

Running Head: OVERHEARERS USE ADDRESSEE BACKCHANNELS

For the definitive version, please refer to: Tolins, J. & Fox Tree, J. E. (2016). Overhearers use addressee backchannels in dialog comprehension. *Cognitive Science* 40(6), 1412–1434.

Overhearers Use Addressee Backchannels in Dialog Comprehension

Jackson Tolins and Jean E. Fox Tree

University of California Santa Cruz

## Abstract

Observing others in conversation is a common format for comprehending language, yet little work has been done to understand dialogue comprehension. We tested whether overhearers use addressee backchannels as predictive cues for how to integrate information across speaker turns during comprehension of spontaneously produced collaborative narration. In Experiment 1, words that followed specific backchannels (e.g. *really*, *oh*) were recognized more slowly than words that followed either generic backchannels (e.g. *uh huh*, *mhm*) or pauses. In Experiment 2, we found that when the turn after the backchannel was a continuation of the narrative, specific backchannels prompted the fastest verification of prior information. When the turn after was an elaboration, they prompted the slowest, indicating that overhearers took specific backchannels as cues to integrate preceding talk with subsequent talk. These findings demonstrate that overhearers capitalize on the predictive relationship between backchannels and the development of speakers' talk, coordinating information across conversational roles.

**Keywords:** Dialogue; Comprehension; Backchannels; Overhearers; Narrative

### Overhearers Use Addressee Backchannels in Dialog Comprehension

Dialogue, as a form of joint action, necessarily involves the active collaboration of both the current speaker as well as actively participating listeners, to whom the talk is directed (Clark, 1996). These active listeners are called *addressees*. Addressees are not, however, the only type of listeners, as dialogues may also potentially involve other people who hear the conversation but do not participate directly. These non-participating listeners are called *overhearers*. Overhearing the dialogues of others is a common experience, from the audiences of political debates and television dramas to eavesdropping on conversations at work or at the bus stop. Despite overhearing being a common format for language comprehension, and even language acquisition (Akhtar, 2005; Akhtar, Jipson, & Callanan, 2001), little empirical research has been conducted to understand the processes involved in third party dialogue comprehension.

Studying dialogue comprehension reveals some unique features of talk produced in interactive contexts. Overhearers have been shown to perform better when listening to dialogues compared to listening to monologues produced by an isolated speaker (Fox Tree, 1999). This advantage of understanding dialogue may be because dialogues provide more information than monologues, for example by encouraging more perspectives to be discussed (Fox Tree & Mayer, 2008), by encouraging speakers to present information in a more optimal way (Branigan, Catchpole, & Pickering, 2011), or ensuring that talk is clearly expressed. In all these explanations, addressee participation, when present, steers the course of the developing talk, making the speaker's contributions more generally comprehensible (see also Bavelas, Coates, & Johnson, 2000; Fox Tree & Clark, 2013; Schober & Brennan, 2003). These explanations of the effect on overhearer comprehension rely on addressees'

influencing the speaker's talk, rather than relying directly on addressees' own contributions. In the present study we tested the degree to which overhearers take account of addressee participation: To what degree was dialogue understood as a collaborative activity between two interactants by third party observers, and how did this influence comprehension?

In a typical conversation, both interactants take turns speaking and contributing. However, even in more strictly regimented interactions, such as in storytelling where the roles of speaker and addressee are entrenched over a larger sequence, addressees participate through *backchannels*, short responses to speakers' talk (Goodwin, 1986; Oreström, 1983; Schegloff, 1982; Yngve, 1970). Backchannel communication – variously referred to as acknowledgement tokens (Jefferson, 1984), reaction/reactive tokens (Clancy, Thompson, Suzuki, & Tao, 1996; Wilkinson & Kitzinger, 2006), response tokens (Gardner, 2001), and accompaniment signals (Kendon, 1967) – subsumes a broad range of expressions, including verbal responses such as *mhm*, *oh*, and *really*, vocal responses such as laughter and gasps, and facial expressions and other embodied displays.

