

## **Darren Gergle**

Human Computer Interaction Institute  
School of Computer Science  
Carnegie Mellon University  
Pittsburgh, PA – USA

### **Vision:**

As described in the workshop call, a major challenge for CSCW researchers is in developing robust methods for measuring the benefits of technologically mediated collaboration. My research at Carnegie Mellon University is geared towards answering this challenge and is guided by the premise that the design of technologies to support collaborative work requires a theoretically based understanding of both media and task features.

Clark and colleagues posit that successful interactions rely on a foundation of common ground and shared mutual understanding [3]. The process of reaching this state of common ground is called grounding. Together, Clark & Brennan [4] have developed a theory that elaborates how different communication media affect the grounding process. They set forth eight interrelated constraints of media that may be absent, reduced or otherwise altered by a given technology: copresence, visibility, audibility, cotemporality, simultaneity, sequentiality, reviewability, and revisability. Their theory states that when a media alters the availability of one of these affordances, users will have to put more time and effort into grounding their conversations. A good deal of empirical work has shown that collaboration media do indeed impact the grounding process [10,11,14,15].

While many of these studies compared face-to-face groups with groups in *distributed* conditions (e.g., email, chat, telephony), Clark & Brennan's constraints are also useful for investigating technological support in *collocated* collaborative environments. Consider a system that captures group conversations in real-time, converts the speech to text, and displays it for the group to see on a local display. In this case, users can now review the prior discourse, use visual copresence of the discourse to point and gesture at, and can revise and edit previous statements for archival purposes. However, changing normally ephemeral information into a persistent form may alter group communication in several ways. Formulation and production costs may increase since the individuals know their words will be recorded and displayed for all to see. The frequency of dissenting opinions may decrease with the added social pressures that are part of a "reviewable" conversation. Finally, the distribution of contributions may change across the group as a whole since the typical imbalance may be made more salient to all members of the group.

While Clark & Brennan's theory is useful for understanding the potential merits of technology for collocated collaboration, there are two major limitations to their work. The first is that their theory applies primarily to language and there is no similar theory for understanding the interaction of media and tasks for *physical actions* that are much more common in collocated interaction. A second limitation is their description of tasks and how they interact with media constraints. Rather than simply stating that media interacts with task, a more detailed description of how *sub-features* of the media interact with *sub-features* of the task environment and objects is needed. In its current state, Clark & Brennan's theory does not provide a nuanced enough account of the variability in the nature of tasks or technologies in order to generate strong theoretical predictions about performance. To demonstrate this, consider the following scenarios:

First, consider a medical setting where scrub nurses need to provide instruments to surgeons in real time. The names of the instruments are typically shared and domain specific (e.g., mosquito clamps or suture needles), the task typically occurs in pre-scripted sequential stages (e.g., anesthetize the patient, prep the surgical area, begin surgical procedure, post-op closure and cleaning), and the pairs have likely been through several hundred repetitions. Now compare this with the case of a collocated novice and expert jointly repairing a car engine. Here the task information is asymmetric (e.g., the expert knows more), common language for the task objects may or may not be established (i.e., the pair needs to ground their conversational terms), the task goals are not clear and there is no strict order to the subtasks, the dependencies between the actions of the pairs can fluctuate from being loosely to tightly-coupled over the

duration of the task. Note that without a sufficiently nuanced description of media and task both of these collaborative tasks may look quite similar: the dyads are collocated, they both need to communicate to one another about their task, and they are performing a joint physical task. However, a more detailed investigation reveals that they are very different both in group structure, task features, social structure and environment.

My agenda for the future is to elaborate how particular task features vary across a more basic underlying set of media features. For example, collocated visual evidence may be more or less useful depending on the particular features of the shared visual space (e.g., field-of-view), the nature of the task (e.g., tightly vs. loosely-coupled), or the features of the task objects (e.g., easily discriminable vs. highly similar objects).

## **Experiences & Challenges:**

In order to inform the development of such a framework my colleagues and I have engaged in a series of decompositional lab studies where we strive to identify the key features of tasks and technologies in order to predict what technologies will be best suited for given tasks (see [9] for a more complete description; [6] for a detailed example). This approach uses a stylized puzzle paradigm based on a referential communication task in which Helper / Worker pairs collaborate on the construction of computationally shared visual puzzles. In this paradigm a Helper describes to a Worker how to construct a puzzle so it matches a target configuration. This research paradigm allows the systematic examination of collaboration by providing three major sources of data: Task performance data, conversational and communicative efficiency data, and sequential data that allow the exploration of interactions between communication content and behavioral actions.

