

Conversation, Gaze Coordination, and Beliefs About Visual Context

Daniel C. Richardson,^a Rick Dale,^b John M. Tomlinson^c

^a*Cognitive, Perceptual & Brain Sciences, University College London*

^b*Department of Psychology, University of Memphis*

^c*Psychology Department, University of California, Santa Cruz*

Received 26 September 2008; received in revised form 14 April 2009; accepted 6 June 2009

Abstract

Conversation is supported by the beliefs that people have in common and the perceptual experience that they share. The visual context of a conversation has two aspects: the information that is available to each conversant, and their beliefs about what is present for each other. In our experiment, we separated these factors for the first time and examined their impact on a spontaneous conversation. We manipulated the fact that a visual scene was shared or not and the belief that a visual scene was shared or not. Participants watched videos of actors talking about a controversial topic, then discussed their own views while looking at either a blank screen or the actors. Each believed (correctly or not) that their partner was either looking at a blank screen or the same images. We recorded conversants' eye movements, quantified how they were coordinated, and analyzed their speech patterns. Gaze coordination has been shown to be causally related to the knowledge people share before a conversation, and the information they later recall. Here, we found that both the presence of the visual scene, and beliefs about its presence for another, influenced language use and gaze coordination.

Keywords: Discourse; Eye movements; Gaze coordination; Common ground

1. 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 their exchange came not only from the words they spoke but from

Correspondence should be sent to Daniel C. Richardson, Cognitive, Perceptual & Brain Sciences, University College London, Gower Street, London WC1E 6BT, UK. E-mail: dcr@eyethink.org

the rich body of knowledge they shared (Clark, 1996; Clark & Marshall, 1981). This included a conversation that morning (about the guests coming to dinner), other specific knowledge (about one guest's tastes), and the current visual scene (Penny standing in front of the stove and John in front of a particular drawer). With this knowledge, John recognized Penny's reference to the tool within his reach that would allow her to crush some garlic into a casserole, although not too much.

In this paper, we examine one aspect of this shared knowledge: beliefs about 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? We addressed this issue using a novel paradigm in which participants had a conversation while looking at the same visual scene—a natural situation that might arise when people talk during a movie or lecture, or discuss a document over the telephone. We recorded both participants' gaze patterns and quantified the degree to which they were coordinated. This gives an object measure of conversants' joint attention which corresponds to the success of their communication (Richardson & Dale, 2005). We measured changes in this coordination as we manipulated—on a trial-by-trial basis—what participants saw, and what they believed each other could see.

The ability to reason about others' beliefs and perspectives has been called a "theory of mind." Its emergence in children and other primates has been studied extensively, but its role in adult behavior such as language use is less well understood. Reasoning about others' intentions, desires, and perspectives is not exclusive to human beings, but around 3–4 years, children advance to a level that other species never reach (Kaminski, Call, & Tomasello, 2008). They are able to correctly predict what happens when other people have false beliefs about the world (Wimmer & Perner, 1983). Even for adults, however, reasoning about others' false beliefs is not automatic (Apperly, Riggs, Simpson, Chiavarino, & Samson, 2006) or without cognitive cost (Apperly, Back, Samson, & France, 2007). Our reasoning about others' mental states rarely escapes the pull of our own knowledge and beliefs (Epley, Keysar, Van Boven, & Gilovich, 2004; Keysar, Barr, Balin, & Brauner, 2000; Mitchell, Robinson, Isaacs, & Nye, 1996; Nickerson, 1999; Van Boven & Loewenstein, 2003). The struggle to accurately apprehend others' mental states stands in contrast to the ease and fluidity of many social interactions, collaborations, and forms of "joint action" (Galantucci & Sebanz, 2009). This raises the question—to what degree do these forms of interaction require that we consider other people's beliefs and perspectives? A prime example is a conversation between two people. Do we take a "fast and frugal" (Gigerenzer & Goldstein, 1996) approach to considering the other's perspective, or do we invest the cognitive effort into computing each other's visual perspective (Clark, 1996)? Here, we addressed that question by examining the role of conversants' beliefs about each others' visual context during a spontaneous conversation.

In our experiment, two people had an extended conversation about contentious political issues. The situation was analogous to John and Penny discussing the political views of their dinner guests while looking at the meal they were about to serve. Things in the shared visual scene were not the content of the conversation, but provided a visual context. In this case,

when it is not required by the task, will conversants be swayed by each others' visual perspective? We manipulated what the conversants could see as they talked, and what they thought each other could see, examining how these beliefs determined gaze coordination.

