Coordinating spatial referencing
using shared gaze

Mark B. Neider
University of Illinois at Urbana-Champaign, Urbana, Illinois

Xin Chen
Stony Brook University, Stony Brook, New York

Christopher A. Dickinson
Appalachian State University, Boone, North Carolina

and

Susan E. Brennan and Gregory J. Zelinsky
Stony Brook University, Stony Brook, New York

To better understand the problem of referencing a location in space under time pressure, we had two remotely located partners (A, B) attempt to locate and reach consensus on a sniper target, which appeared randomly in the windows of buildings in a pseudorealistic city scene. The partners were able to communicate using speech alone (shared voice), gaze cursors alone (shared gaze), or both. In the shared-gaze conditions, a gaze cursor representing Partner A's eye position was superimposed over Partner B's search display and vice versa. Spatial referencing times (for both partners to find and agree on targets) were faster with shared gaze than with speech, with this benefit due primarily to faster consensus (less time needed for one partner to locate the target after it was located by the other partner). These results suggest that sharing gaze can be more efficient than speaking when people collaborate on tasks requiring the rapid communication of spatial information. Supplemental materials for this article may be downloaded from http://pbr.psychonomic-journals.org/content/supplemental.

Many collaborative human activities require that people attain joint attention on an object of mutual interest (e.g., Baldwin, 1995; Baron-Cohen, 1995; Clark & Wilkes-Gibbs, 1986). This often requires *spatial referencing*—the communication and confirmation of an object's location—and this spatial referencing component of a joint attention task is often time critical. Seconds matter when one person from a search-and-rescue team needs corroboration from another or when two agents monitoring a dynamic environment need to reach consensus on a potential threat.

As with communication more generally, coordinating a joint activity such as spatial referencing can be analyzed using grounding theory (Brennan, 2005; Clark, 1996; Clark & Brennan, 1991; Clark & Wilkes-Gibbs, 1986; Gergle, Kraut, & Fussell, 2004), which proposes that collaborators monitor and coordinate their behavior to minimize the collective effort expended in joint action. During coordination, different communication modalities and the strategies they enable (e.g., pointing or speaking) incur different costs and benefits (Clark & Brennan, 1991). Communicating a referent in space is relatively easy when two collaborators are copresent; one need only point to the target and call out "Right there!" But although pointing cues are effective for spatial referencing (Brennan, 2005; Hanna & Brennan, 2007), both finger and gaze pointing can be imprecise (Pechmann & Deutsch, 1982; Schmidt, 1999). A target must be triangulated on the basis of information from another's hand or face, with precision decreasing as the angle of offset or the distance increases (Gibson & Pick, 1963; Pusch & Loomis, 2001). Even more problematic, such deictic (pointing) gestures require the partners to be visually copresent in order to see where each other is orienting, which is often impractical or impossible.

Using speech to communicate spatial information, on the other hand, can be difficult and time consuming, because individuals often lack a shared reference frame to specify precise target coordinates (Logan, 1995; Logan & Sadler, 1996). Rather, they must construct impromptu coordinate systems and use potentially ambiguous referring expressions that take several speaking turns to succeed (e.g., Partner A: "Building on the right of the—the church it's—it's like on the right side of it with the darkened windows." Partner B: "Which one?" See also Brennan, 2005; Carlson & Logan, 2001; Garrod & Anderson,

M. B. Neider, mneider@uiuc.edu

1987). Although early studies of remotely mediated interaction concluded that communication is substantially more efficient when partners can speak (for a discussion, see Brennan & Lockridge, 2006), conversation unfolds in spoken turns, and these turns take time to unfold. Grounding theory therefore predicts potentially large speech costs when coordinating spatial referencing.

