What eye believe that you can see: Conversation, gaze coordination and visual common ground (DSiJA)

Daniel C. Richardson
University of Reading
dcr@eyethink.org

Rick Dale
University of Memphis
radale@memphis.edu

John M. Tomlinson Jr
UC Santa Cruz
otomlins@ucsc.edu

Herbert H. Clark
Stanford University
herb@psych.stanford.edu

Abstract

We can only share information because of how much we share already. Conversation is supported by the common ground between us, such as the beliefs and the visual context that we have in common. It has been shown that both of these components determine how we communicate, yet it is not clear how they interact. In a new paradigm, we separated the fact that a visual scene was shared or not and the *belief* that a visual scene was shared or not. We quantified the effects of these factors upon joint attention by measuring the coordination between conversants' eye movements. Participants had a conversation about a controversial topic, such as the Iraq war. The discussion was first framed by four short videos of actors espousing tendentious views. Participants discussed their own views while they looked at either a blank screen or pictures of the four actors. Each believed (correctly or not) that their partner was either looking at a blank screen or the same images. We found that both the presence of the visual scene and beliefs about its presence for another influenced participants' discussion and the coordination of their joint attention.

Introduction

"Can you pass me the thingy for the whatsit?" Penny asked John. It is hard to imagine a more vacuous sentence. Yet to John, it was a precise instruction. He replied, "Not too much". The content of this communication came not from the words spoken as much as the rich body of knowledge that John and Penny shared. This is termed their common ground (Clark, 1996). A conversation that morning (about the guests coming to dinner), specific knowledge that they shared (concerning one guest's tastes) and their current visual context (Penny standing in front of the stove and John in front of a particular drawer) restricted her reference to a tool that was within John's reach that would allow Penny to crush some garlic into a casserole, although not too much.

In this paper, we examine one aspect of common ground: the shared visual context. What role does this information play in the production and comprehension of spontaneous dialog? Would Penny have spoken the same oblique phrase, and would John have understood it, if he had been facing the other way? Clark and Marshall (1981) argued that we interpret ambiguous references using the co-presence heuristic. Only items seen by both conversants are considered as possible referents. This claim has been tested in various 'reference game' studies with mixed results. A speaker refers to an object which is in one or both of the participants' sight, and as a consequence there are sometimes changes in the listener's eye movements (Hanna, Tanenhaus & Trueswell, 2003) and the speaker's manner of reference (Haywood, Pickering & Branigan, 2005), and sometimes not (Keysar, Barr & Brauner, 2000; Horton & Keysar, 1996). Language does more than lead people to objects, however. It can describe people, ideas and opinions that are abstract or simply absent (Spivey & Richardson, in press). In contrast to reference game studies that have a speaker, a listener and a reference to an object, our experiment examined the role of visual context when two people have an extended conversation. The situation was analogous to John and Penny discussing the political views of their dinner guests for that evening, while looking at the meal they are to serve. Objects in the shared visual scene were not the content the utterances, but, in a more germane sense, provided a visual context for the discussion. Whether or not conversants chose to incorporate this shared visual information becomes a more interesting question, since the constraints of a referencing task do not demand that they do so.

There are two ways that visual context could play a part in a conversation (Keysar 1997). First, it provides information. The sight of certain objects or people may relieve the burden of memory or lexical access for a speaker, and help disambiguate language sounds or structure for a listener. Certainly, in speech production (Griffin & Bock, 2000; Meyer, Sleiderink, & Levelt, 1998) and comprehension

(Cooper, 1974; Kamide, Altmann, & Haywood, 2003; Knoeferle & Crocker, 2006; Richardson & Matlock, 2007; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995) eye movements to a visual scene are closely linked, moment by moment, to linguistic processes. Second, if a conversant believes that her conversational partner can see certain things, then she can interpret utterances in relation to the common ground, and make remarks relying on the fact that this information will be available to help her listener. Beliefs about shared knowledge influence speech. Speakers change their descriptions of locations in New York (Isaacs & Clark, 1987) or famous faces (Fussell & Kraus, 1992) depending upon their estimation of how much relevant knowledge the listener might have. But what of beliefs about shared visual context?

