THE UNIVERSITY of EDINBURGH

Edinburgh Research Explorer

Referring and gaze alignment: accessibility is alive and well in situated dialogue

Citation for published version:
Bard, E, Hill, R, Arai , M & Foster, ME 2009, Referring and gaze alignment: accessibility is alive and well in situated dialogue. in N Taatgen & H van Rijn (eds), Proceedings of CogSci 2009. Cognitive Science Society, pp. 1246-1251.

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Publisher's PDF, also known as Version of record

Published In:
Proceedings of CogSci 2009

Publisher Rights Statement:
© Bard, E., Hill, R., Arai , M., & Foster, M. E. (2009). Referring and gaze alignment: accessibility is alive and well in situated dialogue. In Taatgen, N., & van Rijn, H. (Eds.), Proceedings of CogSci 2009. (pp. 1246-1251). Cognitive Science Society.

OPEN ACCESS

Referring and Gaze Alignment: Accessibility is Alive and Well in Situated Dialogue

Ellen Gurman Bard (ellen@ling.ed.ac.uk)
Linguistics and English Language, University of Edinburgh, Edinburgh EH8 9LL, UK

Robin Hill (r.l.hill@ed.ac.uk)
Human Communication Research Centre, University of Edinburgh, Edinburgh EH8 9LW, UK

Manabu Arai (manabu.arai@ed.ac.uk)
Human Communication Research Centre, University of Edinburgh, Edinburgh EH8 9LW, UK

Abstract

Accessibility theory (Ariel, 1988; Gundel, Hedberg, & Zacharski, 1993) proposes that the grammatical form of a referring expression depends on the accessibility of its referent, with greater accessibility permitting more reduced expressions. From whose perspective is accessibility measured? Recent experiments (Bard, Hill, & Foster, 2008; de Ruiter & Lamers, submitted) using a joint construction task suggest that the speaker's view often determines referential form. Two objections to these results would neutralize accessibility predictions in many real-world situations. First, objects in shared visual space may be so salient that all will be highly accessible and reference to them in whatever form cannot fail (Smith, Noda, Andrews, & Jucker, 2005). Second, since joint action demands joint attention, the listener's and speaker's view of what is accessible should seldom differ. We use cross-recurrence analysis of interlocutors' gaze to show that neither objection applies. Gaze is not always well aligned. Dyads whose referring expressions ignored listeners' needs did not coordinate attention well. Dyads referring cooperatively coordinated attention better and in a way linked to the elaboration of their referring expressions.

Keywords: reference, accessibility, corpus experimental studies, pragmatics, situated dialogue

Introduction

The question "How shall a thing be called?" (Brown, 1958) still engages anyone who deals with human or machine language production. One very wide-ranging approach (Ariel, 1988, 1990, 2001) attempts to key elaboration of the form of referring expressions to how difficult the producer of the expression estimates it will be to access the referent concept, discourse entity, or extra-linguistic object. Expressions introducing entities deemed completely unfamiliar to the audience should be maximally detailed, as in, for example indefinite NPs including modifiers of various kinds ('a former Republican senator from strongly democratic Massachusetts'). Expressions of intermediate accessibility might be definite NPs, deictic expressions, or personal pronouns in that order. Expressions making reference to a single most immediately mentioned entity in focus can be as minimal as so-called clitics, unstressed and all but deleted pronouns ('/z/ in the garage'). Accessibility theory offers a unified framework for predicting how forms of referring expressions will respond to givenness, discourse focus, and inferrability from local scenarios. Accessibility ought to include effects of any conditions which might draw attention to the correct referent. Our research asks whose attention it is that determines referential form, and whether, in situations where both a speaker and a listener are present, there is any point in attempting to distinguish between them.