A potentially better way to communicate spatial information combines deictic cues, such as cursor movements, with speech (e.g., Brennan, 2005). Cursors are efficient tools for communicating spatial information, since they are immune to the triangulation errors that limit the usefulness of finger pointing or line-of-sight estimates. When used with cursors, speech would presumably serve mainly an alerting function in a spatial-referencing task, enabling one person to easily tell another when to use the deictic information from the cursor. Because the spatial position indexed by a cursor is essentially ambiguous (Jacob, 1995), sometimes indicating an intentional attempt to point and other times being entirely spurious (Brennan, 2005), speech would offer a means of resolving this ambiguity.

One particularly interesting type of cursor indicates, moment by moment, where another person is looking in a display. Such gaze cursors (e.g., Velichkovsky, 1995) are preferable to manual cursors in that they do not require a flat surface, a mouse, or a free hand to move them, making them better suited for mobile teams faced with time-critical decisions. Recently, we implemented gaze cursors bidirectionally, creating a fully collaborative shared gaze system that enabled two partners to be mutually aware of where the other was looking (Brennan, Chen, Dickinson, Neider, & Zelinsky, 2008; see also Carletta et al., 2010). The task was collaborative visual search; pairs of remotely located partners searched together for an "O" target among "Q" distractors, and the first person to find the target could conclude the search for both (no consensus was required). We found that partners searching using only shared gaze were nearly twice as efficient as solitary searchers, and significantly more efficient than searchers using only speech (a shared gaze benefit). Remarkably, using speech with shared gaze was substantially less efficient than using shared gaze alone; in that simple task, the costs associated with speaking outweighed any additional coordination benefits (supporting a speech cost hypothesis; Brennan et al., 2008).

Our present goal is to better understand how remotely located people coordinate their behavior in a more complex collaborative task, one requiring not only locating a spatial target, but also reaching consensus on it. Efficient consensus requires that the partner who finds the target first communicate its location to the other. We examined this using three communication conditions—shared gaze only (SG), speech only (shared voice, SV), and shared gaze plus speech (SG+V)—and a no-communication (NC) condition in which pairs prevented from communicating still needed to (both) locate the target. Our predictions follow directly from grounding theory's principle of least collaborative effort (Clark & Wilkes-Gibbs, 1986), which predicts that partners should coordinate their individual behavior to minimize their joint effort. In the case

of the SG condition, this means monitoring and using the shared gaze cursors. We therefore expected to replicate the benefit for shared gaze reported by Brennan et al. (2008). In the case of the SV and SG+V conditions, grounding theory predicts that the number of speaking turns, and hence task inefficiency, should decrease with increased reliance on shared gaze. We therefore expected the number of words and speaking turns to be high in the SV condition, relative to when the partners could combine shared gaze and speech. Finally, although speaking can incur a cost (Brennan et al., 2008), the fact that a consensus task requires input from both partners may introduce a new alerting role for speech in coordination with shared gaze, as in one partner's explicitly drawing the other's attention to the target. To the extent that such an alerting benefit may outweigh a speech cost, we expected to find faster consensus times in the SG+V condition than in the SG condition.

METHOD

Participants

Thirty-two undergraduates (16 pairs) from Stony Brook University participated for research credit. All had normal or corrected-to-normal vision by self-report.

Apparatus and Stimuli

Paired participants were seated in separate rooms, and each wore an eyetracker (Eyelink II, SR Research) equipped with a head restraint that allowed for normal speaking and controlled viewing distance (112 cm). Synchronized computers ensured that each participant saw the same stimulus and performed the same task. Computers were connected via an ethernet hub, enabling the bidirectional exchange of gaze signals in the SG and SG+V conditions. This involved displaying the x, y coordinates from each partner's eyetracker as a gaze cursor (a 1.7° yellow ring) superimposed over the other's display. An estimated 24 msec was needed to obtain a fixation position from one partner and to display the corresponding gaze cursor on the other's monitor, on the basis of 500-Hz gaze position sampling and a 100-Hz monitor refresh rate. Partners viewed an identical computer-generated city scene (28° × 21°, 800 × 600 pixels; see Figure 1). A sniper target, represented by a single red pixel, appeared randomly in one of the building windows on each trial.

