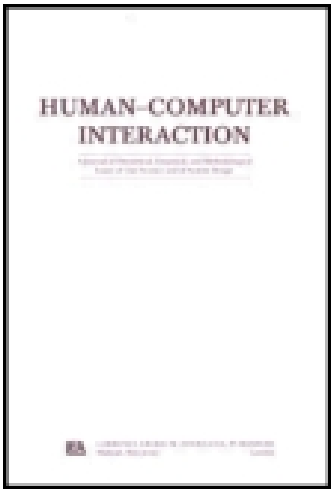


Human-Computer Interaction
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/hhci20

Using Visual Information for Grounding and Awareness in Collaborative Tasks

Darren Gergle [a], Robert E. Kraut [b] & Susan R. Fussell [c]
[a] Northwestern University
[b] Carnegie Mellon University
[c] Cornell University
Accepted author version posted online: 27 Mar 2012.Published online: 12 Dec 2012.

To cite this article: Darren Gergle , Robert E. Kraut & Susan R. Fussell (2013) Using Visual Information for Grounding and Awareness in Collaborative Tasks, Human-Computer Interaction, 28:1, 1-39

To link to this article: http://dx.doi.org/10.1080/07370024.2012.678246

HUMAN–COMPUTER INTERACTION, 2013, Volume 28, pp. 1–39

ISSN: 0737-0024 print / 1532-7051 online
DOI: 10.1080/07370024.2012.678246

Taylor & Francis
Taylor & Francis Group

Using Visual Information for Grounding and Awareness in Collaborative Tasks

Darren Gergle,[1] Robert E. Kraut,[2] and Susan R. Fussell[3]
[1]Northwestern University
[2]Carnegie Mellon University
[3]Cornell University

When pairs work together on a physical task, seeing a common workspace facilitates communication and benefits performance. When mediating such activities, however, the choice of technology can transform the visual information in ways that impact critical coordination processes. In this article we examine two coordination processes that are impacted by visual information—situation awareness and conversational grounding—which are theoretically distinct but often confounded in empirical research. We present three empirical studies that demonstrate how shared visual information supports collaboration through these two distinct routes. We also address how particular features of visual information interact with features of the task to influence situation awareness and conversational grounding, and further demonstrate how these features affect conversation and coordination. Experiment 1 manipulates the immediacy of the visual information and shows that immediate visual feedback facilitates collaboration by improving both situation awareness and conversational grounding. In Experiment 2, by misaligning the perspective through which the Worker and Helper see the work area we disrupt the ability of visual feedback to support conversational grounding but not situation awareness. The findings demonstrate that visual information supports the central mechanism of conversational grounding. Experiment 3 disrupts the ability of visual feedback to support situation awareness by reducing the size of the common

Darren Gergle is an associate professor in the Department of Communication Studies at Northwestern University and the Department of Electrical Engineering and Computer Science at Northwestern University; his research interests include small-group communication in face-to-face and mediated environments and the design and study of tools to support distributed collaboration. **Robert E. Kraut** is the Herbert A. Simon Professor of Human–Computer Interaction and Social Psychology at Carnegie Mellon University; he conducts research on the design and impact of computer-mediated communication systems. **Susan R. Fussell** is an associate professor in the Department of Communication and the Department of Information Science at Cornell University; her research interests include interpersonal communication in face-to-face and computer-mediated contexts, online communities, and the dynamics of collaboration in work teams and organizations.

CONTENTS

1. INTRODUCTION
2. THE ROLE OF VISUAL INFORMATION IN SUPPORTING COLLABORATION
   2.1. Situation Awareness
   2.2. Conversational Grounding
   2.3. The Impact of Technological Mediation on the Availability of Visual Information
   2.4. Overview of Experiments
3. EXPERIMENT 1: VISUAL DELAY STUDY
   3.1. Method
      The Puzzle Task
      Participants and Procedure
      Measures and Statistical Analysis
   3.2. Results and Discussion
      Task Performance
      Communication Processes
4. EXPERIMENT 2: ROTATION STUDY
   4.1. Method
      Experimental Manipulations
      Statistical Analysis
   4.2. Results and Discussion
      Task Performance
      Communication Processes
5. EXPERIMENT 3: FIELD OF VIEW STUDY
   5.1. Method
      Experimental Manipulations
      Statistical Analysis
   5.2. Results and Discussion
      Task Performance
      Communication Processes
6. GENERAL DISCUSSION
   6.1. Theoretical Implications
   6.2. Practical Design Implications
   6.3. Limitations and Future Directions
7. CONCLUSION

viewing area. The findings suggest that visual information independently supports both situation awareness and conversational grounding. We conclude with a general discussion of the results and their implications for theory development and the future design of collaborative technologies.

1. INTRODUCTION

Recent structural changes in organizations, such as the rise of large multinational corporations, together with technological advances, such as the widespread availability

of the Internet, have contributed to increases in distributed work practices mediated by telecommunication technologies. However, distributed work is generally more difficult and less successful than comparable work in collocated settings (Olson & Olson, 2000; Whittaker, 2003). Part of this problem stems from developing collaboration tools and work practices without a thorough understanding of the ways groups coordinate their activities in collocated environments. In this article, we focus on one aspect of collocated collaboration: the use of shared visual information to provide communicative cues. We consider two major purposes that this visual information can serve in collaborative work—awareness and communication—and examine how the particular nature of the visual information and the requirements of the task affect communication processes and ultimately influence collaborative performance.

Many researchers hypothesize that visual information plays a central role in coordinating collaborative work. Although early research posited that seeing other people's faces during conversation was critical for successful coordination (e.g., Daft & Lengel, 1986; Short, Williams, & Christie, 1976), many empirical studies failed to support this claim (see Nardi & Whittaker, 2002; Williams, 1977, for reviews). More recently, researchers have noted the importance of dynamic visual information about the objects and activities in a work environment (Kraut, Fussell, & Siegel, 2003; Monk & Watts, 2000; Nardi et al., 1993; Whittaker, Geelhoed, & Robinson, 1993; Whittaker & O'Conaill, 1997). This approach has identified a range of conditions under which visual information is valuable. For example, viewing a partner's actions facilitates monitoring of comprehension and enables efficient object reference (Daly-Jones, Monk, & Watts, 1998), providing visual information about the workspace improves information gathering and recovery from ambiguous help requests (Karsenty, 1999), and varying the field of view of a coworker's environment influences performance and shapes communication patterns in partner directed physical tasks (Fussell, Setlock, & Kraut, 2003).

In previous research we proposed a decompositional framework for understanding how visual information affects collaboration (Kraut et al., 2003). We suggested that the degree to which visual information will improve performance in any particular situation depends both on technological choices and the task the group is performing. Technological choices influence the amount, quality, and utility of the visual information exchanged. For example, instructors provide better guidance on a robot construction task when using a scene-oriented camera with a wide-angle view of the work area than when using a head-mounted camera that shows a narrow, dynamic view of the work area (Fussell et al., 2003). Task features also influence whether visual information improves performance (Whittaker et al., 1993). For example, visual feedback helps collaborators more when they are working with objects that are difficult to describe than when they are working with objects that are easy to describe (Gergle, Kraut, & Fussell, 2004b; Kraut, Gergle, & Fussell, 2002). Similarly, the value of shared visual space, and whether it should be tightly or loosely coupled to a partner's actions, can be dependent upon the stage of the task (Ranjan, Birnholtz, & Balakrishnan, 2007). These studies, and others like them (Clark & Krych, 2004; Velichkovsky, 1995), demonstrate the need for a more nuanced theoretical understanding of the

precise functions that visible information serves in collaboration (see Whittaker, 2003; Whittaker & O'Conaill, 1997, for discussions).

Our approach is based on two psychological theories that help explain the role of visual information in collaborative work. First, situation awareness theory[1] (Endsley, 1995; Endsley & Garland, 2000) holds that visual information helps pairs assess the current state of the task and plan future actions. For example, a teacher watching over a student's shoulder might intervene to provide timely instructions because she can see from the calculations that the student has completed the first stage of the problem and is ready to receive instructions for the second stage. In this case, the visual feedback supports the instructor by providing evidence regarding the state of the joint activity. The second theoretical framework is grounding theory (Clark, 1996; Clark & Marshall, 1981; Clark & Wilkes-Gibbs, 1986), which maintains that visual information can support the conversation surrounding a joint activity by providing evidence of common ground or mutual understanding. For example, a teacher may clarify her verbal instruction after seeing the student's calculations because she notices that the student misunderstood something she said. In this case the visual information provides evidence of mutual understanding (or lack thereof), and the instructor can use it to determine when she needs to refine her statements to provide further clarification. Together these theories predict that when groups have access to the correct visual information they will coordinate their work better because they can monitor the state of the task, deliver instructions and clarifications in a more timely fashion, and refer to objects and actions more efficiently.

Although visual information is thought to influence both situation awareness and conversational grounding, most empirical research has failed to distinguish these conceptually distinct coordination mechanisms. Doing so is important for developing theoretical models of the role visual information plays in collaborative work and for building systems that provide the right type of visual information for the task at hand. We address these theoretical lacunae by observing pairs performing a tightly coupled collaborative task in the laboratory where we can rigorously control both features of the visual environment and features of the task. In addition to the theoretical contributions of this work, we also aim to reveal how different features of a visual communication technology can affect conversation and coordination. In our paradigm, participants jointly complete visual block puzzles, in which one person, the "Helper," directs a partner, the "Worker," on how to arrange the blocks on a computer screen. We report the results of three experiments that use this paradigm to disentangle the independent effects of situation awareness and conversational grounding. Experiment 1 manipulates the immediacy of the visual information and shows that immediate visual feedback facilitates collaboration by improving both situation awareness and conversational grounding. Experiment 2 disrupts the ability of visual feedback to support conversational grounding by misaligning the perspective

[1]Situation awareness serves as the underlying theoretical construct of interest. However, a related construct is Gutwin and Greenberg's (2002) notion of workspace awareness which can be considered a "specialization of situation awareness, one that is tied to the specific setting of the shared workspace" (p. 417).

through which the Worker and Helper see the work area. The findings suggest that conversational grounding is a central mechanism supported by visual information. Experiment 3 disrupts the ability of visual feedback to support situation awareness by reducing the size of the common viewing area. Together these findings suggest that visual information independently supports situation awareness as well as conversational grounding.

2. THE ROLE OF VISUAL INFORMATION IN SUPPORTING COLLABORATION

In this section we present a brief overview of situation awareness theory (Endsley, 1995) and grounding theory (Clark & Marshall, 1981; Clark & Wilkes-Gibbs, 1986), focusing on the ways that visual information improves collaborative performance via these mechanisms.

2.1. Situation Awareness

According to situation awareness theory, visual information improves coordination by giving actors an accurate view of the task state and one another's activities. This awareness allows accurate planning of future actions (Endsley, 1995). For example, Nardi and colleagues (1993) described how a scrub nurse on a surgical team uses visual information about task state to anticipate the instruments a surgeon will need. If the nurse notices that the surgeon has nicked an artery during surgery, she can prepare cauterization and suture materials and have them ready before the surgeon asks for them.

It is important to note that the visual information does not need to be identical for all group members for it to support situation awareness, as long as it allows them to form an accurate view of the current situation and appropriately plan future actions (Bolstad & Endsley, 1999). For example, two fighter pilots can converge on a target aircraft, even if one of them uses the visual line of sight and the other uses radar to "see" the target. However, if the differing displays lead them to form different situational representations, then their performance is likely to suffer. For example, if visual sighting allows one pilot to distinguish between friendly and enemy aircraft but the radar fails to support this discrimination for the other pilot, then the two fighters are unlikely to coordinate their attack purely on the basis of the situation awareness provided by the visual information (Snook, 2000).

2.2. Conversational Grounding

According to grounding theory, visual information improves coordination by supporting the verbal communication surrounding a collaborative activity. Grounding

theory states that successful communication relies on a foundation of mutual knowledge or common ground. Conversational grounding is the process of establishing common ground. Speakers form utterances based on an expectation of what a listener is likely to know and then monitor that the utterance was understood, whereas listeners have a responsibility to demonstrate their level of understanding (Brennan, 2005; Clark & Marshall, 1981; Clark & Wilkes-Gibbs, 1986). Throughout a conversation, participants are continually assessing what other participants know and using this knowledge to help formulate subsequent contributions (Brennan, 2005; Clark & Marshall, 1981; Clark & Wilkes-Gibbs, 1986).