The types of backchannels addressees provide, and the degrees of comprehension and agreement displayed (Bangerter & Clark, 2003; Brunner, 1979), are an integral part of the process of *grounding* in conversation, through which speakers and addressees add information to a conversation in a cumulative way (Clark & Schaefer, 1986). In story telling dialogue, backchannels may also take on a role of *co-tellings*, taking part in the production of the narrative (Bavelas et al., 2000; Norrick, 2010). In this context, backchannels have effects on both the global qualities of the narrative, as well as on how the discourse develops turn by turn (Tolins & Fox Tree, 2014). Backchannels have been analyzed as having two distinct

functions: they function reactively, displaying uptake of the prior talk, and they function proactively, steering the dialogue in a particular direction (Tolins & Fox Tree, 2014).

We propose that backchannels are not only important to conversational addressees, but also to conversational overhearers. While overhearers are not active participants, as language users they have experience engaging in dialogues themselves. As such they may be sensitive to the role backchannels play in discourse. This reconceptualization of overhearers, as not simply secondary or reduced-access listeners, widens the notion of what comprehending dialogue entails. Comprehending dialogue differs from comprehending monologue on a number of dimensions; importantly, one of these dimensions is the collaborative nature in which the talk is jointly produced by speakers and addressees. In the current experiments, we manipulated the presence of different types of backchannels in response to speakers' story telling. Comprehending talk produced collaboratively, integrating information from speaker and addressee rather than solely comprehending a single speaker's talk, suggests that the verbal responses addressees provide should influence overhearers' comprehension of speakers' talk in two-party dialogue.

### **Addressees and Overhearers**

There are a variety of ways that people can participate in a conversation (Goffman, 1981). Although there is typically only one speaker at a time, there are many different types of listeners. Addressees are *ratified listeners*, people to whom the speakers direct their talk, and from whom speakers receive backchannel responses. Overhearers can be *ratified* or *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). 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* (Clark & Schaefer, 1987; Dynel, 2011; Goffman, 1981; Schober, 1998). The child listening to parents can be either a bystander or an eavesdropper, depending on the parents' knowledge. The speakers whose communication was tested in our studies knew they were being recorded for potential future comprehension experiments, thus making our experimental participants ratified overhearers of the bystander sort, whom 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). While the category of side participation is not monolithic, with varying degrees to which other listeners may be considered privy to the conversational common ground, in comparison to addressees, non-active listeners are treated by other speakers as being less capable of capitalizing on information previously established within a conversation (Wilkes-Gibbs & Clark, 1992). This reduced access causes overhearers to 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 they fare worse in story retelling (Clark & Wilkes-Gibbs, 1986; Kraut, Lewis, & Swezey, 1982; Schober & Clark, 1989). Through the development and use of mutually shared knowledge, active interactants develop partner-specific conceptual pacts, which are collaboratively established and maintained agreements regarding how to refer to particular objects (Brennan & Clark, 1996; Brown-Schmidt, 2009). The partner-specific nature of these conceptual pacts can create a barrier to overhearers' comprehension, at times even by design (Clark & Schaefer, 1987). However, engagement in the interaction can also sometimes lead to

addressees' underperforming overhearers, such as in detection of evasion (Bly, 1993).

Because addressees are vested in co-constructing meaning with a speaker, they often fail to spot evasive language that overhearers detect.

These studies compared overhearers and addressees as different types of listeners. Other research has focused on different formats of overhearing, for example contrasting overhearing monologue with dialogue. In studying what overhearers glean from a dialogue, prior researchers have generally focused on the propositional content of the dialogues, such as the number of perspectives presented, whether a perspective was metaphorical or geometrical, the number of details, the number of substantive addressee contributions, the liveliness of descriptions, and the way the concepts were refined over time (Branigan et al., 2011; Fox Tree & Mayer, 2008; Schober & Clark, 1989). Some researchers have gone beyond propositional content, testing the role of more versus fewer dialogue coordination devices, such as discourse markers and backchannels (Branigan et al., 2011; Fox Tree, 1999; Fox Tree & Mayer, 2008), or other spontaneous phenomena such as disfluencies and fillers (Fox Tree & Mayer, 2008). We focus in further on the role of particular backchannels on the processing of subsequent and prior information in a conversation. Investigating the number of backchannels as a group (Fox Tree & Mayer, 2008) may obscure the role of different types of backchannels in overhearers' comprehension of dialogue. As overhearers listen to others' dialogue, responses from addressees may inform overhearers' expectations about upcoming talk (Tolins & Fox Tree, 2014). We explore how it is that backchannels may function in this proactive manner in the following section.

