

---

# Visualizing Language Use in Team Conversations: Designing through Theory, Experiments, and Iterations

**Gilly Leshed**

Information Science  
Cornell University  
301 College Ave.  
Ithaca, NY 14850 USA  
gl87@cornell.edu

**Dan Cosley**

Information Science  
Cornell University  
301 College Ave.  
Ithaca, NY 14850 USA  
drc44@cornell.edu

**Jeff Hancock**

Information Science  
Cornell University  
301 College Ave.  
Ithaca, NY 14850 USA  
jth34@cornell.edu

**Geri Gay**

Information Science  
Cornell University  
301 College Ave.  
Ithaca, NY 14850 USA  
gkg1@cornell.edu

**Abstract**

Effective teamwork skills are important for team collaboration. One way to potentially help people develop these skills is to visualize elements of their language use during team conversations. There are several challenges in designing such visualizations, such as how to balance attention between the conversation and the visualization and how much normative guidance to provide about appropriate behaviors. We discuss the design space around these questions in the context of GroupMeter, a web-based chatroom augmented with visualizations of language use. We use prior theoretical and empirical work to generate and critique potential answers to these design questions, then describe how the interface evolved, and how our answers sometimes changed, over a series of prototypes we deployed in experimental studies. We conclude by summarizing the lessons from our experience that could be used by designers of collaborative feedback systems.

**Keywords**

Teamwork, visualization design, linguistic analysis, behavioral experiment, persuasive computing.

**ACM Classification Keywords**

H.5.3 Information Interfaces and Presentation: Group and Organization Interfaces – CSCW.

---

Copyright is held by the author/owner(s).

CHI 2009, April 4 – 9, 2009, Boston, MA, USA

ACM 978-1-60558-246-7/09/04.

## Introduction

Teams are sometimes ineffective not because they lack the right tools to coordinate, collaborate, or communicate, but because their members don't always have the teamwork skills necessary for effective collaboration. A key aspect of effective teamwork is therefore developing behaviors that support the team interaction process. Recent work shows that analysis of language use can reveal some teamwork-relevant behaviors. For example, using self-references ('I', 'me') is associated with involvement in the interaction [4], frequent agreement ('yes', 'ok') is associated with passivity [24], and the use of justification words (e.g., 'because') indicates task focus [37].

These findings suggest that knowing what language to use during a conversation can be an important skill for effective teamwork. Further, presenting visualizations of linguistic behavior to team members can potentially stimulate reflection and guide team members to adopt effective teamwork behaviors. With these goals in mind, we designed GroupMeter, a chat system that visualizes linguistic metrics generated from conversations, such as frequency of emotion words, level of agreement, and overall participation level.

Systems that present this kind of linguistic information pose a number of design challenges: (1) When, during the team interaction, should awareness information be presented? (2) How should attention to feedback and conversation be balanced? (3) Should group or individual level feedback be displayed? (4) What kinds of feedback should be computed and how should they be interpreted? and, (5) How much normative guidance should be provided in a given context?

In this paper, we explore the design space around these questions in the context of the GroupMeter system. We present our initial answers to these questions, driven by theoretical and empirical work in human-computer interaction, social psychology, and cognitive science. We then describe how GroupMeter's user interface—and our answers—evolved over a series of prototypes and experiments.

Unlike previous descriptions of GroupMeter [24] [25], which looked at specific versions of the system and focused on behavioral experiments, our goal here is to present a case study of how its design changed over time as the research evolved. As HCI scholars and designers, we all face difficult design decisions, trying to settle conflicting goals and balance design tradeoffs. Tools developed in research settings often must deal with extra constraints: they need to account for theory and prior work, and meet research goals while demonstrating ecological validity. We hope that our experience helps other designers building similar systems and facing similar challenges. We also hope that the more general story of how theory and practice shaped our designs will be a useful case study for the HCI design and research community as a whole.

## The GroupMeter System

To ground the discussion of GroupMeter's design goals and evolution, we start with a high-level description of the system's design and architecture. The basic design of GroupMeter consists of a web-based system in which groups communicate through chat to perform tasks while receiving feedback about aspects of their language use through a visualization that appears near the chat window. The visualization represents linguistic features that are automatically extracted from the chat

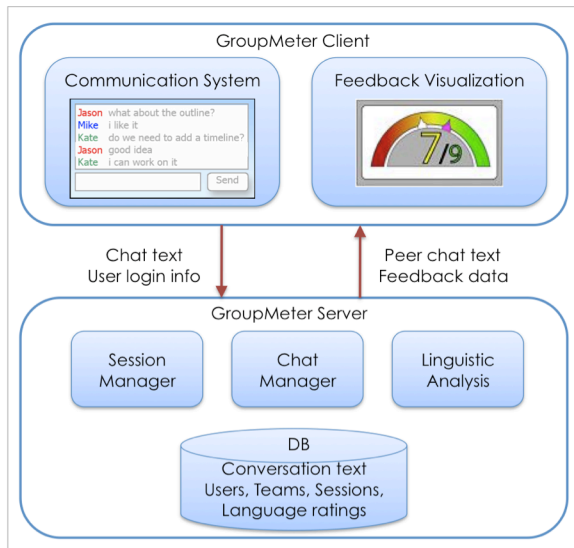


Figure 1. The GroupMeter software architecture.

and change dynamically as the conversation progresses. The linguistic features are generated using a technique based on Pennebaker's Linguistic Inquiry and Word Count (LIWC) [33], counting percentages of words in the chat text that fall into categories such as positive and negative emotion, self-references, and justification words. The choice of a chat communication medium over speech enables near real-time computation of these linguistic features.