All reference game studies of visual common ground and conversation have confounded the belief that visual information is shared with the fact of it being shared. Our experiment separated these factors for the first time. A visual image was either present or absent for both conversants, and both believed that it was either present or absent for their partner. We hypothesized that both the presence of the visual context for an individual and the belief in its presence for another would influence their conversation. We captured the success of this joint activity by measuring the coordination between conversants' eye movements as they talked and looked at a shared display.

Gaze coordination and conversation

The temporal dynamics of gaze coordination are intertwined with discourse. In the first quantification of gaze coordination, Richardson and Dale (2005) recorded the eye movements of speakers talking spontaneously about a TV show while looking at pictures of its cast members. These speeches were played back to listeners who were looking at the same display. Cross-recurrence analysis (Zbilut, Giuliani, & Webber, 1998), measured the degree to which speaker and listener's eye positions overlapped at successive time lags (see below for an explanation). From the moment a speaker looked at a picture, and for the following six seconds, a listener was more likely than chance to be looking at that same picture. The overlap between speaker and listener eye movements peaked at about 2000ms. In other words, two seconds after the speaker looked at a cast member, the listener was most likely to be looking at there too. The same eye movement coupling when two participants had a live spontaneous dialog and looked at the same images (Richardson, Dale & Kirkham, 2007). On this occasion, gaze recurrence peaked at 0ms, presumably representing the average of each conversant acting as speaker and then listener. This coordination was achieved in virtue of the knowledge conversants shared. Gaze recurrence was increased when conversants heard the same (rather than different) encyclopedia passages about Salvador Dali prior to discussing one of his paintings. Closer gaze recurrence appears to facilitate communication. When pictures in a display flashed in time with the speakers' fixations, it caused listeners' eye movements to follow the speakers' more closely. Consequently, listeners answered comprehension questions faster than those who had seen a randomized sequence of flashes (Richardson & Dale, 2005). Since gaze recurrence is causally connected to what conversants know and remember, we predicted that it would reveal effects of what they see and what they believe each other can see.

Visual context for self and for others

Our conversants watched four actors give their views on a contentious topic and then discussed the topic between themselves. We manipulated two factors. First, the actors could either be present on the screen for each conversant, or absent, replaced by an empty two by two grid. We termed this as the visual context *for self*. Second, each participant was told prior to the discussion that the actors were either present or absent on the screen of their conversational partner. We termed this the visual context believed *for other*. We refer to the combinations of these conditions by stating the visual context *for self* followed by believed *for other*. For example, *present - absent* refers to the condition in which both participants could see a visual scene, but believed that their partner could not. To be clear: the two conversants were in a symmetrical situation, always looking at the same thing as each other and always believing the same thing as each other.

Conversants had beliefs that were factually incorrect in the present-absent and absent-present conditions. We inculcated these beliefs by a slight deception. At the start of each conversation, both participants read the words

'You are participant B'. They were then shown, for example, a blank screen and told, 'Participant A is looking at a blank screen. Participant B should still be looking at the pictures. Please say, 'yes' if this is true'. Both participants, seeing a blank screen and believing themselves to be participant B, said 'yes'. They also heard each other saying yes, and interpreted this as their partner, participant A, confirming that looked at a blank screen.

Our first, straightforward prediction was that the presence of a visual context for self would increase gaze recurrence, since during speech production and comprehension relevant visual objects are fixated. We did expect some recurrence between gaze patterns even when the screen was empty, however. During language comprehension and memory tasks, empty locations of a screen can be systematically fixated when a reference is made to items or events that were previously there (Altmann, 2004; Hoover & Richardson, in press; Richardson & Spivey, 2000; Richardson & Kirkham, 2004; Spivey & Geng, 2001). The more contentious question is what will be the effect of the visual context that is believed for others. We put forward three possibilities: (1) visual context is ignored, and so there will be no effect of beliefs about it, (2) visual context is exploited, and so the belief that more of it is shared will increase gaze coordination (3) visual context is compensated for, and so the belief that it differs between conversants will increase gaze coordination. These possibilities are not exhaustive, but there is support for each in the literature.

