Persistence Matters: Making the Most of Chat in Tightly-Coupled Work

Darren Gergle[1], David R. Millen[2], Robert E. Kraut[1], Susan R. Fussell[1]

[1]Human Computer Interaction Institute
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA 15213 USA

[2]IBM T.J. Watson Research Center
One Rogers Street
Cambridge, MA 02142

{dgergle+@cs.cmu.edu; David_R_Millen@us.ibm.com; robert.kraut, susan.fussell@cmu.edu}

ABSTRACT
How much history of the dialogue should a chat client include? Some chat clients have minimized the dialogue history to deploy the space for other purposes. A theory of conversational coordination suggests that stripping away history raises the cost of conversational grounding, creating problems for both writers and readers. To test this proposition and inform design, we conducted an experiment in which one person instructed another on how to solve a simple puzzle. Participants had chat clients that showed either a single conversational turn or six of them. Having the dialogue history helped collaborators communicate efficiently and led to faster and better task performance. The dialogue history was most useful when the puzzles were more linguistically complex and when instructors could not see the work area. We present evidence of participants adapting their discourse to partially compensate for deficits in the communication media.

Categories and Subject Descriptors: H.5.3 [Information Interfaces and Presentation]: Group and Organization Interfaces – collaborative computing, computer-supported cooperative work.

General Terms: Design, Experimentation, Human Factors, Performance, Theory.

Keywords: Persistence, text chat, shared visual space, computer-mediated communication, empirical studies, language, and communication.

INTRODUCTION
It is increasingly common for people to collaborate by jointly viewing a dynamic work area, while communicating via real-time text chat. Using collaboratories, scientists at research labs scattered around the world discuss rapidly changing visual data from expensive instruments monitoring solar storms, network traffic, biological and geological samples [20]. In massively multi-player online role playing games (MMORPG), playgroups separated by hundreds of miles can join together to attack a common foe. To formulate joint tactics, they move through a virtual world observing their own environment and what others are doing, while they exchange chat messages that scroll across their display. At NORAD, air defense officers evaluate threats to the North American airspace by monitoring displays that show the distribution of aircraft over North America, while simultaneously text chatting with air traffic controllers at airports around the country. Collaborative software that integrates visual information and text-based chat for tightly-coupled interactions has been deployed for medical teams, research teams, design teams, performing artists, students and their teachers, among others [13,16].

What should the design of these communication systems look like? Consider the simple question of how much dialogue history a chat client should display. Even on large computer screens, space is scarce. Space devoted to dialogue history or linguistic persistence takes away from the visual work area (e.g., the virtual world in a MMORPG or data visualization in scientific research). In an attempt to mimic the transient nature of spoken conversation, Viegas and Donath created a text-based chat client in which participants can see only a single utterance for each "speaker", which fades with time [22]. Apple's iChat client [15] spends screen space on avatars of the participants, limiting the space available for displaying the history of the dialogue. These designs contrast with other commercial designs, such as AOL's Instant Messenger or IBM Lotus Sametime®, which provide a resizable window, and show as many utterances as fit.

The standard way to make design decisions like these is through user testing and iterative design. Designers base their initial design on formal task analysis, prior practice or intuition, and then modify their designs based on user testing. We argue that, for designs involving computer-mediated communication, existing theory can provide good guidance for design. In this paper, we explore the way that Clark's contribution model of communication and Brennan's model of language as hypothesis testing can provide guidance for the design of computer-mediated communication systems.

BACKGROUND LITERATURE

Grounding in Collaboration

When people work together to solve a problem, they contribute different perspectives. In order to coordinate their activities, they need a common set of goals and a shared language to discuss them. Work by Herb Clark describes the collaborative process by which conversational partners work together to develop this shared understanding [3]. This process is based on building shared knowledge or common ground. Common ground is comprised of the mutual knowledge, beliefs, attitudes, and expectations of the conversational partners [3,5], and the process of reaching common ground is referred to as grounding. Brennan extended this model by proposing that speakers continually form and test hypotheses about what a conversational partner knows at any moment both to plan utterances and to revise them after they have been delivered.

Media Differences in Grounding

Clark and Brennan argue that different communication media have features that change the costs of grounding. For example, the media may change the time speakers have to plan an utterance, the evidence from which speakers can infer a listener's state of understanding, or the listener's ability to provide feedback to show understanding or ask for clarifications [4]. Two sources of common ground that are often affected by media and provide important evidence to interlocutors in a dyad are linguistic copresence and visual copresence.

