# Visualizing Real-Time Language-Based Feedback on Teamwork Behavior in Computer-Mediated Groups

**Gilly Leshed[1], Diego Perez[4], Jeffrey T. Hancock[1,2], Dan Cosley[1,2], Jeremy Birnholtz[1,2], Soyoung Lee[3], Poppy L. McLeod[2], Geri Gay[1,2]**

| [1]Information Science | [2]Communication Dept. | [3]Operations Research | [4]Natural Language Group |
|---|---|---|---|
| Cornell University | Cornell University | Cornell University | Microsoft Corporation |
| Ithaca, NY | Ithaca, NY | Ithaca, NY | Redmond, WA |

[gl87, jth34, drc44, sl426, jpb277, plm29, gkg1]@cornell.edu          diegope@microsoft.com

## ABSTRACT

While most collaboration technologies are concerned with supporting particular tasks such as workflows or meetings, many work groups do not have the teamwork skills essential to effective collaboration. One way to improve teamwork is to provide dynamic feedback generated by automated analyses of behavior, such as language use. Such feedback can lead members to reflect on and subsequently improve their collaborative behavior, but might also distract from the task at hand. We have experimented with GroupMeter – a chat-based system that presents visual feedback on team members' language use. Feedback on proportion of agreement words and overall word count was presented using two different designs. When receiving feedback, teams in our study expressed more agreement in their conversations and reported greater focus on language use as compared to when not receiving feedback. This suggests that automated, real-time linguistic feedback can elicit behavioral changes, offering opportunities for future research.

## Author Keywords

Feedback visualization, teamwork, linguistic analysis, peripheral displays, CMC, CSCW.

## ACM Classification Keywords

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

## INTRODUCTION

Growing interest in computer-supported group work [10] has motivated research on developing communication and collaboration technologies that support many aspects of group work: fostering informal interaction (e.g., [2]), creating awareness of colleagues' presence and behavior (e.g., [44]), and supporting shared tasks such as writing or meeting (e.g., [48]). An implicit assumption in this work is that simply pro-

viding these tools is sufficient to improve computer-mediated collaboration. However, unless teams are given guidance on the basics of effective collaboration skills, the tools may offer little benefit [42]. For instance, giving a design team a screen sharing tool in order to help them come up with a single design proposal will not necessarily help them to develop the consensus-building skills needed to complete the assignment collaboratively.

Teamwork can be a powerful tool for learning [41,43] and for accomplishing tasks [20]. However, if people are expected to successfully work in teams, they should be given tools to adopt appropriate interpersonal skills to overcome challenges and to achieve effective team processes and outcomes [5,35]. These skills are often taught by providing team members with feedback on their own behavior, along with guidance on how certain changes in their behavior might improve group outcomes [34]. Our high-level research goal, therefore, is to understand how collaboration technology can illuminate social processes and behaviors within teams, allowing members to reflect and learn to become better collaborators.

One of our specific goals is to design computer-mediated environments that both focus attention on group process and support the team task, thus enriching teamwork experiences. Below we describe our underlying principles for providing automated feedback to teams using peripheral visualizations as the team goes about completing its tasks. We then present GroupMeter, a research platform proposed by Leshed et al. [27] that applies these principles and used for the study of teamwork behaviors, language, and feedback. We conclude by presenting a distributed group-based user study of GroupMeter, focusing on four main questions:

RQ1: Does automated feedback about language behavior cause people to reflect on their use of language?

RQ2: Do people modify their communication patterns when provided with automated feedback?

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

RQ4: How do people experience the use of different feedback designs?

## FEEDBACK ON TEAMWORK PROCESS

Bosworth suggested that providing feedback on teamwork behaviors in practice situations is a key element in teaching collaborative teamwork skills, allowing team members to process and reflect on their own experiences [4]. McLeod et al. [34] found that feedback on interpersonal teamwork behaviors, such as how much members were participating or focusing on the task, helped balance participation rates among team members and also shift the focus from the task toward also considering some socio-emotional aspects of teamwork (see also [29]).

Traditionally, the source for group process feedback has been either mentors—teachers, managers, or expert observers [34]—or group members through peer assessment procedures such as interviews or questionnaires [8]. However, asking people to provide feedback has drawbacks. In particular, feedback from peers, supervisors, or instructors, can be obtrusive and expensive [29,34].

