Addressee Backchannels Influence Overhearers' Comprehension of Dialogue

Jackson Tolins (jtolins@ucsc.edu)

Jean E. Fox Tree (foxtree@ucsc.edu)
Psychology Department, University of California Santa Cruz
Santa Cruz, CA 95064 USA

Abstract

We tested whether overhearers made use of the relationship between specific (e.g. *really*, *oh*) and generic (e.g. *uh huh*, *mhm*) backchannels and speakers' talk in online dialogue comprehension. In Experiment 1 we found that words that followed specific backchannels were recognized more slowly than words that followed either generic backchannels or pauses. In Experiment 2 we found that the type of backchannel and the discourse relationship between the speaker's subsequent and previous turn predicted overhearer's recall of words that preceded backchannels and pauses. When the turn was a continuation of the narrative preceding the test point, specific backchannels resulted in faster responses. When the turn was an elaboration of the narrative preceding the test point, specific backchannels resulted in slower responses. We conclude that overhearers make use of the predictive relationship between listener backchannels and speakers' discourse in comprehending dialogue.

Keywords: Dialogue; Comprehension; Backchannels; Overhearers; Narratives; Elaborations; Continuations

Introduction

Observing others in interaction is a common means of communication, from political debates and television dramas to overhearing conversations at work or at the bus stop. According to the *collaborative theory of language use* (Clark, 1996), all active participants play a role in shaping a conversation, no matter whether they contribute multi-turn utterances, short utterances like *uh huh*, facial expressions, or postural responses. While early understanding of how the collaborative theory affected third party dialogue comprehension demonstrated both enhanced and decreased performance for overhearers of interactions (Fox Tree, 1999), increasing evidence has shown that overhearers view interactive dialogue holistically. For example, children used overheard conversation to learn novel words (Akhtar, Jipson, & Callanan, 2001) and adults used information created through dialogue to enhance success in referential tasks (Branigan, Catchpole, & Pickering, 2011; Fox Tree, 1999). In the current experiments, we demonstrate that non-interruptive addressee responses also influence overhearers' comprehension of speakers' talk in two-party dialogue.

Speakers, Addressees, and Overhearers

In our studies, we treat *speakers* as the conversational participants who produced the principal narratives in our experimental materials. We treat *addressees* as listeners of the narratives, who in our cases contributed backchannels such as *uh huh* to the ongoing dialogue. *Overhearers* are listeners who do not actively participant in the talk.

Addressees can make significant contributions to the development of ongoing talk. Speakers actively monitor their audience and dynamically adjust their speech accordingly (Clark & Krych, 2004). Rather than being passive recipients of information, listeners actively participate in the construction of ongoing dialogue, even when they are only using backchannels rather than full turns at talk. The type of backchannels provided may change the course of the dialogue, for example, influencing how well stories are told (Bavelas, Coates, & Johnson, 2000). Backchannels can also influence the way narratives develop turn by turn, leading people to make predictions about what will happen next in a conversation (Tolins & Fox Tree, submitted). At the level of discourse construction, when addressees provide generic backchannels such as *mhm* and *uh huh*, displaying attention and understanding, participants expected speakers to continue their subsequent talk with new events, or *continuations*. However, when addressees respond with context specific backchannels such as *oh* and *really*, it was expected that the next turn would be an explanation of the prior talk, or *elaboration*. That is, the type of backchannel the addressee displayed may modulate not only global features of narration such as quality and story structure (Bavelas et al., 2000), but also the utterance-by-utterance development of the speaker's talk (Tolins & Fox Tree, submitted).

In addition to speakers and addressees, any particular conversation may have a number of other individuals involved. For example, a conversation could have *ratified* and *unratified* side listeners, such as an audience listening in on an interview (ratified) or a child listening to parents through a door (unratified; Goffman, 1981). Similarly based on participation status, listeners can be divided into *addressees* and *overhearers* (Schober & Clark, 1989), with addressees participating in the conversation and overhearers not participating. Overhearers can be further divided into those whom the conversational participants are aware of, *bystanders*, and those whom the conversational participants are not aware of, *eavesdroppers* (cf Clark & Schaefer, 1987; Schober, 1998). The speakers whose communication was tested in our studies knew they were being recorded for potential future comprehension experiments, thus making our listeners ratified overhearers of the bystander sort, which we will henceforth refer to as overhearers.