Linguistic copresence is the mutual involvement that the interlocutors have of the conversation up to the present point, which allows them to infer what each other knows of the conversation. In a synchronous medium, like a telephone conversation, speakers can assume that references mentioned in recently preceding turns are jointly known to listeners and to themselves. Consider a case where one person (a helper) is describing to another (a worker) how to position pieces in a jigsaw puzzle. If the helper told the worker in one speaking turn, "Take the dark blue piece and put it in the center of the work area," the helper could say in the next turn, "Take a red piece and put it next to the blue one". The helper can use the phrase "the blue one", with its definite article and pronoun, confident that the worker would understand it because the antecedent had occurred so recently in the linguistically copresent dialogue. However, if the antecedent had occurred further back in the spoken dialogue, the speaker would not necessarily make this assumption given the ephemeral nature of spoken language. Indeed, the speaker may not remember which objects had been previously mentioned. In contrast, written dialogue makes previous speaking turns visible and provides linguistic history as a reminder and as evidence to infer that the listener would probably understand the reference. In fact, McCarthy and Monk [18] show that interlocutors are more likely to make reference to speaking turns further back in the dialogue when they have a larger dialogue history available.

Visual copresence is the visual environment mutually known to be available to the interlocutors. If a pair were constructing the same puzzle side-by-side, the helper could use a deictic reference and say, "Take the red piece and put it next to that one," because he knew that both he and the worker had the blue piece visually present in the work area, and he could infer that the worker would properly decode this reference. In contrast, the helper would be unlikely to use this deictic expression if giving instruction over a telephone because he would not know what was in the worker's field of view at the time of the utterance. Previous research by Kraut, Gergle and Fussell [17] has shown that conversations about linguistically complex objects are more efficient when interlocutors have visual copresence. Both speakers and listeners change their conversational strategies to be more explicit when visual copresence is absent.

The key insight from this discussion is that efforts required in coordinating and maintaining coherence throughout the discourse and the ease with which pairs can form common ground are critically dependent on the features of the media.

The Value of Persistence and Shared Visual Space

In this paper, we examine the way that two resources for grounding—a persistent dialogue history and a shared visual space—interact with features of a collaborative task and with each other to influence task and conversational performance. Our goals are to answer practical design questions about conditions under which a dialogue history is valuable and more theoretically informed questions about the ways different conversational resources trade-off in the grounding process. Although we examine the theoretical questions in the context of a text chat system, we believe our results inform more general comparisons among a wider range of communication media with different persistence qualities.

Persistence in Communication

Recent work on persistence of text chat emphasizes its value for facilitating social awareness and knowledge sharing. According to this view, an archived conversation preserves the history of a project and is a "boon to asynchronous interaction…" (Erickson et al., 1999) [9,19].

Besides providing a resource for long-term information sharing, persistence also plays a critical role for grounding in real-time chat conversations. Herring hinted at this when she suggested, "Without textual persistence… CMC would no doubt be more interactionally incoherent and a great deal more limited in its uses" [14]. However, she did not specify the mechanisms by which textual persistence plays a role in conversational coherence.

As we outlined previously, dialogue persistence in a chat environment supports linguistic copresence, making it easier for interlocutors to know at any instant what information has recently been shared by the pair. Thus, it provides an external representation of information likely to be in common ground and jointly understood. Increasing the amount of history should increase the salience of shared information and decrease memory load for the conversational partners. When the dialogue history is reduced (e.g., to only one or two turns), pairs will be less certain what prior aspects of the conversation are remembered and understood. Therefore, pairs should be less likely to refer to prior aspects of the conversation [18,21]. They might compensate by putting more idea units in a message, which may result in slower performance.

In addition to changing the grounding process, a larger dialogue history provides a means for pairs to parallelize communication and action. Clark and Marshall [6] describe conversation as a sequence of offerings and acceptances in which speakers offer an incremental contribution to the dialogue and listeners provide evidence about their understanding. For example, in the puzzle scenario previously described, helpers generally offer descriptions of a puzzle piece, and elaborate and clarify the descriptions until they have evidence that the worker has accepted the description, either through a verbal response (e.g., "OK") or through an action (e.g., moving the piece). By using the dialogue history as a buffer, a pair can minimize delays produced by slow typing [2]. For example, a helper can type instructions while the worker is performing actions.

A handful of researchers have investigated how dialogue history changes grounding and task performance [2,7,12,18]. Work by McCarthy & Monk is most relevant. They examined dialogue history and shared visual space in a referential communication task. They manipulated dialogue history by giving subjects a chat window with either 6 lines or 30 lines (of up to 10 words each) of history. Their research found no influence of these differences in conversational resources on task performance. Their task—redesigning a banking hall—may not have been sensitive enough. They did find, though, that the larger dialogue history enabled the pairs to reference utterances further back in the discussion.