We suggest an alternative approach: using automated analysis of communication content to provide feedback. In many settings, teamwork consists mostly of conversations and meetings—often conducted through CMC tools that can capture the conversation. In these cases, automated analysis of group conversations could reveal much about a group's collaborative processes. This, in turn, could be presented to teams to help them improve their skills and performance. Prior to deployment, however, an automated technique must demonstrate (a) analyses of linguistic features that correspond to actual desirable collaborative behaviors, and (b) feedback presentation that causes participants to reflect on and change their behavior.

## LINGUISTIC ANALYSIS OF TEAM CONVERSATIONS

Recent work suggests that automated linguistic analysis can achieve these two criteria. For instance, Latent Semantic Analysis (LSA) of team conversations predicted team performance scores on a flight simulation task [17,19]. In that task, teams where members stated more facts and acknowledged their peers tended to perform better. Further, machine-learning approaches have shown promise in collaborative learning settings in identifying how learners construct arguments that build on others' contributions [23].

Others have investigated using word-level analysis in group situations, such as using the Linguistic Inquiry and Word Count tool (LIWC; [38]), a dictionary-based approach to analyzing language features. For example, higher rates of agreement words (e.g. *yes*, *agree*, *ok*) can predict low peer ratings on Bales and Cohen's [1] scale for evaluating teamwork practices [27]; the use of individual versus collective pronouns ('I' vs. 'we') can indicate a sense of group belonging [7]; and the use of reasoning expressions can indicate focus on the group task [46].

These findings suggest that a system that analyzes language in team conversations can be used to provide effective near-real time feedback to team members, causing them to reflect on their conversational practices. Moreover, we reasoned that such feedback might lead them to change their behavior in beneficial ways.

## VISUALIZING LINGUISTIC INDICATORS AS FEEDBACK

A general strategy for making collaboration behavior visible is to use *social proxies*, graphical representations of users that depict their presence and activities [14]. For example, Babble [15] represents individuals in a discussion forum as colored dots around a circle that represents the group, placing individuals closer to or farther from the center based on their participation in the discussion.

Systems such as Chat Circles [47] and visualizations of turn taking based on audio input [11,24] together with eye-gaze patterns [26] have shown that social proxies can stimulate reflection on and influence collaborative practice. Visualizing each member's participation relative to other group members, for example, has been shown to prompt dominant contributors to become aware of their behavior [11,12] and attempt to achieve more balanced participation [26]. Similarly, augmenting synchronous CMC environments with visualizations of contribution level and amount of group agreements led users to perceive the group experience as more positive and effective, and to perform better [21,22].

These latter studies highlight the potential of using finer-grained behaviors, such as indicators of agreement obtained through linguistic analysis, as a source of feedback to help group members attend to, reflect on, and change their communicative and collaborative behaviors. We thus posed the following research questions:

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

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

### Peripheral Feedback Visualizations

Despite its value, feedback cannot be the star of the show. A long-recognized principle of teamwork effectiveness is that group members must attend both to their task and to social matters (e.g., [20]). It is thus important, in designing a system to provide group process feedback, to ensure that team members are able to focus on their task as well.

Our approach is to provide dynamic feedback in the periphery of the team activity interface. This allows team members to maintain awareness of the social process and how their behaviors are tied to it, while still focusing on the primary team activity. It is important that the peripheral visualization effectively conveys awareness information [6,40] in ways that minimize cognitive load and do not significantly impact users' performance on the primary task [30].

Still, our aim is to stimulate reflection on and awareness of language use and teamwork behavior, leading to a possible tension between attending to ambient feedback visualizations and accomplishing a task. Another key research question,

therefore, is how to create designs that are cognizant of this balance between task and process:

RQ3: Do people find feedback about their language use distracting?

## GROUPMETER

We implement our visualizations of linguistic feedback using GroupMeter, a platform designed to help teams reflect on their group behaviors and support research into the relationships between communication behaviors, feedback, and teamwork. GroupMeter was conceptualized in [27] based on a study with a peer feedback intervention, but was not fully designed and implemented. We present here a working prototype developed to carry out the study presented later.