GroupMeter is implemented using a web client-server architecture (Figure 1). The server manages sessions that specify the names of team members, linguistic metrics

to calculate, and the feedback interface to display. Team members log in to the website and communicate through the chat. As they talk, a chat manager on the server monitors the conversation and sends the chat text to the linguistic analysis module for processing. After analyzing the communication, the server sends the computed metrics to the front end, which presents them through the visualization specified by the session. The modular structure allows elements to be replaced easily: changing the linguistic computation leaves the visualization design unaffected, and vice versa.

### Theory-Informed Design Decisions

As stated in the introduction, tools designed to raise awareness of social behaviors by visualizing language use in a collaborative activity pose a number of ques-

tions. Here, we discuss the questions we faced during the design process and present possible solutions based on theories and prior empirical work from disciplines including human-computer interaction, social psychology, and cognitive science. We use a prototype developed early in the project (Figure 2) to help ground the discussion and illustrate some of the choices to be made.

#### 1. When, during the team interaction, should awareness information be presented?

According to Bales' Interaction Process Analysis [1], effective teamwork involves balancing task-related with socio-emotional behaviors. While it is easy to focus on behaviors aimed at completing the task, being aware of socio-emotional behaviors, such as language use and word choice, helps maintain the team's well-being, and is therefore important for effective teamwork [31].

A key question then, is when to present information during the team interaction process to help members become aware of socio-emotional elements of teamwork. One option, based on Gersick's punctuated equilibrium model of group development [13], is to provide information about such behaviors at transition points in the team's life span. At such points, team members will be most willing and able to reflect on their behavior and change them in subsequent team sessions. For instance, visualizing participation patterns between tasks was shown to lead to more equal distribution of participation in face-to-face settings [6] [32].

On the other hand, consistent with Kluger and DeNisi's Feedback Intervention Theory [24], an ongoing, dynamic visualization enables individuals to see more clearly how their behaviors are linked to the feedback

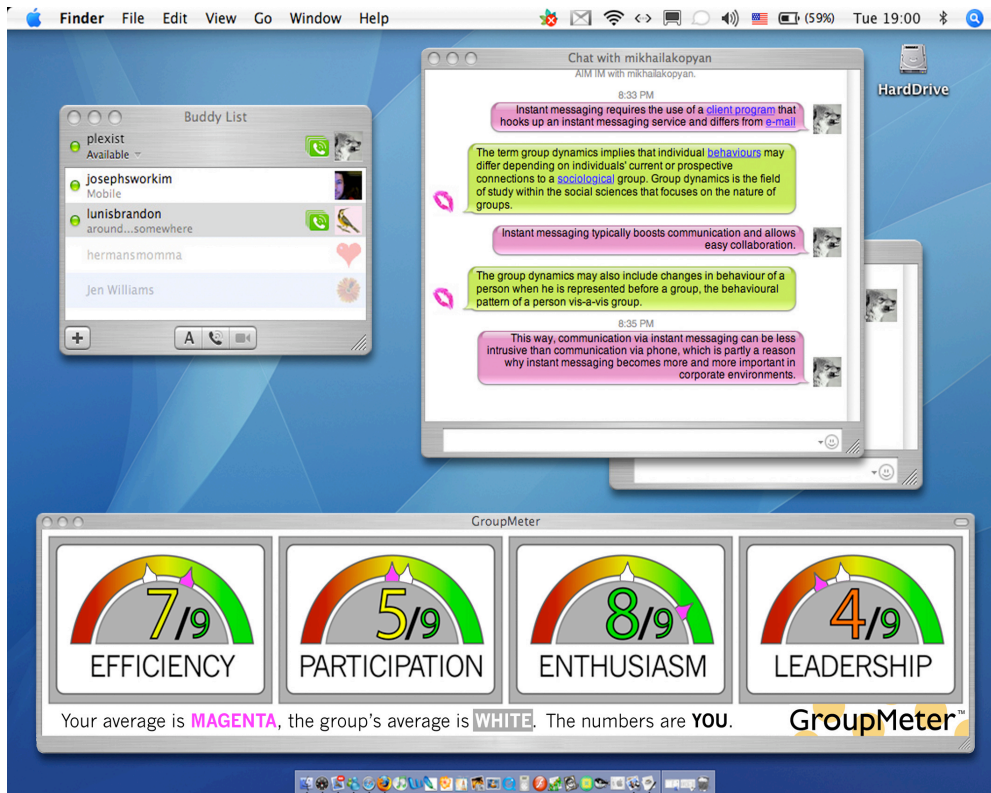


Figure 2. An early prototype of GroupMeter. It interfaces with an external, off-the-shelf chat software and presents feedback dials in a separate window.

they receive. This creates a *self-monitoring tool* for team members to monitor, review, and modify their behaviors as needed in real-time [11]. Other systems to support ongoing awareness of social behaviors have used a number of dynamic displays, projecting representations of behavior on individual's desktops [7] [15], a wall [6], a table [2], and recently on members' cell phone screens [20].