### **Verbal Backchannels in Dialogue**

Rather than being passive recipients of information, addressees actively participate in the construction of ongoing dialogue, even when they are only producing backchannels. Backchannels serve interactional functions at different levels within an interaction, such as indicating continued attention on the part of the addressee, displaying comprehension, and presenting the addressee's stance on the speakers' talk (Bangerter & Clark, 2003; Brunner, 1979; Gardner, 2001). The types of responses addressees make, and the degree of understanding they display in response to speaker talk, provide structure to the joint activity, allowing the speaker to make appropriate next contributions (Bangerter & Clark, 2003).

These different responsive functions have been used to categorize distinct types of backchannels, which display distinct reactions to the prior talk. *Continuers*, such as *mhm* and *uhuh*, indicate that the prior talk has been understood, allowing the speaker to continue (Goodwin, 1986; Schegloff, 1982). Continuers have also been labeled *generic* backchannels, highlighting that the meaning of these responses are context generic – they display attention and comprehension but do not rely on the content of the previous talk for meaning (Bavelas et al., 2000; Stivers, 2008). Their use as displays of comprehension does not require specific uptake of the speaker's prior turn. *Assessments* include expressions that mark prior talk as discourse-new or newsworthy, such as *really* and *oh*, and expressions of affective stance towards the content of the speaker's talk, such as *wow* or *gross* (Gardner, 2001; Heritage, 1984; Norrick, 2010). Assessments display understanding on the part of the addressee and also provide information on addressees' affective or informational stance towards the particular content of the prior talk. As such, this category has also been labeled *specific* backchannels, highlighting that both informational and affective responses are context specific – their roles are tied to the conversational context in which they occur (Goodwin,

1986; Kraut, Lewis, & Swezey, 1982; Yngve 1970; Stivers, 2008). In collaborative narration, for example, specific backchannels allow the addressee to take on the role of co-narrator, indicating not only understanding of the story as told, but taking part in its production through the use of context appropriate affective displays (Bavelas, et al., 2000). These functionally distinct categories of backchannels have also been found to exist in spontaneous dialogue across various languages (Clancy, et al., 1996; Ward & Tsukahara, 2000).

While the majority of studies exploring backchannels have focused on their role as reactions to the prior talk (Clancy, et al., 1996; Gardner, 2001), backchannels may also be viewed as proactive, taking part in the construction of the dialogue (Bangerter & Clark, 2003; Tolins & Fox Tree, 2014). Addressee responses of specific and generic backchannels have distinct effects on the development of speaker talk and narration. Because specific backchannels respond to the content of the current speaker's speech, such as demonstrating disgust, sorrow, or surprise at appropriate points in a story (Bavelas, et al., 2000), they are likely to be explicitly responded to by speakers, ratifying their stance or incorporating them into the narrative (Norrick, 2010b; 2012). Addressees' production of specific versus generic backchannels can change the course of a narrative (Bavelas, et al., 2000). In an experiment in which addressees were distracted by a secondary task (counting the number of *t*'s in the speaker's talk), addressees made fewer contributions to the conversation and produced fewer specific backchannels. In response, speakers told stories that were of poorer quality with less climactic endings.