Shared Visual Space

Just as a dialogue history provides a visible representation of elements of the conversation in common ground, a shared visual space provides a representation of the elements of the physical environment in common ground. In the puzzle task, for example, a helper can see what objects are visible to the worker and can refer to these by efficient deictic expressions (e.g., "take that one"). In addition, the helper can see the actions that the worker performs in response to an utterance and use this visual information as evidence about whether the worker understood the utterance or not. Using these mechanisms, the shared visual space is a resource for grounding that makes the conversation more efficient [1,10,17].

In addition, people performing a joint task can use a shared visual space as evidence about the evolving state of the task in relation to an end goal. For example, in a puzzle task with a shared visual space, helpers can see when workers have finished a subtask (e.g., selecting or positioning a piece), without the workers having to announce this explicitly. Among other benefits, this information about the state of the task aids the helper in planning how to proceed towards the goal, what instructions to give next and when to give them, and how to repair incorrect actions [8,11].

Hypotheses

The following hypotheses summarize this discussion:

H1. Dialogue history, which provides a representation of the elements of the prior conversation that are in common ground, should make grounding more efficient and improve performance in a referential communication task.

H2. Shared visual space, which provides a representation of the elements of a visual environment that are in common ground, should make grounding more efficient and improve performance in a referential communication task.

H3. Shared visual space, as a tool for grounding, should have stronger effects on conversational efficiency than dialogue history. A shared visual space makes visible information that a conversational partner would not otherwise be exposed to. In contrast, a dialogue history only provides reminders of information that both members of a dyad had previously been exposed to.

H4. The effects of dialogue history and shared visual space on grounding, however, should occur only when the linguistic task is difficult. When the task is easy, interlocutors should place all the information they need into a current utterance, without reference to the history of the dialogue or to the external visual environment.

H5. Similarly, trade-offs should occur between dialogue history and shared visual space in grounding. In particular, because dialogue history serves as a reminder of aspects of a conversation that interlocutors have been previously exposed to, it should be valuable as a resource for grounding primarily when other sources are weak (i.e., in the absence of a shared visual space).

METHOD

Overview

We conducted an experiment to examine the influence of dialogue history and shared visual space for conversational and task coordination. Participant pairs played the role of Helper and Worker in the puzzle game introduced by Kraut et al. [17]. The Helper directed the Worker on how to complete a simple jigsaw puzzle so it matched a target. They communicated by text chat. The text chat showed either one speaking turn or six. The Helper could either see what the Worker was doing in real time or could not see the Worker. The pieces were either easy or difficult to describe.

Independent Variables

Chat Persistence: We varied the size of the dialogue history available to the pairs. The *Persistence* condition allowed 12 lines of visible history (or approx. six conversational turns) (Figure 1a.). However, in the *No Persistence* condition, the chat interface only showed 2 lines of history (or approx. one conversational turn) at any given time (Figure 1b.). In both interfaces, multi-line entries could be created using a 'Shift-Enter' keystroke, and messages could be sent either by pressing the 'Enter' key or by clicking on or tabbing to the 'Send' button.

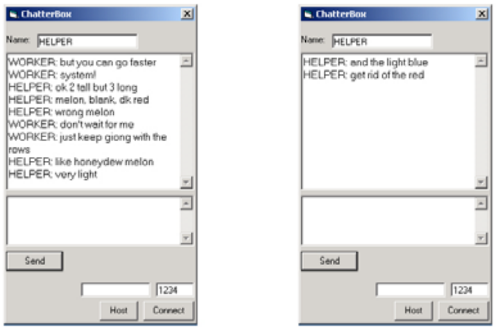

Figure 1. Text-chat application with a.) Persistence (left) and b.) No Persistence (right).

Shared Visual Space: We manipulated whether or not the participants viewed the same work area. The displays were programmed as shared Visual Basic applications. In the Shared Visual Space condition (*SVS*), Helpers could see the Worker's work area in real-time, while in the No Shared Visual Space (*No SVS*) condition they could not.

The basic structure of the Worker and Helper displays can be seen in Figure 2. The Worker's display (Figure 2a.) contained a staging area on the right where eight pieces for the puzzle were stored and a work area on the left where the Worker constructed a four-piece solution. The Helper's display (Figure 2b.) contained the target puzzle on the right, holding the goal state. The left-hand side of the Helper's display either contained the view of the Worker's work area (*SVS* condition) or was black (*No SVS* condition).

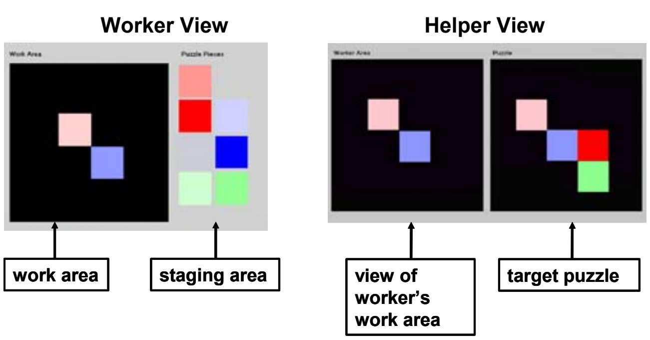

Figure 2a). Worker (left) and 2b). Helper (right) displays.

