyre, R. M., & Salas, E. (1995). Measuring and managing for team performance: Lessons from complex environments. In R. A. Guzzo, & E. Salas (Eds.), *Team effectiveness and decision making in organizations* (pp. 9-45). San Francisco, CA: Jossey-Bass.

McKinney, D., & Denton, L. F. (2006). Developing collaborative skills early in the CS curriculum in a laboratory environment. *SIGCSE '06: Proceedings of the 37th SIGCSE Technical Symposium on Computer Science Education,* Houston, TX. 138-142.

McLeod, P. L., Liker, J. K., & Lobel, S. A. (1992). Process feedback in task groups: An application of goal setting. *Journal of Applied Behavioral Science, 28*(1), 15-41.

McLeod, P. L., Baron, R. S., Marti, M. W., & Yoon, K. (1997). The eyes have it: Minority influence in face-to-face and computer-mediated group discussion. *Journal of Applied Psychology, 82*(5), 706-718.

McLeod, P. L., & Kettner-Polley, R. B. (2004). Contributions of psychodynamic theories to understanding small groups. *Small Group Research, 35*(3), 333-361.

McLeod, P. L., & Kravec, N. G. (2008). Reward distribution in anonymous computer-supported task groups: The effects of social comparison. *Paper Presented at the Annual Meeting of the International Communication Association,* Montreal, Quebec, Canada.

Mehl, M. R., & Pennebaker, J. W. (2003a). The social dynamics of a cultural upheaval: Social interactions surrounding September 11, 2001. *Psychological Science, 14*, 579-585.

Mehl, M. R., & Pennebaker, J. W. (2003b). The sounds of social life: A psychometric analysis of students' daily social environments and natural conversations. *Journal of Personality and Social Psychology, 84*(4), 857-870.

Mercer, N. (1996). The quality of talk in children's collaborative activity in the classroom. *Learning and Instruction, 6*(4), 359-375.

Montoya-Weiss, M. M., Massey, A. P., & Song, M. (2001). Getting it together: Temporal coordination and conflict management in global virtual teams. *The Academy of Management Journal, 44*(6), 1251-1262.

Morris, C. G. (1966). Task effects on group interaction. *Journal of Personality and Social Psychology, 4*(5), 545-554.

Mousavi, S. Y., Low, R., & Sweller, J. (1995). Reducing cognitive load by mixing auditory and visual presentation modes. *Journal of Educational Psychology, 87*(2), 319-334.

Nanas, N., & Vavalis, M. (2008). A "bag" or a "window" of words for information filtering? *SETN '08: Proceedings of the 5th Hellenic Conference on Artificial Intelligence,* Syros, Greece. 182-193.

Navon, D., & Gopher, D. (1979). On the economy of the human-processing system. *Psychological Review, 86*(3), 214-255.

Newman, M. L., Groom, C. J., Handelman, L. D., & Pennebaker, J. W. (2008). Gender differences in language use: An analysis of 14,000 text samples. *Discourse Processes, 45*(3), 211-236.

Niederhoffer, K. G., & Pennebaker, J. W. (2002). Linguistic style matching in social interaction. *Journal of Language and Social Psychology, 21*(4), 337-360.

Oakley, B., Felder, R. M., Brent, R., & Elhajj, I. (2004). Turning student groups into effective teams. *Journal of Student Centered Learning, 2*(1), 9-34.

Orlikowski, W. J. (1996). Learning from notes: Organizational issues in groupware implementation. In R. Kling (Ed.), *Computerization and controversy: Value conflicts and social choices* (2nd ed., pp. 173-189). San Diego, CA: Academic Press.

Orlikowski, W. J., & Gash, D. C. (1994). Technological frames: Making sense of information technology in organizations. *ACM Transactions on Information Systems, 12*(2), 174-207.

O'Rourke, J. F. (1963). Field and laboratory: The decision-making behavior of family groups in two experimental conditions. *Sociometry, 26*(4), 422-435.

Overbeeke, K., Djajadiningrat, T., Hummels, C., Wensveen, S., & Frens, J. (2003). Let's make things engaging. In M. A. Blythe, A. F. Monk, K. Overbeeke & P. C. Wright (Eds.), *Funology: From usability to enjoyment* (pp. 7-17). Norwell, MA: Kluwer Academic Publishers.

Paulus, T. M. (2005). Collaboration or cooperation? analyzing small group interactions in educational environments. In T. S. Roberts (Ed.), *Computer-supported collaborative learning in higher education* (pp. 100-124). Hershey PA: Idea Group Inc.