### *Assessing Task Performance*

Across the stylized puzzle studies, we've consistently found that shared visual space improves performance as measured by task completion time. Participants complete the task much faster when they share a visual space (or are visually co-present). This paradigm also allows us to examine interactions with the features of the shared space. For example, performance is enhanced when visual spaces are tightly synchronized (i.e., the helper views workers' activities in real time) or when the proportion of shared visible space increases (e.g., field-of-view).

In addition to facilitating a more detailed exploration of the features of shared visual space, this paradigm allows a more detailed understanding of how the value of visual information changes depending on attributes of the task. Collaborative tasks differ widely along a variety of dimensions that might plausibly affect the ways in which visual information is used. In these studies the value of the shared visual space is dramatically increased when the task objects or environment are dynamic. Similarly, when the visual complexity of the task objects or the environment increases there is a greater benefit to having shared visual space [6,7]. Our experimental design provides a straightforward measure of performance in the time it takes to complete the puzzles. Interactions between task and media can be examined by assessing performance times across a range of experimental conditions.

### *Assessing Conversational and Linguistic Efficiency*

In addition to exploring task completion times, capturing the communications and actions of the pairs allows us to examine the efficiency of the conversational content. When shared visual space is available pairs can rely on more efficient and less ambiguous visual evidence, increasing the overall conversational efficiency of the pair. This efficiency can be seen in several ways. First, when pairs share full visual co-presence the overall number of linguistic tokens needed to successfully complete the task decreases. Second, the pairs are more likely to use efficient linguistic terms such as deictic expressions (e.g., "this one") when they share visual co-presence. Finally, in several studies we have found that the presence of visual information impacts how partners distribute their initiative and conversational effort. Helpers take more initiative to determine when Workers appear to have understood referring expressions when they have full visual co-presence whereas Workers take this responsibility when there is no visual co-presence [6].

### *Assessing the Sequential Structure of Physical Actions and Spoken Discourse*

Our recent work has explored using sequential analysis of this same task to uncover the ways in which the general structure of communication changes when technological mediation is introduced. A detailed sequential analysis of the communicative content reveals that pairs with a shared workspace were less likely to explicitly verify their actions with speech. Rather, they relied on visual information to provide the necessary communicative and coordinative cues [5]. This work provides a quantitative method for demonstrating the way in which actions and language interact and unfold over the duration of a communication episode, and how these sequences vary according to the presence of shared visual information. At a theoretical level, this work extends previous analyses of the effects of media on interpersonal communication by providing a richer understanding of the way that physical actions and language are integrated to perform joint actions and ground communication.

### *Benefits and Drawbacks*

There are several benefits to using a referential communication paradigm. First, referring to objects, people and prior discourse is a basic requirement for a number of communication exchanges. Second, it provides a highly structured task that is replicable yet sensitive to task and media manipulations. Third, there are several levels of analysis that can be performed on the data, and as you drill down you tend to get more detailed information. For example, the raw performance times give you some evidence as to the speed or overall efficiency with which communication happens in different mediated conditions. However, this tells you little about the details of how the communication changes. Information about *how* the pairs are more efficient can be gained through an exploration of the conversational content and efficiency calculations based on coded linguistic transcripts. This takes more effort and the analysis becomes more time consuming, but the tradeoff is richer detail. Finally, details of the sequential structure of the behaviors and language provide the most compelling examples of how technological mediation can change communication and performance. However, this involves a major investment of time and effort as it requires transcription, behavioral coding, and rather advanced statistical analyses to determine sequential trends (see [5] for details). In addition, without theoretical guidance, researchers can easily find themselves swimming in data with no real idea of where to begin and where to go.

In addition to the time constraints of such analyses, there are certainly some drawbacks to this experimental paradigm. First, a referential communication task typically assumes an asymmetric relation between the Helper and the Worker and the information they hold. This may be problematic when applying this technique to collocated collaborative tasks where the participants are equally contributing to a physical task and their work is highly interdependent. Further research needs to address this concern. Another challenge is that the external validity of such tasks is often challenged. While we believe that the base actions observed in referential communication tasks are the building blocks of successful communication, it is useful to support these lab studies with more realistic tasks and field studies.

In addition to my work using referential communication tasks, I've used game theoretic constructs such as social dilemmas to better understanding grounding and trust development in mediated communications (for examples see [1,2]; and for a review see [8]). I've developed a series of Stasser tasks (i.e., hidden profile tasks) for exploring information sharing and decision making in groups supported with large scale displays and interactive media displays. I've also recently been involved in examining ways to explore both how interdependencies of tasks in small group activities play into the assessment of technological tools. In addition, I've begun developing a more detailed theory of how task selection plays a critical role in measurable outcomes of lab and field studies of mediated communication. This work involves integrating task taxonomies such as McGrath's [12] and Mennecke & Wheeler's [13] into a useful way of labeling experimental tasks for the assessment of technologically mediated systems.