Overhearers understand talk differently from direct addressees. Because they are unable to participate in the process of grounding, they have reduced access to the jointly maintained mutual knowledge (Garrod & Anderson, 1987). As such, they fare worse on a number of tasks compared to addressees who are able to provide their conversational partners with feedback. For example, they are less accurate in matching speakers' descriptions to abstract objects, and also fare worse in story retelling (Clark & Wilkes-Gibbs, 1986; Kraut, Lewis, & Swezey, 1982; Schober & Clark, 1989). Through the use of common ground and mutually shared knowledge, active interactants develop partner-specific conceptual pacts (Brennan & Clark, 1996; Brown-Schmidt, 2009) that often create a barrier to overhearers' comprehension – although sometimes it can lead to overhearers' outperforming addressees, such as in detection of evasion (Bly, 1993). Because addressees are vested in co-constructing meaning with a speaker, they fail to spot evasive language that overhearers detect.

The approach we take to overhearers varies from that taken by prior researchers. Prior researchers have focused on either global features of the dialogue, for example number of discourse markers used or number of perspectives presented, or more specifically on the relationship between speakers and overhearers (Schober & Clark, 1989), rather than focusing on the role that all active interactants play in providing useful information to the overhearer. We suggest that as overhearers listen to dialogue, responses from the addressee may be informative, cueing the overhearer to interpret the speaker's talk in a particular way or allowing the overhearer to make predictions as to what the speaker's next talk will likely be (Tolins & Fox Tree, submitted).

Specific and Generic Backchannels

Backchannels serve an interactional function in conversation. They are used, among other things, to indicate continued attention on the part of the listener, display comprehension, or present the addressee's stance on the speakers' talk (Bangerter & Clark, 2003; Brunner, 1979). Backchannels can be verbal, such as *yeah*, *oh*, *okay*, *uh huh* or *mhm*, or visual, such as facial expressions, nods, and gestures (Bavelas et al., 2000). In the current report, we focus on verbal backchannels.

Verbal backchannels have been broken down into two main categories: those that display continued attention, such as *mhm*, and those that indicate a listener's assessment of preceding talk, such as *oh no!* (Goodwin, 1986). Assessment backchannels can communicate affective responses to the content of the current speaker's speech, such as demonstrating disgust or sorrow at appropriate points in a story telling. They can also be informational, indicating that what the speaker just said was new for the listener. Affective/informational, or *specific*, backchannels have been analyzed as relying more heavily on the specific conversational context in which they occur than attention-indicating, or context *generic*, backchannels (Bavelas et al., 2000; Goodwin, 1986).

In a previous study that paired qualitative analysis of spontaneous story telling in conversation with an experimental test of the inductively derived hypotheses, we found evidence that listener backchannels, whether generic or specific, shape the unfolding narrative (Tolins & Fox Tree, submitted). While both types of backchannels ground the speaker's developing story and add to the shared common ground of the interactants, specific backchannels, as displays of surprise or discourse newness, act as requests for elaboration from the speaker. Following generic backchannels, speakers continued on to a new discourse event, whereas following specific backchannels, the speaker's next utterance presented an elaboration or explanation of the event being responded to.

Given the influence that backchannels have on speakers' ongoing talk, and in particular the relationship between generic and specific backchannels and the speaker's following utterance, it is possible that overhearers may rely on listener communication as cues for comprehension, allowing them to predict how the discourse will develop next. However, if the context of dialogue comprehension leads overhearers to imagine themselves as direct addressees, they may comprehend the speaker's talk as though it was directed towards themselves and become distracted by responses from the listener.

The Current Investigation

We compared how specific and generic backchannels influenced third party comprehension. In two experiments we compared overhearers' comprehension of speakers' talk before and after the two different types of backchannels, as well as before and after pauses in which no addressee feedback was provided. We explore three distinct hypotheses.