1.1. Gaze coordination and conversation

The gaze coordination of two conversants is causally related to the success of their communication (Richardson & Dale, 2005) and the degree to which they share knowledge (Richardson, Dale, & Kirkham, 2007). In a first quantification of gaze coordination (Richardson & Dale, 2005), eye movements of speakers were recorded as they talked about a TV show and looked at pictures of its cast members. These monologs were played back to listeners looking at the same display. Cross-recurrence analysis (Zbilut, Giuliani, & Webber, 1998) showed that from the moment a speaker looked at a picture, and for the following 6 s, the listener was more likely than chance to be looking at that same picture. When two people had an interactive dialog and looked at the same images, gaze was also coupled for a span of around 6 s (Richardson et al., 2007).

We refer to this coupling of gaze patterns as *gaze coordination*. The implication is not that conversants had the explicit aim to synchronize their eye movements, but that gaze patterns become aligned due to the joint activity of conversation and commonalities between the processes of speech comprehension and production. In this sense, gaze is only one aspect of behavior that is coordinated during conversation, like accents (Giles, Coupland, & Coupland, 1992), body movements (Condon & Ogston, 1971; Shockley, Santana, & Fowler, 2003), and syntactic structures (Branigan, Pickering, & Cleland, 2000; Dale & Spivey, 2006). These acts of coordination serve many purposes, such as making sure that our conversation flows easily and intelligibly (Garrod & Pickering, 2004).

Indeed, closer gaze coordination 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. Listeners were later able to answer comprehension questions faster than those who had seen a randomized sequence of flashes (Richardson & Dale, 2005). In conversation, gaze coordination was achieved in virtue of the knowledge that conversants share. It was increased when the two partners heard the same (rather than different) encyclopedia passages about Salvador Dali prior to discussing one of his paintings (Richardson et al., 2007).

In our previous experiments, and in the current study, we measure the degree to which people are looking at the same thing at the same time. One could refer to this as gaze alignment or gaze co-occurrence, but we chose the term "coordination" to highlight the fact that it appears to be causally connected to what people know going into a conversation, and what they recall of information that was spoken. Like other levels of coordination during conversation—posture, gesture, accent, and word choice—gaze coordination could be intentional or automatic, or generated by conversational processes that are on a continuum somewhere in between. We leave this question open here (but for discussions, see Pickering & Garrod, 2009; Shockley, Richardson, & Dale, 2009). Our claim is only that gaze coordination is a measure of the "joint activity" (Clark, 1996) of language use, and in the present

experiment, we use it to reveal the effects of what conversants see and what they believe each other can see.

1.2. *Beliefs about visual context*

Visual context can play a part in a conversation in two ways (Keysar, 1997). First, it can provide perceptual information that is useful in speaking and listening. 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 (Knoeferle & Crocker, 2006; Richardson & Matlock, 2007; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995), eye movements to a visual scene are closely linked to linguistic processes (Spivey & Richardson, 2008). Second, if people believe their conversational partners can see certain things, they can interpret utterances in relation to the common ground, and make remarks relying on the fact that this information will be available to their partners. Speakers change their descriptions of locations in New York (Isaacs & Clark, 1987) and famous faces (Fussell & Krauss, 1992) depending upon how much relevant *knowledge* they believe the listener has. But what of *beliefs* about shared visual context?

This question has been investigated at the sentence level in several ‘reference game’ studies, in which one participant in a pair refers to an object that can be seen by one or both of them (Hanna, Tanenhaus, & Trueswell, 2003; Haywood, Pickering, & Branigan, 2005; Horton & Keysar, 1996; Keysar et al., 2000). In these studies, however, the two influences of visual context are bound together: the belief that visual information is shared is tied to the fact of it being shared. Our experiment separated these factors for the first time and studied them at the level of an extended conversation. 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.

1.3. *Visual context and gaze coordination*

In each trial of this experiment, two participants saw videos of four actors espousing views on a contentious topic (Fig. 1), and then discussed the topic between themselves. We manipulated two factors, at random, on a trial-by-trial basis. First, as they talked, participants saw either pictures of the four actors, or an empty 2×2 grid. We termed this as the visual context *for self*. Second, the participants were told before the discussion that the actors were either present or absent on the screen of their 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* (Fig. 2). *Present-absent* refers to the condition in which both participants could see a visual scene but believed that their partner