Because people's production of language in conversation is largely spontaneous and unconscious [26], we initially decided to present the linguistic information dynamically—making people aware of their behavior in the moment—rather than at intermissions between tasks. In Figure 2, as team members modify their behaviors intentionally or unintentionally, the indicators on the dials and the numbers change dynamically.

## 2. How should attention to feedback and conversation be balanced?

Dynamic visualizations, however, risk distracting team members from the task. Thus, we had to consider ways to stimulate awareness of the team's socio-emotional behaviors without damaging task performance. We therefore decided to display the information about language use at the periphery of the interface. Such *peripheral displays* promote awareness of background information, are not part of a primary activity, and do not overload users with too much cognitive effort [29]. Peripheral displays have been widely used to present awareness information in collaborative settings (see a review in [18]).

Presenting awareness information in a peripheral display requires individuals to divide their attention between primary and secondary information sources. This involves allocating cognitive resources to several information sources simultaneously, as well as shifting between them smoothly [39]. Visual design can support or hinder these shifts; for instance, in Figure 2, the feedback meters are physically separate from the chat window. This reflects an initial design goal to make GroupMeter work with many communication tools; however, the visual distance between the chat window

and visualization could make it harder to move back and forth between them.

Both the issues of timing and location call out a key design challenge around managing people's attention: while evaluating the various designs we created for the feedback visualization, we always sought to monitor the tension between awareness of teamwork behaviors and distraction from the team task and conversation.

### *3. Should group or individual level feedback be displayed?*

Another design factor that needs to be considered when presenting awareness information to a group is the level at which this information is aggregated and publicized: should behavioral information be kept private, at the individual level, compared to an aggregate, or made publicly available to the group?

The prototype in Figure 2 presents each individual's behavior privately (magenta markers on dials), along with an aggregate average for the group (white markers). Keeping individuals' information private can reduce potential stress caused by exposing behaviors publicly, thus avoiding embarrassment and discomfort [35]. Further, designs that lead to people comparing themselves against each other might encourage competition and negative interpersonal processes such as low trust, low coordination of effort, and attempts to mislead others [19]. For instance, people might act in ways that advance their own agenda rather than the team process (e.g., people might talk more to increase their participation meter whether or not their current role or task demands high levels of participation).

An alternative is to make everyone's individual-level information available to all team members. That is, every member sees the feedback information of every other team member, rather than just seeing their own information or seeing an aggregate of the team's behavior. Based on social comparison theory [10], presenting public information about individuals allows members to interpret the feedback about their own behaviors in comparison to the behaviors of others. Further, based on functional leadership theory [16], a strong team member or leader can observe others' behavior and intervene actively to motivate and direct others to change their behaviors [21].

Providing public individual-level feedback also supports social translucence [8]. 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. Publicly showing each individual's feedback can increase accountability and, as suggested by Festinger [10], affords explicit comparisons.

We decided that the advantages of publicly presenting individual information, including social comparison, leadership facilitation, and social translucence, outweighed the concerns users may have with having analyses of their behaviors made public. Still, the design needed to negotiate this tradeoff, and as we will see later, aspects of the design did affect whether people reflect on their language use and whether they changed their language for the good of the team or simply to manipulate the visualization.

*4. What kinds of feedback should be computed and how should they be interpreted?*

Another important design question was to select which data to present and how much to process the data. In principle, systems can compute an enormous amount of information about language use—LIWC, for instance, has over 70 linguistic categories [33]. To reduce potential distraction and the complexity of the interface, we chose to focus on linguistic metrics that correlate with the language used by people perceived to be good team members. In an earlier study, we identified several measurable aspects of language use associated with peer ratings of teamwork behaviors, including participation, friendliness, and task-focus [24]. For example, peer ratings correlated positively with overall contribution (measured by word count) and frequency of achievement-oriented terms, and negatively with frequencies of emotional terms and agreements.

Once we chose linguistic metrics such as word count, self-references, and agreements, which are related to peer-rated collaborative behaviors, the next question was how to process and present them. There is a tradeoff between presenting the behaviors in raw form versus attempting to map them onto higher-level collaborative concepts such as leadership and enthusiasm [12], as shown in Figure 2.

Mapping raw data onto higher-level concepts has some intuitive value; for instance, “high leadership” might be more meaningful to people than “uses many self-references.” However, these kinds of mappings put much of the burden of interpretation on the system designers, reducing the flexibility of the system. Less interpretation-laden representations allow users to develop multiple understandings of what the system is for

and how to use and experience it in different contexts [14]. Further, with high-level feedback it might be hard for team members to understand how their language use connects to the visual display and what they can do to change their behavior [22].

Thus, rather than mapping raw linguistic features such as self-references onto constructs such as leadership, we chose to simply help people become aware of their own and each other’s linguistic behaviors, and let them decide based on the task and the context of the conversation what those behaviors mean. As we will discuss later, users sometimes had trouble understanding what the linguistic metrics we chose to visualize meant, leading them to conflicting interpretations or to ignoring the metrics. Based on people’s reactions to early versions, over time we presented fewer metrics, focusing on metrics that people found more meaningful and that impacted teamwork more reliably.