Backhannels may not provide any information to overhearers. Although they affect how speakers tell stories (Bavelas et al., 2000) and what overhearers think will happen if they put themselves into the speaker's place (Tolins & Fox Tree, submitted), they may not have any effect on ovehearers' listening in on a conversation that they are not participating in or contributing towards. They are small words of seemingly little consequence that might be easily tuned out. We will call this the *tuned out hypothesis*.

Listening to two people talk may be more cognitively demanding than listening to one. That is, overhearers may find listener responses distracting in that listener responses may cause overhearers to set up two models in their heads, one of the speaker's communicative processes, and another of the listener's. Under this hypothesis, overhearers may find specific responses more distracting than generic. Because they are more informative, specific responses demand increased processing. We will call this the *distraction hypothesis*.

In contrast, listener responses may help overhearers coordinate information from both the speaker and the listener, with specific and generic backchannels serving as distinct cues. Overhearers may use specific backchannels as cues to how the next utterance should be integrated with the discourse content of the previous turn (Tolins & Fox Tree, submitted). By hypothesis, the type of backchannel will indicate whether the next utterance will present a discourse new event or re-present a discourse old event in a new way. When a specific backchannel predicts an elaboration, overhearers should be prepared to update information from the prior turn. We will call this the *coordination hypothesis.*

We tested the effects of backchannels on the processing of information after the backchannels with a word monitoring technique (Marslen-Wilson & Tyler, 1980). We tested the processing of information prior to the backchannels with a semantic verification technique (Fox Tree & Schrock, 1999).

According to the tuned out hypothesis, no effect of backchannels will be observed with either technique. If they are tuned out as unimportant for overhearers, responses to target words should be similar regardless of what listeners say between speakers' turns.

According to both the distraction and the coordination hypotheses, hearing a specific backchannel should slow word monitoring (Marslen-Wilson & Tyler, 1980) in comparison to hearing a generic backchannel or a pause. Specific backchannels should distract more than generic backchannels, slowing monitoring more for specifics. Alternatively, specific backchannels could also cause overhearers to maintain access to the content of the previous turn, as they allow the overhearer to predict that the next speaker's talk will likely involve elaborative information on the same content. This divided focus between prior and subsequent talk would slow word monitoring after specific backchannels in comparison to generic backchannels and pauses.

According to the distraction hypothesis, specific backchannels should also slow semantic verification. That is, distraction will play a role in both paying attention to upcoming talk as well as recalling prior talk.

In contrast, according to the coordination hypothesis, specific backchannels should speed or slow semantic verification dependent on the next speaker turn. This hypothesis suggests that hearing a particular backchannel allows overhearers to predict what type of next turn the current speaker will have, in relation to the information of the turn to which the backchannel responds, what we will call the *discourse relationship*. If specific backchannels allow overhearers to make predictions as to the discourse level relationship between two turns, then there should be an interaction between backchannel type and relationship type. Based on a prediction that elaborating or updating discourse-old information with new information interferes with access to the old, we make two distinct predictions. If overhearers maintain access to the previous utterance following hearing a specific backchannel, as they expect to update this information, words from prior discourse should be recognized faster compared to when these same words are followed by generic backchannels or pauses when the speaker's next turn is a continuation, where no updating occurs. In contrast, when an overhearer hears a specific backchannel, and the next turn does present elaborative information, writing over the previous turn, access to this turn should be reduced.

Experiment 1 –Word Recognition

We tested how well overhearers monitored for words following specific versus generic backchannels as well as in comparison to the same talk with the backchannels replaced with pauses. Target words all occurred in the next turn following the backchannel of interest.

Method

Participants. 89 students from the University of California, Santa Cruz, participated in exchange for course credit.