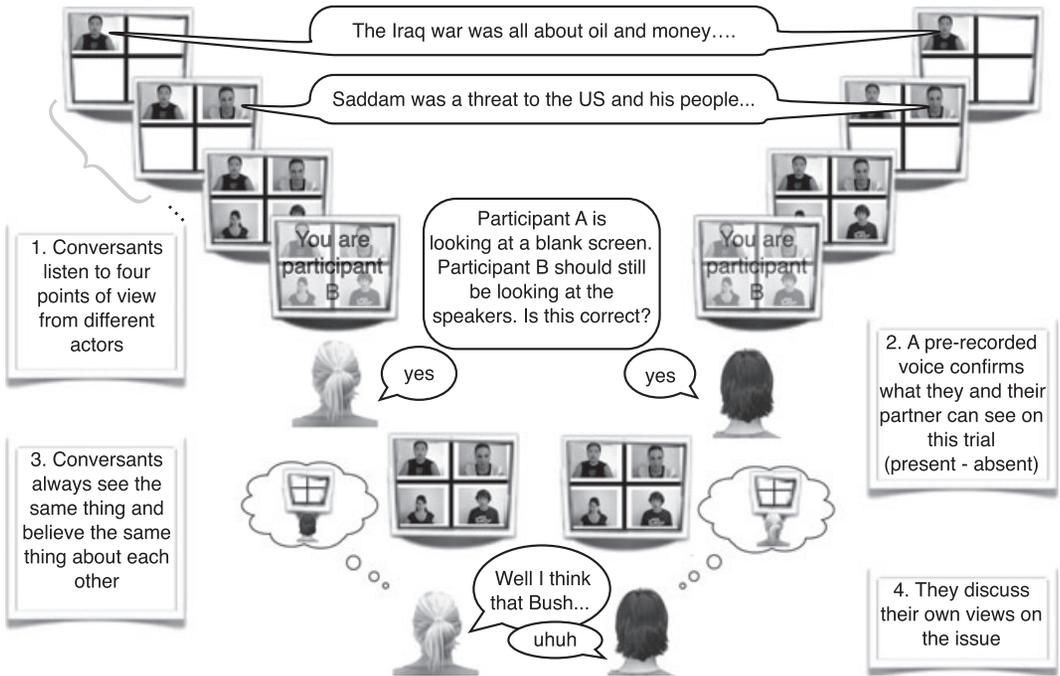


Fig. 1. Schematic of the trial design.

could not. The two participants were always in the same experimental condition: They saw the same image and had the same belief about what was being shown on their partner's screen.

Consequently, participants had beliefs that were factually incorrect in the *present-absent* and *absent-present* conditions (see Fig. 1 for an example). 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 he/she looked at a blank screen.

Our first prediction was that the presence of a visual context *for self* would increase gaze recurrence, since people tend strongly to fixate relevant visual objects during speech production and comprehension (Griffin & Bock, 2000; Meyer et al., 1998; Tanenhaus et al., 1995). But we also expected some recurrence between gaze patterns even when the screen was empty. During language comprehension and memory tasks, people systematically fixate on empty locations in a screen when they hear or make a reference to items that were previously there (Altmann, 2004; Hoover & Richardson, 2008; Richardson & Kirkham, 2004; Richardson & Spivey, 2000; Spivey & Geng, 2001). But what will participants do when a visual context is not present for them, but is believed to be present *for their partners*, or vice versa? We put forward three (nonexhaustive) possibilities: (a) visual context is ignored, and

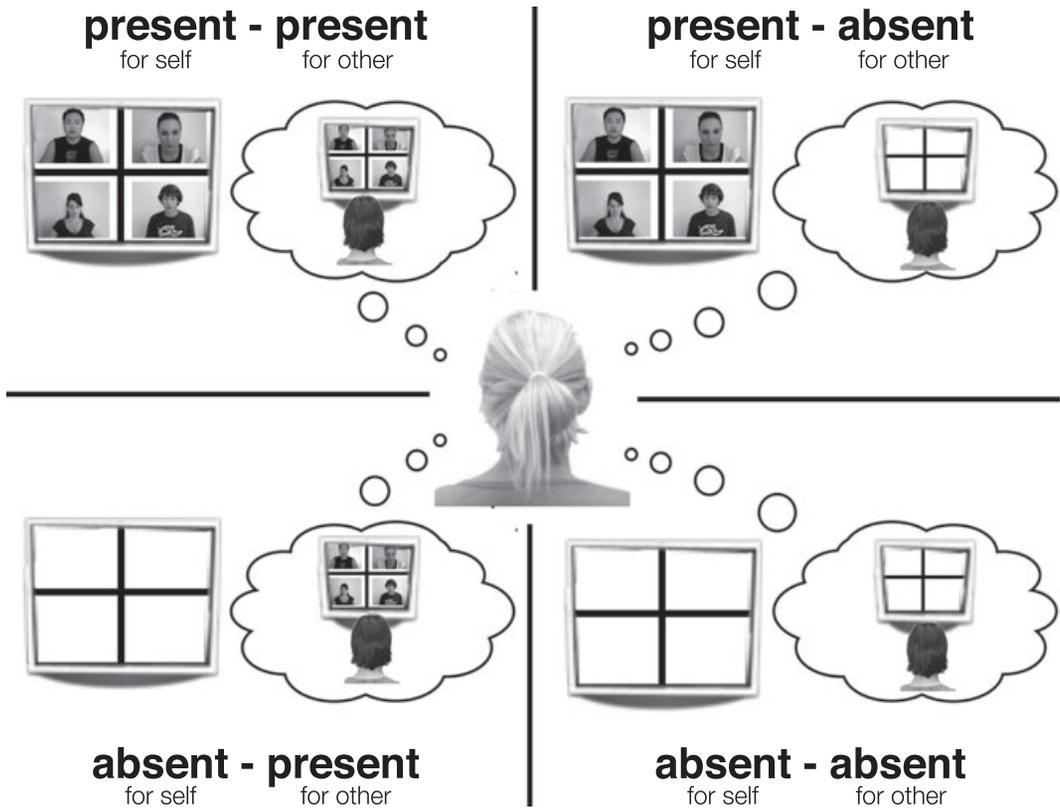


Fig. 2. The experimental conditions.