Clark and Marshall (1981) proposed three major factors that allow speakers to anticipate what a partner knows: *community co-membership*, *linguistic co-presence*, and *physical co-presence*. Because of community co-membership, members of a professional group, for example, can use technical jargon with each other that they could not use with outsiders. Because of linguistic co-presence, one party in a conversation can safely use a pronoun to refer to a person previously mentioned in the conversation. Because of physical co-presence, one person can point to an object in their shared physical environment and refer to it using the deictic pronoun "that" if she believes the other is also privy to the object and her gesture.

Shared visual information helps communicators to establish common ground by providing evidence from which to infer another's level of understanding. This evidence can be given off deliberately (e.g., as in a pointing gesture) or as a side effect of proper performance of the desired action, provided both parties are aware of what one another can see. For example, when responding to an instruction, performing the correct action without any verbal communication provides an indication of understanding, whereas performing the wrong action, or even failing to act, can signal misunderstanding (Gergle, Kraut, & Fussell, 2004a).

Shared visual information can support conversational grounding at two distinct phases of the communication process: the *planning stage* and the *acceptance stage* (Clark & Schaefer, 1989). During the planning stage, in which speakers formulate their utterances (Levelt, 1989), visual information provides cues to what a listener is likely to understand. In our puzzle paradigm, Helpers need to refer to puzzle pieces so that Workers can identify them easily. If Helpers can see the work area and are aware that the Worker can also see it, they can make use of the mutually available visual information during reference construction. For example, when describing a plaid piece they can use an efficient expression such as "the one on the left" rather than a lengthier description of the patterns on a particular piece (e.g., "the piece with the red stripe on the left and the white box in the lower right corner with a green stripe running through it"). Similarly, they can reduce verbal ambiguity by using the phrase "the *dark* red one" when both dark and light red pieces are mutually visible.

During the acceptance stage, speakers and hearers mutually establish that they have understood the utterance well enough for current purposes (Clark & Wilkes-Gibbs, 1986). In our puzzle paradigm, Helpers can use visual feedback from the Worker's performance to monitor whether the Worker has understood the instructions. This visual feedback is efficient because with it the Worker does not need to

explicitly state his or her understanding (see, e.g., Doherty-Sneddon et al., 1997; Gergle et al., 2004a). It is also less ambiguous than verbal feedback. Clark and Krych (2004) demonstrated that when shared visual information was available, pairs spent approximately 15% less time checking for comprehension (see also Doherty-Sneddon et al., 1997). We show that not only does the availability of the shared visual information matter, but the particular form of the shared visual information can have a differential influence on coordination and communication.

In addition to the communication stages in which grounding occurs, Clark and Brennan (1991), and more recently Convertino and colleagues (Convertino, et al., 2008; Convertino, Mentis, Rosson, Slavkovic, & Carroll, 2009), have highlighted two primary forms of common ground that are important for collaborative work. The first is grounding and coordination on the content or subject of the work. The second is grounding and coordination on the processes and procedures that underlie coordinated interaction. The latter focuses on developing a shared understanding of how things are done and on understanding the "rules, procedures, timing and manner in which the interaction will be conducted" (Convertino et al., 2008, p. 1638). It is important to note that this form of grounding is related to and relies upon, but is ultimately distinct from, situation awareness. It focuses on the joint "knowing how," yet these actions require an accurate awareness of the task state.

In most real-world settings, visual feedback provides evidence of both the current state of a task and of a listener's degree of comprehension. As a result it is often difficult to empirically distinguish the routes through which visual information improves collaborative performance. The experiments reported next are designed to demonstrate that visual information improves performance on collaborative tasks by independently supporting both situation awareness and conversational grounding.

2.3. The Impact of Technological Mediation on the Availability of Visual Information

Although visual information can generally improve collaborative task performance via situation awareness and conversational grounding, the benefit provided in any particular situation will likely depend on the technology used and the characteristics of the collaborative task. For designers and engineers creating technologies to provide visual information through telecommunications, the goal is often to make a collaborative environment as similar as possible to the gold standard of physical collocation (for a critique, see Hollan & Stornetta, 1992). In attempting to achieve this goal, however, they must trade off features that affect the utility of the visual information, such as the field of view and who controls it, delays, alignment of perspective, degree of spatial resolution, frame rate, and the level of synchronization with a voice stream. These different features of the communication media change the costs of grounding and achieving situation awareness (Clark & Brennan, 1991; Kraut, Fussell, Brennan, & Siegel, 2002), but how do we know which of these features need to be reproduced in order to provide the benefits of a collocated environment?

Our digital puzzle paradigm provides a method for decomposing the visual space in order to better understand the impact various features of the visual space have on collaborative performance. We examine the impact of particular visual features such as delay, perspective, field of view, and view control, in addition to distinguishing between the coordination mechanisms of situation awareness and conversational grounding.

2.4. Overview of Experiments

We present a series of three experiments intended to disentangle the effects of visual information on conversational grounding and situation awareness. As shown in Figure 1, the experiments manipulate different features of the visual environment.

Experiment 1 manipulates the temporal immediacy of the visual information, with the Helper seeing the Worker's work area immediately, after a delay, or not at all. The results are consistent with the hypothesis that immediate visual feedback helps collaborative performance by improving both situation awareness and conversational grounding. However, this manipulation does not distinguish between these two mechanisms, because delay disrupts both situation awareness and grounding.

In Experiment 2, the participants' spatial perspectives are misaligned, which disrupts the ability of visual feedback to support conversational grounding. This misalignment makes it difficult for pairs to use a common spatial vocabulary. In this case, if visual feedback improves collaborative performance, it does so primarily through situation awareness.

Finally, Experiment 3 manipulates the size of the field of view and who controls the view. Both manipulations disrupt the ability of visual feedback to support situation awareness by reducing the size and likely availability of a common viewing area. As a result, the Helper has difficulty keeping track of the puzzle layout as it is being constructed, but this manipulation interferes less with the pairs' ability to develop a common vocabulary to describe the pieces and their placement. If visual feedback improves collaborative performance when the Helper can see only a small shared

FIGURE 1. Overview of studies and experimental manipulations.

	Features of Visual Environment				Task Features
	Immediacy	Perspective Alignment	Field of View	Field of View Control	Linguistic Complexity (Plaids vs. Solids)
Experiment 1: Visual delay study	**X**				**X**
Experiment 2: Rotation study	**X**	**X**			**X**
Experiment 3: Field of view study			**X**	**X**	**X**

field of view, it does so primarily through benefits to conversational grounding that manifest in a pair's ability to easily refer to puzzle pieces.

3. EXPERIMENT 1: VISUAL DELAY STUDY

Experiment 1 introduces the research paradigm, demonstrates the value of visual information for improving performance in a collaborative task, and examines the way its value varies with the nature of the task. If visual information improves either situation awareness or conversational grounding, pairs who have visual feedback should perform better in the puzzle experiment, completing the puzzles more quickly.

At the technological level, this experiment examines how delays in the availability of the visual feedback—of the sort introduced by video compression or network lags—are likely to undercut its value. Prior work by Krauss and Bricker (1967) established that small auditory delays could have a negative impact on communication. Research by Cohen (1982) and by O'Conaill, Whittaker, and Wilbur (1993) showed that simultaneous delays in the audio and video channels also harm conversation (e.g., pairs are less interactive). However, these studies did not establish that visual delay by itself influences the conversation. It is likely that delays in visual updating will reduce the value of visual information. Collaborators in face-to-face settings use visual information to precisely time when they will provide new information and to change speech midsentence in response to their partner's gaze (Boyle, Anderson, & Newlands, 1994) or behavior (Clark & Krych, 2004). In other research, we varied the availability of the visual feedback on a continuous range between 60 ms and 3,300 ms (Gergle, Kraut, & Fussell, 2006). We found that communication breakdowns occur when the delay is greater than 950 ms. This study used a delay of 3,000 ms, a number chosen to ensure the delay was well above the threshold found for disrupting collaborative performance and discourse.

In this experiment, we manipulate the linguistic complexity of the task objects either by using simple primary colors that have high codability and high discriminability or by making the task objects tartan plaids, which have low codability and low discriminability[2] (illustrated in Figure 3). Shared visual information should have the most benefit when these properties are low as the pairs must use language to describe what are essentially difficult to describe objects, and the expressions they use will be less efficient and more ambiguous with a heightened likelihood of misunderstanding.

Along with the differences in task performance, we expect to see differences in the ways that pairs adapt their discourse structure to make use of the visual information. If the visual information benefits task performance through situation awareness, Helpers who receive visual feedback should more quickly introduce instructions after a Worker has completed the previous task. In addition, they should more readily

[2] *Discriminability* refers to how easy it is to linguistically differentiate one object from the others based on its visual features, whereas *codability* refers to how easy it is to initially describe or name an object (Hupet, Seron, & Chantraine, 1991).

identify errors or deviations from the optimal solution path and quickly correct these problems. They should also add more spoken contributions that aim to make their partner aware of the state of the task. If visual information benefits task performance by facilitating conversational grounding, participants should spend less time requesting and giving confirmation that they have understood their partners' utterances (Brennan, 1990, 2005; Clark & Krych, 2004; Fussell, Kraut, & Siegel, 2000). In addition to this, the *principle of least collaborative effort* (Clark & Wilkes-Gibbs, 1986) suggests that pairs should change the structure of their discourse in order to expend the least amount of effort for the group as a whole (Kraut, Gergle, et al., 2002). Therefore, both the Helpers and Workers should be influenced by the presence of visual feedback, even though only the Helpers see different views.

The following hypotheses summarize this discussion:

H1: A collaborative pair will perform their task more quickly when they have a shared view of the work area.
H2: A collaborative pair will perform their task more slowly as the linguistic complexity of the task increases.
H3: A shared view of the work area will have additional performance benefits when the linguistic complexity of the task objects increases.
H4: Delay in transmission of the shared visual information will decrease the benefit of a shared view of the work area.

3.1. Method

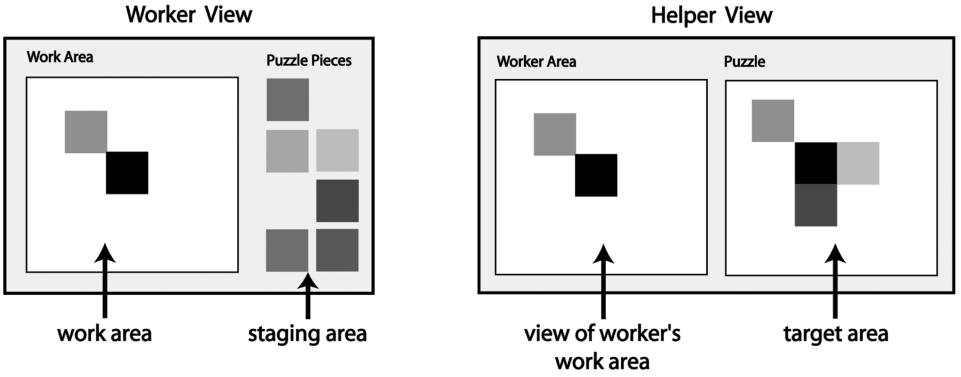
The Puzzle Task

Pairs of participants were randomly assigned to play the role of Helper or Worker. The Helper described a configuration of puzzle pieces to the Worker, who assembled the puzzle to match the target area (Gergle et al., 2004a; Gergle, Millen, Kraut, & Fussell, 2004; Kraut, Gergle, et al., 2002). The puzzle solutions consisted of four blocks selected from a larger set of eight. The Worker's goal is to correctly place the four solution pieces as quickly as possible to match the target solution that the Helper is viewing.

Figure 2 illustrates the Worker's screen (left) and Helper's screen (right). The Worker's screen consisted of a staging area on the right-hand side in which the puzzle pieces were shown and a work area on the left-hand side in which he or she constructed the puzzle. The Helper's screen showed the target solution on the right and a view (if available) of the Worker's work area on the left. We manipulated two factors: (a) whether the Helper viewed the same work area as the Worker and, if so, how quickly the visual information was delivered and (b) the linguistic complexity required to describe the puzzle pieces.