Ariel's (2001) notion of accessibility depends on what the speaker supposes is the case, not on what is genuinely easier for the listener. Opinions differ on how firmly speakers' suppositions are based on evidence about listeners' genuine states, both in design of referring expressions and in other aspects of behaviour. While accessibility of referring expressions was more sensitive to the knowledge of the listener than was clarity of articulation (Bard & Aylett, 2004), speakers' tendencies to match nomenclature to listeners' history or current situation are quite variable (Brennan & Clark, 1996; Horton & Gerrig, 2002, 2005a; Horton & Keysar, 1996; Keysar, Lin, & Barr, 2003). Though speakers may construct careful models of their interlocutors (Brennan & Clark, 1996), they may be unwilling or unable to recall, or deploy any such model in a timely fashion (Bard et al., 2000; Bard & Aylett, 2004; Horton & Gerrig, 2002, 2005a, 2005b; Horton & Keysar, 1996). It may be much easier to adopt a global account of a situation rather than constructing an incremental evidence-contingent plan: for example, when speakers can see the eye track of their interlocutors during a shared task, their search patterns may differ from those they follow without this cue (Bard et al., 2007; Brennan, Chen, Dickinson, Neider, & Zelinsky, 2007), but when the listener's eye track indicates an error, they may fail to make individually contingent responses (Bard et al., 2007).

Two recent experiments have explored factors that make speakers more or less sensitive to their listeners' knowledge. Both studies used a joint physical and visual task which makes it possible to vary participants' knowledge and responsibilities. Figure 1 illustrates the task.

In the Joint Construction Task two players cooperate to construct a tangram in a shared workspace represented on their yoked screens. Each trial offers a new target tangram using a non-exhaustive selection from the same set of coloured geometric shapes. Each player can manipulate the component shapes or partly built tangrams by mouse actions, but two parts can be joined together only if they are

moved by different players. Anything grasped by both will break. Poor constructions have to be discarded. Extra pieces and extra time are needed to rebuild them. Because each player can act on the parts, the activity of grasping or moving the named object adds a haptic or praxic modality to spoken forms. Even 'hovering' the mouse over a part without grasping it offers a chance to gesture towards it.
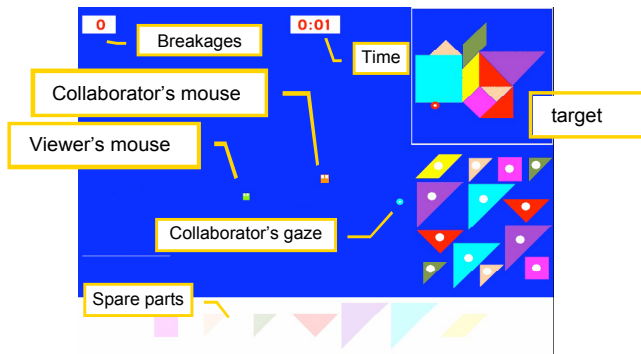

Figure 1. Joint Construction Task screen as a trial begins. Players' screens are yoked. A player sees his/her own mouse cursor and may or may not see the other's cursors.

How speakers framed references to the parts was affected by global perspectives and access to one another's actions. In Bard, Hill and Foster (2008), which shared a corpus with the present research, any dyad worked under one of two role-assignment conditions: either one player was to manage the task and the other to assist or no roles were assigned. Though players were always to cooperate, Managers and Assistants had different tasks: one planned and the other followed where s/he was led. Mouse cursors could be cross-projected. With the collaborators' cursors visible (the orange 'mouse' in Figure 1), a player could see the other's mouse 'hovering' over a stationary tangram part. In this study either both mouse cursors were cross-projected or neither was. In De Ruiter and Lamers (submitted), asymmetric conditions were also included and players were reminded before each new trial who would see what.

Both studies found conditions in which speakers seemed curiously insensitive to how easily a listener might discover a referent. Since moving objects attract attention, they should also attract referring expressions of greater accessibility. Pointing is associated with shorter, less detailed referring expressions and pointing to closer targets has an even stronger effect (Kranstedt, Lücking, Pfeiffer, Rieser, & Wachsmuth, 2006), so touching and moving should have a very marked effect. This was the case: a larger proportion of deictic expressions (*this square; these, mine*) – of middle accessibility – than of other forms coincide with mouse-referent overlap (Foster et al., 2008).