Pedersen, E. R., & Sokoler, T. (1997). AROMA: Abstract representation of presence supporting mutual awareness. *CHI '97: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems,* Atlanta, GA. 51-58.

Pennebaker, J. W., Francis, M. E., & Booth, R. J. (2001). *Linguistic inquiry and word count: LIWC2001*. Mahwah, NJ: Erlbaum.

Pennebaker, J. W., & King, L. A. (1999). Linguistic styles: Language use as an individual difference. *Journal of Personality and Social Psychology, 77*(6), 1296-1312.

Pennebaker, J. W., & Lay, T. C. (2002). Language use and personality during crises: Analyses of mayor Rudolph Giuliani's press conferences. *Journal of Research in Personality, 36*(3), 271-282.

Pennebaker, J. W. (2002). What our words can say about us: Toward a broader language psychology. *Psychological Science Agenda, 15*, 8-9.

Pennebaker, J. W., Mehl, M. R., & Niederhoffer, K. G. (2003). Psychological aspects of natural language use: Our words, our selves. *Annual Review of Psychology, 54*, 547-577.

Plaue, C., Miller, T., & Stasko, J. (2004). Is a picture worth a thousand words?: An evaluation of information awareness displays. *GI '04: Proceedings of the 2004 Conference on Graphics Interface,* London, Ontario, Canada. 117-126.

Poole, M. S. (1981). Decision development in small groups I: A comparison of two models. *Communication Monographs, 48*(1), 1-24.

Poole, M. S. (1983a). Decision development in small groups II: A study of multiple sequences in decision making. *Communication Monographs, 50*(3), 206-232.

Poole, M. S. (1983b). Decision development in small groups III: A multiple sequence model of group decision development. *Communication Monographs, 50*(4), 321-341.

Poole, M. S., & DeSanctis, G. (1989). Use of group decision support systems as an appropriation process. *HICSS'89: 22nd Annual Hawaii International Conference on System Sciences: Emerging Technologies and Applications Track,* Kailua-Kona, HI. 149-157.

Poole, M. S., & Roth, J. (1989a). Decision development in small groups IV: A typology of group decision paths. *Human Communication Research, 15*(3), 323-356.

Poole, M. S., & Roth, J. (1989b). Decision development in small groups V: Test of a contingency model. *Human Communication Research, 15*(4), 549-589.

Postmes, T., Spears, R., Sakhel, K., & de Groot, D. (2001). Social influence in computer-mediated communication: The effects of anonymity on group behavior. *Personality and Social Psychology Bulletin, 27*(10), 1243-1254.

Ranganathan, A., Campbell, R. H., Ravi, A., & Mahajan, A. (2002). ConChat: A context-aware chat program. *IEEE Pervasive Computing, 1*(3), 51-57.

Reeves, B., & Nass, C. (1996). *The media equation: How people treat computers, television, and new media like real people and places*. Stanford, CA: CSLI.

Reimann, P., & Zumbach, J. (2003). Supporting virtual learning teams with dynamic feedback. *International Conference on Computers in Education 2003. A Conference of the Asia-Pacific Chapter of the Association for the Advancement of Computing in Education (AACE),* Hong Kong.

Richardson, D. C., Spivey, M. J., Barsalou, L. W., & McRae, K. (2003). Spatial representations activated during real-time comprehension of verbs. *Cognitive Science: A Multidisciplinary Journal, 27*(5), 767-780.

Roberts, T. S. (Ed.). (2005). *Computer-supported collaborative learning in higher education*. Hershey, PA: Idea Group.

Rogers, R. D., & Monsell, S. (1995). Costs of a predictable switch between simple cognitive tasks. *Journal of Experimental Psychology: General, 124*(2), 207-231.

Rousseau, V., Aube, C., & Savoie, A. (2006). Teamwork behaviors: A review and an integration of frameworks. *Small Group Research, 37*(5), 540-570.

Rummel, N., & Spada, H. (2005). Learning to collaborate: An instructional approach to promoting collaborative problem solving in computer-mediated settings. *The Journal of the Learning Sciences, 14*(2), 201-241.

Saavedra, R., & Kwun, S. K. (1993). Peer evaluation in self-managing work groups. *Journal of Applied Psychology, 78*, 450-462.

Salas, E., Dickinson, T. L., Converse, S. A., & Tannenbaum, S. I. (1992). Toward and understanding of team performance and training. In R. W. Swezey, & E. Salas (Eds.), *Teams: Their training and performance* (pp. 3-29). Norwood, NJ: Ablex Publishing.