The basic design of GroupMeter consists of a web-based system in which groups communicate through a chatroom to perform their tasks and see feedback visualizations based on their conversation. The front end consists of an AJAX-driven HTML chat with feedback visualizations presented using Adobe Flash. A server manages the chat sessions and parses the chat text to provide the feedback metrics. An administrative UI supports management of user accounts, group creation, feedback metrics and sessions. GroupMeter's modular approach allows easy modification of variables such as the linguistic metrics presented and the specific design of the visualizations.

The chatroom text is processed using a dictionary-based word count tool based on LIWC [38]. LIWC counts what percentage of words in a block of text fall into various categories such as emotion words, self-references, and assents. Unlike advanced natural language processing techniques (e.g., [17,23]), LIWC can produce linguistic markers in near-real time. Further, these markers can serve as measures of conversation style and social behavior [37]. For example, the use of pronouns requires a shared understanding of their referent between the speaker and listener [9].

In the current study we chose to present feedback on *word count* and *percentage of agreement words* (LIWC's *assent* category modified for chat communication style), based on two sets of previous research findings. First, [27] showed that the amount of contribution to discussion is positively associated with peer-rated teamwork behaviors. Second, they also showed that critically addressing the conversation without passively agreeing was related to positive collaborative behaviors (see also [21]).

### Designing the Visualizations

We tested two designs for feedback visualization. Our main design goal for both was to give group members a clear notion of how their linguistic behaviors compare to those of other group members. This allows individuals to compare themselves to specific others in the group as opposed to an abstract aggregate (see social comparison theory, [16]), as well as allowing them to reflect on both their own and their teammates' behavior (e.g., [34]).

At the same time, we wanted to be sure the design did not encourage competitive behavior. Hence it was important for us to show individuals as being part of a larger group – all of this through aesthetically pleasing visualizations that "tell a story about the data" [45, p. 177]. We also wanted to be ambiguous to some extent with respect to behavioral norms, hoping to make room for users' own interpretations relevant to their specific context [18]. Finally, to minimize distraction, we designed the visualizations with the principle of glance-ability [31], aiming for easy visibility of relevant feedback information.

### *Design for Effectiveness: Bar Graphs*

In the first visualization, feedback is presented as horizontal bars, one for word count and one for percentage of agreement words. The bars change in length based on each participant's linguistic behavior (Figure 1a). We chose to present each linguistic indicator as an aggregate stacked bar to minimize the possibility



**Figure 1. The GroupMeter chatroom with two feedback designs: (a - top) bar graphs, (b - bottom) school of fish. Length of bars and fish size and position represent team members' word count and proportions of agreement word use.**

that people would process the visualization in a competitive way. An alternate display using a standard bar chart might cause people to try to have the highest bar. Further, showing individuals' behavior as an aggregate bar emphasizes the idea of being part of a group. We located the bars at the bottom of the chat box as this allowed us to place them on-screen without the need for scrolling. The intent of the bars' design and placement is to provide ambient, unobtrusive feedback that can be efficiently processed at a glance.

*Design for Attention: School of Fish*
The second visualization (Figure 1b) consists of an animated school of fish with each member represented by a colored fish. We hoped that the school of fish would instill the feeling of being part of a group. The fish start in a circular formation, all at the same size and equidistant from the center. When a participant uses more agreement words, their fish moves closer to the center. Fish increase in size when a participant speaks more. Again, the shape of the visualization dictated its placement. Locating it below the chat box would have required users to scroll to see it; placing the visualization to the right of the chat box allowed it, like the bars, to be continually visible. The purpose of the lively nature of the fish was to draw attention to the feedback information represented by their position and size.

Our final research question, then, relates to how these different visualizations of the same feedback information affect users' experience:

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

## EXPERIMENT OVERVIEW
These visualizations allowed us to address each of our four research questions in the experiment described below. We wanted to assess: the capability of GroupMeter's feedback visualizations to trigger reflection on language use and word choice (RQ1); people's actual use of language beyond their perceptions and reflections when provided with these visualizations (RQ2); whether one or the other visualization was more distracting (RQ3); and users' felt experiences with each visualization (RQ4).

## Method
*Participants.* Twenty-five undergraduate students (9 females) from a Human-Computer Interaction course in a large university in the United States received course credit for their participation in the study. As HCI students, these participants are likely sensitive to interface design issues and we sought their opinions about our designs.