Bard et al. (2008) found that while all dyads appropriately used more deictic forms to introduce concurrently moving parts than stationary parts, players' roles determined how they adjusted referring expressions to hovering. A hovering mouse is an appropriate pointer or 'demonstrator' with (demonstrative) deictic expressions only when the other player can see it. An invisibly hovering mouse can only make a private gesture. It cannot make the referent easier for the listener to identify. The No-role players observed this restriction: Only visible gestures increased the rate of deictic initial mentions. The Manager-Assistant players, however, used more deictic expressions when performing public (mouse visible) or private (mouse hidden) gestures. Any player could easily keep in mind whether mouse cursors were hidden in a trial, because the partner's mouse was missing from his screen. Manager-Assistant pairs were simply more speaker-oriented than No Role pairs.

De Ruiter and Lamers (submitted) also found that use of deictic expressions was speaker-sensitive. Concern for the listener should curtail use of deictic expressions whenever the listener could not see the speaker's mouse cursor. Yet, despite reminders, what listeners could see had no effect. Instead speakers used fewer deictics when *they* could not see the listener's mouse.

While each study indicates conditions where referring expressions follow a speaker-centric definition of referent accessibility, their findings may not be to the point. The results might be uninterpretable because they come from situations where the visible objects and the task might bring accessibility to the same ceiling for both speaker and listener.

First, observing that dyads viewing the same film often made initial mentions in highly accessible form (*It...*), Smith, Noder, Andrews, and Jucker (2005) suggested that shared experience of the referents could make them salient for speaker and listener. If so, speaking about shared visual experiences would provide few opportunities to refer to an inaccessible referent. Any choice of form will be acceptable (Gundel et al., 1993), because references to salient objects are unlikely to fail.

Second, if a visible object were not compelling in itself, the coordination required for a joint physical task might assure that speakers and listeners were genuinely attending to the same objects at the same time. A speaker's gesture, private or public, might be irrelevant if players are working in a coordinated fashion.

Both objections suggest that dialogue situated in realistic contexts with visible referents and cooperative activities is the wrong domain for distinguishing speaker and listener knowledge. In fact, they suggest that it is the wrong place to look for any orderly relationship between cognitive accessibility and form of referring expression. The present paper asks whether this is the case.

To make the required tests, we must use an indicator of focussed attention outside of linguistic expressions. The players' gaze provides such a measure. People look at what they are talking about (Griffin, 2004; Meyer, van der Meulen, & Brooks, 2004), at what they hear mentioned (Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995), and at what they anticipate will be mentioned (Altmann & Kamide, 2004). If both players' gaze was largely and consistently aligned, that is, focussed on the same objects,

while referring expressions began, we can be reasonably certain that their situation kept visible objects highly and equally accessible to both. If, on the other hand, agreement in attention is not consistently high, but instead varies with referential behaviour, then we may have evidence that accessibility of referring expressions is worth studying even in situated dialogue and that apparently selfish speaking has consequences.

Corpus Collection And Coding

Task

The Joint Construction Task or JCT (Carletta et al., under revision) offers to two collaborating players a model tangram, geometric shapes for reproducing it, a work area, a counter for breakages, a set of replacement parts, and a clock measuring elapsed time. The players' task is always to construct a replica of the model tangram as quickly, as accurately, and as cheaply in terms of breakages as possible. An accuracy score is provided at the end of each trial.

Participants manipulate objects by left-clicking and dragging them around the screen or by right-clicking and rotating them. Each player sees his/her own mouse cursor change colour when it successfully grasps a part.

Carefully timed collaboration is important. Any part or partially constructed tangram 'held' by both players will break and must be replaced from the spare parts store (Figure 1) to complete the trial. Objects can be joined only if each is held by a different player. Moving an object across another breaks both. Objects join permanently wherever they first meet. Inadequate constructions can be purposely broken and rebuilt from spare parts, incurring a cost in both parts and time.

As Figure 1 shows, the viewer's and the collaborator's current mouse positions are represented by an orange cursor and a green cursor respectively. A blue circle marks the collaborator's current gaze position. A mouse cursor is changes colour while it grasps an object.

Apparatus

Each participant sat approximately 40cm from a separate 19 inch CRT display in the same sound-attenuated room. Participants faced each other, but direct eye contact was blocked by the two projection computers between them. Participants were eye-tracked monocularly via two SR-Research EyeLink II head-mounted eye-trackers. Head-worn microphones captured speech. Continuous audio and video recordings provided a full account of locations and movements of individual parts, constructed objects, and cursors. Composite Camtasia videos recorded movements and audio.