(1) Ignoring visual context. Listeners can seem strikingly egocentric. They ignore, in the first instance at least, the fact that a speaker's visual perspective differs from their own (Keysar, Barr & Horton, 1998). In a reference game, an array of objects was placed between the participant and a confederate (Keysar, et al, 2000). Some of the objects were occluded so that only the participant could see them. For example, the participant could see three candles of different sizes, but the confederate could only see the larger two. When the participants were asked for 'the smallest candle', they were more likely to look at the very smallest candle. Since it could not be seen by the confederate, it could not have been the intended referent. Therefore, mutual knowledge is a non-existent or partial constraint upon speech comprehension. In our case, conversants are not even directing each other to pick up objects but discussing current affairs. The prediction from these results is that when we manipulate the visual context that is believed for others it will have no effect on behaviour.

(2) Exploiting visual context. Subsequent work has suggested that the 'partial constraint' of mutual knowledge can become dominant with slightly different participants or circumstances. Native speakers of Mandarin come from a culture that has a greater focus on other people during social interactions (Markus & Kitayama, 1991). In the same reference game, they almost never failed to take into account the speaker's visual perspective (Wu & Keysar, 2007a). However, even English speaking participants are not insensitive to common ground constraints. When two possible candidates for 'the smallest candle' were on display, listeners immediately fixated the candle that was in the visual common ground, and ignored the one that was blocked from the speaker's view (Hanna, Tanenhaus & Trueswell, 2003). When a speaker begins to ask a question about an object, a listener is more likely to fixate those that are hidden from the speaker's view (Brown-Schmidt, Gunlogson & Tanenhaus, in press). Speakers will use names for objects that they believe to be known to the listener (Isaacs & Clark, 1987; Metzing & Brennan, 2003), though they sometimes overestimate the degree that the knowledge is shared (Wu & Keysar, 2007b). These results suggest that common ground will be used if it is available. This hypothesis predicts a main effect of the belief condition: whether or not the visual context is present *for self*, gaze coordination will increase if it is believed to be present *for others*.

(3) Compensating for visual context. Our third hypothesis is that there will be an interaction between the visual context for self and the visual context that is believed for others. When conversants believe there is a difference between what they see and what their partners can see, they will seek to redress the imbalance. Speakers use more gestures when they are describing a toy that listeners have not played with before (Gerwing & Bavelas, 2004) or a location within a picture that they have not seen (Holler & Stevens, 2007). Speakers produce better explanations when they believe their listeners do not have access to a diagram

(Bromme, Jucks & Runde 2005), and provide disambiguating information when it is not present in the visual common ground (Haywood, Pickering & Branigan, 2005). This suggests that when conversants believe there is a mismatch between their visual context and their partners' in our task, they will boost their efforts to establish common ground, leading to better gaze coordination. This hypothesis predicts that gaze coordination will be highest in the *present-absent* and *absent-present* conditions.

Methods

Participants

112 undergraduates from the University of California, Santa Cruz took part in exchange for course credit. The data from 19 participants was discarded due to failures in calibration, resulting in 37 dyads with two usable data series.

Apparatus

Each participant sat in a cubicle in a reclining chair, looking up at an arm-mounted 19" LCD 60cm away with a *Bobax3000* remote eye tracker mounted at the base. They wore a headset with a boom mic. The experimenter controlled when the participants could hear the stimuli, each other's voices, or the experimenters voice. For each participant, an iMac calculated gaze position approximately 30 times a second, presented stimuli and recorded data. A third Apple Mac computer synchronized the trials and data streams from the iMacs and saved an audio-video record of what was seen, heard and said during the experiment, superimposed with gaze positions.

Design

Four different opinion pieces were written for each of eight contentious topics (the iraq war, vegetarianism, drugs in sport, UCSC professors, UCSC campus expansion, violence in video games, online social networks, and gay marriage). The opinion pieces were delivered straight to camera by sixty four different actors, producing movies that varied from 8 to 20 seconds in length. For each pair of participants, the topics were randomly allocated across the four experimental conditions.