*Procedure.* Figure 2 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.

To engender a feeling of distributed chatroom collaboration, we 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 text) 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 *word count* and *agreements* measures were then described for both visualizations.

For each task, the team was given five minutes to brainstorm ideas for solving a problem, and then five 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, we felt that the risk of carryover effects was smaller than it would be for visualizations. We therefore kept the task order constant for all groups and randomized 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 (e.g., "The visualization caused me to choose words differently as the task progressed"), distraction by the visualizations (e.g., "I was paying attention to the feedback"), and open-ended questions to capture their experience (e.g., "What did you like about the fish visualization?").

We 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.
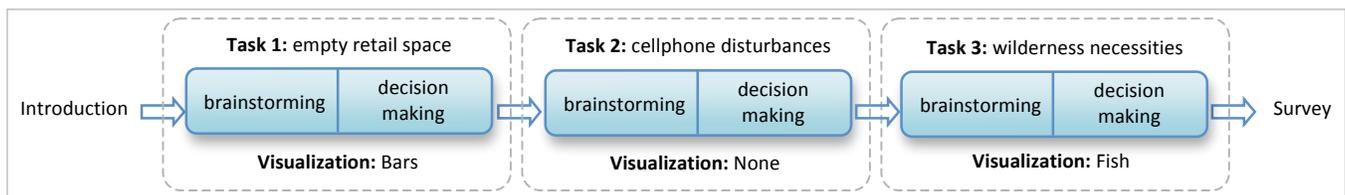


**Figure 2. An example session. Task order remained constant while visualization order was randomized across groups.**

**Data Analysis**

Our data consisted of numerical and open-ended responses to the survey and the transcripts of the chatroom conversations. To examine communication patterns within teams, we coded each chat entry with one of six codes representing a behavioral move that we believe relate to the nature of the tasks and the feedback provided through the visualizations.

- *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."

We excluded entries about study administration, troubleshooting, digressive remarks, and fillers. Two coders coded the transcripts independently, 1269 entries in total, assigning each entry the most prominent code. Inter-coder agreement was 80%; disagreements were resolved by discussion.

Our data were hierarchical in nature, with observations nested within participants and participants nested within teams. We 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 our statistical tests.

**RESULTS**

**Awareness of Language Use**

Since RQ1 was framed in terms of participants' perceptions, we 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 (Cronbach's α = 0.77): "I fo-

cused 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 3a).

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 the 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. We therefore used the coded chat transcripts as behavioral data in answering this question. Based on [33], we 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. We therefore divided the transcripts into two segments, according to the 5-minute subtask periods during which they were typed.

Figure 4 shows the relative proportions of each coded entry type by condition and segment. The data for the bar charts appears as percentages in Table 1. For each entry type, we 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, we 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 4a).* We were 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.
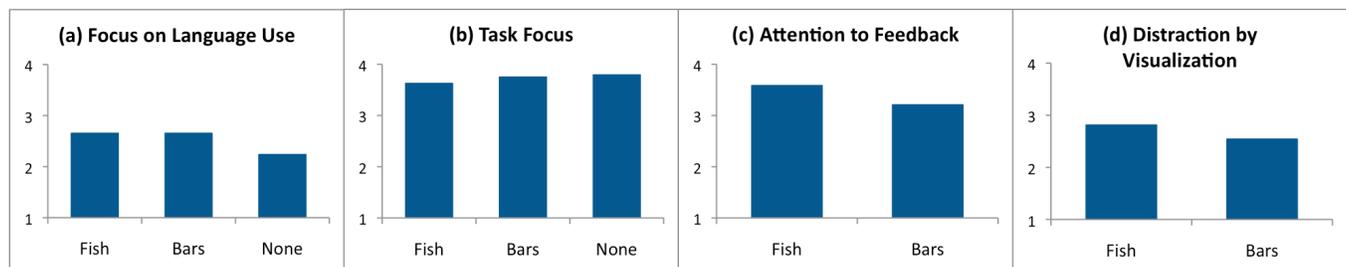


Figure 3. User responses to survey questions: (a) Focus on language use, (b) Focus on task,
(c) Attention paid to feedback visualizations, (d) Distraction by feedback visualization. Responses were on a 5-point scale.