Each participant was seated in a separate room in front of a computer with a 21-in. display. They communicated over a high-quality, full-duplex audio link with

FIGURE 2. Experiment 1: The Worker's view (left) and the Helper's view (right).



no delay. The experimental displays for the Worker and Helper were written as two communicating Visual Basic and C++ programs.

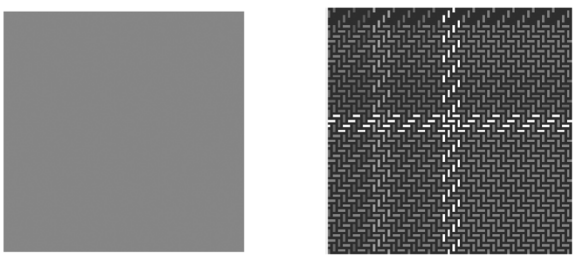
Immediacy of Visual Feedback (Immediate vs. Delay vs. None). In the immediate visual feedback condition (*immediate*), the Helper's view of the work area was identical to the Worker's work area, with no perceptible delay. In the delayed condition (*delay*), the Helper saw the Worker's work area with a 3-s delay. In the no visual feedback condition (*none*), the Helper's view was solid black.

Linguistic Complexity (Solid Pieces vs. Plaid Pieces). We manipulated linguistic complexity by providing pairs with different types of puzzle pieces. The pieces were either lexically simple primary colors (e.g., red, yellow, orange, etc.) or more complex visual patterns (e.g., tartan plaids) that required negotiating a common name (i.e., they were not part of a previously shared lexicon and required the pairs to ground on the referential descriptions; see Figure 3).

Participants and Procedure

Participants consisted of 12 pairs of undergraduate students. They received $10 for their participation and were randomly assigned to play the role of Helper

FIGURE 3. Example of solid piece (left) and plaid piece (right).



or Worker. The immediacy of the visual feedback and the visual complexity were manipulated within pairs, whereas linguistic complexity was a between-pair factor. Each pair participated in six blocks of four trials. They completed 24 puzzles in approximately 1 hr.

Measures and Statistical Analysis

The pairs were instructed to complete the task as quickly as possible, so task performance was measured as the time it took to properly complete the puzzle. Because the vast majority of the puzzles were solved correctly and differences in error rates among conditions were negligible, we focus on completion time as our primary measure of task performance.

The analysis is a mixed-model regression analysis in which Block (1–6), Trial (1–4), and Immediacy of the Visual Feedback (Immediate, Delayed, None) are repeated within-subject factors, and Linguistic Complexity (Solid or Plaid) is a between-pair factor. We include two-way and three-way interactions in the analysis. Because each pair participated in 24 trials (six conditions by four trials per condition), observations within a pair were not independent of each other. Pair, nested within Linguistic Complexity, is modeled as a random effect.

3.2. Results and Discussion

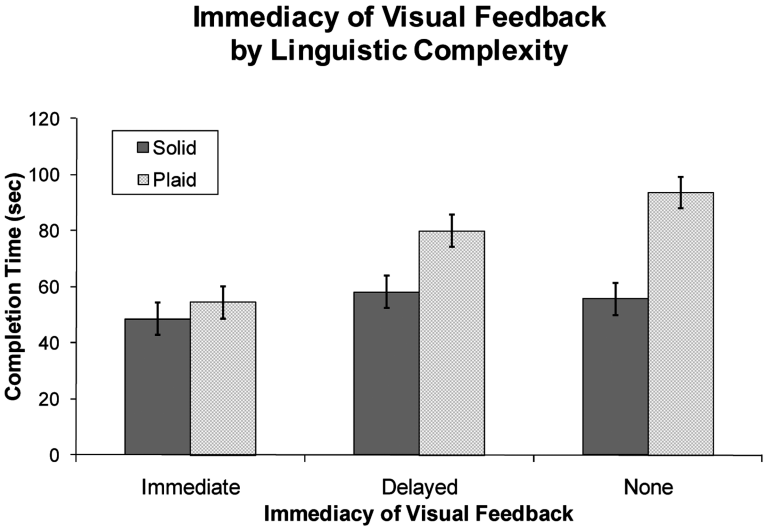
Task Performance

Immediacy of Visual Feedback. Consistent with Hypothesis 1, a shared view of the work area benefited performance. The pairs were approximately 30% faster at completing the puzzles when they received immediate visual feedback ($M = 51.27$ s, $SE = 4.12$) than when they did not receive any visual feedback ($M = 74.63$ s, $SE = 4.03$), $F(1, 266) = 47.43$, $p < .001$. Consistent with Hypothesis 4, a 3-s delay in updating the shared visual information considerably reduced its benefits. The delayed visual feedback ($M = 69.04$ s, $SE = 4.12$) was only 7% faster than when the pairs did not receive any visual feedback, $F(1, 266) = 2.71$, $p = .10$.

Linguistic Complexity. Linguistic complexity substantially increased completion times. The pairs were approximately 30% faster in trials where the pieces were solid colors ($M = 53.95$ s, $SE = 5.04$) than when they were more complex plaids ($M = 76.0$ s, $SE = 5.04$), $F(1, 10) = 9.62$, $p = .011$. This finding is consistent with Hypothesis 2, which states when the linguistic complexity of the task increased, the pairs were slower to complete the task.

The visual information had greatest benefit in the linguistically complex condition. This supports Hypothesis 3, which states that a shared view of the work area will have additional benefits as the linguistic complexity of the task objects increases—for the interaction, $F(1, 266) = 66.40$, $p < .001$. A detailed examination of this interaction reveals that the pairs took much longer to complete the task for plaids than for solid

FIGURE 4. Shared visual space by linguistic complexity on task completion time (all figures show LS means ±1 standard error).



colors when there was no shared visual space available, $F(1, 266) = 22.05$, $p < .001$, $d = .58$. They also took longer to complete the task for plaids than for solids when the shared view was delayed; however, the effect size was smaller, $F(1, 266) = 7.26$, $p < .008$, $d = .33$. This difference all but disappeared when the shared visual space was immediately available, $F(1, 266) = 0.64$, $p = .426$, $d = .10$ (see Figure 4 for details).

Communication Processes

Previous work has detailed how discourse structure changes when shared visual information is available. Immediate shared visual information about the workspace results in lower rates of spoken discourse because the communicators rely instead on more efficient visual information (Boyle et al., 1994; Brennan, 1990, 2005; Clark & Krych, 2004; Daly-Jones et al., 1998; Fussell et al., 2000; Kraut, Gergle, et al., 2002). Visual information is also useful for supporting efficient referring expressions (Clark & Krych, 2004; Fussell et al., 2000; Kraut, Gergle, et al., 2002). It provides evidence of understanding (conversational grounding) as well as unambiguous information about the state of the task (situation awareness). We previously demonstrated this by showing that visual information is used in place of verbal acknowledgments of understanding (a vital part of the grounding process) as well as in place of explicit verbal statements that a task had been completed (a vital component of situation awareness; Gergle, Millen, et al., 2004; Kraut, Gergle, et al., 2002; Kraut, Miller, & Siegel, 1996). The following excerpts show the differences in these acknowledgment processes when visual feedback is provided (Figure 5a) and when it is not (Figure 5b).

FIGURE 5. (a) Immediate visual feedback and plaid pieces. (b) No visual feedback and plaid pieces.

1. **H:** the first one is gray, gray lines on the top and brown lines on the left
2. **W:** *[moves correct piece]*
3. **H:** put it on the right middle corner
4. **H:** yeah perfect
5. **H:** uh take it up slightly
6. **H:** and the second one is uh two blue vertical bands
7. **H:** a lot of light gray err light blue lines
8. **W:** *[moves correct piece]*
9. **H:** take it half a block down
10. **H:** to . . . yeah.

(a)

1. **H:** the last one is
2. **H:** the, it has two light blue : ah : big stripes going up the sides with . . .
3. **H:** with a like vertical royal blue up the middle like
4. **W:** it just has . . .
5. **H:** the background is royal blue
6. **W:** does it just have one, one
7. **H:** just one royal blue up the middle
8. **W:** *[moves correct piece]*
9. **W:** I got it
10. **H:** and it has two hash marks going through the middle horizontally
11. **W:** yeah, I got it
12. **H:** yeah, that goes directly to the left of the the:that last one I just told you
13. **W:** ok, done

(b)

In line 1 of Figure 5a, the Helper begins by generating a description of a puzzle piece. The Worker demonstrates her understanding of the intended referent by moving a piece into the shared view (line 2) and not bothering to produce a verbal acknowledgment. Contrast this with the case when there is no shared visual information available. The Worker becomes more active in negotiating a shared understanding (Figure 5b, lines 4 and 6), and he provides explicit confirmation that he understands the intended referent by declaring, "I got it" (line 9). These excerpts demonstrate how the pairs use the shared visual information to support the coordination mechanism of conversational grounding.

Yet these same excerpts also demonstrate the use of visual information to support situation awareness. When visual information is available, the Helper uses it to determine when a piece is correctly placed and it is time to start on the next. Returning to Figure 5a, once the Worker moves the correct piece into the shared display area (line 2), the Helper instantly provides the next instruction describing where to place the active piece. This same trend can be seen again at lines 6 and 9, when the Helper describes the piece, the Worker places it in the work area, and the Helper immediately instructs the Worker where to place it. In contrast, without the visual feedback the Helper must rely upon the Worker's explicit declaration that he has finished a subtask, and this may require persuasion before the Helper is convinced that a subtask is complete. Note in Figure 5b the Worker explicitly declared that he had completed the instruction (line 13). It is also important to note here that the linguistic evidence is more ambiguous than the visual information. For example, the first "I got it" on line 9 could indicate that the Worker had understood the Helper or that he had obtained the piece. The Helper continues to describe the piece, until the Worker

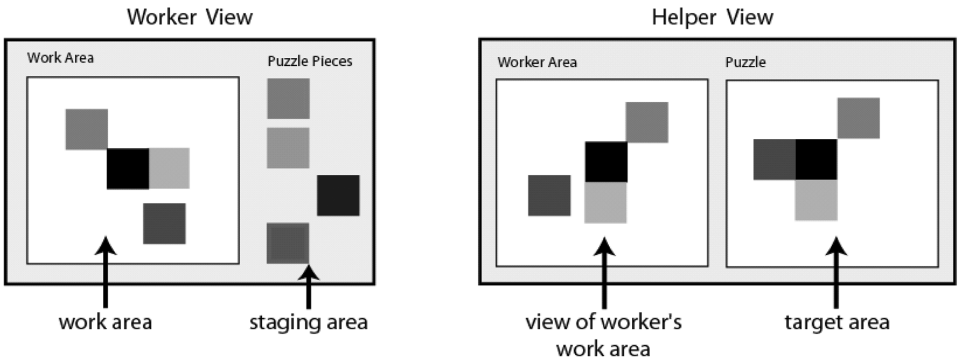
follows up again and says, "Yeah, I got it" on line 11. Only at this point does the Helper describe where to place the piece.

These excerpts provide qualitative demonstrations of visual information being used to support both conversational grounding and situation awareness. Either or both of these mechanisms could account for the performance benefits we found in this experiment and in prior studies. It is the goal of the following experiments to demonstrate in a more controlled fashion how different features of the shared visual space can differentially influence these two independent coordination mechanisms.

4. EXPERIMENT 2: ROTATION STUDY

Whereas the first experiment suggested that visual information could potentially serve two roles in a collaborative task, the second experiment was designed to differentiate the use of visual information for situation awareness and conversational grounding. It does so by manipulating the display so that the Helper and Worker see the visual information from different perspectives. The Helper's display and target area were rotated so that they were different from the one the Worker saw (see Figure 6). After the rotation the Helper and Worker no longer had a common reference point from which to describe object locations and discuss spatial features of the objects. The rotated views forced the pairs to negotiate their shared spatial perspective and shift their reference frame (Schober, 1993, 1995). Whether using a speaker-centric or an object-centric reference frame, rotating the Helper's view of the work area will cause difficulties for the Helper and Worker in agreeing upon a description of some of the objects and their relative positions. For the visual information to support conversational grounding, the Helper and Worker need similar views of the task and environment so that they can use the same language to describe it. For example, in the rotated condition, the Helper's use of an object-centric description

FIGURE 6. Rotated view.