*5. How much normative guidance should be provided in a given context?*

How much to interpret the data is related to another question concerning context: Should the system model contextual factors such as roles and tasks and provide normative goals for language use based on these contexts?

Contextual factors that feed into what normatively counts as good or poor behavior include factors at the individual level, such as members’ gender, personality, and skills; at the group level, such as its size, hierarchical structure, roles, and development phase; and exogenous factors, such the task assigned to the group and its characteristics, the reward structure, and cultural norms of using language. Based on goal-setting

theory [28], introducing such factors and the norms that build off of them allows team members to easily understand how they are expected to act upon seeing the visualization.

However, presenting normative goals may restrict the range of settings where the system can be applied. An alternative is to leave contextual factors out of the system, allowing for greater control by the group to appropriate the technology for its own purposes [34]. For instance, if a system visualizes the extent of agreement expression, team members can construct an interpretation of the feedback based on the task and the phase of the group development. They might see high agreement as beneficial when attempting to reach consensus, and as detrimental when critically discussing and negotiating solutions. Thus, we decided that GroupMeter should minimize assumptions about contextual factors and that its design should be mindful about what kinds of norms it suggests.

One important observation is that the graphical representation itself can imply norms. For instance, the green-red colors on the dials shown in Figure 2 imply certain norms to be achieved by team members—toward the green and away from the red. And even without the colors, a meter display might suggest that it should be filled up—that high values on the meters equate to “good” behavior. As we discuss later, despite our goal to keep the graphic representation open to interpretation of behavioral norms, some of our designs led to suggesting interpretations of normative behaviors that were not always effective in the contexts they were used.

## Principles Meet Practice: Co-evolution Through Use

We now discuss how both the interface design and our answers evolved as we deployed prototypes in experiments. Some changes were motivated by technical issues. Others were driven by results from user studies, controlled lab experiments in which visualizations were compared against each other or against chat use without any visualization. We present the discussion chronologically in an effort to clarify how our thinking evolved with insights we gained from each study.

### *Version 1: Unobtrusive bar-charts*

Figure 3 presents the first deployed version of GroupMeter. We abandoned the original design of a standalone feedback window, instead choosing to integrate the visualization with a custom-built chatroom. This allowed us to avoid the technical difficulties of interfacing with other systems and to create a more integrated, aesthetically unified experience. We chose to implement the client as a chat window embedded in a web browser. Together with a graphic designer, we created a new design for GroupMeter that included the chat window and bar meters providing feedback. In this design, 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.

The feedback visualization in version 1 consists of horizontal bar charts, one for each linguistic metric. The bars change their length based on each team member’s behavior on the linguistic metric. In this version, the interface presents feedback about team members’ overall contributions measured by their word count, their proportion of references to self (‘I’, ‘me’), and their use of emotion-laden words. In a previous study

we found these linguistic metrics to be associated with peer-rated teamwork behaviors such as participation, friendliness, and task-focus [24]. The linguistic analysis that feeds into the bars operates every one minute, based on the text entered by each member in the past five minutes, creating a moving average. This supported the goal of presenting feedback dynamically, but without constant change that might be distracting.

The design of the bar charts and their location below the chat window was intended to make the visualization subtler than the original meter dials. We hoped the visualization would be less distracting and thus more suitable as a peripheral display.

We chose to present each feedback dimension as an

aggregate stacked bar to accomplish our goal of presenting each individual's behavior while reducing the possibility that people would process the visualization in a competitive way. An alternate display using a clustered bar chart could potentially cause people to meticulously compare the length of their bars to others'—and attempt to increase their bars' length. Finally, showing how individuals' behaviors accumulate to an aggregate bar emphasizes the idea of being part of a group or team unit.

Finally, we did not provide either normative instructions or benchmarks in the interface about “appropriate” behaviors; based on the design arguments given earlier, we simply presented the data and let people interpret it as they wished.

We deployed version 1 in a lab experiment with 88 users, who worked together in 3- and 4-member teams to complete a decision making task. Half of the teams used GroupMeter with the bar-chart visualization. Because we were interested in seeing what interpretations people come up with for the feedback on their own, we did not tell people how the metrics were computed. The other half of the teams used a version of GroupMeter with the chatroom alone and no visualization.

We found that the general idea of visualizing language use stimulated reflection on teamwork behaviors, but that compared to the control group, people did not change their communication patterns in response to the feedback visualization. We interviewed our users, finding that the location and unobtrusive design of the bars might have led them to focus more on the task and not think much about how their word choice would affect the bar lengths.