Materials. Stimuli for both experiments were selected from a previously recorded audio corpus of spontaneous conversations. Interactants were asked to speak for 12 minutes with the topic of bad roommate experiences as the starting point. From this corpus, 30 short audio clips were selected, varying in length from 25 to 85 seconds. Audio stimuli were selected in which the voices of both participants in the interaction were heard prior to the target backchannel, so as to allow participants a chance to hear the addressee's voice prior to the point in which the addressee provided the critical feedback. Ten of the audio clips contained an authentic generic backchannel token, 10 contained an authentic specific backchannel token, and 10 contained an authentic pause in the speakers' talk. For each stimulus item, a set was created using digital splicing through Praat (Boersma & Weenink, 2005), with the critical backchannel or pause removed and replaced with a token from the other two categories. Because the generic backchannels were generally reduced in length and in volume compared to the specific backchannels, the audio pairs were edited so that the onset of the backchannels and the onset of the next turn at talk following the backchannel were matched to within four milliseconds. Similarly, pauses were created by taking white noise from elsewhere in the audio recording for the same conversation and replacing the backchannel tokens. In regards to the discourse relationship across the turns surrounding the target backchannel, 16 were continuations and 14 were elaborations.

The audio manipulation resulted in 30 triads of stimuli, with each triad presenting the exact same audio except in the critical location, which consisted of a specific backchannel, a generic backchannel, or a pause. The generic backchannels of interest included 12 *mhm*s, 7 *uh huh*s, and 11 *yeah*s. For the specific backchannels there were 10 *oh*s, 11 *really*s, and 9 from a more varied category of responses including, for example, *gee*, *whoa*, and *wow*.

Target words were identified for each triad, and consisted of unique content words found in the following turn at talk by the speaker. Thus, there were no further addressee responses between the critical backchannel location and the target word. Target words were identified from a variety of word categories and varied in length from 1 to 4 syllables (average = 2.1). Target words followed the critical backchannel location by 1 to 15 words, with an average distance of 8 words (average time = 2.15 *s*). From the same corpus we also selected 15 filler stimuli and 4 training stimuli. Filler trials did not contain targets, preventing participants from adopting a strategy of immediate response.

Design. Three lists were created that contained equal numbers of specifics, generics, and pauses, as well as equal numbers of authentic and manipulated audio clips, with no stimuli created from the same audio being presented together in the same list. Both filler and target stimuli contained a number of noncritical backchannels, so it was unlikely that participants adopted a strategy of listening for backchannels to predict a word's occurrence.

Procedure. Instructions were presented on screen. After reading the instructions, participants were given four practice trials before starting the task. Each trial consisted of the presentation of a target word to monitor for, followed by the presentation of an audio conversation. Participants first saw a centered fixation point for 500 ms, followed by the presentation of the target word for 3500 ms. After the presentation of target word, the screen was cleared and the audio clip started. Participants pressed a reaction button as soon as they heard the target word. If they did not hear the target word they pressed no button. Participants were randomly assigned to one of the three list conditions, each consisting of 30 target and 15 filler trials.

Results

Reaction time was measured from the onset of the target word to when the participant pressed the button on the reaction pad. Participants who failed to respond to at least two-thirds of the critical trials were not considered on task and were dropped from the analysis (9 total). One item was abandoned because the target word was phonetically similar to a word earlier in the conversation, causing the majority of participants to respond prematurely. Latencies longer than three standard deviations from the mean were removed (27 data points in total), leaving an average of 24 critical trials responded to for the 80 remaining participants. Remaining latencies were analyzed with a 2 (discourse relation: continuation or elaboration) x 3 (backchannel type: specific, generic, pause) repeated measures ANOVA.

There was no main effect of discourse relation, $F(1, 78) = 0.65$ $p = .42$, and no interaction, $F(2,77) = .94$ $p = .40$. There was a main effect of backchannel type, $F(2,77) = 7.07$, $p = .002$. Post hoc comparisons with Bonferonni adjusted alpha levels revealed that recognition of words following specific backchannels, $M = 1203$ *ms*, $SD = 442$, was slower than recognizing words following generic backchannels, M = 1056, SD = 349, mean difference = 147 *ms*, $SE = 55$, $p = .016$, 95% CI = [22,271]. Similarly, words following specific backchannels were recognized more slowly than words following pauses, $M = 1012$, $SD = 347$, mean difference = 190 *ms*, $SE = 51$, $p = .002$, 95% CI = [62, 319]. Response latencies for words following generic backchannels and pauses did not differ, $p > .05$. See Figure 1 for a summary of the results.