Note. **The Helper's view of the work area and the target are rotated 90° clockwise when presented in the Helper's view of the worker's work area (right).**

such as "the white cross in the upper left" may no longer accurately correspond to the Worker's view. Similarly, in the rotated condition it is more difficult to use efficient speaker-centric spatial descriptions such as "to the left." In this way, we expect the rotations to interfere with the role shared visual information plays in conversational grounding.

Yet, although the rotation of the Helper's view is likely to degrade the Helper and Worker's ability to ground their conversation, it should not degrade the Helper's ability to maintain situation awareness. Because we rotated the Helper's view of both the work area and the target area, he or she could still compare the work area to the target state and assess whether the Worker has performed actions correctly. For example, the Helper could easily assess when the Worker had placed a piece in the correct relative position and when the Worker needed the next instruction. The following hypothesis summarizes this reasoning:

H5: Pairs will perform the task more quickly when they share spatial perspective (i.e., when their view of the work area is aligned rather than rotated) because the visual information supports conversational grounding.

Alternatively, one should note that if the visual information is primarily used for situation awareness, having a shared spatial perspective should have little additional influence on task performance.

Experiment 2 also manipulated the immediacy of the visual feedback in a similar manner to that of Experiment 1. Visual feedback was either (a) continuous with immediate updating, or (b) updated only when the Worker sent snapshots of the current state by pressing a Send Image button. In addition to replicating the hypotheses examined in Experiment 1, we expected several additional interactions between the alignment of the visual space, immediacy of the visual feedback, and linguistic complexity. In particular, if rotating the Helpers' view of the work area was likely to harm conversational grounding by limiting the ability to use and monitor spatial descriptions, it would be especially important for the Helper to have rapid visual feedback to correct any misunderstandings the pairs might develop. Therefore, if the rotation degrades conversational grounding and not simply situation awareness, we should expect the following:

H6: An immediately available view of the work area will have additional performance benefits when the views between the partners are rotated.

However, we would not expect the degree of similarity between viewpoints to impact performance equally for both levels of linguistic complexity. Because the pairs likely have a shared lexicon to describe the solid pieces but not the complex plaid pieces, the inability to ground references should be affected more in the latter condition. Visual rotation is especially likely to interfere with a pair's ability to agree

upon referring expressions for the plaid pieces, as rotation requires the pairs to first establish a shared perspective from which to make reference to piece attributes. Most participants described the plaids by mentioning their detailed features (e.g., "white stripe on the right"), and when the pieces were rotated, some of these spatial descriptors were no longer the same for the Helper and Worker. In contrast, when describing the solid pieces, the visual information could easily be used to describe the object referent (e.g., "the red one"). Therefore, we expected an interaction whereby the rotated views would impact performance more for the plaid pieces than for the solid pieces.

H7: An identical perspective on the work space will have additional performance benefits when the linguistic complexity of the objects increases.

4.1. Method

Participants consisted of 32 pairs of undergraduate students. They received an hourly payment of $10 for their participation in the study and were randomly assigned to play the role of Helper or Worker. The Immediacy of Visual Feedback and the Viewspace Alignment (the spatial symmetry between the Helper and Worker views) were manipulated within pairs, whereas the Linguistic Complexity was a between-pair factor. Each pair participated in four blocks of six trials each.

Experimental Manipulations

Linguistic Complexity (Solid Pieces vs. Plaid Pieces). As in the first experiment, the pieces were either solid colors (e.g., red, yellow, orange) or more complex visual patterns (tartan plaids).

Immediacy of Visual Feedback (Immediate vs. Snapshot). Either the shared view was immediately available to the Helper (*immediate condition*), or the Worker had to manually choose when to send back an image of the work area to the Helper (*snapshot condition*).

Viewspace Alignment (Aligned vs. Rotated). The views were either identical between the Helper and Worker displays or rotated offsets of one another (see Figure 6). In the rotated condition, the view that the Helper saw was randomly flipped in the vertical or horizontal direction and then rotated 45, 90, or 135°. The target puzzle the Helper saw was also transformed the same way, so that the Helper's view of the work area and target had the same orientation. For example, with a 90° rotation, when the Worker placed a puzzle piece to the right of another, the Helper might see the two pieces as aligned one on top of the other. We used the same geometric transformation for all trials for a single pair of subjects.

Statistical Analysis

The performance analysis uses completion time as the dependent variable. It is a mixed-model regression analysis in which Block (1–4), Trial (1–6), Field of View Alignment (Aligned or Rotated), and Immediacy of the Visual Feedback (Immediate or Snapshot) are repeated, and Linguistic Complexity (Solid or Plaid) is a between-pair factor. We include two-way and three-way interactions in the analysis. Because each pair participated in 24 trials (four conditions by six trials per condition), observations within a pair were not independent of each other. Pair, nested within Linguistic Complexity, is modeled as a random effect.

In addition to performance metrics, we also performed a strategic content analysis to illustrate systematic discourse phenomena. The last trial of each experimental block was transcribed, and two independent coders blind to condition and unfamiliar with the goals of the study then scored two theoretically informative verbal behaviors.

4.2. Results and Discussion

Task Performance

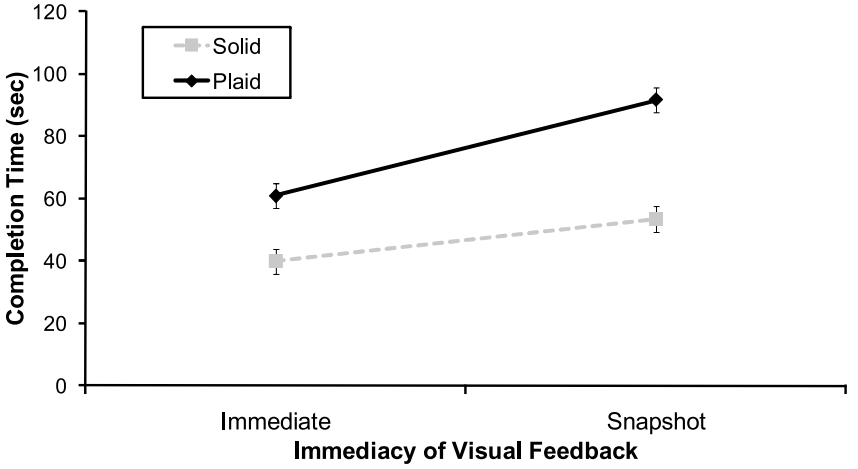
Immediacy of Visual Feedback. As in Experiment 1, an immediate shared view of the work area benefited performance (in support of Hypothesis 1). The pairs were approximately 30% faster at completing the puzzles when they had an immediately available shared visual space ($M = 50.45$ s, $SE = 2.85$) than when the Worker had to send back snapshots ($M = 72.53$ s, $SE = 2.85$), $F(1, 721) = 118.80$, $p < .001$.

Linguistic Complexity. Also consistent with the results from Experiment 1, the linguistic complexity of the pieces significantly increased completion time (consistent with Hypothesis 2). The pairs were more than 35% faster in the trials when the pieces were solid colors ($M = 46.67$ s, $SE = 3.76$) than when they were more complex plaids ($M = 76.0$ s, $SE = 3.76$), $F(1, 30) = 31.0$, $p < .001$.

The immediate visual feedback had the greatest benefit when the puzzles consisted of the plaid pieces (see Figure 7; for the interaction, $F(1, 721) = 17.89$, $p < .001$). A detailed examination of this interaction revealed that although immediate visual space improved performance when the pieces were solid colors, $F(1, 721) = 22.25$, $p < .001$, $d = .35$, the benefits were even greater when the pieces were linguistically complex plaids, $F(1, 721) = 114.44$, $p < .001$, $d = .80$. In other words, when grounding requirements increased for the pieces in the puzzle, the shared visual space was more crucial (providing additional support for Hypothesis 3).

Field of View Alignment. Results from the manipulation of view alignment also suggest that the visual information is used for conversational grounding (supporting Hypothesis 5). The pairs were more than 55% faster when the views were aligned ($M = 37.07$ s, $SE = 2.85$) than when they were reflected and rotated ($M = 85.91$ s, $SE = 2.85$), $F(1, 721) = 581.44$, $p < .001$. The pairs took longer when they had to further ground the terms used to describe the spatial arrangement of the pieces.

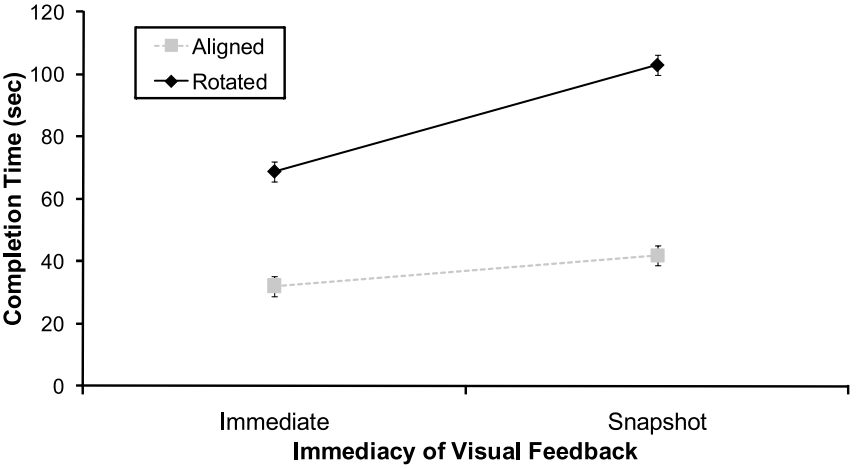
FIGURE 7. Immediacy of the visual feedback by linguistic complexity (LS means ±1 standard error).



Also consistent with the reasoning that visual information was supporting conversational grounding (Hypothesis 6), the visual feedback had greatest benefit in the rotated condition; for the interaction, $F(1, 721) = 36.30$, $p < .001$. Although the availability of the shared visual space improved performance when the environments were aligned, $F(1, 721) = 11.88$, $p < .001$, $d = .25$, it was even more beneficial when the workspaces were rotated, $F(1, 721) = 143.23$, $p < .001$, $d = .89$ (see Figure 8).

Although the difference in the time it took groups to complete the task when rotated views were involved appeared greater for plaids than for solids, the difference was not significant, $F(1, 721) = 2.32$, $p = .13$. Therefore, we failed to find support for

FIGURE 8. Immediacy of the visual feedback by field of view alignment (LS means ±1 standard error).



the hypothesis that an identical perspective on the workspace would have additional benefits when grounding requirements increased (Hypothesis 7). It is unclear whether this was an issue of experimental power or a true lack of finding, and this needs to be explored in future studies.

In this experiment, one could argue that there is a potential conflation whereby the pairs could react to the spatial rotation by sending less visual information. For example, the Worker may decide that there is no use in sending back "false" information and as a result may reduce the number of times she sends visual feedback to her partner. Although this is a reasonable conjecture, the data suggest otherwise. The Workers increased the number of visual images they sent back in both the rotated condition ($M_{aligned} = 4.77$, $SE = 0.52$ vs. $M_{rotated} = 11.72$, $SE = 0.52$), $F(1, 384) = 312.28$, $p < .001$, and in the linguistically complex condition ($M_{solid} = 6.90$, $SE = 0.69$ vs. $M_{plaid} = 9.59$, $SE = 0.69$), $F(1, 384) = 7.65$, $p = .009$, in an attempt to compensate for the grounding problems that occured.

Communication Processes

In a second stage of analysis we examined the conversations for evidence of adaptations that coincided with our theoretical propositions. We transcribed the spoken behaviors and applied a targeted coding scheme designed to draw out two discourse elements of interest: spatial deixis and verbal alignments.