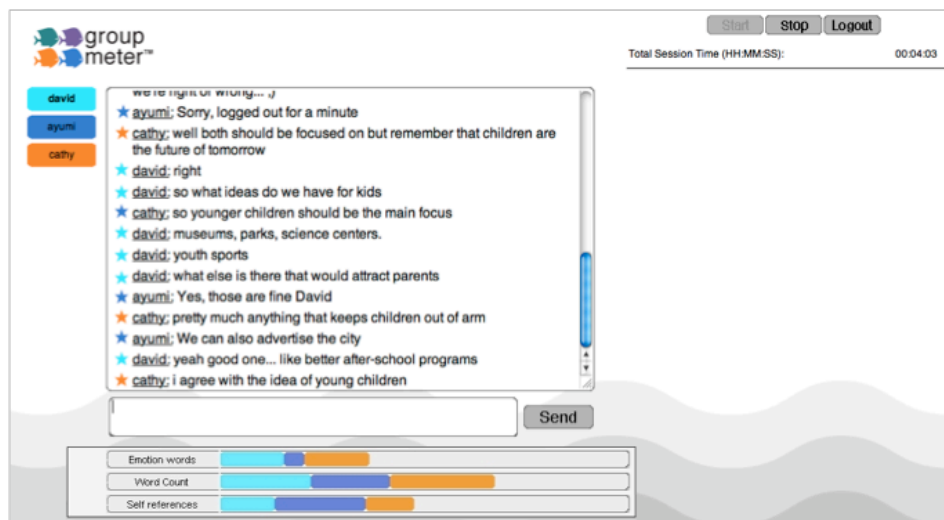


Figure 3. Version 1 of GroupMeter, using a custom-built web-based chat with feedback presented as stacked bar charts at the bottom.



Our choice of linguistic metrics based on [24] was not always congruent with users' perceptions of how collaborative behaviors are mapped into language use. The emotion words metric was particularly hard to interpret, since it was not obvious which words fall under this category, leading many users to ignore it. This suggests that explaining how a linguistic metric is computed is essential for making a link between behavior and its output measure. Finally, not providing norms led to conflicting interpretations of what counts as good or poor behavior. For instance, one participant understood the *self-references* bar such that high levels of it are undesirable:

*"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 said:

*"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."*

In a second experiment with 25 participants, teams completed a brainstorming task and saw the bar chart visualization presenting information about proportion of agreement words and word count (see [25]). In this experiment we told participants what language metrics will be visualized and how they are computed. Again, the bars made participants aware of their use of language as compared to not seeing any visualization, and at the same time were considered unobtrusive: *"the bars were just there"* and *"could be ignored if wanted."*

In this study, users changed their language use in response to the bars, expressing more agreement with their team members compared to not seeing any visualization. Despite the divergent thinking required from the brainstorming task and our explicit avoidance of normative guidelines, our decision to represent agreement as a bar might have implicitly guided this behavior: an embodied view of linguistic representations [23] assumes that people interpret "long" as better than "short", and agreeing more makes the bar longer.

#### *Version 2: Playfulness and the fish metaphor*

Our next design was aimed at both trying to reduce the implicit norms of bar charts and to see how a more visible and aesthetically pleasing visualization would affect people's use of GroupMeter. We therefore chose a more playful, abstract design using the metaphor of a school of fish. We found the school of fish inspiring because it symbolizes "togetherness", breaks from conventional forms of data presentation, and has a natural and serene connotation that "tells a story about the data" [36]. Fish have also been used to represent activity by other researchers seeking to persuade people to engage in physical activity [27] and in a shared display to represent workplace activity [9].

In this visualization, colored fish represent individual team members, matching members' colors in the chat window, as shown in Figure 4. The fish start in a circular formation, all at the same size and equidistant from the center. We placed the fish visualization to the right of the chat window so that, like the bars in version 1, it would be visible without the need to scroll the page.

The visualization is animated, dynamically changing the size of the fish to represent conversational activity

(measured by word count) and their distance from the center to represent agreement with the group (measured by proportion of agreement words). We chose the circular form to better convey a sense of unity and community, as well as providing a natural mapping for “teamness” as measured by agreement: the more team members agree, the closer to each other their fish appear in the visualization. Again, despite our intention not to design for a specific context or to encourage certain behaviors, this decision implied the norm that more agreements are desirable—since the unity of fish closer together could easily be interpreted as preferable to fish scattered all over.

This visualization was studied in comparison to the bar chart visualization [25]. The fish visualization seemed to encourage users to reflect on their communication behaviors, and was referred to by participants as “cute”

and “fun to watch”. Communication patterns also changed compared to when not seeing any visualization: teams seeing the fish expressed more agreement with each other, but at the cost of conducting less discussion. This suggests that our design decisions might have caused people to respond to the visualization in ways that sabotage effective teamwork.

However, to notice changes in the visualizations, users had to constantly monitor them, distracting them from the conversation. Self-reports of distraction and chat about the fish during the task itself also implied that we had gone too far toward engagement, disturbing people’s ability to balance task and process.

### *Version 3: Glanceability, history, and guidance*

The goal of version 3 was to refine the feedback visualization to address the problems we found in version 2. One primary interface change was to improve the visualization’s glanceability [30] by enriching the display with a history view [17]. Historical information can be important for understanding changes in social behaviors within a collaborative space [38]. This enables users to consider their behavior not in isolation, but in relation to the trend they see in the past [5]. Similar to [3], 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 with the trend of the feedback over the conversation. We expected that this would help people balance attention between the primary conversation about the task and the peripheral feedback display.