## **Workshop Goals:**

My primary reason for attending this workshop is to find a group of colleagues with which to discuss the future development of evaluation methods for collaborative technologies. A secondary reason is to gain exposure to other techniques and methods of evaluation and assessment currently in use. It would be beneficial to come out of this workshop with a collection of tasks and their advantages and disadvantages for assessing systems designed to support collocated collaboration—particularly those with an eye towards

assessing systems designed to support collaborative physical tasks. Finally, I believe this will be a useful audience for discussing some of my ideas regarding the way that task features can severely interact with and modify the value of technological interventions. This is a topic that lately has received a good deal of attention, yet there have been few systematic treatments of the topic in the literature.

## Bio:

Darren Gergle is a PhD student in the Human-Computer Interaction Institute (HCII) at Carnegie Mellon University. He is advised by Professor Robert E. Kraut and works closely with Susan R. Fussell exploring shared visual space and its effect on communication and discourse. He received a Master's degree in Human-Computer Interaction from the University of Michigan. His work is currently supported by an IBM PhD Fellowship and he has worked with the Collaborative User Experience (CUE) group at IBM T.J. Watson Research in Cambridge, MA. He has authored several long papers in the areas of collaborative shared visual spaces, mediated communication environments, and large scale displays at ACM CHI, CSCW and DIS conferences. He has also coauthored a top-selling book on web usability with Morgan Kaufmann Publishers.

For additional information see:

<http://www.cs.cmu.edu/~dgergle>

## References:

1. Bos, N., Gergle, D., Olson, J. S., & Olson, G. M. (2001). Being there versus seeing there: Trust via video. In *Proceedings of CHI 2001, Extended Abstracts*, pp. 291-292. New York: ACM Press.
2. Bos, N., Olson, J. S., Gergle, D., Olson, G. M., & Wright, Z. (2002). Rich media helps trust development. In *Proceedings of CHI 2002*, pp. 135-140. New York: ACM Press.
3. Clark, H. H. (1996). *Using language*. Cambridge University Press.
4. Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In L.B. Resnick, R.M. Levine, & S.D. Teasley (Eds.), *Perspectives on socially shared cognition*, 127-149. Washington, DC: APA.
5. Gergle, D., Kraut, R. E., & Fussell, S. R. (2004). Action as language in a shared visual space. To appear as a full paper in *Proceedings of Computer Supported Cooperative Work (CSCW 2004)*. New York: ACM Press.
6. Gergle, D., Kraut, R.E., & Fussell, S.R. (In Press). Language Efficiency and Visual Technology: Minimizing Collaborative Effort with Visual Information. To appear in *Journal of Language and Social Psychology*.
7. Gergle, D., Millen, D., Kraut, R.E., & Fussell, S.R. (2004). Persistence Matters: Making the Most of Chat in Tightly-Coupled Work. In *Proceedings of CHI 2004*, pp. 431-438. New York: ACM Press.
8. Kollock, P. (1998). Social dilemmas: The anatomy of cooperation. *Annual Review of Sociology*, 24, 183-214.
9. Kraut, R. E., Fussell, S. R., Brennan, S. E., & Siegel, J. (2002). Understanding effects of proximity on collaboration: Implications for technologies to support remote collaborative work. In P. Hinds & S. Kiesler (Eds.), *Distributed work* (pp.137-162). Cambridge, MA: MIT Press.
10. Kraut, R. E., Miller, M. D., & Siegel, J. (1996). Collaboration in performance of physical tasks: Effects on outcomes and communication. In *Proceedings of CHI '96*, pp. 57-66. New York: ACM Press.
11. McCarthy, J. C., Miles, V. C., & Monk, A. F. (1991). An experimental study of common ground in text-based communication. In *Proceedings of CHI 91*, pp. 209-214. New York: ACM Press.
12. McGrath, J. E. (1984). *Groups: Interaction and Performance*. Englewood Cliffs, NJ: Prentice-Hall, Inc.
13. Mennecke, B. E., & Wheeler, B. C. (1993). Task matters: Modeling group task processes in experimental CSCW research. In *Proceedings of the Twenty-Sixth Annual Hawaii International Conference on System Science*, Vol. IV, pp. 71-80.
14. Veinott, E. S., Olson, J., Olson, G. M., & Fu, X. (1999). Video helps remote work: Speakers who need to negotiate common ground benefit from seeing each other. In *Proceedings of CHI 99*, pp.302-309. New York: ACM Press.
15. Vertegaal, R. (1999). The GAZE groupware system: Mediating joint attention in multiparty communication and collaboration. In *Proceedings of CHI 99*, pp. 294-301. New York: ACM Press.