To examine the benefits of shared visual information for conversational grounding, we report on the pairs' use of spatial deixis. These are concrete spatial phrases and terms such as "at the top right," "in front of," "left of," "next to," "above," and so on, that require a degree of shared understanding about the visual space or object features. They are used during the grounding process to efficiently identify distinct referential attributes or spatial positions. If the shared visual space is supporting conversational grounding, then we would expect heightened use of these terms. However, when the visual space is misaligned, these efficient terms can no longer be used and more ambiguous and less efficient expressions are needed. Thus, if the rotation interferes with the ability of the pairs to effectively ground their references, we should see a reduction in the use of these terms that rely upon a shared visual representation.

A second category called verbal alignments (Anderson et al., 1991) was used to capture utterances that provide situation awareness such as whether one is ready to proceed, to describe the state of the task, or to update a partner on availability (e.g., "Helper: Ready for the next one?" "Worker: Did it . . . what's next?" or "Worker: OK, it's there."). In other words, verbal alignments are used to keep the pairs situationally aware. They capture the state of the task in words and describe how things are proceeding. If the pairs were suffering from a lack of situation awareness, we would expect to see an increase in these terms in order to compensate for the lack of visual support. In this study, where the visual manipulation is theorized to primarily influence conversational grounding, we should expect no difference in their use.

Two independent coders blind to the goals of the study applied this scheme to a complete transcription of the dialogue. Interrater reliability was very good for both

categories and was stable across the coding sessions (for spatial deixis, Cohen's κ = .94; for verbal alignments, Cohen's κ = .84).

Results show that when the pairs experienced misaligned spatial perspectives, efficient spatial terms such as "to the right of it" were no longer sufficient to ground references to the objects or space. This is demonstrated by a decrease in the use of relative spatial terms when the pairs performed in the rotated condition, $F(1, 84) = 4.38$, $p = .039$ (as shown in Figure 9a). Yet the pairs still retain the benefits of shared visual space for situation awareness even when the views are misaligned. This is shown by the fact that there is no evidence of an increase in the production of verbal alignments, as shown in Figure 9b, $F(1, 83) = .072$, $p = .79$.

The excerpts presented in the following figures provide qualitative examples of the grounding problems encountered when the shared visual space was rotated. In Figure 10a, when the workspace was aligned the Helper produces a sequence of unambiguous, mutually comprehensible spatial descriptions (e.g., "one the left side," "on top of the black one," etc.). She is confident that understanding is being achieved as is evidenced by her continual progression of directives.

Contrast this with the case when the views were rotated. Use of efficient spatial descriptors was no longer possible, and the pairs had to adapt by using less efficient, more verbally ambiguous descriptions. Figure 10b shows how the Helper compensates for the lack of alignment by using imprecise and ambiguous terms such as "circle round back the other way," "move that around there," and "back the other direction," (lines 3, 4, 11) rather than the concrete spatial descriptions seen in the previous example. It helps that the visual feedback is continuous and immediate, which facilitates situation awareness, which the pairs increasingly rely upon. The pairs then proceed in directives and checks of the task in a tightly coupled fashion, such as "keep going, keep going, keep going, . . . " (lines 17–23). When the feedback was not continuously available, the pairs had to proceed in a much slower lock-step fashion when trying to confirm that the directives were understood. As a result, the Helper

FIGURE 9. (a) Rate of spatial deixis terms by field of view alignment. (b) Rate of verbal alignments by field of view alignment.

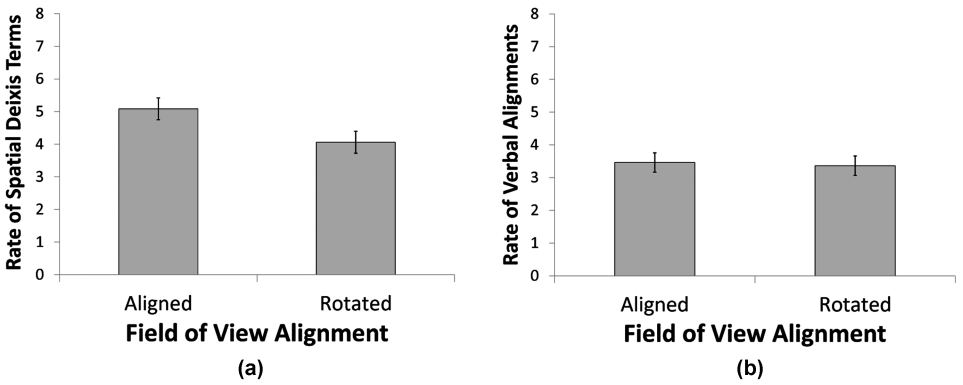


FIGURE 10. (a) Aligned, immediate, and plaid pieces. (b) Rotated, immediate, and plaid pieces.

1. **H:** alright, um take the main black one
2. **H:** and stick it in the middle
3. **W:** *[moves and places correct piece]*
4. **H:** take the one-stripe yellow
5. **H:** and put it on the left side
6. **W:** *[moves and places correct piece]*
7. **H:** uh yeah, that's good
8. **H:** take the um two stripe one
9. **H:** and put it on top of the black one
10. **W:** *[moves and places correct piece]*
11. **H:** and take the half shaded one
12. **H:** and put it diagonally above the one that you just moved to the right
13. **W:** *[moves and places correct piece]*
14. **H:** yup, done.

(a)

1. **W:** *[moves correct piece]*
2. **H:** right cool
3. **H:** alright, go back the other way
4. **H:** circle round back the other way
5. **H:** alright cool
6. **H:** keep the :: keep the two black ones without any colors on them
7. **W:** ok
8. **H:** and now the one with the white and the yellow
9. **W:** *[selects and moves correct piece]*
10. **H:** move that, move that around there
11. **H:** move it back the other direction
12. **H:** alright, keep that there
13. **H:** and the one that looks like a road
14. **W:** *[selects correct piece]*
15. **H:** move that around the little axis thing
16. **W:** *[moves correct piece]*
17. **H:** keep going
18. **H:** keep going
19. **H:** keep going
20. **H:** keep going
21. **H:** keep going
22. **H:** keep going
23. **H:** keep going
24. **H:** alright, cool
25. **H:** wait, let me check this out
26. **H:** alright, we're done

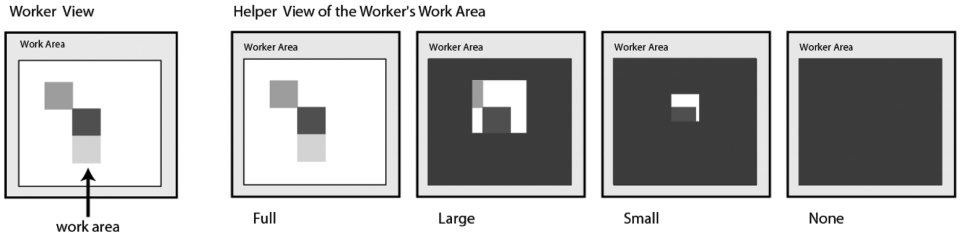
(b)

had to wait for the Worker to transmit an image before she could confirm or elaborate, leading to slower descriptions and overall performance.

Although the rotated visual display harmed the Helper's ability to ground their spatial descriptions, it did not seem to harm their ability to track task progress. As previously mentioned, the use of verbal alignments was relatively stable across conditions.

To summarize, the pattern of results from Experiment 2 is consistent with the interpretation that visual information improves task performance by supporting conversational grounding. The visual information needs to be both temporally and spatially synchronized between people performing the task in order to achieve this result. When this did not exist, the pairs suffered in their ability to effectively and efficiently ground their spatial references. On the other hand, if the pairs were simply using the visual information for situation awareness and not for grounding, the rotations should not have made the task more difficult nor affected the use of efficient grounding terms such as spatial deictic expressions. Furthermore, the snapshot manipulation should not have accentuated the performance drop in the rotated condition. If the rotated views were affecting situation awareness, we should have seen an increase

FIGURE 11. Field of view.



Note. **Given the Worker's view on the left, the four Helper views on the right demonstrate the corresponding view onto the work area (full, large, small, none).**

in the use of verbal alignments used to compensate for any disruption that would have occurred to tracking the state of the task.

5. EXPERIMENT 3: FIELD OF VIEW STUDY

Although the second experiment demonstrated that shared visual information serves a critical role for conversational grounding, the manipulation had little influence on the pairs' ability to maintain situation awareness. Experiment 3 was designed to determine whether visual information can also improve task performance via situation awareness.

In this study, we varied how much of the shared work area could be seen. The Helper had one of three sizes of a shared view (full, large, or small) or no shared view (see Figure 11). Compared to the full display, partial fields of view should degrade the Helpers' situation awareness (i.e., their knowledge of the overall puzzle layout and task state) but should not interfere as much with conversational grounding (i.e., their ability to use efficient vocabulary to describe puzzle pieces in ways their partner can understand). A small field of view that provides a view of the puzzle piece should suffice for grounding piece descriptions. As a result, the pairs should complete the puzzle more quickly as they go from having no shared visual information to having a small shared viewing area. This benefit should increase when the pieces are linguistically complex plaids. However, a narrow field of view, in comparison to a wider field of view, should not make a difference for grounding piece descriptions but should make it more difficult for the Helper to track the overall progress of the puzzle. Therefore, if the pairs also get faster as they go from a small shared field of view to a larger one, or from a larger one to a full shared view of the workspace, the more likely explanation for these performance improvements is the additional impact on situation awareness.[3] Finally, we should see a greater benefit from the availability

[3]As evidenced in the previous studies, the lack of shared visual space should affect both situation awareness and conversational grounding.

of visual information for the linguistically complex plaid pieces when going from no shared visual information to a small amount of shared visual information. However, as the field of view grows larger, the benefits should be equal for the plaids and solid colors, provided the visual information is primarily supporting situation awareness and not providing additional benefits for conversational grounding.

The following hypotheses summarize this reasoning:

H8: Pairs will perform the task more quickly when larger shared fields of view are available because the additional visual information supports situation awareness.

H9: The benefit of a larger field of view will not interact with the linguistic complexity of the task because pairs are primarily using the visual information for situation awareness.

A second condition varied how the view of the work area was controlled: automatic, manual worker controlled, or manual helper controlled. With automatic view control, the field of view was centered on the Worker's mouse pointer. In this case, when the Worker grabbed a piece it was guaranteed to be in the Helper's view. The other two conditions featured forms of manual control. In the Worker-controlled condition, the Worker used the mouse to grab an outlined window frame, indicating the area of shared view, and then either manually positioned the frame within the work area or moved the pieces over the frame to "show" them to their partner (see Figure 11). In the Helper-controlled condition, the Helpers controlled the window with their cursor. This allowed Helpers to refresh their awareness of the puzzle layout at their own pace, by moving the window around the work area. However, there were relatively few significant differences resulting from this manipulation. For ease of exposition we only report on the findings that are statistically significant and provide important theoretical insight.

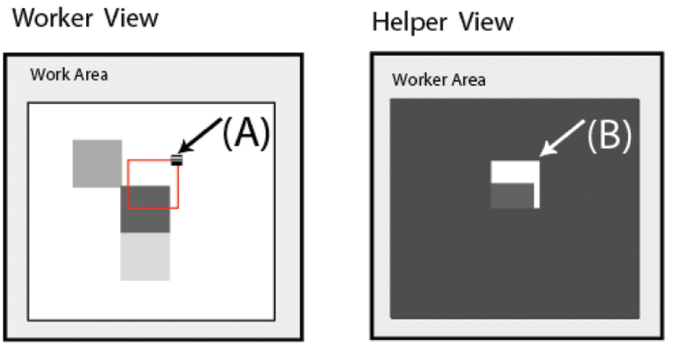
5.1. Method

Participants consisted of 24 pairs of undergraduate students who were randomly assigned to play the role of Helper or Worker. They received payment of $15 to $20 based on their performance. The Field of View Size and Linguistic Complexity manipulations were within pairs, whereas the Field of View Control manipulation was between pairs. Each pair participated in eight blocks of four trials (32 puzzles) in a session that lasted approximately $1\frac{1}{2}$ hr.

Experimental Manipulations

Linguistic Complexity (Primary Pieces vs. Plaid Pieces). The same pieces were used as in the prior two studies. The pieces were lexically simple, easy-to-describe solid colors (e.g., red, yellow, orange, etc.), or more linguistically complex tartan plaids.

FIGURE 12. Field of view control in the manual worker condition.