**Table 1. 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 |

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*).

We did find a statistically significant main effect of *Segment* ($F(1,72)=64.94$, $p<0.001$). As expected, more ideas were presented during the brainstorming segment.

*Agreement Entries (Figure 4b).* 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 we 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, we found a main effect of *Segment*, with participants using a higher proportion of agreement entries during the decision-making segment than during 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 4c).* For disagreement entries, we did not find a main effect of *Condition* ($F(2,48)=2.00$, *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, we 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 our finding on the increase of agreement entries from the brainstorming to the decision-making segment; both changes suggest that groups attempted to reach consensus.
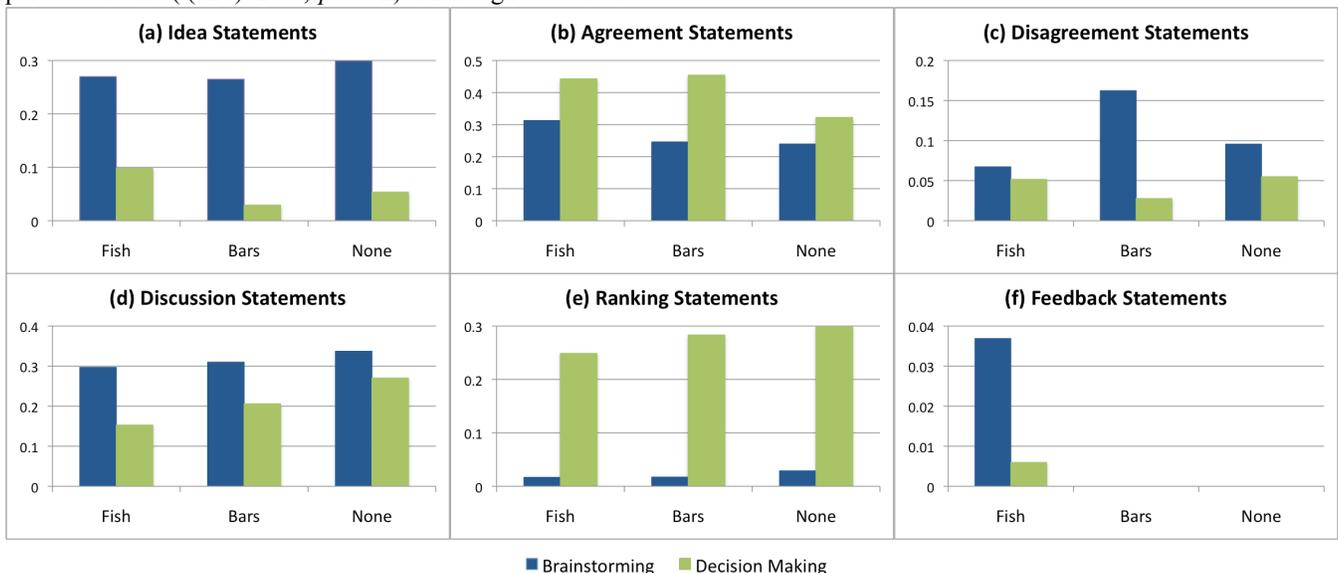


**Figure 4. Proportions of each entry type by condition and segment. For each condition/segment pair, its bars sum to 100% across the six entry types, as shown in Table 1.**

*Discussion Entries (Figure 4d).* 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. This calls for additional investigation, but we 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 4e).* We 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 4f).* As one indicator of the extent to which different visualizations changed team interaction, we 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. We will present further evidence relative to RQ3 and RQ4 below.

*Total Conversational Activity.* Finally, we 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 us to address 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 the 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, we found little evidence of gaming the system in our reading and coding of the transcripts.

### Automated Analysis Captures Communication Patterns
An important assumption in our approach is that automated linguistic analysis can effectively capture communication patterns. To assess this assumption, we 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, we could not simply examine correlations given the non-independence in the group data. Instead, we 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, we addressed RQ3 by asking participants to evaluate their task focus in each of the tasks, using the 5-point item "I remained focused on the task throughout the exercise" (Figure 3b). 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, we asked participants to rate the item "I was paying attention to the feedback" (Figure 3c). 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. People also found them marginally more distracting (Figure 3d), based on an aggregate score of two 5-point items (Cronbach's $\alpha$ = 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, our results provide several answers. First, 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, we were interested in participants' qualitative experiences with and opinions about the visualizations (RQ4). We asked participants in open-ended questions to tell us 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 our users experienced the fish visualization as engaging and even enchanting (as defined by [32]), 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 for. They liked the unobtrusiveness of the bars, but also criticized their lack of excitement.