Schler, J., Koppel, M., Argamon, S., & Pennebaker, J. W. (2006). Effects of age and gender on blogging. *AAAI Spring Symposium on Computational Approaches for Analyzing Weblogs,* Stanford University, CA. 199-205.

Schmidt, F. L., & Hunter, J. E. (1996). Measurement error in psychological research: Lessons from 26 research scenarios. *Psychological Methods, 1*(2), 199-223.

Schön, D. (1983). *The reflective practitioner: How professionals think in action*. London, UK: Temple Smith.

Scissors, L. E., Gill, A. J., Geraghty, K., & Gergle, D. (2009). In CMC we trust: The role of similarity. *CHI '09: Proceedings of the 27th International Conference on Human Factors in Computing Systems,* Boston, MA. 527-536.

Scott, E., van der Merwe, N., & Smith, D. (2005). Peer assessment: A complementary instrument to recognize individual contributions in IS student group projects. *The Electronic Journal of Information Systems Evaluation, 8*(1), 61-70.

Sengers, P., & Gaver, W. W. (2006). Staying open to interpretation: Engaging multiple meanings in design and evaluation. *DIS '06: Proceedings of the 6th Conference on Designing Interactive Systems,* University Park, PA. 99-108.

Sexton, G. A. (1988). Cockpit-crew systems design and integration. In E. L. Wiener, & D. C. Nagel (Eds.), *Human factors in aviation* (pp. 495-526). San Diego, CA: Academic Press.

Shami, N. S., Leshed, G., & Klein, D. (2005). Context of use evaluation of peripheral displays (CUEPD). *Proceedings of INTERACT 2005,* Rome, Italy. 579-587.

Shaw, M. E. (1981). *Group dynamics: The psychology of small group behavior* (3rd ed.). New York: McGraw-Hill.

Short, J. A., Williams, E., & Christie, B. (1976). *The social psychology of telecommunications*. New York: John Wiley & Sons.

Sillars, A., Shellen, W., McIntosh, A., & Pomegranate, M. (1997). Relational characteristics of language: Elaboration and differentiation in marital conversations. *Western Journal of Communication, 61*(4), 403-422.

Slatcher, R. B., & Pennebaker, J. W. (2006). How do I love thee? Let me count the words: The social effects of expressive writing. *Psychological Science, 17*(8), 660-664.

Smith, H., Cooper, A., & Lancaster, L. (2002). Improving the quality of undergraduate peer assessment: A case for student and staff development. *Innovations in Education and Teaching International, 39*, 71-81.

Smith, M. A. (1999). Invisible crowds in cyberspace: Mapping the social structure of the usenet. In M. A. Smith, & P. Kollock (Eds.), *Communities in cyberspace* (pp. 195-218). New York: Routledge.

Soller, A., Martinez, A., Jermann, P., & Muehlenbrock, M. (2005). From mirroring to guiding: A review of state of the art technology for supporting collaborative learning. *International Journal of Artificial Intelligence in Education, 15*, 261-290.

Sommerville, I. (1992). *Software engineering* (4th ed.). Reading, MA: Addison-Wesley.

Springer, L., Stanne, M. E., & Donovan, S. S. (1999). Effects of small-group learning on undergraduates in science, mathematics, engineering, and technology: A meta-analysis. *Review of Educational Research, 69*(1), 21-51.

Stahl, G. (2006). *Group cognition: Computer support for building collaborative knowledge*. Cambridge, MA: MIT Press.

Stanton, N. A. (1994). Explorations into hypertext: Spatial metaphor considered harmful. *Innovations in Education and Teaching International, 31*(4), 276-294.

Stevens, M. J., & Campion, M. A. (1994). The knowledge, skill, and ability requirements for teamwork: Implications for human resource management. *Journal of Management, 20*(2), 503-530.

Sundström, P., Ståhl, A., & Höök, K. (2007). In situ informants exploring an emotional mobile messaging system in their everyday practice. *International Journal of Human-Computer Studies, 65*(4), 388-403.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*(2), 257-285.

Swezey, R., W., & Salas, E. (1992). Guidelines for use in team-training development. In R. W. Swezey, & E. Salas (Eds.), *Teams: Their training and performance* (pp. 219-245). Norwood, NJ: Ablex.

Tang, J. C., Lin, J., Pierce, J., Whittaker, S., & Drews, C. (2007). Recent shortcuts: Using recent interactions to support shared activities. *CHI '07: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems,* San Jose, CA. 1263-1272.

Tarnow, E. (2000). Self-destructive obedience in the airplane cockpit and the concept of obedience optimization. In T. Blass (Ed.), *Obedience to authority: Current perspectives on the Milgram paradigm* (pp. 111-124). Mahwah, NJ: Lawrence Erlbaum.