Note. **In this condition the Worker had to manually select the shared view indicator by clicking on its corner as shown in (A) and position it within the work area, while (B) presents the corresponding Helper view. (Color figure available online.)**

Field of View Size. We varied how much of the shared work area could be seen. The Helper could see the *full* area, a *large* area (equivalent to the size of four puzzle pieces), a *small* area (equivalent to the area of a single puzzle piece), or nothing (*none*). Figure 11 shows the corresponding levels.

Field of View Control. For the small and large views, we also varied how the shared view was controlled. In the *automatic* condition, the subview automatically followed the Worker's cursor when it was in the work area. In the *manual helper* control condition, the Helper controlled where they wanted to look by moving their cursor to the appropriate space. In the *manual worker* control condition, the Worker had to position the view over the work area (as shown in Figure 12).

Statistical Analysis

Completion time is the primary performance measure. A first stage of analysis tested the influence of the Field of View Size and Linguistic Complexity, using a mixed-model regression analysis in which Block (1–8), Trial (1–4), Field of View size (None, Small, Large, Full), and Linguistic Complexity (Solid or Plaid) are repeated within-pair factors. Pair is modeled as a random effect.

A second analysis examined the Field of View Control manipulation, which is a between-subjects factor. This manipulation affected only a subset of the trials because in the cases where there was a full field of view or no shared visual information there was not a subview to control. This analysis relies on a mixed-model regression analysis in which Block (1–8), Trial (1–4), Field of View (Small or Large), and Lexical Complexity (Solid or Plaid) are repeated within-pair factors, and Field of View Control (Auto, Manual worker, or Manual helper) is a between-pair factor. Pair, nested within field of view control condition, is modeled as a random effect.

In addition to these performance metrics, we also perform a strategic content analysis to illustrate systematic discourse phenomena. The final trial of each experimental block was transcribed, and two independent coders blind to condition and unfamiliar with the goals of the study then coded the verbal behaviors.

5.2. Results and Discussion

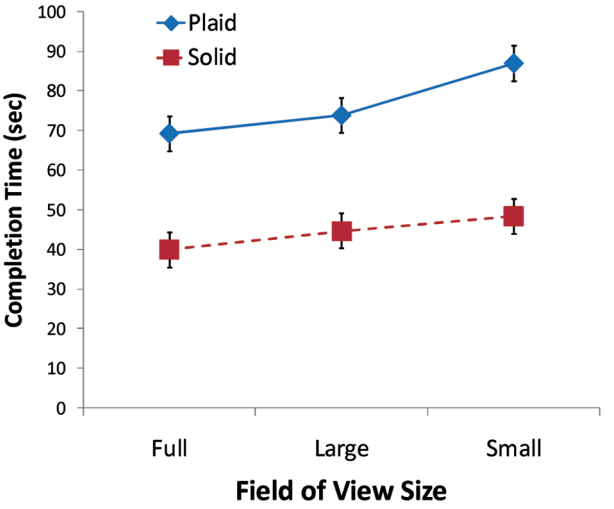
Task Performance

Field of View Size. As the proportion of viewable space increased, there was a large decrease in the time it took the pairs to complete the task: for the linear contrast, $F(1, 707) = 340.11$, $p < .001$; Full = 54.5 s (4.5), Large = 59.1 s (4.4), Small = 67.6 s (4.4), and None = 92.4 s (4.5); all contrasts, $p < .05$. The largest improvement in performance was going from the no visual space condition to the small view condition, at 27%. This was substantially larger than the 12.6% improvement going from a small to large field of view or the 7.8% improvement in going from the large to full field of view. This initial pattern suggests that the use of visual information may have a greater impact for conversational grounding than for situation awareness. However, the fact that performance improved when going from a small to large view and a large to full view suggests that situation awareness also improved task performance (support for Hypothesis 8).[4] This is in contrast to the notion that the larger field of view should have little impact on performance if conversational grounding is the primary mechanism at play.

Linguistic Complexity. Similar to the first two studies, the manipulation of linguistic complexity had a significant influence on completion time. The pairs were approximately 40% faster in the trials with solid colors ($M = 50.9$ s, $SE = 4.33$) than they were in trials with plaids ($M = 85.9$ s, $SE = 4.32$), $F(1, 707) = 593.14$, $p < .001$. Contrary to the expectation that there would be no additional benefits for the plaid pieces with an increased field of view, there was a significant interaction. There was greater benefit derived from a larger view space when the pieces were linguistically complex, $F(1, 707) = 16.21$, $p < .001$. This interaction was driven by the increased difference for the plaid pieces between the small and large field of view conditions, $F(1,707) = 5.39$, $p = .02$, as seen in Figure 13. As expected, there were no additional benefits for the linguistically complex plaid pieces when comparing the difference between the large and full field of view conditions, $F(1,707) = .001$, $p = .99$. Thus, these results only partially support Hypothesis 9 and suggest that the large field of view may provide some support for grounding spatial references, in addition to making it easier for the Helper to maintain situation awareness. We return to this issue in the discussion.

[4]There is another form of conversational grounding that occurs with respect to the references used for piece placement. The possibility that these differences are due to grounding for a different purpose (i.e., piece placement) is addressed with the targeted content analysis presented in Section 5.2.

FIGURE 13. Field of view size by linguistic complexity on completion time. (Color figure available online.)



Field of View Control. There were no main effects of the field of view control on time to complete the puzzle. There was, however, a significant interaction between field of view control and the linguistic complexity of the puzzle pieces, $F(2, 343) = 5.58$, $p = .004$. The automatic condition was found to provide greater benefit for the plaid pieces than the two manual conditions: for the contrast, $F(1, 343) = 10.32$, $p < .001$. This is in part because an automated view of what the Worker is currently working on—as a side effect of the view being yoked to their cursor—provides visual information about which piece the Worker had currently selected (or not selected). In grounding theory, this finding would be predicted based on the reduced cost for monitoring comprehension and understanding. The automatic viewfinder provides immediate visual evidence of understanding.

Communication Processes

To help clarify the conclusions drawn from the performance data, and provide insight into the ways in which the pairs made use of the varying fields of view, we further examined the conversations for evidence of dialogue adaptations. We transcribed the spoken behaviors and applied the same targeted coding scheme as in Experiment 2 in order to draw out the use of spatial deixis and verbal alignments.

To demonstrate the benefits of shared visual information for situation awareness, we should see the opposite patterns to those found in Experiment 2. Spatial deixis terms are primarily used during the grounding process to efficiently identify distinct referential attributes or spatial positions. If the increased field of view is primarily supporting situation awareness, then we should expect no change in the use of these terms across levels of the field of view. Verbal alignments, on the other hand, should increase in production if the pairs are missing the situation awareness provided by

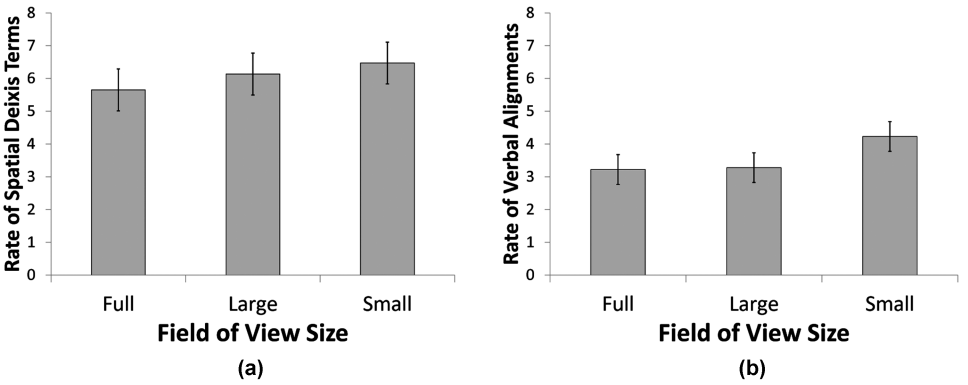
a full view of the work area. Recall that verbal alignments capture explicit verbal utterances that describe readiness, when one is moving on to the next task, checking on the state of the task, and so on. In other words, they are used to keep the pairs apprised of the task state and partner's availability. If the pairs suffer from a lack of situation awareness when the field of view is reduced, we should expect an increase in the use of these dialogue acts.

Two independent coders blind to the goals of the study applied the coding scheme to a transcription of the dialogue. Interrater reliability was very good for both categories and was stable across the coding sessions (for spatial deixis, Cohen's $\kappa = .96$; for verbal alignments, Cohen's $\kappa = .88$).

Results from the content analysis show that when the field of view was reduced, the pairs retained the benefits for conversational grounding: There was no evidence of an influence on the production of spatial deixis terms. Although Figure 14a may at first glance appear to demonstrate a slight increase, it is not statistically significant, $F(2, 108) = 0.56, p = .57$. However, consistent with the proposal that field of view will influence situation awareness, the reduced field of view did have a significant impact on the production of verbal alignments. As shown in Figure 14b, the pairs increased their production of verbal alignments when the field of view was narrowed, $F(2, 108) = 3.10, p = .04$. In the reduced views the visual evidence was no longer sufficient for maintaining situation awareness, and the pairs compensated by increasing their production of verbal alignments. This difference was primarily driven by the change between the large and small views, $F(1, 108) = 4.9, p = .02$, suggesting that the larger view showed enough of the shared space to not hinder awareness.

The excerpts presented in the following figures provide examples of the awareness problems encountered when the field of view was reduced. When the pairs had a smaller view of the work area, their ability to track the task state and intervene at appropriate times was diminished. We believe this stems from an inability to gather information about the surrounding environmental context and to assess the current task state in a timely fashion. As a result, the pairs increase their

FIGURE 14. (a) Rate of spatial deixis terms. (b) Rate of verbal alignments.



production of utterances intended to make explicit the task state and his or her availability.

Figures 15a and 15b contrast several examples from the small field of view with a typical example from the large field of view condition. The excerpts in Figure 15a are drawn from different pairs in the small field of view condition. In Example 1, the Worker selects a piece and provides verbal feedback updating the Helper about the status of her actions with "ok, got it" (line 4) and also states when she has completed an action, "ok, done" (line 7). Contrast this with a very typical example of the performance in the large view condition shown in Figure 15b. Here the Worker simply performs the actions without verbally updating his partner (a typical pattern of performance in the large and full view conditions). It is important to note that in Example 1 the pair still retains the ability to use more efficient spatial deixis (e.g., ". . . next to," "one the right," and "left of center").

Examples 2 through 4 all demonstrate additional ways in which both the Worker and Helper use verbal alignments to keep one another apprised of the task state and their actions (e.g., Helper: "you got it?" Worker: "one going down . . . ok, I got it" and Worker: "ok, one-three, done"). It should also be noted that these verbal alignments are initiated by both the Helper, (e.g., Helper: "Did you get it?") as well as the Worker

FIGURE 15. (a) Excerpts from small field of view condition. (b) Excerpts from large field of view condition.

Example 1

1. . . .
2. **H:** they are right next to each other and a little bit left of center
3. **W:** *[selects correct piece]*
4. **W:** ok, got it
5. **H:** alright, put that directly next to, on the right – to the first one
6. **W:** *[correctly positions piece]*
7. **W:** ok, done.
8. **H:** umm, now you need another blue . . .

Example 2

1. **H:** you got it?
2. **H:** and at the very right?
3. **W:** yeah

Example 3

1. **H:** and also one going down the left
2. **W:** one going down . . . ok, I got it

Example 4

1. **H:** umm, the one I start with is [in position] one-three
2. **H:** alright?
3. **W:** ok
4. . . .
5. **H:** one yellow stripe across
6. **W:** ok, one-three, done

(a)

1. **H:** umm, it's the light pink
2. **H:** goes in the middle
3. **W:** *[correctly positions piece]*
4. **H:** and then the light purple
5. **H:** goes diagonally to the top right
6. **W:** *[correctly positions piece]*
7. **H:** yeah
8. **H:** and then red goes to the left
9. **W:** red?
10. **H:** yeah
11. **W:** *[correctly positions piece]*
12. **H:** cool

(b)

FIGURE 16. Excerpt from small field of view and manual worker control condition.

1. . . .
2. **H:** that goes directly below the dark blue one.
3. **W:** okay.
4. **H:** can you pan across please?
5. **W:** alright *[simultaneously moves the view portal across the entire puzzle]*
6. **H:** um yeah, it looks good.