Procedure and Design

Partners participated in a time-critical spatial consensus task or a no-communication paired-search task as part of four between-subjects conditions (4 participant pairs per condition, randomly assigned). In the SG condition, the partners could see each other's gaze cursor in real time. In the SV condition, they could speak to each other through an audio channel. In the SG+V condition, they could communicate using both gaze cursors and speech. In the NC condition, both participants had to locate the target independently. See the supplemental materials for representative videos from the three communication conditions.

Participants were told that they were testing a police-training simulator and that their task was to reach consensus on a sniper target as quickly as possible. To do this, they were instructed to find the target and to manually press a button while looking at it. Feedback regarding one partner's buttonpress was not provided to the other, but they were free to communicate this information (conditions permitting). Spatial referencing was operationalized as both partners reaching consensus on the target's location. A trial ended with a win if this occurred within 30 sec; otherwise, the trial ended with the virtual sniper winning. To increase time pressure and engagement, a gunshot sounded every 3 sec, beginning 6 sec into the trial, and con-

Figure 1. What a participant might see in a representative shared gaze trial. Note that the yellow gaze cursor would typically be moving over the scene, reflecting their partner's changing gaze position. No cursor was displayed in the SV and NC trials. The placement of the target, a single red pixel (the size is exaggerated in the figure for illustrative purposes), varied in position from window to window and building to building. To view this figure in color, please see the online issue of the journal.

tinued for a possible eight shots total. Partners were given identical instructions and were informed as to their communication condition, but they did not communicate prior to starting the experiment. There were 10 practice trials followed by 50 experimental trials.

RESULTS

Error rates appear in Table 1 and indicate better performance in each communication condition relative to the NC condition [$t(12) \geq 2.6$, $p < .05$]. In the NC condition, both partners found the target in the allotted 30 sec

Table 1
Error Rates (%) As a Function of Condition and Task Phase

	Condition			
	SG	SG+V	SV	NC
Consensus phase	9.5	6.0	15.8	40.0
Search phase	19.0	24.0	16.7	19.0
Total task	27.5	30.0	32.5	59.0

Note—Consensus phase errors correspond to cases in which only one partner found the target and pressed a button before the trial timed out. Search phase errors correspond to cases in which neither partner found the target and pressed a button before the trial timed out. Search phase and consensus phase errors sum to the total task error, which corresponds to cases in which one or both partners failed to fixate the target and press a button before the trial timed out. Since the participants could not communicate in the NC condition, there was technically no consensus phase in that condition, just a period between one person finding the target and the trial timing out. SG, shared gaze; SG+V, shared gaze plus speech; SV, shared voice; NC, no communication.

only 41% of the time; win rates with communication were ~70% and did not differ among conditions [$F(2,9) = 0.26$, $p = .77$].