Participants, design and materials

Sixty-four Edinburgh University students, paid to participate, were paired into 32 same-sex dyads who had never met before. Four further dyads were discarded because of technical failures. Each pair participated in 8 experimental conditions produced by the factorial manipulation of 3 communication modalities: speech, gaze (each player's current eye-track cross-projected onto the other's screen twice within each 42 ms cycle), and mouse cursor (also cross-projected). Participants could always see their own mouse cursor. Without additional communicative modalities they saw only the moving parts. Gaze and mouse conditions were pseudo-randomised following a latin square. Speech was allowed in the first four presented conditions for half the dyads and in the second four for the rest. Only conditions with speech are analyzed here.

In 16 dyads one participant was randomly designated as manager and the other as assistant. The manager was asked to maintain "quality control" in speed, accuracy, and cost, and to signal the completion of each trial. The assistant was to help. The members of the remaining dyads were assigned no roles but otherwise had the same working instructions. Trials ended when one player signalled by pressing the keyboard spacebar and the other confirmed. An accuracy score reflecting similarity between the built and the target tangrams then appeared across the built tangram.

Each dyad reproduced 16 different target tangrams, 2 per condition. No tangram resembled a nameable object. Each contained 11 parts. At each trial onset, the same set of 13 parts (see Figure 1) was available, 2 copies of each of 6 shape-colour combinations and a single yellow parallelogram. The parts appeared in 4 different layouts counterbalanced across experimental items. The unused pieces differed from trial to trial.

Table 1 Accessibility of Referential Form Coding Scheme

Level	Definition	Examples
0	Indefinite NP	*a purple one* *one of the nearest* *blue pieces*
	Bare nominal	*pink one* *triangles*
1	Definite NP	*the red bit* *the other purple one*
2	Deictic NP	*those two little kids.*
	Deictic Possess }Pron	*these* *mine*
3	Other Pronouns	*it*
	Clitic/inaudible.	*-/z/*

Coding referring expressions

Dialogues were transcribed orthographically. Each referring expression was time-stamped for start and end points. Then, each expression referring to any on-screen object was coded with a referent identifier linking it to the object. Coders used

video and audio tracks, drawing on any material within a dialogue to determine the referent of any expression. All referring expressions were coded for accessibility on the scale given in Table 1. This system modestly expands a system applied to an earlier corpus of task-related dialogues (Bard & Aylett, 2004). Based on linguistic form, it yields negligible disagreement between coders.

Results

Cross-recurrent gaze

The cross-recurrence analysis used by Richardson, Dale and colleagues (Richardson & Dale, 2005; Richardson, Dale, & Kirkham, 2007) was applied to players' eye tracks. This technique (Zbilut, Giuliani, & Webber, 1998) measures both absolutely and nearly simultaneous entrained activity. The regions of interest (ROI) for gaze were both fixed (the clock, penalty counter, target tangram, spare parts store) and dynamic (the movable parts and tangrams under construction). Fixations on blank areas of the background, looks off-screen and blinks were excluded. Each player's gaze was located at increments of 20ms before being pooled into bins of 200ms. With one player's gaze location as a reference, the other's gaze at each bin before and after was checked for a percentage match in ROI. A temporal window of ±4000ms centered on the expression's onset. The likelihood of overlap between participants' eye movements was therefore examined when they lagged each other by up to four seconds.

To assure that gaze patterns reflected the status of an individual expression, expressions were used only if the 4s prior to onset contained no other referring expression. Some 936 expressions met this criterion. Table 2 shows their distribution by Role Assignment and Accessibility. All trials with speech provide the data, collapsed over levels of gaze and mouse cross-projection.