Beyond the global features of narratives, such as abstract or descriptive language or more climactic stories, specific and generic backchannels make distinct contributions to narratives as they develop turn by turn, shaping the directly subsequent unfolding of the story

(Norrick, 2010b, 2012; Tolins & Fox Tree, 2014). While both types of backchannels ground speakers' developing stories, specific backchannels are more likely to be explicitly incorporated into the story-telling activity (Norrick, 2010b, 2012). In a previous analysis of spontaneously produced dialogue, we found that storytellers respond to specific and generic backchannels in systematic ways in the next turn, with specific backchannels acting as requests for elaboration on the previously established discourse event. Generic backchannels, in contrast, were more likely to elicit a new discourse event in the following talk (Tolins & Fox Tree, 2014). In an experiment testing the predictive power of these relations, participants were asked to read dialogues up to a target backchannel and contribute suggestions as to how they believed the discourse would develop next. Holding the content of the story constant, when participants read generic backchannels on a transcript, they expected speakers to continue their subsequent talk with some discourse new event, or *continuation*. However, when they read the same transcript with a specific backchannel instead, they were more likely to contribute an explanation of the prior talk, or *elaboration* (Tolins & Fox Tree, 2014).

### **Current Studies: Addressee Backchannels in Overhearer Comprehension**

Backchannels play a key role in the collaborative production of talk. As demonstrated, different categories of backchannels have distinct influences on the types of discourse development following their production. Within the activity of producing talk, then, it appears clear that addressees actively ground talk in distinct ways, displaying various degrees of uptake of a speaker's talk and, by doing so, contributing to the ongoing creation of meaning and the production of joint action (Brennan, Galati, & Kuhlen, 2010). It remains uncertain, however, whether these minimal addressee contributions are of interest or value

for those engaged in observational, third party comprehension, such as overhearers. Should overhearers make use of backchannels as a part of third party dialogue comprehension, it is further unknown whether this would involve a reactive backchannel function of highlighting prior talk or a proactive/predictive function of connecting addressee responses to discourse developments in the next turn.

Overhearer comprehension of talk produced in dialogue might be influenced by either of these two main functions ascribed to addressee backchannels (reactive and proactive). For either of these roles to influence overhearer comprehension, third party comprehension would entail capitalizing on the collaborative nature of dialogically produced talk. If backchannels play a role in comprehension for non-participating listeners, this would demonstrate that listening to dialogue is not the same as listening to monologue. We are not merely suggesting that two interlocutors contribute to a dialogue, and so overhearers pay attention to both. Instead, we propose that an addressee's backchannels during a speaker's production of talk influences how that speaker's talk is processed.

In support of this collaborative view, sensitivity to mutually shared knowledge, or the common ground of two interlocutors, has been demonstrated in third party comprehension. Sixteen-month-olds who observed an interaction looked longer when addressees violated assumptions of common ground, for example by responding to "Can you give me the key?" by offering a key from privileged ground rather than a key in common ground (Martin & Vouloumanos, 2013). That is, the toddlers made use of conversational common ground even when not actively engaged in the interaction themselves. Support for this perspective also comes from studies contrasting overhearing full dialogues and halfalogues, in which only one speaker's talk is available to an overhearer (Emberson, Lupyan, Goldstein, & Spivey, 2010).

Overhearers listening to halfalogues were more distracted, suggesting that they used interactive cues to predict such things as speaker turn exchanges and contribution types (Emberson, et al., 2010).

To test how various verbal backchannels and pauses may influence overhearers' comprehension of surrounding talk, we made use of two reaction time paradigms involving listening to spontaneously produced conversation. We tested processing of information after the backchannels or pauses with a word monitoring technique (Marslen-Wilson & Tyler, 1980). Word monitoring has been used in many prior studies to measure spontaneous speech comprehension (Fox Tree, 1995; 2001; Fox Tree & Schrock, 1999; Tomlinson & Fox Tree, 2011). Here, the task captures the attentional and integrative processes following target backchannels up to the test point at which the speaker produces the monitored word (Fox Tree, 1995; 2001). We tested the processing of information prior to the backchannels and pauses with a word verification technique, which has also been used in prior studies of spontaneous speech comprehension (Fox Tree & Schrock, 1999). The verification task captures the maintenance of activation of prior discourse information during processing of subsequent information (Fox Tree & Schrock, 1999). Pairing the two techniques allows us to more precisely determine how addressee backchannels influence overhearer comprehension of surrounding talk.