The total time taken by both partners to fixate the target and to press a button on winning trials appears in Figure 2. These total response times (RTs) averaged 2.15 sec faster in the SG condition than in the SV condition [$t(9) = 2.3$, $p < .05$] and 3.27 sec faster in the SG+V condition than in the SV condition [$t(9) = 3.5$, $p < .01$]. Partners could communicate target locations without using words (as they did in collaborative search without consensus; Brennan et al., 2008), but in the present task, adding speech to shared gaze did not slow performance [$t(9) = 1.2$, $p = .26$] (unlike in Brennan et al., 2008). Data from the NC condition are shown for completeness, but direct comparisons to the communication conditions are not possible because of large differences in error rates.
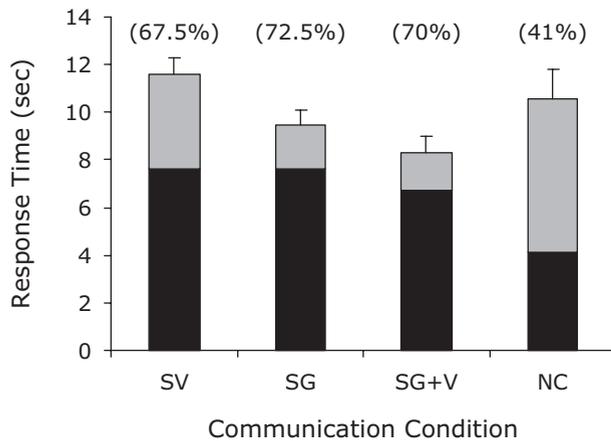
To better understand the contributions of shared gaze and speech to coordination, we divided the total RTs for all correct trials into search and consensus phases. The search phase reflects the time needed for the first partner (A) to locate the target; sharing gaze would benefit search if partners divided labor (e.g., one searching left, the other right; Brennan et al., 2008). The consensus phase captures any difficulty communicating a spatial referent and consists of the time needed for the second partner (B) to acquire the target once it was located by Partner A. This by-phase analysis distinguishes two kinds of errors (see Table 1): the target being missed by both partners

Figure 2. Manual response times for win trials, grouped by condition. The total height of each bar indicates the average total time needed for both partners to find the target and to press a button. Bar stacking represents the two components of this total task time: the average time needed for the first person of a pair (A) to press a button after finding the target (search time, black) and the average time needed for the second person (B) to do the same (consensus time, gray). Error bars indicate one standard error of the total task time mean, and the value above each bar indicates the task win rate corresponding to that condition. Note that differential win rates make it impossible to directly compare response times in the communication conditions with those in the no-communication (NC) condition (e.g., the relatively short NC search time likely results from the most difficult trials being excluded because of time-out errors). SV, shared voice; SG, shared gaze; SG+V, shared gaze plus speech.

(a search error) or by one partner (a consensus error). If speech serves an alerting function when combined with deictic information from shared gaze, consensus errors should be less frequent than search errors in the SG+V condition. No difference in error type would be expected in the SV condition. These predictions were confirmed; search errors outnumbered consensus errors in the SG+V condition [$t(3) = 3.64$, $p < .05$] but not in the SV condition [$t(3) = 0.24$, $p = .83$]. In the SG+V condition, when one partner found the target, so did the other.

With regard to timing, a shared gaze benefit in the present search-and-consensus task would establish that gaze affords a more efficient medium for communicating spatial information than does speech alone. We found large shared gaze benefits in the consensus phase; in the SG and SG+V conditions, RTs averaged 2.25 sec shorter than those in the SV condition [$t(9) \geq 2.83$, $p < .05$]. Consensus was reached no faster in the SG than in the SG+V condition [$t(9) = 0.31$, $p = .76$]. This pattern demonstrates efficient spatial referencing with shared gaze. Search phase times did not reliably differ among communication conditions [$F(2,9) = 0.47$, $p = .65$]. Partners searching complex displays while communicating did not simply divide the search labor spatially, a pattern different from what Brennan et al. (2008) found for simple displays.

Figure 3 quantifies this shared gaze advantage during the consensus phase on a fixation-by-fixation time scale. In the SV condition, Partner B took an average of 10 fixations to acquire the target after it was detected by Partner A, on the basis of a 90% criterion. These fixations suggest substantial spatial referencing difficulty with speech alone. In contrast, in the SG+V condition, Partner B was able to acquire the target within only 2 fixations of Partner A at the same 90% criterion. Having shared gaze alone yielded an intermediate result. With shared gaze, particularly in combination with speech, the spatial-referencing problem essentially vanished; after Partner A located the target, the gaze cursor revealed its location to Partner B.