Table 2 Distribution of Analyzed Referring Expressions by Level and Roles of Speakers

Roles Assigned	Coded Accessibility			
	0	1	2	3
No role	46	168	152	57
Manager-Asst	89	162	187	75

Figures 2a and 2b show cross-recurrent gaze. Colours indicate form of referring expression. Hollow points represent real cross-recurrent gaze, reflecting the time lag between partners' fixations on a ROI. Filled points represent randomized cross-recurrent gaze. This baseline is generated by randomly reordering one player's gaze records and running a cross-recurrence analysis with the other player's real record. The baseline provides identical distributions of fixation locations, reflecting the probability that both individuals' focuses of visual attention will overlap purely by chance. The baseline is therefore typical of the situation

in which the measurements were made, while the correctly ordered recurrence figures express both the situation and the timing. Significant differences between corresponding real and random curves indicate gaze coordination beyond chance.

Our first question is whether gaze coordination between interlocutors is in fact always high. The second is whether any discovered patterns in gaze coordination relate in an orderly way to referential behaviour, either in terms of speakers' tendencies to respect listeners' needs or in terms of referential form.
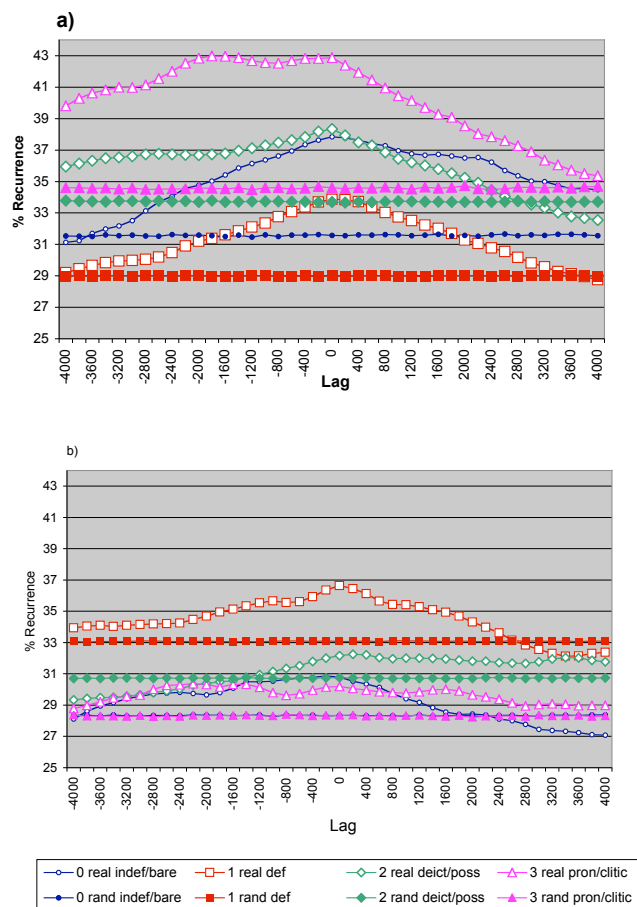

Figure 2. Real and random cross recurrent gaze for a) No Role dyads b) Manager-assistant dyads

In fact, Gaze coordination was not always high. Though the peaks of real cross-recurrence curves (30-43% coinciding gaze) are certainly above chance for an array with up to 15 regions of interest, alignment of gaze is inconsistent. Figure 2 suggests a major difference between the two role conditions – in a direction predictable from their use of referring expressions. ANOVAs were run by lag, role assignment, and accessibility with referring expressions as the random variable. Manager Assistant pairs (2b), who had used referring expressions in a speaker-centric way, had less absolute overlap in gaze (F(1, 904) =

4.66, $p < .04$) and less difference between real temporal patterns and the randomized baseline $F(40, 36160) = 1.55$, $p < .02$) than No role pairs, who had respected listeners' needs.

Second, there are robust differences in coordinated gaze in exactly the directions that referring expression usage suggests. Not only was gaze better coordinated for the listener-sensitive No Role dyads, but the coordination can be predicted from form of referring expression. Speakers who match referring expressions to listeners' needs should use less elaborate expressions for objects that listeners are already attending to (because they are already accessible) and more elaborate expressions for those they need to look for.

To test for this pattern, we examined the shape of the real cross-recurrence curves. Though Figure 2a has clear peaks of aligned gaze at 0 lag (simultaneous), neither figure shows symmetrical cross-recurrence: one side of the tent-like shape is usually higher than the other. By subtracting recurrent gaze where the speaker looked before the listener (positive lags in Figure 1) from recurrent gaze where the listener looked before the speaker (the corresponding negative lags), and averaging the results, we can characterize aligned attention for each of the four accessibility levels within each role type. In Figure 3, negative values mean that listeners followed speakers and positive values mean that listeners preceded speakers.