Lexical Complexity: We varied the ease with which lexical tokens could be generated and used to describe the pieces by changing whether the blocks were static and easy to describe solid colors (e.g., red) or plaids that were difficult to describe and required more effort for grounding. The pieces were chosen randomly for each experimental condition from a palette of easy to describe *Solids* or hard to describe *Plaids*.

Participants and Procedure

Sixteen pairs of professionals and students from the Cambridge, MA area participated in the research. They were selected to have typing skills and significant prior use of chat or instant messaging software. They were paid for their involvement and the average group took approximately 1 hour to complete the study. The participants were randomly assigned to play the role of Helper or Worker, and the groups were gender balanced across conditions. The design was a 2 × 2 × 2 mixed model design where Shared View Space (*SVS* vs. *No SVS*) was a between-subjects factor, while Chat Persistence (*Persistence* vs. *No Persistence*) and Lexical Complexity (*Solid* vs. *Plaid*) were within-subjects manipulations. Eight pairs participated in the *SVS* condition and eight in the *No SVS* condition, and the levels of Chat Persistence and Task Complexity were counter-balanced across trials. Pairs solved four puzzles within each experimental condition for a total of 16 puzzles.

Apparatus

The Helper and Worker were each seated in front of identical desktop computers with 18-inch IBM ThinkVision L170p LCD displays driven by nVidia GeForce 2 MX200 video cards. Since the pairs were often discussing subtle color differences, great care was taken to calibrate the displays.

Measures

Task performance

The pairs were instructed to complete the task as quickly as possible. We used time to complete a puzzle as the primary measure of task performance. Because almost all puzzles were solved correctly, error rates were a less useful indicator of task performance.

Process Measures

In order to better understand how the pairs performed in different conditions, we explored several features of conversational structure. The first part of this analysis looked primarily at the conversational efficiency of the communication. The length of utterances, total number of words, and other quantitative measures examined the coarse communication differences between conditions. The second part of these data explored the conversational structure in the various communication conditions. We performed a qualitative exploration of the ways in which the form of the conversation changed across the different features of media.

Statistical Analysis

We used a mixed model analysis of variance to predict puzzle completion time. Chat Persistence (*Persistence, No Persistence*), Lexical Complexity (*Solid, Plaid*), Block (1-4), and Trial (1-4) were repeated factors, and Shared Visual Space (*SVS, No SVS*) was a between-pair factor. We included all 2-way and 3-way interactions in the analysis. Because each pair participated in 16 trials, observations within a pair were not independent of one another. Pairs, nested within Shared Visual Space, were modeled as a random effect. Although these analyses were full-factorial analyses of covariance, for reasons of space, we focus primarily on the influence of the experimental manipulations.

RESULTS

Task Performance

Lexical complexity. The manipulation of Lexical Complexity had a large impact on the speed with which the pairs could solve the puzzles. Pairs were substantially faster in the trials in which the puzzle pieces were Solids than when they were Plaids (104.7 sec vs. 184.6 sec; $F(1,226)=221.24$, $p<.001$).

Shared visual space. Consistent with *H2*, a Shared Visual Space had a very large impact on the speed with which the pairs could solve the puzzles. The pairs were over twice as fast when they had the shared view than when they did not (92.3 sec vs. 197.0 sec; $F(1,14)=53.65$, $p<.001$).

Chat persistence. Consistent with *H1*, dialogue history had a small but reliable impact on performance. Pairs were faster when they had the longer dialogue history available than when they had the last two utterances available (136.5 sec vs. 152.8 sec; $F(1,226)=9.19$, $p=.003$). Consistent with *H3*, the effect of dialogue history on performance was substantially smaller than the effect of the shared visual space.

Interaction between media characteristics and lexical complexity. Consistent with *H4*, the impact of both shared visual space and chat persistence on performance was larger when the verbal task was more complex. The Shared Visual Space × Lexical Complexity interaction showed that the shared visual space was more useful when the blocks were difficult to describe Plaids rather than easy to describe Solids (for the interaction $F(1,226)=22.56$, $p<.001$). Similarly, the Chat Persistence × Lexical Complexity interaction (see Figure 3; for the interaction $F(1,226)=11.55$, $p<.001$) showed that the longer dialogue history only improved performance for the difficult to describe plaids (Persistence: 167.3 sec vs. No Persistence: 201.8 sec; $F(1,226)=20.67$, $p<.001$) and not for the easy to describe solids (Persistence: 105.7 sec vs. No Persistence: 103.7 sec; $F(1,226)=.07$, ns). Together these results suggest that features of the communication channel that improve grounding have the greatest impact for lexically complex tasks.