To better understand the contribution of shared gaze over speech alone, we analyzed partners' speech under SV and SG+V conditions. Utterances were transcribed from the start until the end of each trial (see the supplemental materials for representative transcripts of spoken exchanges in winning SV and SG+V trials and in losing trials). Speech-only exchanges were lengthy, averaging 16.5 words per trial (including, on average, over 2.4 speaking turns). Of these trials, 90% contained spatial or scenic description, usually produced by the first partner to find the target. In contrast to those in SV trials, the exchanges in SG+V trials were terse, averaging only 4.5 words per trial [$t(6) = 3.67$, $p = .01$] and taking, on average, only one speaking turn [$t(6) = 2.44$, $p = .05$]. Moreover, only 37.9% of the SG+V trials had any descriptive content at all [different from the SV trials at $t(6) = 3.83$, $p < .01$], and their descriptive content was highly abbreviated (e.g., "top blue building, yeah"). Even more telling, 35.9% of the SG+V trials contained only deictic references without any descriptive content ("up here" or simply "mmhmm"). Speech clearly filled different functions in these conditions. Without shared gaze, speech was the only medium available to communicate target location, hence the longer exchanges and greater reliance on descriptive content. With shared gaze, speech served more of an alerting function, a means for one partner to draw the other's attention to the gaze cursor and target.

To obtain further evidence of an alerting benefit for speaking, we partitioned the SG+V trials into zero, one, and two speaking turns and compared consensus times for these trials with those in the SG and SV conditions. Trials with no turns were completely silent, and one-turn trials included purely deictic utterances in which one partner verbally alerted the other while gazing at the target. For the one-turn SG+V trials, grounding theory predicts shorter consensus times than for SG trials, because of the potential for one partner to alert the other to attend to their gaze cursor. The opposite pattern was predicted for two-turn SG+V trials; consensus should be achieved more slowly than for SG trials, because of the introduction of speech costs in the SG+V condition. The prediction for consensus times in the no-turn SG and SG+V trials is complicated by the fact that the SG+V trials had the potential for verbal alerting, whereas the SG trials did not, meaning that these conditions do not afford partners the same coordination strategies. Consensus times should therefore depend on how accurately partners monitored each other's gaze cursors.