Figure 3 Order of interlocutors' gaze with different kinds of referring expressions: Mean recurrent gaze rates with the negative lags (listener first) less those on the positive lags (speaker first) by dyad role assignments.

The relationship between referential form and gaze coordination was significant only for No Role dyads, whose use of reference had been sensitive to listeners' needs: the more accessible a form's referents should be, the stronger the tendency for the speakers' gaze to follow where the listeners' led (Linear mixed-effects regression: Coefficient = 2.91, $p = .03$; Spearman's *rho*: Coefficient = .10, $p = .03$ for No Role dyads; *n.s.* for Manager-Assistant dyads).

Discussion

In this paper, we attempted to discover whether situated dialogue precludes real differences in accessibility of referents between speaker and listener. If it did, several consequences could follow.

First, situated dialogue would not be a suitable domain for distinguishing the perspectives of speaker and listener, who would have in common ground whatever they jointly act on. Co-presence would be more than a shortcut for estimating shared knowledge; it would be a full account of shared attention and shared accessibility. Actions designed to render entities more or less accessible would be irrelevant.

By the same token, accessibility would become an irrelevant notion for situated language. Where every entity under discussion is very accessible, no externally driven accessibility differences motivate choice of referring expression. Any differences in form from one expression to the next could not reflect the situation's demands, which are few and constant.

Finally, we would have to conclude that apparent misbehaviour on the part of interlocutors in keying referential form to their own private gestures was not so much misbehaviour as irrelevant variation in expression.

We have found instead that situated dialogue does not coerce interlocutors' attention into a common pattern. Shared attention is variable and keyed to reference.

First, speakers who accommodate expressions to their private gestures, the Manager-Assistant dyads, do not entrain each other's attention well. Why they fail is not entirely clear. Their gaze patterns differ in other ways' from their No Role counterparts'. They seem to spend more time planning, for they spend more time looking off screen ($F(1, 501) = 14.31$, $p < .001$) or at the clock ($F(1, 501) = 6.39$, $p = .012$), though not enough more (2-3%) to make coordinated on-screen gaze impossible. They may simply be taking too individual a view of their own roles to consider the importance of the other player to their joint success.

Second, where speakers coordinate gesture and referring expression according to listeners' needs, they do coordinate attention in exactly as accessibility theory would predict. More accessible forms accompany a greater tendency for the speaker's attention to follow where the listener's has focussed. No corresponding trend appears in the 'misbehaving' group.

The results confirm the role of accessibility in determining form of referring expressions even in co-present joint action. They also make it plain that poor referential practice accompanies poor alignment of attention. Finally, they suggest that even in joint physical action, there are limits to the power of co-presence.

Acknowledgments

Jean Carletta for the experiment and analysis systems, to the reference coders, and to JP de Ruiter for helpful discussions.

References

Altmann, G. T. M., & Kamide, Y. (2004). Now you see it, now you don't: mediating the mapping between language and the visual world. In J. Henderson & F. Ferreira (Eds.), *The integration of language, vision and action.* (pp. 347–386): Psychology Press.

Ariel, M. (1988). Referring and accessibility. *Journal of Linguistics, 24*, 65-87.

Ariel, M. (1990). *Accessing Noun-Phrase Antecedents*. London.: Routledge/Croom Helm.

Ariel, M. (2001). Accessibility theory: An overview. In T. Sanders, J. Schilperoord & W. Spooren (Eds.), *Text representation: Linguistic and psycholinguistic aspects.* (pp. 29-87). Amsterdam: John Benjamins.

Bard, E. G., Anderson, A. H., Chen, Y., Nicholson, H. B. M., Havard, C., & Dalzel-Job, S. (2007). Let's you do that: Sharing the cognitive burdens of dialogue. *Journal of Memory and Language, 57*(4), 616-641.