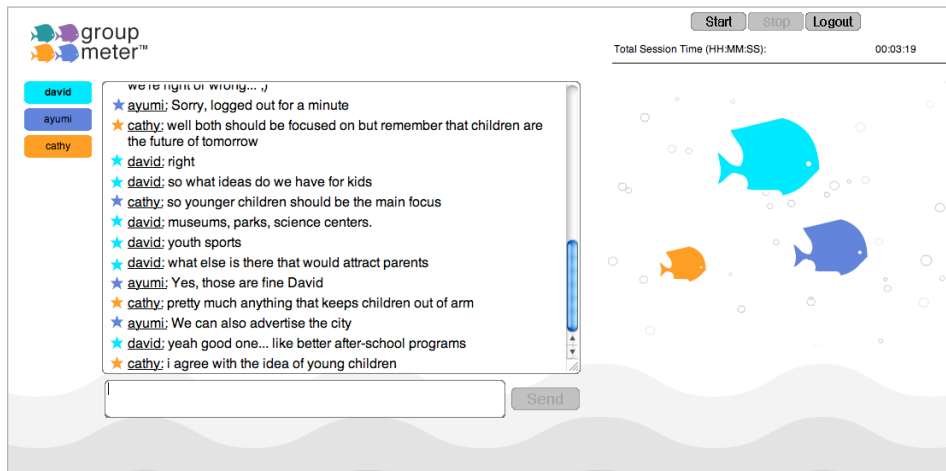


Figure 4. Version 2 of GroupMeter: feedback is visualized as fish in a circular position, changing size and distance from the center.

We represented the history as trails of bubbles, as shown in Figure 5. To simplify the history view and the number of feedback dimensions represented by the visualization, here the visualization presents only one linguistic metric: fish move higher on the vertical axis based on the proportion of agreement words people used. With this design we also wanted to avoid implying that a team that agrees with one another, represented by closer fish, is more desirable. Every minute, the system calculates the linguistic metric, moves the fish to its new position on the vertical axis, and leaves a bubble behind it in its previous location. As a result, the bubble trails appear on the horizontal axis, giving an impression of the fish swimming from left to right.

Because the display can only show ten minutes of feedback, a button at the bottom-left of the visualization opens a window with the full history view. This allowed us to experiment with another design consideration:

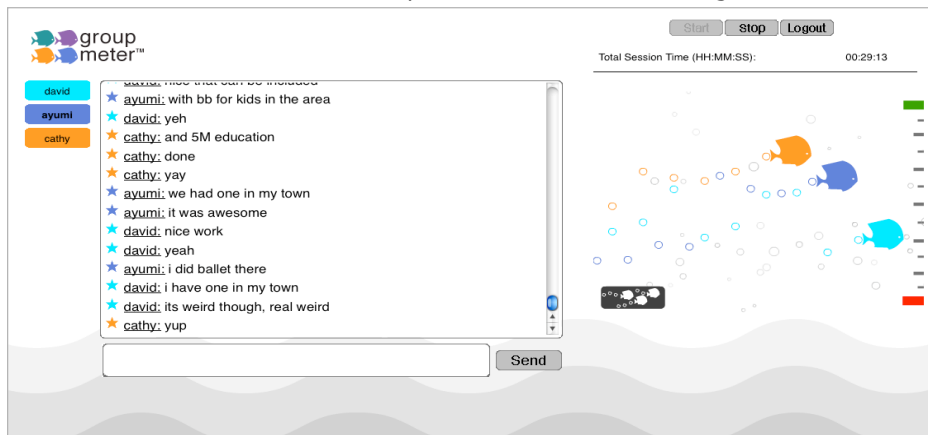


Figure 5. Version 3 of GroupMeter: fish move up and down in response to one feedback dimensions and leave bubble trails behind them as they update their position. A ruler on the right with a green-red marks serve as directive guidelines.

the timing of the feedback. In our next deployment of the GroupMeter interface, we assigned two tasks with a pause between them (similar to [8][35]), in which participants opened the full history view and were encouraged to reflect on their teamwork and linguistic behaviors. This procedure provided both real-time display of feedback and explicit periods for reflection.

We also made a major change in this version in our thinking about providing normative goals. Initially, we attempted to avoid designing for a single interpretation of behavior, with the goal of allowing teams to develop their own meanings of the feedback. However, without guidance, people using version 1 developed mixed interpretations of what levels of self-references are appropriate, making it more difficult to see whether GroupMeter was really making a difference. Further, in version 2 users tended to agree more with implicit guidance (e.g., bring the fish closer together)—going against our views of appropriate behavior of divergent thinking in a brainstorming task.

To explore whether providing normative guidance would drive behavior changes in a certain direction, we provided teams both with cues in the visualization and with explicit instructions on appropriate linguistic behavior before using the system in our next experiment. The instructions encouraged people in some groups to agree more, and in others to agree less. To reinforce the instructions, we added to version 3 a ruler at the right of the visualization (Figure 5). The ends of the ruler are red and green; the green marker can be placed either at the top or the bottom of the ruler.

This last version of GroupMeter was deployed in an experiment with 123 users. In a 2X2 design, 3-member

teams were first given instructions that either encouraged them to agree more or less with each other, and then completed two brainstorming tasks, either seeing the visualization or not. Between the tasks, teams receiving feedback reviewed the full history visualization and were prompted to reflect on it, while the other teams completed a filler survey.

Our results show that providing behavioral guidance stimulated change in language use in response to the visualization, although the changes were asymmetric. In particular, seeing the visualization induced users to be more agreeable when instructed to do so. However, when instructed to be more argumentative, users seeing the visualization did not express less agreement. This may again be because of the notion of embodied representations of language [23]: people tend to perceive up as more—and more as better. Because the fish move up in response to more agreements, asking people to agree less, and thus to move their fish down, works against a natural cognitive representation. Perhaps reversing the display, such that the fish move up with *less* agreement, would stimulate a more critical and less agreeable conversation pattern.