there will be no effect of beliefs about it, (b) visual context is exploited, and the belief that more of it is shared will increase gaze coordination, (c) visual context is compensated for, and the belief that it differs between conversants will increase gaze coordination.

1.3.1. Ignoring a partner's visual context

Listeners and speakers can seem strikingly egocentric (Keysar, 2007). In one reference game experiment, for example, a participant could see three candles of different sizes, but a confederate could only see the larger two (Keysar et al., 2000). When the confederate asked for “the smallest candle,” participants were 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. It was concluded that a consideration of the speaker's visual perspective did not constrain the initial stages of the listener's comprehension. Speakers do invest effort into avoiding ambiguity in their speech, by using intonation (Kraljic & Brennan, 2005), pronouncing new words clearly (Bard et al., 2000), or supplying atypical information (Brown & Dell, 1987), but in each case they appear to be relatively insensitive to the needs of a particular listener. In the case of our experiment, listeners are not being directed to objects in front of them but are hearing about current affairs. Speakers would have to keep track of what their partners

could see on each particular trial to be sensitive to the listeners' needs. The suggestion from these results is that when we manipulate the visual context that is *believed for others* it will have little effect on behavior.

1.3.2. Exploiting shared visual context

There are circumstances in which listeners and speakers seem very sensitive to what each other can see (Nadig & Sedivy, 2002; Wu & Keysar, 2007). For example, when they hear the beginnings of a question, listeners are more likely to fixate objects that are hidden from the speakers' view (Brown-Schmidt, Gunlogson, & Tanenhaus, 2008). Speakers will use names for objects that they believe to be known to the listener (Horton & Keysar, 1996; Metzger & Brennan, 2003). These results suggest that common ground will be used if it is available, and predict a main effect of the belief condition: Whether the visual context is present *for self*, gaze coordination will increase if it is believed to be present *for others*.

1.3.3. Compensating for mismatched 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 context (Haywood et al., 2005). Although earlier studies suggested that speakers do not change how they refer to objects as a function of their listener's knowledge (Brown & Dell, 1987), Lockridge and Brennan (2002) found that in a more interactive conversation, when the listener is also a naive participant, speakers do try and compensate for differences in knowledge. Therefore, when conversants believe there is a mismatch between their visual context and their partners' in our task, perhaps they will try and compensate for this, by using information that is mutually known, and confirming mutual understanding with back-channel responses (Schegloff, 1982). These efforts to increase shared knowledge will lead to higher gaze coordination, just as more shared information about Dali led to higher gaze recurrence (Richardson et al., 2007). This hypothesis predicts that gaze coordination will be highest in the *present-absent* and *absent-present* conditions.

2. Methods

2.1. Participants

A total of 112 undergraduates from the University of California, Santa Cruz, took part in exchange for course credit. The data from 19 were discarded due to failures in calibration, resulting in 37 usable dyads.

2.2. Apparatus

Each participant wore a headset with a boom mic and sat reclined in a cubicle looking up at a 48 cm LCD positioned 60 cm away with a *Bobax3000* remote eye tracker mounted at the base. The experimenter controlled when participants could hear the stimuli, each other's voices, and the experimenter's voice. An iMac calculated gaze position of each participant approximately 100 times a second, presented stimuli, and recorded data. A third Apple Mac computer synchronized the trials and data streams and saved an audio-video record of what was seen, heard, and said during the experiment, superimposed with gaze positions.

2.3. Stimuli

Four polemics were filmed 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). They were delivered straight to camera by 32 different actors, producing 8–20 s movies.

2.4. Procedure

Participants were introduced to each other and taken to adjacent cubicles for a brief calibration routine. The experimental conditions were presented in a randomized order across eight trials. On each trial, movies were shown one at a time in each quadrant of a 2×2 grid (Fig. 1). Location and order of presentation were randomized, but the same for each participant within a pair. 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. "You are participant B" appeared on the screen for both and they heard a prerecorded voice saying, "Please discuss these issues."

Fig. 2 represents the different experimental manipulations that were introduced at this stage of the trial. In the *absent-absent* and *present-present* conditions, 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.