Interactions among communication media. Features in the communication media that improve grounding compensate for each other. Consistent with *H5*, the Shared Visual Space × Chat Persistence interaction shows that the larger chat history improved performance only when the pair had no shared visual space (see Figure 4; for the interaction $F(1,226)=12.34$, $p<.001$). There was no reliable difference between persistence conditions when the Helper could see

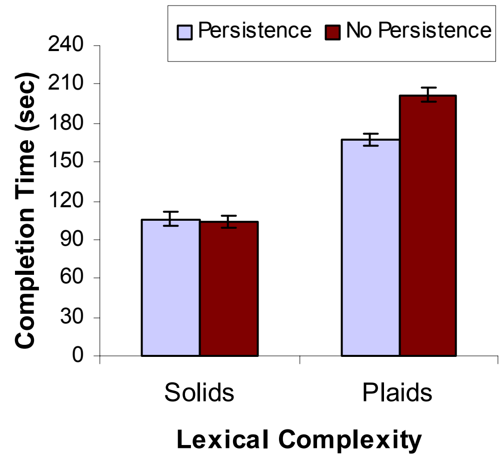

Figure 3. Completion Time by Chat Persistence and Lexical Complexity.

the work area (Persistence: 93.6 sec vs. No Persistence: 91.0 sec; $F(1,226)=.12$, ns), but there was when the Helper could not see the area (Persistence: 179.4 sec vs. No Persistence: 214.5 sec; $F(1,226)=21.41$, $p<.001$).

The 3-way interaction for Lexical Complexity × Shared Visual Space × Chat Persistence was not significant (for the interaction, $F(1,226)=2.68$, $p=.11$).

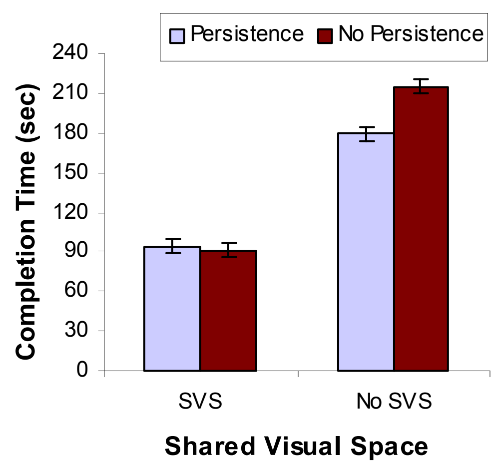

Figure 4. Completion Time by Chat Persistence and Shared Visual Space.

Conversational Efficiency and Qualitative Descriptions

While these data tell us about the speed with which the pairs performed the puzzle task in various media conditions, they tell us little about the way the media changed conversational efficiency and structure. Table 1 presents data on the conversational efficiency. It shows the number of utterances and words per puzzle and number of words per utterance when pairs had a shared visual space or not and when they had a large or small dialogue history.

Table 1. Communication efficiency (p<.01, +p=.11)**

	Shared Visual Space			Chat Persistence		
	SVS	No SVS	P	6-Turn	1-Turn	P
Utterances	10.8	18.9	**	16	13.6	**
Total # of Words	51.8	121.4	**	85.3	87.9	
Wds/Utterance	5.3	6.9	+	5.7	6.49	**

Effect of shared visual space on efficiency and structure. When the pairs *did not* have a shared visual space, they required almost 2.5 times as many words to complete the puzzle (No SVS: 121.4 vs. SVS: 51.8 words per puzzle; $F(1,14)=63.07$, $p<.01$). They also nearly doubled the number of utterances used (No SVS: 18.9 vs. SVS: 10.8 utterances per puzzle; $F(1,14)=11.98$, $p<.01$). While this led to more words per utterance—or larger messages when there was no shared space—the difference was not quite significant (No SVS: 6.9 vs. SVS: 5.3 words per utterance; $F(1,14)=2.90$, $p=.11$).

Why could pairs complete their puzzles with fewer utterance and words when they had a shared visual space? With it, Helpers could infer Workers' comprehension directly from their actions. This allowed them to provide shorter, more incremental descriptions and to cut short descriptions as soon as the Workers gave behavioral evidence of understanding.