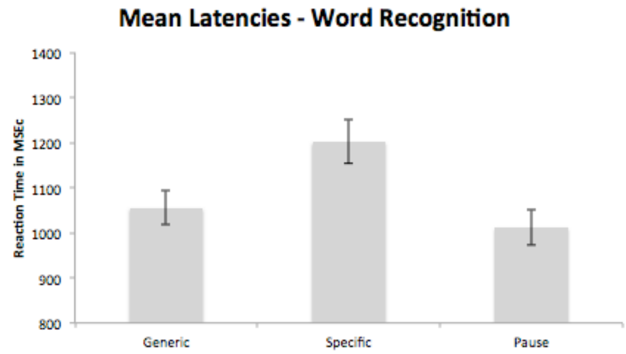

Figure 1: Mean word recognition latencies after generic and specific backchannels, and pauses by discourse relation, (error bars present SE).

Discussion

Listeners' specific backchannels slowed overhearers' identification of subsequent words in a speaker's talk in comparison to generic backchannels and pauses. These data go against the tuned out hypothesis, but support both the distraction and coordination hypotheses. Whether or not the next turn provided elaborative information, as the specific backchannel would predict, specific backchannels caused overhearers to have reduced ability to monitor the next turn. This could be caused through distraction from more informative responses, or because specific backchannels lead overhearers to maintain activation of the content of the prior turn, splitting cognitive resources. Experiment 2 tests these hypotheses.

Experiment 2 – Semantic Verification

We tested how well overhearers recognized words preceding specific versus generic backchannels in comparison to the same talk with pauses instead of backchannels.

Method

Participants. 88 students from the University of California, Santa Cruz participated for course credit.

Materials. The same stimuli from Experiment 1 were used in Experiment 2. Target words were unique content words from the talk preceding the target backchannel. Targets were from a variety of word categories, preceded the target backchannel by an average of 7.13 words (range 1 to 15), and were on average 1.9 syllables in length (range 1 to 3). As before, the 15 filler trials did not contain the target word, but did contain a semantically related word in the turn prior to the target backchannel or pause.

Design. The design was the same as Experiment 1.

Procedure. Participants were randomly assigned to one of the three counter-balanced lists. For each trial, participants listened to the audio clip of the conversation while watching a fixation cross on the screen. At a certain point in each trial, a word replaced the fixation cross. Participants pressed the reaction button as quickly as possible if they remembered hearing the presented word spoken in the conversation. For critical trials, the target word was present in the speaker's turn prior to the critical backchannel or pause. No other listener feedback, besides the backchannel of interest in the backchannel conditions, intervened between the target word and the verification prompt. The visual prompt was displayed on the screen at the onset of the target word in the following turn after the backchannel used in the Experiment 1. This created an average distance of 14.67 intervening words between target and memory probe, (average time = 4.18 *s*). After reading instructions, participants were given four practice trials before starting the task. Reaction times were measured from the onset of the visual display to the button press.

Results

Data from participants who failed to respond to at least two thirds of the critical trials were removed (n = 6). As before, latencies three standard deviations above the mean were also removed (32 data points). Data were then entered into a 2 (discourse relation type) x 3 (backchannel type) repeated measures ANOVA.

The ANOVA revealed a significant interaction between backchannel and discourse relation, $F(2, 80) = 8.02$, $p = .001$, (see Figure 2). Separate one-way repeated measures ANOVAs were run for the different discourse types. Specific backchannels had different effects depending on whether they were followed by continuations or elaborations.

For trials in which the next turn was a continuation, a main effect of backchannel type was found, $F(2, 80) = 4.0$, $p = .02$. Adjusted post hoc comparisons of backchannel type revealed that words prior to specific backchannels, $M = 1243$ *ms*, $SD = 345$, were verified more quickly than words prior to pauses, $M = 1359$, $SD = 484$, mean difference = -116 *ms*, $p = .02$, 95% CI = [-219, -12]. Words prior to specific backchannels were not verified significantly faster than words prior to generic backchannels, $M = 1295$, $SD = 350$, nor was there a significant difference in verification latencies for words prior to generic backchannels compared to pauses, all *p*s > .05.