2.5. Data analysis

The screen was divided up into four regions of interest, each subtending roughly a square of 12° visual angle. Gaze position was expressed as which of these regions (if any) was being looked at every 10 ms. We quantified the gaze coordination between conversants by categorical cross-recurrence. This technique depicts the temporal structure between time

series (Zbilut et al., 1998), and it has been used to capture the entrainment of body sway during conversation (Shockley et al., 2003) and the interrelationships between a child and caregiver's language use (Dale & Spivey, 2006). Here, points of recurrence are defined as the times at which both conversants were 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. The possible points of recurrence were defined as the times at which at least one of the conversants had his/her eyes on the screen. The next step was to lag one of the data streams by 10 ms, so that 0 ms on one data stream was aligned with 10 ms on the other. Again, recurrence percentage was calculated. This represents the degree to which one conversant is looking at the same thing as the other 10 ms later. Since in our experiment the participants had symmetrical roles, we averaged the recurrence for a lag of 10 ms and -10 ms, producing symmetrical plots. A full cross-recurrence analysis consisted of calculating the recurrence for all possible alignments, or lag times, of the two data series (for a more detailed explanation of this application of this technique, see Richardson & Dale, 2005).

A secondary analysis examined the speech of a random subset of 10 pairs (80 conversations in total). We focused on utterances which made any reference at all to 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 analyzed three properties. Firstly, each reference was termed either "visual" if it mentioned visual properties of the actor, or "factual" if it contained only (nonvisual) information relating the actors' viewpoint (e.g., "the guy who said that Iraq was about oil" rather than "the guy in the red shirt who was against the war"). Secondly, all references that occurred at the end of phrases were coded for intonational contours. Of interest were *rising boundary tones*, in which intonation rises in pitch at the end of a sentence (Pierrehumbert & Hirshberg, 1990). They are thought to serve as an implicit request to the listener to confirm the truth status of the speaker's current utterance. This pitch contour might then illicit a back-channel from the listener affirming the speaker's prior utterance (Allen, 1984; Fletcher, Stirling, Mushkin, & Wales, 2002). Lastly, we calculated the proportion of references that did indeed receive a back-channel response such as *ok*, *uh-huh*, *right*. Back-channels had to be directly following or overlapping the reference. These responses are thought to serve the function of coordinating dialog and signaling to the speaker that he/she has been understood (Schegloff, 1982), and hence their frequency serves as an indicator of conversants' efforts to establish mutual understanding.

3. 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*. Conversants appeared to be compensating for differences that they believed existed in their visual context. When the images were absent for conversants, recurrence was higher if they believed their partners could see them. Conversely, when the images were present, recurrence was higher if they

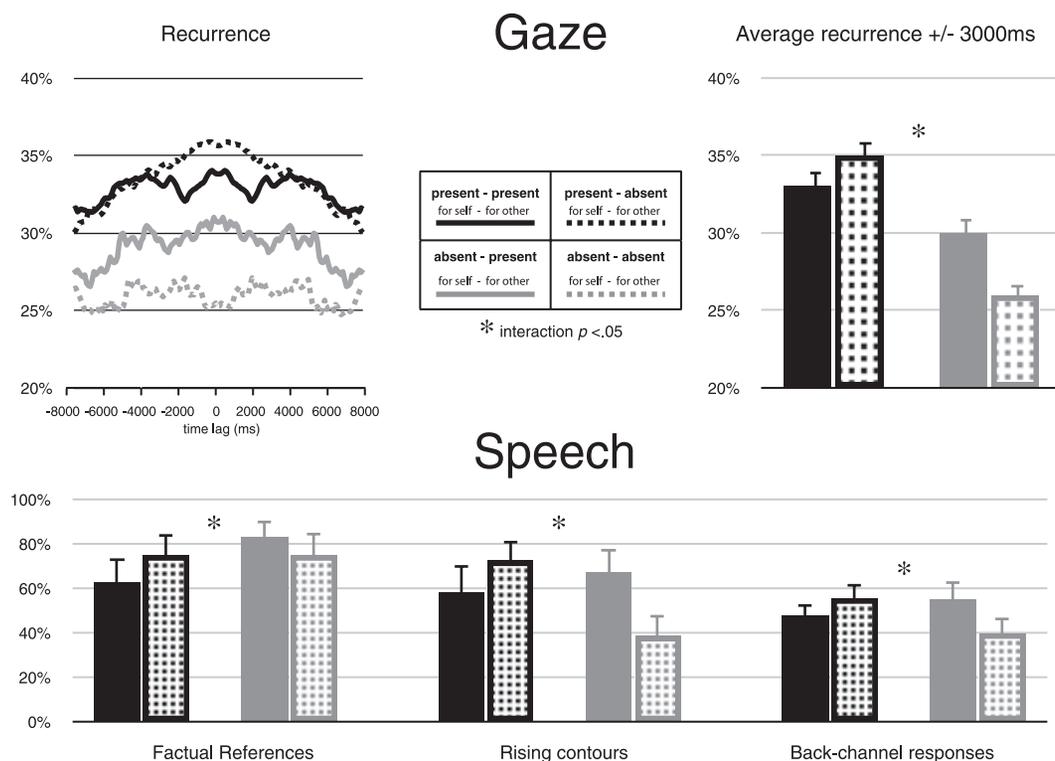


Fig. 3. Results of the gaze and speech analysis.