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

Addressing RQ1, we found that providing visual feedback made users more aware of their language use. 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 [28]. Thus, we 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.

With respect to RQ2, our results show that people changed their behavior when seeing 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 previous research suggesting that people's choice of words in conversation is largely spontaneous, unintentional, and uncontrolled [9,39]. In our study, participants were nonetheless able to effect changes in their communication patterns (e.g., agreeing) in both visualization conditions.

Our 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. We did not explicitly pose normative guidelines, although 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 [13]. If fewer agreements and more discussion are favorable behaviors [21,27], integrating appropriate guidelines for effective teamwork into the design is an important future direction.

Regarding RQ3, our findings demonstrate the complexity in designing feedback visualizations that raise awareness of and reflection about social processes without distracting the team. 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 than the bars, suggesting that given the desired reflection and behavioral change, more subtle visualizations may be sufficient. One limitation of our 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, our 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.

## FUTURE WORK

We view both our design and implementation of GroupMeter, and the experimental results, as successful preliminary efforts in a broader program of understanding the relationships between collaboration, automated linguistic feedback, and visualization design. Extended research is needed to fully understand the ways that technology can be used to illuminate and improve collaborative behavior.

One area for future work lies in understanding the value of specific behavioral changes, and how to provide interpretable feedback that will elicit these changes. For example, brainstorming tasks might call for encouraging people to generate new ideas while discouraging both agreements and disagreements; other tasks might require different metrics and different normative goals. Careful selection of appropriate feedback dimensions is therefore imperative if desirable changes are to result from presenting feedback. More work is also 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.

Another direction to pursue is to examine cross-cultural issues of feedback on language use in computer-mediated communication. For example, the use of singular versus plural first person pronouns in CMC environments differs between Western and Eastern cultures [25]. Providing feedback to team members on their first person pronoun word choice could thus have differing effects depending on their culture. Also, whereas American culture is very task-oriented, European and Asian cultures value the social relationships within the team [36]. This can lead to major differences in how team members accept process feedback that is not directly related to the task at hand.

## CONCLUSIONS

Dynamic feedback visualizations of teamwork process information have the potential to affect the interaction of users in CMC environments. 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. We 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.

Furthermore, our findings show that automated linguistic feedback can be powerful in more than raising awareness of one's own language use – it can also cause people to alter their communication patterns. Our work follows a line of research that shows how automated dynamic feedback in teamwork situations can affect social interactions [3,12,21,22,26]. GroupMeter adds to this work by influencing teamwork behavior through detailed feedback about word choice and production in computer-mediated settings.

Our work also contributes to research around peripheral awareness systems [6,30,31,40]. We demonstrate that the design of such interfaces should carefully consider the balance between focus on the primary task, here the teamwork task, and maintaining awareness of peripheral information, here the nature of team communication. Finally, our findings reveal the power of different visualizations representing the same data to bring about diverse user interpretations and experiences..

## REFERENCES

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

2. Biehl, J.T., Baker, W.T., Bailey, B.P., Tan, D.S., Inkpen, K.M., & Czerwinski, M. (2008). Impromptu: a new interaction framework for supporting collaboration in multiple display environments and its field evaluation for co-located software development. *Proc. of CHI'08*, 939-948.

3. Bergstrom, T., & Karahalios, K. (2007). Seeing More: Visualizing Audio Cues. *Proc. of INTERACT'07*, 29-42.

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

5. Bransford, J. (2000). *How people learn: Brain, mind, experience, and school*. Washington: Nat'l Academy Press.

6. Cadiz, J.J., Venolia, G., Jancke, G., & Gupta, A. (2002). Designing and deploying an information awareness interface. *Proc. of CSCW'02*, 314-323.