For elaborative next turns, a main effect of backchannel was also found, $F(2, 80) = 4.6$, $p = .013$. Adjusted post hoc comparisons revealed that for next turn elaborations, words prior to specific backchannels were verified more slowly than words prior to generic backchannels, ($M = 1507$, $SD = 461$ for specific, and $M = 1366$, $SD = 424$ for generic), mean difference = 156, $p = .018$, 95% CI = [21, 293]. Words prior to specific backchannels were also verified more slowly than words prior to pauses, $M = 1350$, $SD = 374$, mean difference = 157, $p = .035$, 95% CI = [7, 275].

Discussion

Listeners' specific backchannels had opposing effects depending on the discourse relationship across speaker turns. Specific backchannels slowed overhearers' verification of prior words in a speaker's talk when the subsequent talk consisted of elaborative information, and sped overhearers' verification of prior words when the subsequent talk contained a continuation of the narrative. Responses to generic backchannels and pauses were similar. These data go against the tuned out hypothesis. They also go against the distraction hypothesis, as specific backchannels were not uniformly distracting. The data are compatible with the coordination hypothesis.

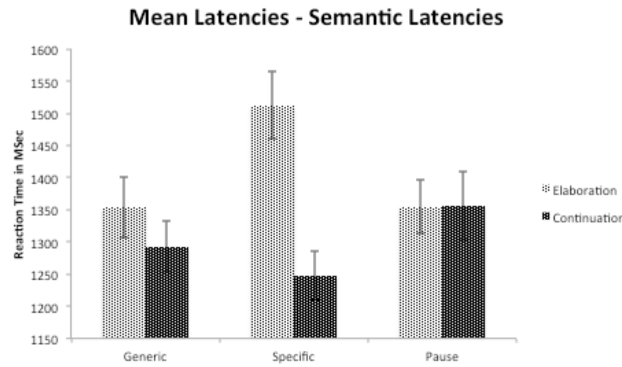

Figure 2: Mean response times in verifying target word prior to generic backchannels, specific backchannels, and pauses, by discourse relation (Error bars present SE).

General Discussion

In Experiments 1 and 2 we found evidence that listeners' specific backchannels affected how overhearers comprehended dialogue. By hypothesis, when overhearers hear a specific backchannel, they expect the next turn to present elaborative information on the content of the previously presented discourse event, causing the previous turn's content to remain active longer in order to facilitate integration. This activation has three effects.

First, specific backchannels' activation of prior talk reduces monitoring of the subsequent talk in comparison to generic backchannels and pauses. Generic backchannels suggest an upcoming continuation (Tolins & Fox Tree, submitted), allowing overhearers to process the next turn without maintaining heightened access to previous talk. Thus, for overhearers a distinct pattern across generic responses from listeners and no responses at all is not visible in our data. Specific backchannels, in contrast, lead overhearers to maintain strengthened access to the previous turn, dividing attention between prior talk and upcoming talk. This divided attention is manifested as slower reaction times to the monitoring of upcoming words.

Second, specific backchannels' activation of prior talk increases access to prior talk when subsequent talk is a discursive continuation. Because an elaboration is expected, overhearers maintain increased access to the content of the prior turn. When, in contrast with this expectation, a continuation is presented this increased access is visible as faster semantic verification of prior content words. With generics and pauses, previous talk is not maintained in heighten activation, and so verification is slower.

Third, specific backchannels' activation of prior talk decreases access to prior talk when subsequent talk is an elaboration. When the elaboration is expected and heard it is integrated with the activated information from the prior turn, updating the overhearers' discourse model. This rapid updating interferes with the recognition of discourse old information, leading to increased latencies in verification.

A number of studies have demonstrated that discourse comprehension involves the integration of information at both local and global levels (Hagoort & van Berkum, 2007). In the context of collaborative dialogue, this integration is modulated turn by turn through attention to responses from active listeners, and the predictive relation between these responses and the unfolding discourse structure. While the role of prediction in comprehension has been previously explored (Federmeier, 2007; Pickering & Garrod, 2013), this is the first study to demonstrate that in the context of dialogue comprehension, overhearers make use of predictive relations across conversational participants.