believed that that their partners could not. Fig. 3 shows how gaze recurrence changed across conditions at different time lags. Recurrence peaked at or around 0 ms and trailed off as the distance between conversants' gaze patterns increased, replicating the pattern observed for dialogs. Following previous analyses (Richardson et al., 2007), differences between conditions were analyzed by averaging recurrence within a window of $\pm 3,000$ ms, to capture the typical periods in which both conversants were acting as speakers and listeners. A 2 (*for self: present/absent*) \times 2 (*believed for other: present/absent*) ANOVA showed a significant interaction between these factors [$F(1, 36) = 4.5, p < .05$], as well as the expected main effect of visual context [$F(1, 36) = 12.4, p < .001$].

What features of the conversants' speech might explain these patterns? One possibility was that higher gaze recurrence was produced by conversants referring more to the people who had been on screen. This was not the case, however. More references were made when the visual context was present *for self*, but the number of references was not influenced by a belief in the presence *for others*. In contrast, several features of those references varied in the same pattern as the gaze recurrence data (Fig. 3). In the *present-absent* and *absent-present* conditions, references were more likely to employ factual properties. 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 measures of conversants' references were analyzed by a 2 (*for self: present/absent*) \times 2 (*believed for other: present/absent*) ANOVA. 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 (all F s < 1).

One concern was that this pattern of results came about because the *present-absent* and *absent-present* conditions involved deception. One possibility is that participants uncovered this deception during their conversation, and so started to talk about precisely what each could see on their screens, thereby inflating gaze coordination. However, none of the participants in debriefing expressed any awareness that they might have had incorrect beliefs about what their partners could see. And our microanalysis of 80 conversations found no occurrences of participants directly discussing their visual displays (e.g., "Can you see the guy on the bottom right?"), in a way that might reveal inconsistencies. As we intended, the images were not the content of the discussion, but provided a visual context, and beliefs about how that visual context was shared appeared to influence participants' speech and gaze patterns.

4. Conclusion

Processing the beliefs of others is not without cognitive cost (Apperly et al., 2007), and often mistakes in language seem to occur because we lapse into expedient egocentrism (Keysar, 2007). For these and a number of other reasons, our experiment might have found no influence of participants' beliefs about each other's visual context. For example, 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 isolated 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. In short, compared to reference game studies, the conversations were more abstract and interactive, the visual context less relevant, and the visual context more fluid.

Under these conditions, conversants could have ignored the visual context of their partners, but they did not. They could have employed a quick and expedient technique of exploiting visual context that was shared, and ignoring it otherwise. Instead they encoded what each other could see on each particular trial (Brennan, 2005). They sought to compensate for their mismatched visual contexts by referring to actors' opinions via a nonvisual means, asking for confirmation that their messages had been understood, and signaling to each other via back-channel responses. Their efforts to establish understanding in these

conditions were reflected in increased gaze coordination. 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 others could still see the actors.

The gaze coordination in our study was produced by the behavior of both participants as they exchanged speaker and listener roles during their dialog. We cannot discriminate therefore if it was their behavior as speakers or as listeners, or both, that produced high gaze coordination in some circumstances. Also, we cannot tell conclusively whether a difference in beliefs about visual context directly changes how conversants allocate their attention, or whether it changes how they speak to each other, which in turn shapes their gaze. While our speech analyses show that changes in speech certainly do occur, emerging evidence shows that beliefs about a partner's visual context alone can change gaze patterns. Richardson, Hoover, and Ghane (2008) reported that participants will look at a set of pictures differently if they believe that another participant is simultaneously looking at the same pictures versus at a set of unrelated symbols. But regardless of the causal route between beliefs and gaze, conversants' sensitivity to each others' visual context provides compelling evidence that natural language use is encircled by social cognition.

As anyone who has suffered a bad dance partner knows, coordination can be a challenge. Yet a range of human activity requires such "joint action" (Sebanz, Bekkering, & Knoblich, 2006). The coordination of joint attention is one such case, and it is essential for prelinguistic learning (Baldwin, 1995) and successful communication (Clark, 1996; Clark & Brennan, 1991). In conversation, it is achieved in virtue of the background knowledge conversants share (Richardson et al., 2007) and their sensitivity to each others' pragmatic constraints (Hanna & Tanenhaus, 2004) and domains of reference (Brown-Schmidt & Tanenhaus, 2008). Here we have shown that conversants are also attuned to what they believe each other can see. They will even coordinate their gaze around an empty screen in the mistaken belief that others can see something. It is the net effect of these multiple partial constraints that restrict the referent of "the thingy for the whatsit" from a universe of objects to a garlic press.