Procedure

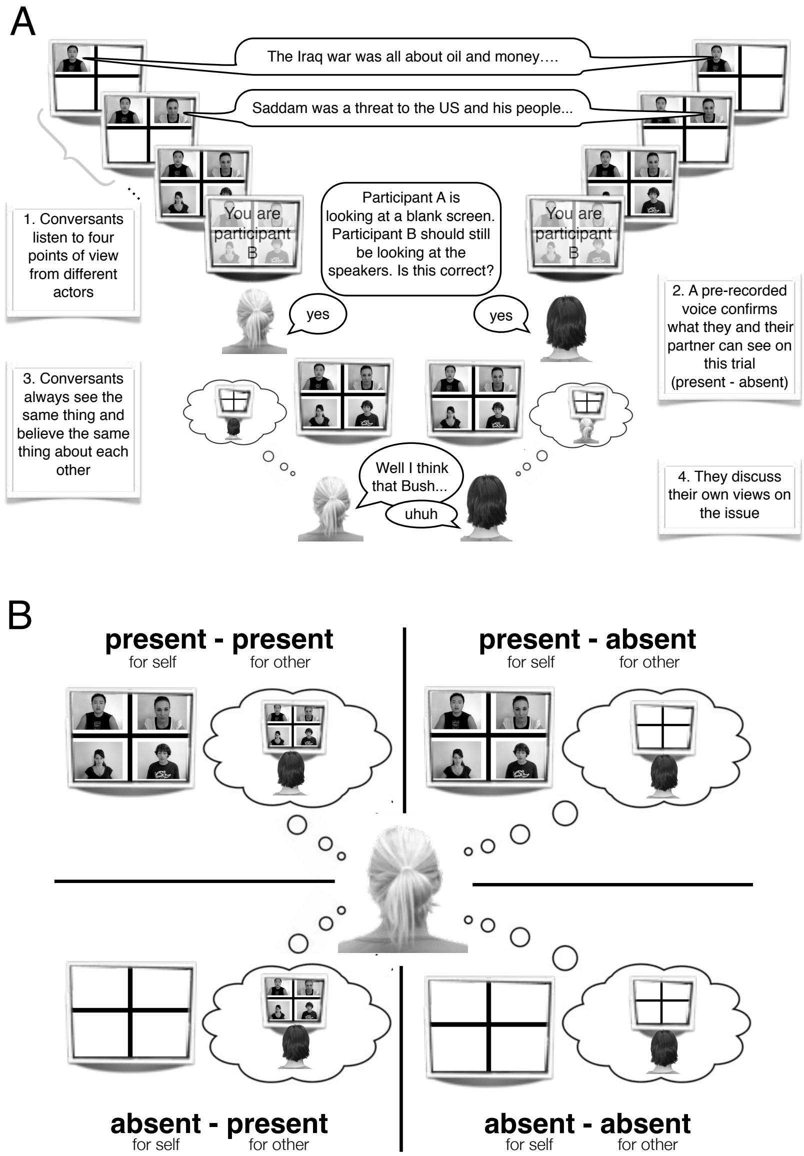
Participants were introduced to each other in the laboratory waiting room. They then sat in adjacent cubicles and underwent a brief calibration routine of roughly a minute. The trial design is shown in Figure 1A. First the movies were shown, one at a time, in each of the quadrants of a 2 x 2 grid. Location and order of presentation were randomized, but were identical for each participant. Each movie ended with a freeze frame which remained on screen. In the two *absent for self* conditions, pictures of the four actors faded from view at the end of presentations, leaving an empty grid. In the *present for self* conditions, the pictures remained in view. The words 'You are participant B' appeared on the screen *for both participants*. The participants then heard a prerecorded voice saying, "Please discuss these issues," and the experimenter activated the audio link between cubicles.

Figure 1B represents the different experimental manipulations that were introduced at this stage of the trial. In the *absent-absent* and *present - present* condition, participants heard "You should both now [be looking at a blank screen / be able to still see the speakers on screen]". In the *absent-present* and *present-absent* conditions they heard "Participant A should [be looking at a blank screen / still be able to see the speakers]. Participant B should [still be able to see the speakers / be looking at a blank screen]". Across all conditions, they were then asked, "Please say 'yes' if this is the case". Once they had affirmed, the experimenter initiated the conversation. Participants typically talked for between one and three minutes before the experimenter decided that the topic had been exhausted, and the trial was terminated.