These predictions are consistent with the pattern of means in Table 1. Consistent with the existence of speech costs, consensus times in the two-turn SG+V trials were
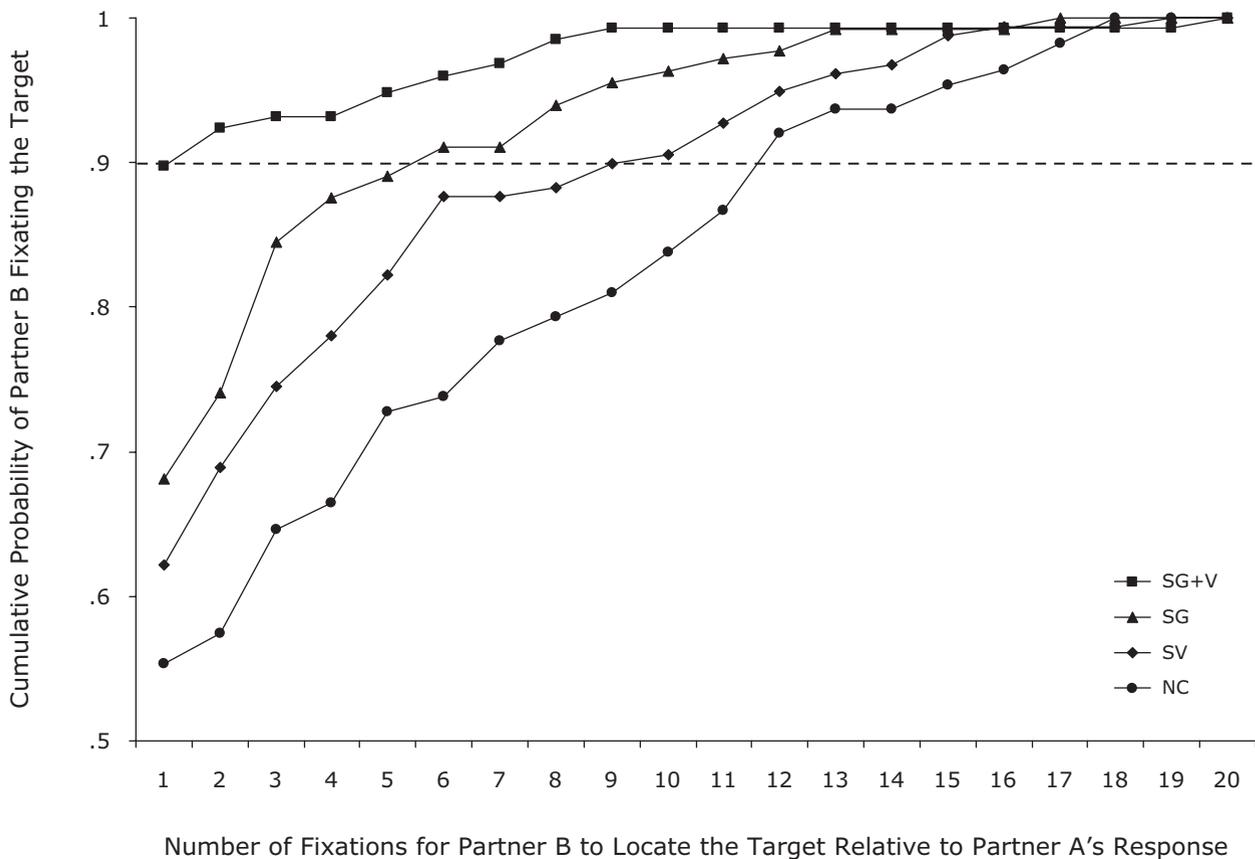
Figure 3. The cumulative probability of the second partner (here, Partner B) acquiring the target following Partner A's response, plotted as a function of fixations by Partner B in the SG+V, SG, SV, and NC conditions. The dashed line indicates the 90% criterion discussed in the text. SG+V, shared gaze plus speech; SG, shared gaze; SV, shared voice; NC, no communication.

numerically longer than those in the SG condition. Consistent with the existence of alerting benefits, consensus times in the one-turn SG+V trials were numerically shorter than those in the SG condition. Mean consensus times were also 1 sec shorter in the silent SG+V trials than those in the SG trials, suggesting that the greater opportunity for coordination in the SG+V condition resulted in more accurate monitoring of gaze cursors. Although none of these contrasts was statistically reliable ($p > .1$), partly because of instability in the data introduced by the turn-taking breakdown, we did find a reliable difference in how predictably consensus unfolded between the SG+V and SG trials with fewer than two speaking turns, with standard deviations nearly five times smaller when the partners sharing gaze could use speech to alert one another than when they could not [$F_{Levene} (5,5) = 10.66, p < .02$]. Consensus time variability was understandably greater in the SG condition because of hit-or-miss monitoring of a partner's gaze cursor; alerting by speaking in the SG+V condition reduced this monitoring variability.

We also compared win rates on silent trials in the SG+V and SV conditions. To the extent that verbal alerting speeds consensus, silence should be a poor strategy in the SG+V condition. This proved to be the case; silence was negatively correlated with winning ($r_z = -.35$,

Fisher's z correlation). Although overall win rates were comparable between the SV and SG+V conditions, silence was never a winning strategy in the SV condition (not a single silent trial was a win) and worked only half of the time in the SG+V condition. Yet despite being an error-prone strategy, no-turn trials were surprisingly common, particularly in the SG+V condition; 32.2% of all of the SG+V trials were silent, compared with only 10.3% of the SV trials [$t(6) = 3.10, p = .02$].