Acknowledgments

The authors would like to thank Herb Clark, Natasha Kirkham, and Michael Spivey for their insightful comments, and Arezou Cavanaugh, Norine Doherty, Jacquelyn Espino, Victor Hernandez, Daniel Janulaitis, and Jonas Nagel for their help with data collection and stimuli design.

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*, B79–B87.
- Apperly, I. A., Back, E., Samson, D., & France, L. (2007). The cost of thinking about false beliefs: Evidence from adult performance on a non-inferential theory of mind task. *Cognition*, *106*, 1093–1108.
- Apperly, A., Riggs, K. J., Simpson, A., Chiavarino, C., & Samson, D. (2006). Is belief reasoning automatic? *Psychological Science*, *17*, 841–844.
- 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* (pp. 131–158). Hillsdale, NJ: Lawrence Erlbaum.
- Bard, E. G., Anderson, A. H., Sotillo, C., Aylett, M., Doherty-Sneddon, F., & Newlands, A. (2000). Controlling the intelligibility of referring expressions in dialogue. *Journal of Memory and Language*, *42*, 1–22.
- Branigan, H. P., Pickering, M. J., & Cleland, A. A. (2000). Syntactic co-ordination in dialogue. *Cognition*, *75*(2), 13–25.
- Brennan, S. E. (2005). How conversation is shaped by visual and spoken evidence. In J. C. Trueswell & M. Tanenhaus (Eds.), *Approaches to studying world-situated language use: Bridging the language-as-product and language-action traditions* (pp. 95–129). Cambridge, MA: MIT Press.
- 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, H. Hesse, & H. Spada (Eds.), *Barriers and biases in computer-mediated knowledge communication—and how they maybe overcome* (pp. 89–118). New York: Springer.
- Brown, M. P., & Dell, G. S. (1987). Adapting production to comprehension: The explicit mention of instruments. *Cognitive Psychology*, *19*, 441–472.
- Brown-Schmidt, S., Gunlogson, C., & Tanenhaus, M. K. (2008). Addressees distinguish shared from private information when interpreting questions during interactive conversation. *Cognition*, *107*(3), 1122–1134.
- Brown-Schmidt, S., & Tanenhaus, M. K. (2008). Real-time investigation of referential domains in unscripted conversation: A targeted language game approach. *Cognitive Science*, *32*(4), 643–684.
- Clark, H. H. (1996). *Using language*. Cambridge, England: 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, I. Webber, & I. Sag (Eds.), *Elements of discourse understanding* (pp. 10–63). Cambridge, England: Cambridge University Press.
- Condon, W., & Ogston, W. (1971). Speech and body motion synchrony of the speaker–hearer. In D. Horton & J. Jenkins (Eds.), *The perception of language* (pp. 150–184). New York: Academic Press.
- 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*(3), 391–430.
- Epley, N., Keysar, B., Van Boven, L., & Gilovich, T. (2004). Perspective taking as egocentric anchoring and adjustment. *Journal of Personality and Social Psychology*, *87*(3), 327–339.
- Fletcher, J., Stirling, L., Mushkin, I., & Wales, R. (2002). Intonational rises and dialog acts in the Australian English map task. *Language and Speech*, *45*(3), 229–252.
- Fussell, S. R., & Krauss, R. M. (1992). Coordination of knowledge in communication: Effects of speakers’ assumptions about what others know. *Journal of Personality and Social Psychology*, *62*(3), 378–391.
- Galantucci, B., & Sebanz, N. (2009). Joint action: Current perspectives. *Topics in Cognitive Science*, *1*, 255–259.
- Garrod, S., & Pickering, M. J. (2004). Why is conversation so easy? *Trends in Cognitive Sciences*, *8*(1), 8–11.
- Gerwing, J., & Bavelas, J. B. (2004). Linguistic influences on gesture’s form. *Gesture*, *4*, 157–195.
- Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review*, *103*(4), 650–669.
- Giles, H., Coupland, N., & Coupland, J. (1992). Accommodation theory: Communication, context and consequences. In J. Coupland & N. Coupland (Eds.), *Contexts of accommodation* (pp. 1–68). Cambridge, England: Cambridge University Press.