Data analysis

We quantified the gaze coordination between conversants by generating categorical cross-recurrence plots. This technique depicts the temporal structure between time series (Zbilut, Giuliani, & Webber, 1998), and has been used to capture the subtle entrainment of body sway during conversation (Shockley, Santana & Fowler, 2003) and the interrelationships between a child and care givers' language use (Dale & Spivey, 2006). In our case, points of recurrence are defined as the times at which both conversants are fixating the same screen quadrant. For each trial, we took the first minute of eye movement data, added up all the points of recurrence and then divided by the total number of possible points to get a recurrence percentage (for a detailed explanation, see Richardson & Dale, 2005). Here, the possible points of recurrence were defined as the times at which at least one of the conversants had their eyes on the screen. The next step was to lag one of the data streams by 10ms, so that 0ms on one data stream was aligned with 10ms on the other. Again, all the points of recurrence were calculated. This represents the degree to which one conversant is looking at the same thing as the other conversant 10ms later. Since in our experiment the participants had symmetrical roles, we averaged the recurrence for a lag of 10ms and -10ms. A full cross-recurrence analysis consisted of calculating the recurrence for all possible alignments, or lag times, of the two data series.

Differences in gaze coordination would be produced by differences in how conversants spoke to each other. Therefore, our gaze analysis was supported by a secondary analysis of the speech of a subset of ten pairs of participants. We coded all eight conversations for each dyad and focused on utterances which made any reference at all to the views or appearance of the actors, as gaze coordination has been found to increase during moments of direct reference to a visual scene (Richardson & Dale, 2005). We counted the number of such references in each conversation and analysed three properties. Firstly, each reference was termed either 'visual' if it made any mention of visual properties of the actor, or 'factual' if it contained only (non-visual) information relating the actors' viewpoint. Secondly, all references that occurred at the end of phrases were coded for intonational contours called *boundary tones*. Of interest were *rising boundary tones*, in which the intonation goes up in pitch at the end of a sentence from either the lower or higher part of a speaker's pitch range (Pierrehumbert & Hirschberg, 1990). They are thought to serve as implicit requests for confirmation from the listen, and hence play a role in establishing information in the common ground (Allen, 1984; Fletcher, Stirling, Mushin & Wales, 2002). Lastly, we calculated the proportion of speakers' references that elicited a back-channel response from the the listener such as *ok*, *uh-huh*, *right*. Back-channels had to be directly following or overlapping the speakers' reference and consist of one word. These responses are thought to serve the function of coordinating dialogue and signaling to the speaker that she has been understood (Schegloff, Jefferson & Sacks, 1977), and hence, like the rising contours, their frequency serves as a indicator of conversants' efforts to establish common ground.



Results

Gaze coordination was increased by the presence of a visual context for the self, and modulated by a belief in the presence of a visual context for the other. When the visual context was absent for conversants, recurrence was higher if they believed their partners could see it. Conversely, when the visual context was present, recurrence was higher if they believed that that their partners could not. Figure 2A shows how gaze recurrence changed across conditions at different time lags. With the exception of the *absent-absent* condition, recurrence peaked at or around 0ms and trailed off as the distance between conversants' gaze patterns increased, replicating the pattern produced by spontaneous conversation that was observed by Richardson, Dale & Kirkham (2007). Following their analysis, differences between conditions were analyzed by averaging recurrence within a window of +/- 3000ms, to capture the typical periods in which both conversants were acting as speakers and listeners. A 2 (*for self: present/absent)* x 2 *(believed for other: present/absent)* ANOVA showed an significant interaction between theses effect ($F(1,36)=4.5$, $p<.05$), as well as the expected main effect of visual context ($F(1,36)=12.4$, $p<.001$). Recurrence in each of the conditions was compared to randomized baselines. Within the critical window, recurrence was higher than chance for the present-present ($t(36)=3.3$, $p<.001$), present-absent ($t(36)=3.7$, $p<.001$) and absent-present ($t(36)=2.8$, $p<.01$) conditions, but not for the absent-absent ($t(36)=1.6$).