Shared Visual Space

(1.1) **H:** okay all blue with 2 vertical darker stripes
H: mostly grey with l bottom red
W: *[moves piece in]*
H: S of blue
W: *[positions piece]*
H: grey with one horizontal white stripe
W: *[moves piece in]*
H: NW of blue
W: *[positions piece]*

No Shared Visual Space

H: the first block we ned has one white stripe at the very top.. with a thinner yelow stripe about a cm below it
W: *[moves piece in]*
W: ok ,got it
H: the other one we need has a thin yellow stripe, then a thick white one 1 cm below, then another thick white one ~3 cm below, then another thin yellow stripe 1 cm below
W: ok. does that have a faint blue stripe in the center?
H: it should be divided down the center.. with a plaid light blue diagonal stripes on the left half, and dark blue solid on the right half
W: *[moves piece in]*
H: do you need more clarification
W: no, ive got the second peice

Excerpt (1.1) demonstrates how the pairs used behavioral action to ground their utterances. On the left is a case where the pair had a shared visual space. The Helper provided the next piece of instruction after noticing that the Worker had moved the correct piece. The Worker gave no verbal acknowledgements. In contrast, on the right, where there was no Shared Visual Space available, the Worker needed to be more explicit about her actions and understanding.

Another thing to notice is when the pairs had a shared view, they produced incremental contributions. The Helper hypothesized the minimal information required and provided it. The shared view space provided evidence about whether or not the Worker understood. If she understood, the Helper simply gave the next directive; however, if she did not, the Helper provided additional details. When the pairs did not have a shared space, they were much more detailed when first identifying a piece.

The shared view also allowed pairs to precisely time their utterances and cut short descriptions as soon as they saw the last utterance was understood. In excerpt (1.2), the Helper describes a piece in short installments. As soon as she received visual confirmation that the Worker had the correct piece, she censored the next description mid-sentence and replaced it with "yah that's right":

(1.2) **H:** *[types]* it has a cross on it sort of centered around the bottom left
H: *[types]* and some green on the right
W: *[moves correct piece into work area]*
H: *[erases prior statement. types]* yah thats right

Figure 5 illustrates that when they had a shared visual space, pairs more tightly integrated text and action, replaced written utterances with actions, and cut short and redefined statements. The black vertical lines indicate the instant a message was sent. The colors preceding the black vertical lines represent the person who issued the utterance (Helper is light; Worker is dark). The colored horizontal bars indicate the turn gap (i.e., the time between utterances), which varies with the time spent formulating and typing the message, the overall size of the message, waiting for the precise time to send the message, and the recency with which a partner has messaged. The top panel in Figure 5 shows the timing in a typical shared visual space trial, while the bottom panel shows the timing in a typical trial without a shared visual space.

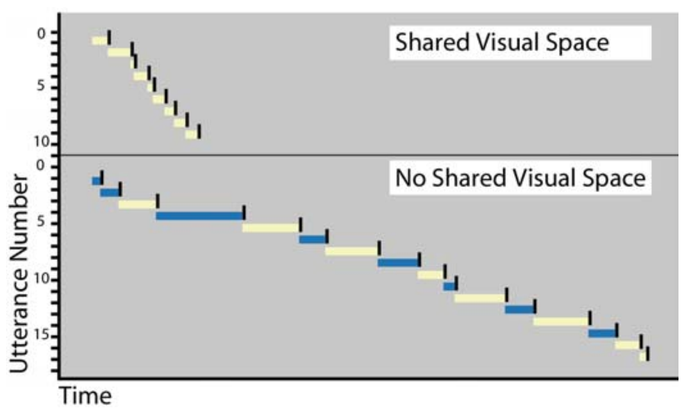

Figure 5. Differences in communication by Helper (light) and Worker (dark) over time for the SVS and No SVS conditions.

It is striking to note the rhythm with which messages were sent when the pairs had shared visual space (top panel). Rather than wait for the Worker to respond, the Helper issued the next statement while the Worker executed the last. The shared visual space provides Helpers with an indication of where the Worker is in the task and allows them to generate their next message ahead of time so it is available as soon as the Worker has selected or positioned the correct piece. When there is no shared space (bottom pane), the rhythm slowed down for two reasons. First, the Helper became more explicit with directions (seen in the aggregate data as larger messages), and secondly, the Worker became explicit in declaring their understanding (seen as increased contributions of the Worker).

Effect of persistence on efficiency and structure. When the pairs had a dialogue history, they generated more utterances (16 vs. 13.6 utterances per puzzle; $F(1,226)=12.84$, $p<.01$), with fewer words per utterance (5.7 vs. 6.5 words per utterance; $F(1,226)=10.68$, $p<.01$) although they used approximately the same number of words overall to solve the puzzles (85.3 vs. 87.9 words per puzzle; $F(1,226)=0.54$, ns). This demonstrated that pairs changed the form of their communication to adjust for the media—they produced smaller messages and sent them more frequently. Doing so allowed actions and text generation to be parallelized, aided memory and cognitive processing, and improved the precision in timing utterances.