(e.g., Worker: "Should I press done?"). This provides additional evidence for a cooperative model of communication in which the Worker adapts their speech patterns for their partner even though they themselves see the full workspace in all conditions.

Finally, there were several examples not captured by our coding scheme where the Helper requested explicit actions to satisfy their need for situation awareness. For example, in Figure 16 the Helper asks the Worker to move their cursor around the workspace in order to provide an updated view of the puzzle state.

To summarize, the pattern of results from this experiment are consistent with the interpretation that visual information improves task performance by supporting situation awareness as well as conversational grounding. The visual information needs to contain a field of view that encompasses not only the current task objects of interest but also a larger view of the workspace to provide an element of situation awareness. When the view was too small and constrained to the task objects, the pairs suffered in their ability to stay apprised of the overall task state as well as being aware of one another's availability. This was demonstrated by the decrease in performance and the associated increase in the production of verbal alignments when the shared field of view was small. However, the field of view did not harm the ability to describe the puzzle pieces or their spatial positioning in the same way that the rotations from the previous experiment did. In fact, the pairs showed little evidence of a change in their production of spatial descriptions with the smaller fields of view. If the workspace views were affecting grounding, we should have seen evidence of a decrease in the use of spatial deixis terms if the shared visual information no longer supported their use.

6. GENERAL DISCUSSION

In this article we presented a series of three experiments investigating the theoretical proposal that visual information serves as a resource for collaboration. Across the studies we established broad support for a cooperative model of communication and demonstrated detailed support for the notion that visual information is a critical resource for both conversational grounding and situation awareness. In addition, we examined how particular features of visual information interact with features of the task to influence both of these coordination mechanisms. Figure 17 presents a recap of the findings from the three experiments and highlights the evidence each provides toward distinguishing between the two theoretical proposals. In the remainder of this

FIGURE 17. Overview of hypotheses, quantitative results, and implications for situation awareness and conversational grounding.

Short Description	Exp 1	Exp 2	Exp 3	General Findings	Impact on Situation Awareness and Conversational Grounding
H1: Pairs perform quicker when they have a shared view	+		+	Pairs exhibit ≈30-40% faster performance when going from no shared visual information to having shared visual information.	Ambiguous results about whether it is situation awareness, conversational grounding, or both that play a role.
H2: Pairs perform slower when the linguistic complexity of the objects increases	+	+	+	Pairs exhibit ≈30-40% faster performance when the linguistic complexity of the task objects increases.	This suggests that when referential grounding is required, the pairs are slower to complete the task. Support for the notion that conversational grounding plays a central role in coordination.
H3: A shared view area will have additional benefits when the linguistic complexity increases	+	+	**partial**	Experiments 1 and 2 demonstrate added benefit to immediately available visual information when the pieces are linguistically complex plaids.	Experiment 2 demonstrates strong evidence consistent with the notion that conversational grounding is a critical mechanism supported by shared visual information. Experiment 3 provides partial support for the notion that situation awareness is also a critical mechanism supported by visual information.
H4: Delay in transmission will weaken the value of a shared view	+	+		Experiments 1 and 2 demonstrate strong support for the hypothesis that a delay in the immediacy of the visual information (in various forms) weakens the value of the visual information.	
H5: Pairs will perform quicker when they share spatial perspective because the visual information supports conversational grounding		+		Pairs were over 55% faster when their views were aligned than when they were rotated.	Provides unambiguous evidence that conversational grounding is a central mechanism supported by shared visual information.
H6: An immediately available view will have additional benefit when the shared views are rotated		+		Pairs gained additional benefit from immediate visual information when the views were misaligned.	
H7: An identical viewpoint onto the work area will have additional benefit when the linguistic complexity of the objects increases		***ns***		While the data were trending in the expected direction, the difference failed to reach a significant level ($p = .13$).	
H8: Pairs will perform quicker when larger shared fields of view are available because the additional visual information supports situation awareness			+	Pairs are ≈27% faster when going from no shared visual information to a small amount, ≈12.6% faster in going from a small view to a large view, and ≈7.8% faster when going from a large to a full view.	This evidence suggests that both situation awareness and conversational grounding play a role. It also suggests that conversational grounding has a greater impact on performance than task awareness in our configuration.
H9: The benefit of a larger shared field of view will not interact with the linguistic complexity of the task because pairs are primarily using the visual information for situation awareness			**partial**	As expected, there was no difference between the Large and Full. However, there was an interaction between the Small and Large field of view sizes.	This evidence provides partial support for the notion that situation awareness plays an independent role in performance. However, the significant interaction in the range between the Small and Large field of views suggests that grounding is also affected by field of view.

section, we discuss the theoretical and practical implications of our findings as well as the limitations and future directions of this work.

6.1. Theoretical Implications

As illustrated in Figure 17, our findings support Hypotheses 1 to 4, and the general notion that shared visual information about the workspace supports communication and coordination. These findings replicate previous work and demonstrate that collaborative pairs perform more quickly and accurately when they share a common view of a workspace (Clark & Krych, 2004; Gergle et al., 2004b; Gergle et al., 2006; Kraut, Gergle, et al., 2002). The pairs were approximately 30% to 40% faster when there was immediately available shared visual information as compared to when none was available. The value of this information, however, depended on the features of the task. Its value increased when the task objects were linguistically complex and not part of the pairs' shared lexicon. However, even a small delay to the transmission of the visual information severely disrupted its value.

Similar to our previous studies, we see that the Workers become much more active in coordinating communication when shared visual information is not available (Gergle et al., 2004b). For instance, they increase their spoken contributions when the Helper does not have a shared view of the workspace. Recall that the Workers see the same basic workspace regardless of condition; thus they are shifting their language use to accommodate to their partner's context. Such behavior is predicted by a cooperative model of communication. Note that it is also easier for the Worker to produce this information than it would be for the Helper to continually query the Worker about her degree of understanding. Grounding theory would also predict this, suggesting that the communication costs are distributed among the partners, and initiative should shift in a way that is most efficient for the pair as a whole.

The studies also provide new evidence that shared visual information benefits collaboration by independently supporting both situation awareness and conversational grounding. In Experiment 2, we examined Hypotheses 5 to 8 and demonstrated the benefit shared visual information has on conversational grounding. By rotating the Helper's view we were able to degrade the ability of the pairs to ground piece descriptions and spatial relations while keeping situation awareness intact. We found that pairs were more than 55% faster when their views were aligned. When the shared view was rotated, the pairs could no longer describe the pieces using their intrinsic spatial properties, nor could they easily describe the location of the pieces using efficient and unambiguous referring expressions such as "to the right of" or "above." Instead, the pairs had to rely on more ambiguous expressions such as "by," "around," or "move it nearby." When this was the case, it was even more critical for the pairs to have real-time visual feedback at their disposal. This demonstrates an independent influence of shared visual information on the ability of the pairs to perform conversational grounding, whereas their ability to track the task state remains intact as evidenced by the fact that there was no change in the production of verbal alignments.

In Experiment 3, our findings support the notion that situation awareness plays a critical coordination role that can be affected by the particular features of the shared visual space. On the whole, we found the pairs were approximately 27% faster when shifting from no visual information to a small shared field of view. We attribute this gain to the fact that the pairs now had visual access to the pieces and used it to support the grounding needed to refer to the puzzle pieces. However, when going from a small shared field of view to a larger field of view, we reasoned that the main benefit would be increased access to the surrounding workspace with added benefits for situation awareness. From a small field of view to a large field of view, the pairs were 12.6% faster, whereas they received an additional 7.8% boost in performance when going from a large view to a full view. These findings suggest that the pairs made use of the increased field of view to maintain a more accurate model of the task state (in support of Hypothesis 8). These claims are further supported by the communication patterns observed. When the field of view was reduced, the pairs could still use efficient deictic references such as "that one," to ground their referential descriptions. However, they had more difficulty confirming the task state, recognizing when actions were completed, and knowing when their partner was available for additional instructions. This is shown by the fact that there is an increase in the use of verbal alignments when the pairs have smaller fields of view—a compensation for the fact that awareness is not being provided via the constrained visual feedback.

Together the findings from Experiments 2 and 3 demonstrate support for a theoretical distinction between the value of shared visual information for grounding and situation awareness. We also extend Clark and Brennan's (1991) hypothesis that different communication features change the cost of achieving common ground and show these features also change the cost of achieving situation awareness. Furthermore, we provide additional evidence that the features of the visual space interact with the particular features of the task (as proposed in Kraut, Fussell, et al., 2002; Kraut et al., 2003).

These results paint a new picture whereby the particular communication media (e.g., video) has a set of important subfeatures (e.g., spatial alignment or field of view), and these subfeatures interact with particular task attributes (e.g., linguistic complexity, discriminability, etc.) in ways that can differentially affect the coordination mechanisms that underlie successful communication and collaborative performance. This more nuanced understanding can be used to inform the development of collaborative systems, particularly those systems that are meant to support tightly coupled collaborative activities that involve joint physical or virtual manipulation of objects that occur simultaneously with spoken communication. In this next section we examine the practical design implications of our findings.

6.2. Practical Design Implications

By identifying the ways in which visual information impacts collaborative behavior, we can begin to make informed design decisions regarding when and how to support visual information in collaborative applications. In this section we describe

several concrete examples of real-world designs and detail how they may be impacted by differential needs for shared visual information.

Our data demonstrate the importance of understanding the task when determining the value of providing visual support. As previously discussed, tasks may vary on several dimensions. The rate of change of the objects within the environment might be quick, as in the case of a rapidly changing world found in a massively multiplayer online role-playing game. In this case, delays to the visual feedback will impact people's ability to maintain updated situational models of the current environment. Conversely, the task state, objects, and environment might change at a relatively slow pace as in the case of a system that supports collaborative 3D architectural planning. For such an application, it may be more suitable to spend effort establishing tools to support conversational grounding (such as remote telepointers or methods for indicating landmarks) so that an architect can easily discuss the details of the model with a client who most likely lacks the domain knowledge to speak in professional architect terms.

In the architecture example just mentioned, there is a disparity in knowledge between the roles of the group members. Here, the architect may have specific domain knowledge that a client lacks. In this case, conversational grounding is likely to be a critical attribute to support interaction. However, some tasks may rely more on shared visual information to support successful situation awareness—as is the case with air traffic control systems. Here an effective domain-specific language exists for the controllers to communicate. This domain language, which is often federally mandated, serves the purpose of ensuring grounded conversation between controllers. What remains is a need to quickly and efficiently establish situation awareness with other controllers and knowing when to pass off control of airplanes transitioning between airspaces. In this case, the shared visual information is used to support situation awareness in a way that is crucial for success. Ensuring that shared visual displays support the formation of situation awareness by making the entities and environmental states visually salient is critical to the design of a successful system.

As with most user-centered designs for collaborative systems, a major first step in the design process is to understand the details of the task, the environment, and the roles and social structures of the group members involved. Once these are understood, then understanding how the proposed applications will impact the availability of shared visual information can be considered in the relative light of the task requirements. Understanding a group's need for particular coordination mechanisms, and then understanding how these mechanisms are impacted by particular technical limitations, underlies the successful implementation of systems to support tightly coupled collaborations.

6.3. Limitations and Future Directions

Maintaining a conceptual distinction between situation awareness and conversational grounding is useful from both theoretical and practical perspectives. Doing so provides insight into how these mechanisms impact collaboration and provides

knowledge that can be applied to future designs. However, although we have provided evidence of the independent existence and influence of these mechanisms, the two can often be difficult to disentangle in real-world tasks as well as in the laboratory. For example, the small field of view in the third study provides benefits for grounding by allowing Helpers to see the piece being manipulated, but it also provides situation awareness over and above what is available in the condition with no shared view.

Part of the difficulty in distinguishing between grounding and situation awareness stems from the fact there are a wide range of grounding activities that take place during a typical collaborative interaction. Some grounding tasks, such as establishing joint reference to an object, are relatively easy to differentiate from awareness activities. This is the case with the majority of the evidence provided in this article. However, there are other grounding activities such as maintaining awareness of shared understanding, where the conceptual clarity is blurred and as a result it is difficult to differentiate grounding activities from awareness activities. In such cases it is important to keep in mind that the particular features of the visual environment may simultaneously provide support for both grounding and situation awareness.