Conversants' speech was also influenced by an interaction between the presence of a visual context for the self and a belief in the presence of a visual context for the other (see figure 2B). Whilst conversants simply made more references overall to the actors when they were on the screen in front of them, our other measures revealed interactions between conditions. In the *present-absent* and *absent-present* conditions, references were more likely to employ factual (non-visual) properties (for example, 'the guy who said that Iraq was about oil' rather than 'the guy in the red shirt who was against the war'). In these conditions, references that came at the end of phrases were more likely to end in a rising contour (as an implicit request for confirmation), and references were more likely to receive a back-channel response from the listener. Our gaze analyses show that these efforts to boost common ground in those conditions did indeed correspond to an increase in gaze recurrence.

Our measures of conversants' references were analysed by a 2 (*for self: present/absent)* x 2 *(believed for other: present/absent)* ANOVAs. There was a main effect of the presence of the visual context *for self* ($F(1,9)=8.3$, $p<.05$) on the number of references made. There were significant interactions between the visual context *for self* and *believed for other* on the proportion of references that mentioned factual properties ($F(1,9)=6.8$, $p<.05$), the proportion of references using rising contours ($F(1,9)=5.3$, $p<.05$), and the proportion of references that received back channel responses ($F(1,9)=6.35$, $p<.05$). No other main effects or interactions were significant across our four measures.

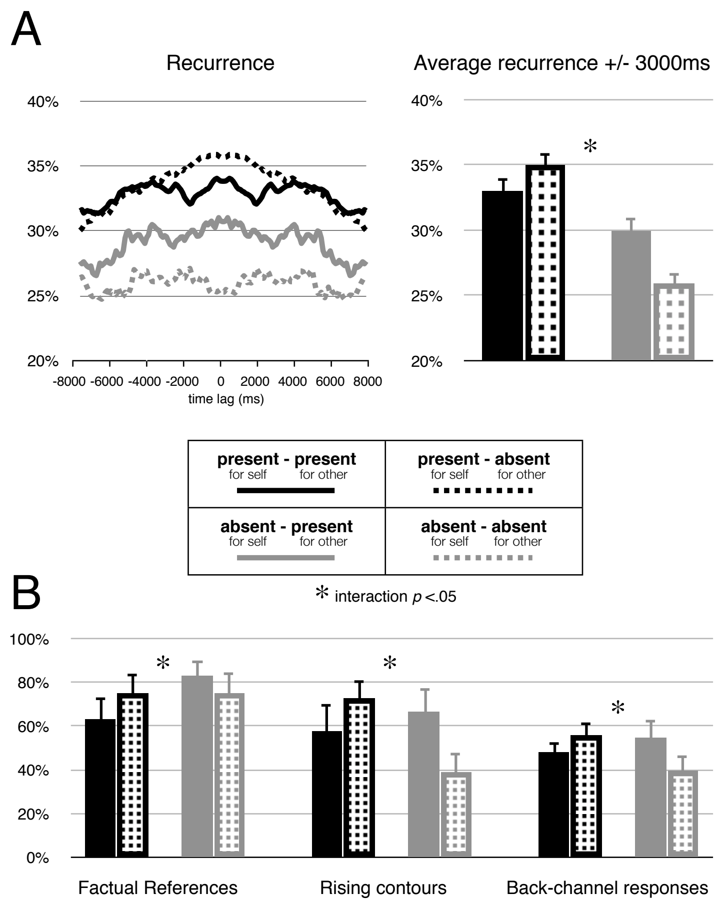


Figure 2. Results of (a) gaze analysis and (b) speech analysis

Conclusion

Reference game studies have given a mixed view of the role of visual common ground in conversation. Sometimes the visual perspective of a speaker has little effect on a listener (Keysar et al 2000), and sometimes it has an immediate constraint on an ambiguous reference to an object (Hanna et al, 2003). For a number of reasons, our experiment might have found no influence of visual common ground. The conversations were not about the actors on display, but concerned politics, sports and campus life. The effect of believing that a conversational partner could see the actors was deconfounded from the fact of them being seen by each conversant. Lastly, the particular circumstances of who saw or was believed to see what changed on a trial by trial basis, demanding that conversants keep track of shifting visual common ground constraints.