Further, our fears about presenting individuals' behavior leading to "gaming the system" came true. Analyzing the team conversations, we found that people sometimes talked not to further the team goals but simply to make their fish move up in the display:

*A: yes yes yes yes! hahaha sorry.*

*A: let's make sure we all use the word "yes" at least once in every comment*

*B: yes, let's do that*

This topic is a hard problem worthy of further research. In particular, researchers need to examine how to design feedback that explicitly links behavior and onscreen representation, but that resists gaming behaviors and guides toward the adoption of behaviors that benefit the team.

## Summary

To summarize our experience, and to call out aspects we hope will be useful for other designers, we present the questions we posed earlier and how our answers changed over the evolution of GroupMeter's design.

*1. When, during the team interaction, should awareness information be presented?* We started by visualizing feedback dynamically so team members could continually monitor it and connect changes in their behavior to changes in the display. Later, we added a history view that allowed for a fuller depiction of how behavior during the conversation unfolds, and introduced pauses between tasks in which teams get a chance to reflect more deeply on their teamwork behaviors. We sense that the combination of dynamic and punctuated feedback was especially useful in raising awareness of unconscious behaviors as word choice.

*2. How should attention to feedback and conversation be balanced?* We used peripheral, glanceable displays that support quick transitions between the conversation and the visualization. With our versions of GroupMeter we examined different levels of attention the visualization attracted, from bar charts that were unobtrusive but easy to ignore to fish that were fun and playful but distracting. Somewhere in between, we hope, other designers will find visualizations that are both stimulating and effective.

*3. Should group or individual level feedback be displayed?* Our versions of GroupMeter visualized all individuals' behavior publicly, facilitating social comparison and leveraging the idea of social translucence. Users reported using the feedback to think about their language use and how it affected the group. However, they also sometimes competed against or played with their team members, changing their language use not for the goal of adopting beneficial communication skills but rather to manipulate the display.

*4. What kinds of feedback should be computed and how should they be interpreted?* We chose to compute metrics of language use that correlate with peer ratings of effective teamwork behaviors. Keeping these metrics at the raw data level instead of mapping them onto high-level concepts enabled users to see a link between their language use and the visualization and to make their own interpretations. Over time we had to adjust our choice of metrics, and provided explicit explanations of the computation behind the visualization to help create the behavior-feedback link.

*5. How much normative guidance should be provided in a given context?* Our initial answer was to give only enough guidance for people to understand the visualization, allowing for flexible appropriation to many contexts. However, we found that this could lead to competing interpretations of normative behaviors and that our graphical designs sometimes encouraged behaviors ineffective in the context in which they were applied (e.g., seeking consensus in a brainstorming task). We later explored how to provide guidance that would drive behavior toward a certain direction, using visual cues and explicit instructions. Our limited success cautions

designers to be attentive to how design details, especially those not thought of, might influence behavior.

## Conclusions

In this paper, we make three main contributions. First, we lay out important questions in the design space for collaborative feedback systems: managing the task-social process balance with timing and positioning of the feedback, deciding whether to display information publicly or privately, choosing which data to display and how much to interpret it, and how much the system should try to account for the context of use and for normative guidelines. Although every system will face unique challenges in its particular context, the issues we call out are likely to apply across a broad range of systems for supporting teams. Designers will need to account for them.

Second, we call attention to studies and theories of cognitive and social behavior that seem relevant for designing collaborative feedback systems. We discuss how we used these theories, prior studies, and our intuition to select reasonable candidate answers to the questions we faced in building GroupMeter. Our design solutions are not the only possibilities. For instance, some groups or cultures might be sensitive to individual criticism and thus prefer computation of group-level feedback. Some circumstances may call for other kinds of feedback than those derived from language analysis. And, designers of tools that are specific to a given context may want to leverage that context. Still, presenting theoretical perspectives on potential solutions opens up the design space for educated exploration and consideration.

Third, we explore how our initial theory-driven answers fare in practice through a series of deployments and experiments using versions of the GroupMeter system. GroupMeter was designed and deployed as a research platform, but it also provides practical lessons for designers. By calling out the design questions, presenting theoretical perspectives on potential solutions and tradeoffs, and reviewing the lessons we learned, we hope other designers can make use of our analysis and experience to make good choices and systems.