- Griffin, Z. M., & Bock, K. (2000). What the eyes say about speaking. *Psychological Science*, 11(4), 274–279.
- Hanna, J. E., & Tanenhaus, M. K. (2004). Pragmatic effects on reference resolution in a collaborative task: Evidence from eye movements. *Cognitive Science*, 28(1), 105–115.
- 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: A Journal of the American Psychological Society/APS*, 16(5), 362–366.
- 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), 4.
- Hoover, M. A., & Richardson, D. C. (2008). When facts go down the rabbit hole: Contrasting features and objecthood as indexes to memory. *Cognition*, 108(2), 533–542.
- Horton, W. S., & Keysar, B. (1996). When do speakers take into account common ground? *Cognition*, 59(1), 91–117.
- Isaacs, E. A., & Clark, H. H. (1987). References in conversations between experts and novices. *Journal of Experimental Psychology: General*, 116, 26–37.
- Kaminski, J., Call, J., & Tomasello, M. (2008). Chimpanzees know what others know, but not what they believe. *Cognition*, 109(2), 224–234.
- Keysar, B. (1997). Unconfounding common ground. *Discourse Processes*, 24, 254–270.
- Keysar, B. (2007). Communication and miscommunication: The role of egocentric processes. *Intercultural Pragmatics*, 4(1), 71–84.
- 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. W. (2006). The coordinated interplay of scene, utterance, and world knowledge: Evidence from eye tracking. *Cognitive Science*, 30(3), 481–529.
- Kraljic, T., & Brennan, S. E. (2005). Prosodic disambiguation of syntactic structure: For the speaker or for the addressee? *Cognitive Psychology*, 50(2), 194–231.
- Lockridge, C. B., & Brennan, S. E. (2002). Addressees' needs influence speakers' early syntactic choices. *Psychonomic Bulletin & Review*, 9(3), 550–557.
- 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.
- Mitchell, P., Robinson, E. J., Isaacs, J. E., & Nye, R. M. (1996). Contamination in reasoning about false belief: An instance of realist bias in adults but not children. *Cognition*, 59, 1–21.
- Nadig, A. S., & Sedivy, J. C. (2002). Evidence of perspective-taking constraints in children's on-line reference resolution. *Psychological Science*, 13(4), 329–336.
- Nickerson, R. S. (1999). How we know—and sometimes misjudge—what others know: Imputing one's own knowledge to others. *Psychological Bulletin*, 125(6), 737–759.
- Pickering, M. J., & Garrod, S. (2009). Joint action, interactive alignment and dialogue. *Topics in Cognitive Science*, 1(2), 292–304.
- Pierrehumbert, J., & Hirschberg, J. (1990). The meaning of intonational contours in the interpretation of discourse. In P. R. Cohen, J. L. Morgan, & M. E. Pollack (Eds.), *Intentions in communication* (pp. 271–310). Cambridge, MA: 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(6), 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: A Journal of the American Psychological Society/APS*, 18(5), 407–413.

- Richardson, D. C., Hoover, M. A., & Ghane, A. (2008). Joint perception: Gaze and the presence of others. In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), *Proceedings of the thirtieth annual conference of the cognitive science society* (pp. 309–314). Austin, TX: Cognitive Science Society.
- Richardson, D. C., & Kirkham, N. Z. (2004). Multimodal events and moving locations: Eye movements of adults and 6-month-olds reveal dynamic spatial indexing. *Journal of Experimental Psychology: General*, *133*, 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*(1), 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. (1982). Discourse as an interactional achievement: Some uses of “uh huh” and other things that come between sentences. In D. Tannen (Ed.), *Analyzing discourse: Text and talk. Georgetown University roundtable on languages and linguistics* (pp. 71–93). Washington, DC: Georgetown University Press.
- Sebanz, N., Bekkering, H., & Knoblich, G. (2006). Joint action: Bodies and minds moving together. *Trends in Cognitive Sciences*, *10*(2), 70–76.
- Shockley, K., Richardson, D. C., & Dale, R. (2009). Conversation and coordinative structures. *Topics in Cognitive Science*, *1*, 305–319.
- Shockley, K., Santana, M. V., & Fowler, C. A. (2003). Mutual interpersonal postural constraints are involved in cooperative conversation. *Journal of Experimental Psychology. Human Perception and Performance*, *29*(2), 326–332.
- Spivey, M. J., & Geng, J. J. (2001). Oculomotor mechanisms activated by imagery and memory: Eye movements to absent objects. *Psychological Research/Psychologische Forschung*, *65*(4), 235–241.
- Spivey, M. J., & Richardson, D. C. (2008). Language embedded in the environment. In P. Robbins & M. Aydede (Eds.), *The Cambridge handbook of situated cognition* (pp. 382–400). Cambridge, England: 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.
- Van Boven, L., & Loewenstein, G. (2003). Social projection of transient drive states. *Personality and Social Psychology Bulletin*, *29*, 1159–1168.
- Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, *13*, 103–128.
- Wu, S., & Keysar, B. (2007). The effect of culture on perspective taking. *Psychological Science: A Journal of the American Psychological Society/APS*, *18*(7), 600–606.
- Zbilut, J. P., Giuliani, A., & Webber, C. L. J. (1998). Detecting deterministic signals in exceptionally noisy environments using cross-recurrence quantification. *Physics Letters*, *246*, 122–128.