Under these conditions, conversants could have ignored the whole issue of what they believed each other could see, but they did not. They could have employed a quick and expedient technique of exploiting visual information when it was believed to be present, and ignoring it otherwise. Instead, they maintained an awareness of what each other could see on each particular trial, when they believed there to be an imbalance between their own view and their partners' they sought to compensate for that difference. They tended to refer to actors' viewpoints via a non-visual route, ask for confirmation that their messages had been understood, and signal understanding to each other via back channel responses. As a result, when they looked at the pictures their eye movements were more tightly coordinated if they believed they were talking to someone who was looking at nothing. Conversely, while looking at an empty screen their eye movements were more closely coordinated if they believed that each other could still the actors.

Coordinating joint attention is essential for successful communication (Clark, 1996; Clark & Brennan, 1991; Schober, 1993). It may even be the basis for pre-linguistic learning (Baldwin, 1995). In spontaneous conversation, people are able to couple their gaze despite ambiguities and disfluencies in the speech stream. This remarkable coordination is achieved in virtue of the background knowledge they share (Richardson, et al., 2007), their sensitivity to each others' pragmatic constraints (Hanna & Tanenhaus, 2004) and moment-by-moment domains of reference (Brown-Schmidt & Tanenhaus, in press). Here we have shown that conversants are also attuned to both their own visual context and what they believe each other can see. They will even coordinate their gaze around an empty screen in the mistaken belief that each other can see something. It is the net effect of these multiple constraints that restrict the referent of "the thingy for the whatsit" from a universe of objects to a garlic press.

References

Allen, K. (1984). The high rise terminal contour. *Australian Journal of Linguistics*, *4*, 19-32.

Altmann, G.T.M. (2004) Language-mediated eye movements in the absence of a visual world: The 'blank screen paradigm'. *Cognition*. 93, 79-87.

Baldwin, D. A. (1995). Understanding the link between joint attention and language. In C. Moore & P. J. Dunham (Eds.), *Joint attention: its origins and role in development*. Hillsdale, NJ: Lawrence Erlbaum.

Bromme, R., Jucks, R. & Runde, A. (2005). Barriers and biases in computer-mediated expert–layperson communication: An overview and insights into the field of medical advice. In R. Bromme, F.W. Hesse & H. Spada (Eds.). *Barriers and biases in computer-mediated knowledge communication—and how they maybe overcome* (pp. 89–118). New York: Springer.

Brown-Schmidt, S. & Tanenhaus, M.K. (in press). "Real-time investigation of referential domains in unscripted conversation: a targeted language game approach," *Cognitive Science*.

Brown-Schmidt, S., Gunlogson, C. & Tanenhaus, M. K. (in press). Addressees distinguish shared from private information when interpreting questions during interactive conversation. *Cognition*.

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. M. Levine & S. D. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 127-149). Washington, DC: APA.

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

Dale, R. & Spivey, M. J. (2005). Categorical recurrence analysis of child language. In *Proceedings of the 27th Annual Meeting of the Cognitive Science Society* (pp. 530-535). Mahwah, NJ: Lawrence Erlbaum.

Dale, R. & Spivey, M. J. (2006). Unraveling the dyad: Using recurrence analysis to explore patterns of syntactic coordination between children and