DISCUSSION

Communicating spatial information is challenging, particularly in visually cluttered environments and under time pressure. Collaborators faced with referencing spatial targets currently rely on lengthy verbal descriptions, or, if they are copresent, pointing. Neither of these methods is optimal for spatial referencing; they are inefficient, imprecise, or require time-consuming spatial transformations before the information can be used to direct attention.

We explored the use of a precise shared gaze cursor as a means of mediating joint attention and spatial referencing between remotely located partners. Using a time-critical task, we found that partners were faster to achieve consensus on a target's location when they could share gaze than

when they could communicate only verbally. By breaking our task into search and consensus phases, we determined that this shared gaze benefit was due almost entirely to faster spatial referencing of the target. With shared gaze, the partner first finding the target did not need to describe its location in detail, which they did need to do with speech alone. Rather, Partner A needed only to alert Partner B to shift his or her gaze to Partner A's cursor, and Partner B often did so with his or her very next saccade. However, this near optimal degree of coordination was only possible in the SG+V condition, in which speech was available to serve this alerting function. In the SG condition, partners had to rely on monitoring each other's gaze cursor for evidence of target detection, resulting in more fixations relative to the SG+V condition (Figure 3).

The fact that this early fixation advantage for the SG+V condition did not translate into shorter task completion times or consensus times (Figure 2) suggests that an alerting benefit of speech may be partly offset by the time that it takes speech to unfold. When speech was limited to serving an alerting function (one-turn SG+V trials), consensus was fast; when speech included scenic descriptions or task-irrelevant details, speech costs were incurred and consensus was slowed (Table 2). These patterns are consistent with grounding theory (Clark & Brennan, 1991) and extend this theory in an important new direction. Previous applications of grounding theory involved coordination costs and benefits almost exclusively in the context of speech; our findings show that near-optimal moment-by-moment coordination is also possible between communication media, as when speech is combined with shared gaze. However, this high level of coordination was not automatic; partner pairs in the SG+V condition failed to exploit speech/alerting benefits on a significant percentage of trials. Identifying optimal coordination strategies (ones that maximize alerting and shared gaze benefits while minimizing speech costs) during each stage of a collaborative task will be an important direction for future work.

In summary, people routinely interact over great distances using a variety of media, and these collaborations are increasingly augmented by technology. Our work describes basic research into how shared gaze and speech are coordinated on a fine-grained scale, work that informs not only the theoretical understanding of human coordination, but also the proper use of this new form of mediated communication. As shared gaze applications become more common, it will be vital to understand the unique role that this medium plays in coordinating behavior. In the context of referencing an object's location, we found that speech can be ambiguous and time consuming, whereas shared gaze is precise and fast.[1] This advantage of shared gaze over speech likely stems from a congruence between spatial attention and the communication medium; gaze cursors represent target location in a form that can be used directly by spatial attention, making the mapping of verbal information to scene elements unnecessary. When moment-by-moment coordination matters, spatial attention can be best directed to a target using a communication medium that includes shared gaze.

REFERENCES

BALDWIN, D. A. (1995). Understanding the link between joint attention and language. In C. Moore & P. J. Dunhams (Eds.), *Joint attention: Its origins and role in development* (pp. 131-158). Hillsdale, NJ: Erlbaum.

BARON-COHEN, S. (1995). The eye direction detector (EDD) and the shared attention mechanism (SAM): Two cases for evolutionary psychology. In C. Moore & P. J. Dunhams (Eds.), *Joint attention: Its origins and role in development* (pp. 41-60). Hillsdale, NJ: Erlbaum.

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

BRENNAN, S. E., CHEN, X., DICKINSON, C. A., NEIDER, M. B., & ZELINSKY, G. J. (2008). Coordinating cognition: The costs and benefits of shared gaze during collaborative search. *Cognition*, **106**, 1465-1477.

BRENNAN, S. E., & LOCKRIDGE, C. B. (2006). Computer-mediated communication: A cognitive science approach. In K. Brown (Ed. in chief), *Encyclopedia of language and linguistics* (2nd ed., pp. 775-780). Boston: Elsevier.