Historically, research pertaining to understanding the status of overhearers has focused on the relationship between the overhearer and the speaker. Overhearers tend to perform worse on particular tasks compared to addressees who can actively participate in the construction of the speaker's talk (Schober & Clark, 1989). Overhearers do better, however, when listening in to a dialogue compared to listening in to a monologue (Fox Tree 1999; Fox Tree & Mayer, 2008). One possible reason that overhearing dialogues leads to better performance than overhearing monologues may be that addressee feedback enhances comprehension. Backchannels may have cued overhearers to predict the type of information in the next turn. Thus, third-party dialogue comprehension does not simply involve becoming another listener. Instead, understanding talk in interaction involves actively coordinating information from all active participants, based on the collaborative, jointly constructed nature of natural conversation.

References

Akhtar, N., Jipson, J., Callanan, M. (2001). Learning words through overhearing. *Child Dev.*, *72*(2), 416-430.

Bangerter, A., & Clark, H. H. (2003). Navigating joint projects with dialogue. *Cognitive Science, 27,* 195-225.

Bavelas, J. B., Coates, L., & Johnson, T. (2000). Listeners as co-narrators. *Journal of personality and social psychology, 79,* 941-952.

Bly, B. (1993). Uncooperative language and the negotiation of meaning (Unpublished doctoral dissertation). Standford University, Standford, CA.

Boersma, P. & Weenink, D. (2005). Praat: doing phonetics by computer (vesion 5.3.42). Retrieved from http://www.praat.org/.

Branigan, H., Catchpole, C., & Pickering, M. (2011). What makes dialogues easy to understand? *Language and Cognitive Processes*, *26*(10), 1667-1686.

Brennan, S. E. & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Exp. Psych., 22,* 1482-1493.

Brown-Schmidt, S. (2009). Partner-specific interpretation of maintained referential precendents during interactive dialog. *Journal of Mem. & Lang., 61,* 171-190.

Brunner, L. (1979). Smiles can be back channels. *Journal of Personality and Social Psychology, 37*(5), 728-734.

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

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

Clark, H. H. & Schaefer, E. (1987). Concealing one's meaning from overhearers. *Journal of Memory and Language, 26,* 209-225.

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

Federmeier, K. (2007). Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology, 44,* 491-505.

Fox Tree, J. (1999). Listening in on monologues and dialogues. *Disc. Proc., 27*, 35-53.

Fox Tree, J. & Mayer, S. (2008). Overhearing single and multiple perspectives. *Disc. Proc., 45*, 160-179.

Fox Tree, J. & Schrock, J. (1999). Discourse markers in spontaneous speech: Oh what a difference an oh makes. *Journal of Memory and Language, 40*, 280-295.

Garrod, S. & Anderson, A. (1987). Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition, 27*, 181-218.

Goffman, E. (1981). *Forms of talk.* Philadelphia: University of Pennsylvania Press.

Goodwin, C. (1986). Between and within: Alternative sequential treatments of continuers and assessments. *Human Studies, 9,* 205-217.

Hagoort, P. & van Berkum, J. (2007). Beyond the sentence given. *Phil. Trans. R. Soc. B, 362*, 801-811.

Kraut, R., Lewis, S., & Swezey, L. (1982). Lisstener responsiveness and the coordination of conversation. *Journal of Pers. and Soc. Psych., 43*(4), 718-731.

Marlen-Wilson, W. & Tyler, L. K. (1980). The temporal structure of spoken language understanding. *Cognition, 8,* 1-71.

Pickering, M. J. & Garrod, S. (2013) Forward models and their implications for production, comprehension, and dialogue. *Behavioral and Brain Sciences, 36,* 377-392.

Schober, M. (1998). How addressees affect spatial perspective choice in dialogue. In: P. L. Olivier & K. P. Gapp (Eds.), *Representation and processing of spatial expressions* (pp. 231-245). Mahwaw, NJ: Lawrence Erlbaum.

Schober, M. & Clark, H. H. (1989). Understanding by addressees and overhearers. *Cog. Psych., 21,* 211-232.

Tolins, J. & Fox Tree. J. E. *Backchannels in Collaborative Narration.* Manuscript submitted for publication.