caregivers in conversation. *Language Learning*, 56, 391-430.
Fletcher, J., Stirling, L., Mushin, I., & Wales, R. (2002). Intonational rises and dialog acts in the Austrialian English Map Task. Language and Speech, 45(3), 229-252.
Gerwing, J., & Bavelas, J. B. (2004). Linguistic influences on gesture's form. *Gesture*, 4, 157–195.
Griffin, Z. M., & Bock, K. (2000). What the eyes say about speaking. *Psychological Science*, 11(4), 274-279.
Hanna, J. E., Tanenhaus, M. K., & Trueswell, J. C. (2003). The effects of common ground and perspective on domains of referential interpretation. *Journal of Memory & Language*, 49(1), 43-61.
Haywood, S.L., Pickering, M.J., & Branigan, H.P. (2005). Do speakers avoid ambiguities during dialogue? *Psychological Science,* 16, 362-366.
Hoover, M. A. & Richardson, D. C. (in press). When Facts Go Down the Rabbit Hole: Contrasting Features and Objecthood as Indexes to Memory. *Cognition*
Holler, J. & Stevens, R. (2007) The effect of common ground on how speakers use gesture and speech to represent size information. *Journal of Language and Social Psychology,* 26, 1-25.
Horton, W. S., and Keysar, B. (1996). When do speakers take into account common ground? *Cognition*, 59, 91-117.
Isaacs, E. A., & Clark, H. H. (1987). References in conversations between experts and novices. *Journal of Experimental Psychology: General* , 116, 26-37
Keysar, B. (1997). Unconfounding common ground. *Discourse Processes*, 24, 253-270.
Keysar, B., Barr, D. J., & Horton, W. S. (1998). The egocentric basis of language use: Insights from a processing approach. *Current Directions in Psychological Sciences*, 7, 46-50.
Keysar, B., Barr, D. J., Balin, J. A., & Brauner, J. S. (2000). Taking perspective in conversation: The role of mutual knowledge in comprehension. *Psychological Science,* 11(1), 32-38.
Knoeferle, P. & Crocker, M (2006). The coordinated interplay of scene, utterance, and world knowledge: evidence from eye tracking. *Cognitive Science*, 30, 481-529.
Markus, H., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review,* 98, 224–253
Metzing, C., & Brennan, S. E. (2003). When conceptual pacts are broken: Partner-specific effects on the comprehension of referring expressions. *Journal of Memory & Language*, 49(2), 201-213.
Meyer, A. S., Sleiderink, A. M., & Levelt, W. J. M. (1998). Viewing and naming objects: Eye movements during noun phrase production. *Cognition*, 66(2), B25-B33.
Pierrehumbert, J. & Hirshberg, J. (1990). The Meaning of Intonational Contours in the Interpretation of Discourse. In P. R. Cohen, J. Morgen, & M. E. Pollack (Eds.), *Intentions in Communications*. (pp. 271-311). Cambridge, Mass.: The MIT Press.
Richardson, D.C & Dale, R. (2005). Looking To Understand: The Coupling Between Speakers' and Listeners' Eye Movements and its Relationship to Discourse Comprehension. *Cognitive Science,* 29, 1045–1060.
Richardson, D.C., Dale, R. & Kirkham, N.Z. (2007) The art of conversation is coordination: Common ground and the coupling of eye movements during dialogue. *Psychological Science*, 18 (5), 407-413
Richardson, D. C. & Kirkham, N.Z. (2004). Multi-modal events and moving locations: Eye movements of adults and 6-month-olds reveal dynamic spatial indexing. *Journal of Experimental Psychology: General*, 133 (1), 46-62.
Richardson, D.C. & Matlock, T. (2007) The integration of figurative language and static depictions: an eye movement study of fictive motion, *Cognition*, 102, 129-138
Richardson, D. C. & Spivey, M. J. (2000). Representation, space and Hollywood Squares: Looking at things that aren't there anymore. *Cognition*, 76, 269-295.
Schegloff, E. A., Jefferson, G., & Sacks, H. (1977). The preference for self-repair in the organization of repair in conversation. *Language*, 53, 361-382.
Schober, M. F. (1993). Spatial perspective-taking in conversation. *Cognition*, 47(1), 1-24.
Shockley, K., Santana, M.V., & Fowler, C. A. (2003). Mutual interpersonal postural constraints are involved in cooperative conversation. *Journal of Experimental Psychology: Human Perception & Performance*, 29(2), 326-332.
Spivey, M. J., & Geng, J. (2001). Oculomotor mechanisms activated by imagery and memory: eye movements to absent objects. *Psychological Research,* 65, 235-241.
Spivey, M. & Richardson, D. (in press). Language embedded in the environment. In P. Robbins and M. Aydede (Eds.), *The Cambridge Handbook of Situated Cognition*. Cambridge, UK: Cambridge University Press.
Tanenhaus, M. K., Spivey Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, 268(5217), 1632-1634.
Wu, S. & Keysar, B. (2007a). Cultural effects on perspective taking. *Psychological Science,* 18, 600-606.
Wu, S., & Keysar, B. (2007b). The effect of communication overlap on communication effectiveness. *Cognitive Science,* 31, 169-181.
Zbilut, J. P., Giuliani, A., & Webber, C. L., Jr. (1998). Detecting deterministic signals in exceptionally noisy environments using cross-recurrence quantification. *Physics Letters,* 246, 122-128.