Figure 6 illustrates how a pair used persistence to parallelize their efforts. The Helper queued their descriptions of pieces in order to maximize group efficiency. The rectangles on the Helper row represent the time spent describing a puzzle piece. The second row shows the time spent reading / comprehending the message (light gray) as well as time spent moving the pieces into the space and acknowledging that actions have been completed (dark gray). When they had no persistence, the Helper typically waited for each message to be read and acted upon by the Worker before issuing the next statement. This was much less efficient than overlapping descriptions and actions.

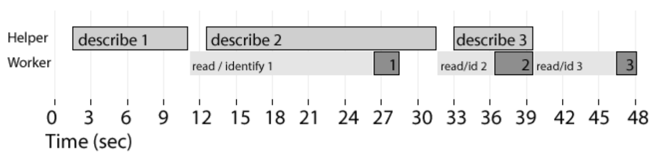

Figure 6. Timeline display of Helper utterances and Worker actions in the Persistence condition.

The dialogue history also helped overcome practical memory limitations. When the pairs had a shorter dialogue history, the Workers frequently needed to re-request information because they forgot the details of a prior utterance or because of an intervening clarification. In excerpt (1.3), the Worker asks the Helper to repeat a placement-related instruction, since it was no longer available in their shared dialogue history:

(1.3) **H:** above that there is a piece where the top lft corner is gray, the top right and btm left are lighter grey, and the bottome right is like purple with a black square in it
W: purple? are you sure it's not brown?
H: yeah, you know what its brown
W: alright, where's it go again?
H: above the center

If a dialogue history had been available, the Worker could have glanced back and found the placement information.

Another strategy commonly seen when persistence was available was the production of shorter sequential contributions. This allowed pairs to cut short surplus descriptions. Notice in excerpt (1.4), the Worker provides a hypothesis about a piece they think is the correct piece. This strategy keeps the pairs from typing more than needed.

(1.4) **H:** fourth has a light blue cross
H: with green on the top
H: and black on the bottom
W: with dark right next to the light blue
H: yes
W: ok, where

Overall, utterances were more tightly integrated on trials in which the pairs had the persistence. This pattern can be seen in Figure 7, which has a steeper slope in the upper panel (a trial with persistence) than in the lower (a trial without persistence).

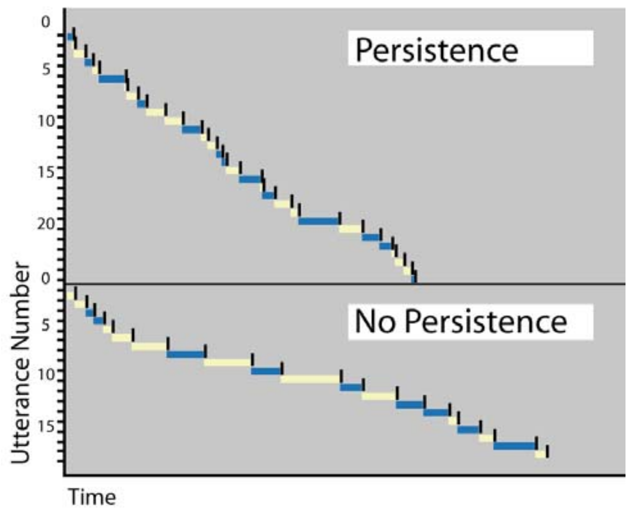

Figure 7. Differences in communication by Helper (light) and Worker (dark) over time for the Chat Persistence conditions.

DISCUSSION

Our findings demonstrate the importance of two resources for conversational grounding—a persistent dialogue history and a shared visual space—and show how these resources interact with task properties to affect conversational efficiency and performance. Persistence and shared visual space were especially helpful for pairs when they were discussing lexically complex material. The results further demonstrate that the value of persistence is greater when pairs do not share visual space, suggesting that for tasks with visual elements, shared visual space is the most efficient grounding mechanism.

The study illustrates the value of basing the design of new media on previously existing communication theory. Clark's theory of conversational grounding [3] was used to generate testable predictions about the effects of persistence and shared visual space on communication and task performance, and these predictions were confirmed in our experiment. Similar predictions about how media properties will shape interaction can be generated in advance for a wide range of communications technologies, potentially shortening the design-evaluate cycle considerably.

The findings have important implications for the design of new media to support collaboration on visual tasks. For such tasks, providing a shared view of the work environment can be highly beneficial. At the same time, we have shown that systems to support collaboration on visual tasks need to reserve sufficient screen space for dialogue history. Thus, designers need to carefully consider the trade-offs between a task's visual and verbal requirements in order to determine the best strategy for allocating screen space. Furthermore, our results suggest that using part of the dialogue box for icons and images, as is done in iChat [15], may have a negative effect on some tasks. When making the conscious decision to limit dialogue history, designers must remain acutely aware of the task and environment into which such designs are placed.