Tollmar, K., Sandor, O., & Schömer, A. (1996). Supporting social awareness @ work: Design and experience. *CSCW '96: Proceedings of the 1996 ACM Conference on Computer Supported Cooperative Work,* Boston, MA. 298-307.

Townsend, A. M., DeMarie, S. M., & Hendrickson, A. R. (1998). Virtual teams: Technology and the workplace of the future. *The Academy of Management Executive, 12*(3), 17-29.

Tuckman, B. (1965). Developmental sequence in small groups. *Psychological Bulletin, 63*(6), 384-399.

Tufte, E. R. (1983). *The visual display of quantitative information*. Cheshire, CT: Graphics Press.

Tufte, E. R. (1990). *Envisioning information*. Cheshire, CT: Graphics Press.

Turner, G. A., & Schober, M. F. (2007). Feedback on collaborative skills in remote studio design. *HICSS'07: 40th Hawaii International Conference on System Sciences,* Waikoloa, Big Island, HI.

Vertegaal, R., Slagter, R., van der Veer, G., & Nijholt, A. (2001). Eye gaze patterns in conversations: There is more to conversational agents than meets the eyes. *CHI '01: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems,* Seattle, WA. 301-308.

Viégas, F. B., & Donath, J. S. (1999). Chat circles. *CHI '99: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems,* Pittsburgh, PA. 9-16.

Viégas, F. B., & Golder, S. D., & Donath, J. (2006). Visualizing email content: Portraying relationships from conversational histories. *CHI '06: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems,* Montréal, Québec, Canada. 979-988.

Viégas, F. B., Wattenberg, M., & Dave, K. (2004). Studying cooperation and conflict between authors with history flow visualizations. *CHI '04: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems,* Vienna, Austria. 575-582.

Wainer, J., & Braga, D. P. (2001). Symgroup: Applying social agents in a group interaction system. *GROUP '01: Proceedings of the 2001 International ACM SIGGROUP Conference on Supporting Group Work,* Boulder, CO. 224-231.

Walther, J. B. (1996). Computer-mediated communication: Impersonal, interpersonal, and hyperpersonal interaction. *Communication Research, 23*(1), 3-43.

Wegerif, R., Mercer, N., & Dawes, L. (1999). From social interaction to individual reasoning: An empirical investigation of a possible socio-cultural model of cognitive development. *Learning and Instruction, 9*(6), 493-516.

Weiser, M., & Brown, J. S. (1996). Designing calm technology. *PowerGrid Journal, 1*.

Whittaker, S., Frohlich, D., & Daly-Jones, O. (1994). Informal workplace communication: What is it like and how might we support it? *CHI '94: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems,* Boston, MA. 131-137.

Wickens, C. D. (1991). Processing resources and attention. In D. L. Damos (Ed.), *Multiple-task performance* (pp. 3-34). London: Taylor & Francis.

Wickens, C. D., & Hollands, J. G. (2000). *Engineering psychology and human performance* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.

Wickens, C. D., & Liu, Y. (1988). Codes and modalities in multiple resources: A success and a qualification. *Human Factors, 30*(5), 599-616.

Wickham, K. R., & Walther, J. B. (2007). Perceived behaviors of emergent and assigned leaders in virtual groups. *International Journal of e-Collaboration, 3*(1), 1-17.

Wiener, E. L., & Nagel, D. C. (Eds.). (1988). *Human factors in aviation*. San Diego, CA: Academic Press.

Wisneski, C., Ishii, H., Dahley, A., Gorbet, M. G., Brave, S., Ullmer, B., et al. (1998). Ambient displays: Turning architectural space into an interface between people and digital information. *CoBuild '98: Proceedings of the First International*

Workshop on Cooperative Buildings, Integrating Information, Organization, and Architecture, Darmstadt, Germany. 22-32.

Zhang, P., Carey, J., Te'eni, D., & Tremaine, M. (2005). Integrating human-computer interaction development into the systems development life cycle: A methodology. *The Communications of the Association for Information Systems, 15*, 512-543.

Zhang, S. (1995). Reexamining the affective advantage of peer feedback in the ESL writing class. *Journal of Second Language Writing, 4*(3), 209-222.

Zumbach, J., Mühlenbrock, M., Jansen, M., Reimann, P., & Hoppe, H. U. (2002). Multi-dimensional tracking in virtual learning teams: An exploratory study. In G. Stahl (Ed.), *Computer support for collaborative learning: Foundations for a CSCL community* (pp. 650-651). Mahwah, NJ: Lawrence Erlbaum.