CARLETTA, J., HILL, R. L., NICOL, C., TAYLOR, T., DE RUITER, J. P., & BARD, E. G. (2010). Eyetracking for two-person tasks with manipulation of a virtual world. *Behavior Research Methods*, **42**, 254-265.

CARLSON, L. A., & LOGAN, G. D. (2001). Using spatial terms to select an object. *Memory & Cognition*, **29**, 883-892.

CLARK, H. H. (1996). *Using language*. Cambridge: Cambridge University Press.

CLARK, H. H., & BRENNAN, S. E. (1991). Grounding in communication. In L. B. Resnick, J. Levine, & S. D. Behrend (Eds.), *Perspectives on socially shared cognition* (pp. 127-149). San Mateo, CA: Morgan Kaufman.

CLARK, H. H., & WILKES-GIBBS, D. (1986). Referring as a collaborative process. *Cognition*, **22**, 1-39.

GARROD, S., & ANDERSON, A. (1987). Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition*, **27**, 181-219.

GERGLE, D., KRAUT, R. E., & FUSSELL, S. R. (2004). Language efficiency and visual technology: Minimizing collaborative effort with visual information. *Journal of Language & Social Psychology*, **23**, 491-517.

GIBSON, J. J., & PICK, A. D. (1963). Perception of another person's looking behavior. *American Journal of Psychology*, **76**, 386-394.

HANNA, J. E., & BRENNAN, S. E. (2007). Speakers' eye gaze disam-

Table 2
Consensus Time (in Milliseconds) for Correct Trials As a Function of Speaking Turns

Speaking Turns	Communication Condition		
	SG	SG+V	SV
0 (silent)	1,756	756	*
1	*	1,150	3,015
2	*	3,222	3,099

*All SG trials, by definition, were silent. None of the winning SV trials were silent. SG, shared gaze; SG+V, shared gaze plus speech; SV, shared voice.

biguates referring expressions early during face-to-face conversation. *Journal of Memory & Language*, **57**, 596-615.

JACOB, R. J. K. (1995). Eye tracking in advanced interface design. In W. Barfield & T. A. Furness (Eds.), *Virtual environments and advanced interface design* (pp. 258-308). New York: Oxford University Press.

LOGAN, G. D. (1995). Linguistic and conceptual control of visual spatial attention. *Cognitive Psychology*, **28**, 103-174.

LOGAN, G. D., & SADLER, D. D. (1996). A computational analysis of the apprehension of spatial relations. In P. Bloom, M. A. Peterson, L. Nadel, & M. Garret (Eds.), *Language and space* (pp. 493-529), Cambridge, MA: MIT Press.

PECHMANN, T., & DEUTSCH, W. (1982). The development of verbal and nonverbal devices for reference. *Journal of Experimental Child Psychology*, **34**, 330-341.

PUSCH, M., & LOOMIS, J. M. (2001). Judging another person's facing direction using peripheral vision. *Journal of Vision*, **1**(3), 288a.

SCHMIDT, C. L. (1999). Adult understanding of spontaneous attention-directing events: What does gesture contribute? *Ecological Psychology*, **11**, 139-174.

VELICHKOVSKY, B. M. (1995). Communicating attention: Gaze position transfer in cooperative problem solving. *Pragmatics & Cognition*, **3**, 199-222.

NOTE

1. Although there are similarities in the information conveyed by gaze and mouse cursors, their value as coordination signals in a time-critical task is different. Mouse movements are slow and intentional; gaze cursor movements are fast and instrumental to the task itself. Gaze cursors can therefore mediate coordination at a finer time scale.

SUPPLEMENTAL MATERIALS

Representative video stimuli and transcripts from this study may be downloaded from http://pbr.psychonomic-journals.org/content/supplemental.