Although we have provided strong evidence concerning the value of persistence and shared visual space, future work is needed to clarify how these grounding mechanisms interact with task properties. First, we studied only two variants of persistence. It is possible that a longer dialogue history would prove even more valuable to collaborators. Second, we studied only one way of integrating the dialogue box with shared visual space. The design space for integrating these elements requires further examination. Third, we studied only two types of visual stimuli; future studies will need to examine how other aspects of the task (e.g., size and shape of puzzle pieces, timing of instructions) influence collaborators' needs for dialogue history and shared visual space as well as the trade-offs between them.

ACKNOWLEDGEMENTS

This work was supported by National Science Foundation Grant No. IIS-9980013. The first author is supported by an IBM PhD Fellowship. We would like to thank Josef Scherpa and Matthew Hockenberry for their contributions.

REFERENCES

1. Brennan, S.E., & Lockridge, C. B. (in preparation). Effects of visual co-presence on referring in conversation.
2. Cherny, Lynn (1999). *Conversation and Community: Chat in a virtual world*. Stanford, CA: CSLI Publications.
3. Clark, H.H. (1996). *Using Language*. Cambridge Univ. 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. Clark, H.H., & Schaefer, E.F. (1989). Contributing to discourse, *Cognitive Science 13*, 259-294.
6. Clark, H.H., & Marshall, C.R. (1978). Reference diaries. In D.L. Waltz (Eds.) *Theoretical Issues in Natural Language Processing – 2*, 57-63. NY: ACM Press.
7. Condon, S., & Čech, C. (2001). Profiling turns in interaction: Discourse structure and function. In *Proceedings of the 34th Hawai'i International Conference on Systems Sciences, (4)*. IEEE Computer Society Press.
8. Daly-Jones, O., Monk, A. & Watts, L. (1998). Some advantages of video conferencing over high-quality audio conferencing: fluency and awareness of attentional focus. *International Journal of Human-Computer Studies, 49*, 21-58.
9. 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". In *Proceedings of the ACM Conference on Human Factors in Computing Systems* (pp.72-76). NY: ACM Press.
10. Fussell, S.R., Kraut, R.E., & Siegel, J. (2000). Coordination of communication: Effects of shared visual context on collaborative work. *Proceedings of CSCW 2000*, 21-30. NY: ACM Press.
11. Gutwin, C., & Greenberg, S. (2002). A descriptive framework of workspace awareness for real-time groupware. *Computer Supported Cooperative Work*, Kluwer Academic Press.
12. Hancock, J.T., & Dunham, P.J. (2001). Language use in computer-mediated communication: The role of coordination devices. *Discourse Processes, 31(1)*, 91-110.
13. Handel, M., & Herbsleb, J.D. (2002). What is chat doing in the workplace? In *Proceedings of CSCW 2002*, 1-10. NY: ACM Press.
14. Herring, S. (1999). Interactional coherence in CMC. In *Proceedings of the 32nd Hawai'i International Conference on System Sciences*. IEEE Computer Society Press.
15. iChat AV. Apple Computer, Inc. Website (Sept., 25, 2003): <http://www.apple.com/ichat/>
16. Isaacs, E., Walendowski, A., Whittaker, S., Schiano, D. J., & Kamm, C. (2002). The character, functions, and styles of instant messaging in the workplace. *Proceedings of CSCW 2002*, 11-20. NY: ACM Press.
17. Kraut, R.E., Gergle, D., & Fussell, S.R. (2002). The use of visual information in shared visual spaces: Informing the development of virtual co-presence. In *Proceedings of CSCW 2002*, 31-40. NY: ACM Press.
18. McCarthy, J., & Monk, A. (1994). Measuring the quality of computer-mediated communication. *Behavior & Information Technology, 13(5)*, 311-319.
19. Nardi, B. A., Whittaker, S., & Bradner, E. (2000). Interaction and outeraction: Instant messaging in action. *Proceedings of CSCW 2000* (pp. 79-88). NY: ACM Press.
20. Olson, G.M., Atkins, D.E., Clauer, R., Finholt, T.A., Jahanian, F., Killeen, T.L., Prakash, A., & Weymouth, T. (1998). The upper atmospheric research collaboratory. *Interactions, May+June,* pp.48-55.
21. Rafaeli, S. (1988). Interactivity: From new media to communication. In R. Pawkins, J. Wiemann, and S. Pingree, Eds., *Advancing communications science: Merging mass and interpersonal processes*, (16), 110-133. Sage Annual Reviews of Communication Research. Beverly Hills, CA: Sage.
22. Viegas, F., & Donath, J. (1999). Chat circles. In *Proceedings of the ACM Conference on Human Factors in Computing Systems* (pp. 9-16). NY: ACM Press.