## References

- [1] Bales, R.F. (1950). *Interaction process analysis: A method for the study of small groups*. MA: Cambridge.
- [2] Bergstrom, T., & Karahalios, K. (2007). Conversation clock: Visualizing audio patterns in co-located groups. *HICSS'07*, 78-86.
- [3] Bezerianos, A., Dragicevic, P., & Balakrishnan, R. (2006). Mnemonic rendering: An image-based approach for exposing hidden changes in dynamic displays. *Proceedings of UIST '06*, 159-168.
- [4] Cegala, D.J. (1989). A study of selected linguistic components of involvement in interaction. *Western Journal of Speech Communication*, 53, 311-326.
- [5] Consolvo, S., McDonald, D.W., & Landay, J.A. (2009). Theory-driven design strategies for technologies that support behavior change in everyday life. *Proceedings of CHI '09*, Boston, MA. 405-414.
- [6] DiMicco, J.M., Hollenbach, K.J., Pandolfo, A., & Bender, W. (2007). The impact of increased awareness while face-to-face. *HCI*, 22(1), 47-96.
- [7] Dourish, P., & Bly, S. (1992). Portholes: Supporting awareness in a distributed work group. *Proceedings of CHI '92*, Monterey, CA. 541-547.
- [8] Erickson, T., & Kellogg, W.A. (2000). Social trans-luence: An approach to designing systems that support social processes. *ACM ToCHI*, 7(1), 59-83.
- [9] Farrell, S. (2001). Social and informational proxies in a fishtank. *CHI '01 Extended Abstracts*, 365-366.
- [10] Festinger, L. (1954). A theory of social comparison processes. *Human Relations*, 7(2), 117-140.
- [11] Fogg, B.J. (2003). *Persuasive technology: Using computers to change what we think and do*. San Francisco, CA: Morgan Kaufmann.
- [12] 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.
- [13] Gersick, C.J.G. (1988). Time and Transition in Work Teams: Toward a New Model of Group Development. *Academy of Management Journal*, 31(1), 9-41.
- [14] Gaver, W.W., Beaver, J., & Benford, S. (2003). Ambiguity as a resource for design. *Proceedings of CHI '03*, Ft. Lauderdale, FL. 233-240.
- [15] Greenberg, S. (1996). Peepholes: Low cost awareness of one's community. *Proceedings of CHI '96*, Vancouver, BC, Canada. 206-207.
- [16] Hackman, J.R., & Wageman, R. (2005). A theory of team coaching. *Academy of Management Review*, 30(2), 269-287.
- [17] Hill, W.C., & Hollan, J.D. (1994). History-enriched digital objects: Prototypes and policy issues. *The Information Society*, 10(2), 139-145.
- [18] Huang, E.M. & Mynatt, E.D. (2003). Semi-public displays for small, co-located groups. *Proceedings of CHI '03*, Ft. Lauderdale, FL, 49-56.
- [19] Johnson, D.W., & Johnson, R.T. (1975). *Learning together and alone: cooperation, competition, and individualization*. Englewood Cliffs, NJ: Prentice-Hall.
- [20] Kim, T., Chang, A., Holland, L., Phillips, J.M., & Pentland, A. (2008). Meeting mediator: Enhancing group collaboration using sociometric feedback. *Proceedings of CSCW '08*, San Diego, CA. 457-466.

- [21] 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.
- [22] Kluger, A.N., & DeNisi, A. (1996). Effects of feedback intervention on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psych. Bulletin*, 119(2), 254-284.
- [23] Lakoff, G., & Johnson, M. (1981). *Metaphors we live by*. Chicago: University of Chicago Press.
- [24] Leshed, G., Hancock, J.T., Cosley, D., McLeod, P.L., & Gay, G. (2007). Feedback for guiding reflection on teamwork practices. *Proceedings of GROUP '07*, Sanibel Island, FL. 217-220.
- [25] 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. *Proceedings of CHI '09*, Boston, MA. 537-546.
- [26] Levelt, W.J.M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.
- [27] Lin, J.J., Mamykina, L., Lindtner, S., Delajoux, G., & Strub, H.B. (2006). Fish'n'Steps: Encouraging physical activity with an interactive computer game. *Proceedings of UbiComp '06*, Orange County, CA, 261-278.
- [28] 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.
- [29] Matthews, T., Dey, A.K., Mankoff, J., Carter, S., & Rattenbury, T. (2004). A toolkit for managing user attention in peripheral displays. *Proceedings of UIST '04*. Santa Fe, NM, 247-256.
- [30] Matthews, T. (2007). *Designing and evaluating glanceable peripheral visualizations*. Unpublished PhD, EECS Department, University of California, Berkeley.
- [31] McLeod, P.L., & Kettner-Polley, R.B. (2004). Contributions of psychodynamic theories to understanding small groups. *Small Group Research*, 35(3), 333-361.
- [32] 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.
- [33] Pennebaker, J.W., Francis, M.E., & Booth, R.J. (2001). *Linguistic inquiry and word count: LIWC2001*. Mahwah, NJ: Erlbaum.
- [34] Poole, M.S., & DeSanctis, G. (1989). Use of group decision support systems as an appropriation process. *HICSS'89*, Kailua-Kona, HI. 149-157.
- [35] Rofé, Y. (1984). Stress and affiliation: a utility theory. *Psychological Review*, 91(2), 235-250.
- [36] Tufte, E.R. (1983). *The visual display of quantitative information*. Cheshire, CT: Graphics Press.
- [37] Turner, G.A., & Schober, M.F. (2007). Feedback on collaborative skills in remote studio design. *HICSS'07*, Waikoloa, Big Island, HI.
- [38] Viégas, F.B., & Golder, S.D., & Donath, J. (2006). Visualizing email content: Portraying relationships from conversational histories. *Proceedings of CHI '06*, Montréal, Québec, Canada. 979-988.
- [39] Wickens, C. D., & Hollands, J. G. (2000). *Engineering psychology and human performance* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.