7. Cassell, J., & Tversky, D. (2005). The language of online intercultural community formation. *JCMC*, 10(2).

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

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

10. Cummings J., Finholt, T., Foster, I., & Kesselman, C. (2008). Beyond being there: A blueprint for advancing the design, development, and evaluation of virtual organizations. *Building Effective Virtual Organizations* workshop.

11. DiMicco, J.M., Pandolfo, A., & Bender, W. (2004). Influencing group participation with a shared display. *Proc. of CSCW'04*, 614-623.

12. 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.

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

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

15. Erickson, T., Smith, D.N., Kellogg, W.A., Laff, M., Richards, J.T., & Bradner, E. (1999). Socially translucent systems: Social proxies, persistent conversation, and the design of "babble". *Proc. of CHI'99*, 72-79.

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

17. Foltz, P.W., & Martin, M.J. (2004). Automated team discourse annotation and performance prediction using LSA. *Proceedings of HLT/NAACL'04*, 97-100.

18. Gaver, W.W., Beaver, J., & Benford, S. (2003). Ambiguity as a resource for design. *Proc. of CHI'03*, 233-240.

19. 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. *Proc. Human Factors and Ergonomics Society*, 424-428.

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

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

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

23. Joshi, M., & Rosé, C.P. (2007). Using transactivity in conversation summarization in educational dialog. *Proc. SLaTE'07*, 53-56.

24. Karahalios, K.G., & Bergstrom, T.(2006).Visualizing audio in group table conversation. *Proc. TableTop'06*, 131-134.

25. Kayan, S., Fussell, S.R., & Setlock, L.D. (2006). Cultural differences in the use of instant messaging in Asia and North America. *Proc. CSCW'06*, 525-528.

26. Kulyk, O., Wang, J., & Terken, J. (2006). Real-time feedback on nonverbal behaviour to enhance social dynamics in small group meetings. *Proc. MLMI'05*, 150-161.

27. Leshed, G., Hancock, J.T., Cosley, D., McLeod, P.L., & Gay, G. (2007). Feedback for guiding reflection on teamwork practices. *Proc. Group'07*, 217-220.

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

29. Losada, M., Sanchez, P., & Noble, E.E. (1990). Collaborative technology and group process feedback: Their impact on interactive sequences in meetings. *Proc. CSCW'90*, 53-64.

30. Maglio, P.P., & Campbell, C.S. (2000). Tradeoffs in displaying peripheral information. *Proc. CHI'00*, 241-248.

31. Matthews, T. (2006). Designing and evaluating glanceable peripheral displays. *Proc. DIS'06*, 343-345.

32. 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.

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

34. 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, 15-41.

35. 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.

36. Olson, G., & Olson, J. (2000). Distance Matters. *Human-Computer Interaction*, 15(2-3), 107-137.

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

38. Pennebaker, J.W., Francis, M.E., & Booth, R.J. (2001). *Linguistic Inquiry and Word Count LIWC 2001*. Erlbaum.

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

40. Plaue, C.M., Miller, T., & Stasko, J.T. (2004). Is a picture worth a thousand words? An evaluation of information awareness displays. *Proc. Graphics Interface'04*, 117-126.

41. Roberts, T.S. (2005). *Computer-supported collaborative learning in higher education*. Hershey, PA: Idea Group.

42. Rummel, N., & Spada, H. (2005). Learning to collaborate: An instructional approach to promoting collaborative problem solving in computer-mediated settings. *Journal of the Learning Sciences*, 14(2), 201-241.

43. Stahl, G. (2006). *Group cognition: computer support for collaborative knowledge building*. Cambridge: MIT Press.

44. Tee, K., Greenberg, S., & Gutwin, C. (2006). Providing artifact awareness to a distributed group through screen sharing. *Proc. CSCW'06*, 99-108.

45. Tufte, E.R., (1983). *The visual display of quantitative information*. Cheshire, CT: Graphics Press.

46. Turner, B., & Schober, M.F. (2007).Feedback on collaborative skills in remote studio design. *Proc. HICSS'07*, 44-53.

47. Viégas, F.B., & Donath, J.S. (1999). Chat circles. *Proc. CHI'99*, 9-16.

48. Weng, C. & Gennari, J.H. (2004). Asynchronous collaborative writing through annotations. *Proc.CSCW'04*, 578-581.