Bard, E. G., Anderson, A. H., Sotillo, C., Aylett, M., Doherty-Sneddon, G., & Newlands, A. (2000). Controlling the intelligibility of referring expressions in dialogue. *Journal of Memory and Language, 42*, 1-22.

Bard, E. G., & Aylett, M. P. (2004). Referential form, word duration, and modeling the listener in spoken dialogue. In J. C. Trueswell & M. K. Tanenhaus (Eds.), *Approaches to studying world-situated language use: Bridging the language-as-product and language-as-action traditions.* (pp. 173-191). Cambridge, MA: MIT Press.

Bard, E. G., Hill, R., & Foster, M. E. (2008). *What tunes accessibility of referring expressions in task-related dialogue?* Paper presented at the Proceedings of the 30th Annual Meeting of the Cognitive Science Society, CogSci2008, Washington, D.C.

Brennan, S. E., Chen, X., Dickinson, C., Neider, M., & Zelinsky, G. (2007). Coordinating cognition: The costs and benefits of shared gaze during collaborative search. *Cognition, 106*(3), 1465–1477.

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*, 1482-1493.

Brown, R. (1958). How shall a thing be called. *Psychological Review, 65*, 14-21.

Carletta, J., Nicol, C., Taylor, T., Hill, R., de Ruiter, J. P., & Bard, E. G. (under revision). Eyetracking for two-person tasks with manipulation of a virtual world. *Behavior Research Methods, Instruments, and Computers*.

de Ruiter, J. P., & Lamers, M. (submitted). What I see is what you get; a heuristic illusion in human communication.

Foster, M. E., Bard, E. G., Guhe, M., Hill, R., Oberlander, J., & Knoll, A. (2008). *The roles of haptic-ostensive referring expressions in cooperative task-based human-robot dialogue.* Paper presented at the Human Robot Interaction, Amsterdam.

Griffin, Z. M. (2004). Why look? Reasons for eye movements related to language production. In J. M. Henderson & F. Ferreira (Eds.), *The integration of language, vision, and action: Eye movements and the visual world.* New York: Psychology Press.

Gundel, J. K., Hedberg, N., & Zacharski, R. (1993). Cognitive status and the form of referring expressions in discourse. *Language, 69*, 274-230.

Horton, W. S., & Gerrig, R. J. (2002). Speakers' experiences and audience design: knowing when and knowing how to adjust utterances to addressees. *Journal of Memory and Language, 47*(4), 589-606.

Horton, W. S., & Gerrig, R. J. (2005a). Conversational common ground and memory processes in language production. *Discourse Processes, 40*(1), 1-35.

Horton, W. S., & Gerrig, R. J. (2005b). The impact of memory demands on audience design during language production. *Cognition, 96*(2), 127-142.

Horton, W. S., & Keysar, B. (1996). When do speakers take into account common ground? *Cognition, 59*, 91-117.

Keysar, B., Lin, S., & Barr, D. J. (2003). Limits on theory of mind use in adults. *Cognition, 89*, 25–41.

Kranstedt, A., Lücking, A., Pfeiffer, T., Rieser, H., & Wachsmuth, I. (2006). Deictic Object Reference in Task-oriented Dialogue. In G. Rickheit & I. Wachsmuth (Eds.), *Situated Communication,* (pp. 155-207). Berlin: Mouton de Gruyter.

Meyer, A. S., van der Meulen, F., & Brooks, A. (2004). Eye movements during speech planning: Speaking about present and remembered objects. *Visual Cognition, 11*, 553-576.

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. (2007). The art of conversation is coordination: common ground and the coupling of eye movements during dialogue. *Psychological Science, 18*(5), 407-413.

Smith, S. W., Noda, H. P., Andrews, S., & Jucker, A. H. (2005). Setting the stage: How speakers prepare listeners for hte introduction of referents in dialogues and monologues. *Journal of Pragmatics, 37*, 1865-1895.

Tanenhaus, M., Spivey-Knowlton, M., Eberhard, K., & Sedivy, J. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268*, 1632.

Zbilut, J. P., Giuliani, A., & Webber, C. L. (1998). Detecting deterministic signals in exceptionally noisy environments using cross-recurrence quantification. *Physics Letters, 246*, 122-1228.