A further way in which the conceptual distinction between grounding and awareness is blurred is evident in the third study when considering the difference between the small and large field of view. Both views support grounding on the piece name, so in this way the difference in the size of the shared view does not differentially influence grounding. However, in our experimental paradigm the pairs also need to ground on location names. An unexpected side effect of our experimental design is that the small field of view may have disrupted grounding on the location names, which then induced initial placement errors. Although we attempted to rigorously control the shared visual space to isolate its influence on situation awareness, in practice this appeared to influence an aspect of grounding as well. Future research needs to work to create a cleaner distinction between situation awareness and grounding as well as explore the ways in which these two coordination mechanisms interrelate in practice.

Another potential drawback of the current work is its use of the stylized puzzle task. The strength of this paradigm is that it allows precise control over the characteristics of the visual space and task along with precise measurements of performance and communication. This level of control has proven useful for providing insight into the interdependencies that exist between language functions and physical actions commonly observed in collaborative physical tasks. However, a possible limitation of this paradigm is that the puzzle task oversimplifies these interdependencies because of the limited range of instructional utterances and worker actions that are possible. However, it is important to note that many more complex, real-world tasks, whether they be remotely instructing the repair of a transformer, jointly building a LEGO house, or simply discussing a journal article with a coauthor located across the globe, comprise the same sorts of object identification-object positioning sequences we have studied here. Thus, we believe that our findings regarding the relationships among base-level actions and language are likely to hold even when tasks involve a much more complex range of activities. However, future research is needed to address the generality of these findings.

7. CONCLUSION

Visual information about a partner and the shared objects that comprise a collaborative activity provides many critical cues for successful collaboration. Visual information impacts situation awareness by providing feedback about the state of a joint task and facilitates conversational grounding by providing a resource that pairs can use to communicate efficiently. Technologies to support remote collaboration can selectively disrupt people's ability to use visual information for situation awareness and grounding, and the extent of this disruption depends in part on task characteristics such as the linguistic complexity of objects. The results clarify basic principles of communication and interaction, and provide insights for the design of future collaborative technologies.

NOTES

Acknowledgments. We thank Daniel Avrahami, Susan Brennan, Gail Kusbit, and the anonymous reviewers for their comments and feedback on drafts of this manuscript. We also thank Lynne Horey, Kathleen Geraghty, and Patti Bao for their work on the content analysis, and research assistants Lisa Auslander, Kenneth Berger, Megan Branning, Sejal Danawala, Darrin Filer, James Hanson, Matthew Hockenberry, John Lee, Gregory Li, Katelyn Shearer, and Rachel Wu for their support in developing the experimental apparatus, collecting the data, and performing behavioral coding.

Support. This research was supported by the National Science Foundation (IIS-#9980013, #0705901, #0953943). Any opinions, findings, and conclusions or recommendations expressed in this article are those of the authors and do not necessarily reflect the views of NSF.

Authors' Present Addresses. Darren Gergle, 2240 Campus Drive, Northwestern University, Evanston, IL 60660, USA. E-mail: dgergle@northwestern.edu. Robert E. Kraut, 5000 Forbes Avenue, Carnegie Mellon University, Pittsburgh, PA 15213, USA. E-mail: robert.kraut@cmu.edu. Susan R. Fussell, 332 Kennedy Hall, Cornell University, Ithaca, NY 14853, USA. E-mail: sfussell@cornell.edu.

HCI Editorial Record. First manuscript received October 17, 2005. Revisions received June 16, 2010, and April 15, 2011. Accepted by Steve Whittaker. Final manuscript received September 8, 2011. — *Editor*

REFERENCES

Anderson, A. H., Bader, M., Bard, E. G., Boyle, E. H., Doherty, G. M., Garrod, S. C., … Weinert, R. (1991). The HCRC map task corpus. *Language & Speech, 34*, 351–366.

Bolstad, C. A., & Endsley, M. R. (1999). Shared mental models and shared displays: An empirical evaluation of team performance. *Proceedings of the 43rd Meeting of the Human Factors & Ergonomics Society*. Houston, TX: Human Factors and Ergonomics Society.

Boyle, E. A., Anderson, A. H., & Newlands, A. (1994). The effects of visibility on dialogue and performance in a cooperative problem solving task. *Language & Speech, 37*, 1–20.

Brennan, S. E. (1990). *Seeking and providing evidence for mutual understanding* (Unpublished doctoral thesis). Stanford University, Stanford, CA.

Brennan, S. E. (2005). How conversation is shaped by visual and spoken evidence. In J. C. Trueswell & M. K. Tanenhaus (Eds.), *Approaches to studying world-situated language use: Bridging the language-as-product and language-as-action traditions* (pp. 95–129). Cambridge, MA: MIT Press.

Clark, H. H. (1996). *Using language*. New York, NY: Cambridge University Press.

Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In L. Resnick, J. Levine, & S. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 127–149). Washington DC: American Psychological Association.

Clark, H. H., & Krych, M. A. (2004). Speaking while monitoring addressees for understanding. *Journal of Memory & Language, 50*, 62–81.

Clark, H. H., & Marshall, C. R. (1981). Definite reference and mutual knowledge. In A. K. Joshi, B. L. Webber, & I. A. Sag (Eds.), *Elements of discourse understanding* (pp. 10–63). Cambridge, UK: Cambridge University Press.

Clark, H. H., & Schaefer, E. F. (1989). Contributing to discourse. *Cognitive Science, 13*, 259–294.

Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition, 22*, 1–39.

Cohen, K. (1982). Speaker interaction: Video teleconferences versus face-to-face meetings. *Proceedings of the Teleconferencing and Electronic Communications*. Madison: University of Wisconsin Press.

Convertino, G., Mentis, H. M., Rosson, M. B., Carroll, J. M., Slavkovic, A., & Ganoe, C. H. (2008). Articulating common ground in cooperative work: Content and process. *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI 2008)*. New York, NY: ACM Press.

Convertino, G., Mentis, H. M., Rosson, M. B., Slavkovic, A., & Carroll, J. M. (2009). Supporting content and process common ground in computer-supported teamwork. *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI 2009)*. New York, NY: ACM Press.

Daft, R., & Lengel, R. (1986). Organizational information requirements, media richness and structural design. *Management Science, 32*, 554–571.

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.

Doherty-Sneddon, G., Anderson, A. O'Malley, C. & Langton, S., Garrod, S., & Bruce, V. (1997). Face-to-face and video mediated communication: A comparison of dialog structure and task performance. *Journal of Experimental Psychology: Applied, 3*, 105–125.

Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors, 37*, 32–64.

Endsley, M. R., & Garland, D. J. (2000). *Situation awareness analysis and measurement*. Mahwah, NJ: Erlbaum.

Fussell, S. R., Kraut, R. E., & Siegel, J. (2000). Coordination of communication: Effects of shared visual context on collaborative work. *Proceedings of the ACM Conference on Computer Supported Cooperative Work (CSCW 2000)*. New York, NY: ACM Press.

Fussell, S. R., Setlock, L. D., & Kraut, R. E. (2003). Effects of head-mounted and scene-oriented video systems on remote collaboration on physical tasks. *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI 2003)*. New York, NY: ACM Press.

Gergle, D., Kraut, R. E., & Fussell, S. R. (2004a). Action as language in a shared visual space. *Proceedings of the ACM Conference on Computer Supported Cooperative Work (CSCW 2004)*. New York, NY: ACM Press.

Gergle, D., Kraut, R. E., & Fussell, S. R. (2004b). Language efficiency and visual technology: Minimizing collaborative effort with visual information. *Journal of Language & Social Psychology, 23*, 491–517.

Gergle, D., Kraut, R. E., & Fussell, S. R. (2006). The impact of delayed visual feedback on collaborative performance. *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI 2006)*. New York, NY: ACM Press.

Gergle, D., Millen, D. E., Kraut, R. E., & Fussell, S. R. (2004). Persistence matters: Making the most of chat in tightly-coupled work. *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI 2004)*. New York, NY: ACM Press.

Gutwin, C., & Greenberg, S. (2002). A descriptive framework of workspace awareness for real-time groupware. *Computer Supported Cooperative Work, 11*, 411–446.

Hollan, J., & Stornetta, S. (1992). Beyond being there. *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI 1992)*. New York, NY: ACM Press.

Hupet, M., Seron, X., & Chantraine, Y. (1991). The effects of the codability and discriminability of the referents on the collaborative referring procedure. *British Journal of Psychology, 82*, 449–462.

Karsenty, L. (1999). Cooperative work and shared context: An empirical study of comprehension problems in side by side and remote help dialogues. *Human–Computer Interaction, 14*, 283–315.

Krauss, R. M., & Bricker, P. D. (1967). Effects of transmission delay and access delay on the efficiency of verbal communication. *Journal of the Acoustical Society of America, 41*, 286–292.

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. 137–164). Cambridge, MA: MIT Press.

Kraut, R. E., Fussell, S. R., & Siegel, J. (2003). Visual information as a conversational resource in collaborative physical tasks. *Human Computer Interaction, 18*, 13–49.

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. *Proceedings of the ACM Conference on Computer Supported Cooperative Work (CSCW 2002)*. New York, NY: ACM Press.

Kraut, R. E., Miller, M. D., & Siegel, J. (1996). Collaboration in performance of physical tasks: Effects on outcomes and communication. *Proceedings of the ACM Conference on Computer Supported Cooperative Work (CSCW 1996)*. New York, NY: ACM Press.

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

Monk, A., & Watts, L. (2000). Peripheral participation in video-mediated communication. *International Journal of Human–Computer Studies, 52*, 933–958.

Nardi, B., Schwarz, H., Kuchinsky, A., Leichner, R., Whittaker, S., & Sclabassi, R. T. (1993). Turning away from talking heads: The use of video-as-data in neurosurgery. *Proceedings of the ACM Conference on Human Factors in Computing Systems (INTERCHI 1993)*. New York, NY: ACM Press.

Nardi, B., & Whittaker, S. (2002). The place of face to face communication in distributed work. In P. Hinds & S. Kiesler (Eds.), *Distributed work* (pp. 83–113). Cambridge, MA: MIT Press.

O'Conaill, B., Whittaker, S., & Wilbur, S. (1993). Conversations over video conferences: An evaluation of the spoken aspects of video-mediated communication. *Human–Computer Interaction, 8*, 389–428.

Olson, G. M., & Olson, J. S. (2000). Distance matters. *Human–Computer Interaction, 15*, 139–178.

Ranjan, A., Birnholtz, J. P., & Balakrishnan, R. (2007). Dynamic shared visual spaces: Experimenting with automatic camera control in a remote repair task. *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI 2007)*. New York, NY: ACM Press.

Schober, M. F. (1993). Spatial perspective-taking in conversation. *Cognition, 47*, 1–24.

Schober, M. F. (1995). Speakers, addressees, and frames of reference: Whose effort is minimized in conversations about locations. *Discourse Processes, 20*, 219–247.

Short, J., Williams, E., & Christie, B. (1976). *The social psychology of telecommunications*. New York, NY: Wiley.

Snook, S. A. (2000). *Friendly fire: The accidental shootdown of U. S. Blackhawks over northern Iraq* (Vol. 280). Princeton, NJ: Princeton University Press.

Velichkovsky, B. (1995). Communicating attention: Gaze position transfer in cooperative problem solving. *Pragmatics & Cognition, 3*, 199–224.

Whittaker, S. (2003). Things to talk about when talking about things. *Human–Computer Interaction, 18*, 149–170.

Whittaker, S., Geelhoed, E., & Robinson, E. (1993). Shared workspaces: How do they work and when are they useful? *International Journal of Man–Machine Studies, 39*, 813–842.

Whittaker, S., & O'Conaill, B. (1997). The role of vision in face-to-face and mediated communication. In K. E. Finn, A. J. Sellen, & S. B. Wilbur (Eds.), *Video-mediated communication* (pp. 23–49). Mahwah, NJ: Erlbaum.

Williams, E. (1977). Experimental comparisons of face-to-face and mediated communication: A review. *Psychological Bulletin, 84*, 963–976.