### **The Attention Hypothesis**

Given the dual nature of the function of backchannels as both responses and cues for what speakers might say next, we propose to contrast two opposing hypotheses. If overhearers focus on the responsive nature of backchannels, this would suggest that they make use of these reactions as cues for what parts of the speakers' talk deserve heightened

attention. As such, specific backchannels may be responded to differently from generic backchannels or pauses. Because they flag prior talk as particularly interesting, worthy of evaluation, or discourse-new (Gardner, 2001; Heritage, 1984), specific backchannels may direct attention back to the remarked-upon prior talk. Because they mark the prior talk as accepted but otherwise not noteworthy, possibly a part of the already established common ground, generic backchannels may draw overhearers away from prior talk and direct attention to subsequent talk. We will call these paired predictions the *attention hypothesis*.

According to the attention hypothesis, specific backchannels will direct overhearer attention to prior talk, causing words from subsequent turns to be recognized more slowly, as measured by the word monitoring paradigm, and words from prior turns to be verified more quickly after specific backchannels than after generic backchannels or pauses, as measured by the word verification task. Also according to the attention hypothesis, generic backchannels will direct overhearer attention to subsequent talk, causing words from subsequent turns to be recognized more quickly and words from prior turns to be verified more slowly after generic backchannels than after specific backchannels or pauses. The attention hypothesis does not predict any interactions between the influence of backchannels on overhearer comprehension and the discourse relation of the surrounding speaker talk.

For comparison to what happens without a backchannel present, we included stimuli with pauses instead of backchannels. The pauses were necessary to indicate a break in the speaker's presentation of ideas, as occurred in the stories told with the backchannels. That is, without pauses, the backchannel-free condition would be heard as a smooth flow of ideas whereas the backchannel-present conditions would be heard as disrupted story tellings. Of course, a pause between speakers' turns could also have meaning (cf. Fox Tree, 2002). But in

the current contexts of face-to-face spontaneous talk, it is not clear what that meaning is, or whether it would be the same in each instance. Given the possibility of nonverbal addressee responses such as nods or smiles, it is not possible to distinguish between pauses indicating a lack of addressee uptake or those in which some other response is provided. As such, while the possibility for other alternatives remain (see General Discussion), for present purposes the pause stimuli serve as a baseline comparison against which any effect of backchannel type on overhearer comprehension can be measured.

### **The Predictive Hypothesis**

Rather than using the responsive nature of backchannels as cues for allotting attention, overhearers might instead make use of the predictive function these backchannels play in steering the discourse. We will call this the *predictive hypothesis*. As systematic cues to how narrative interaction is likely to develop next (Tolins & Fox Tree, 2014), addressee backchannels may provide overhearers with useful information, allowing them to make predictions at the level of the jointly produced discourse and adjust their comprehension accordingly.

In particular, backchannels may act as cues to the type of contribution speakers are likely to make next. Without a backchannel at all, the default next turn in a story is discourse-new. This is because the normal progression in a conversation involves building up information by adding new events onto a background of shared knowledge (Clark & Brennan, 1991; Clark & Schaefer, 1987). Generic backchannels support the production of a discourse-new next turn (Goodwin 1986; Schegloff, 1982), while specific backchannels support elaborative next turns, as has been shown in the story continuation experiment

presented earlier (Tolins & Fox Tree, 2014). The predictive hypothesis proposes that overhearers will use specific backchannels to predict upcoming elaborative next turns.

More specifically, specific backchannels will cue overhearers to the upcoming need to integrate information across turns, causing overhearers to maintain heightened activation to the content of the prior turn in order to facilitate rapid integration. This heightened activation of the content of the prior turn should be visible in both the word monitoring and word verification paradigms. If overhearers maintain access to the previous utterance following a specific backchannel because they expect to update this information, recognition of subsequent talk should be reduced, leading to slower monitoring. Specific backchannels should also affect performance on the word verification task, but in different ways depending on the type of discourse development that follows the specific backchannel. By hypothesis, specific backchannels will cause overhearers to maintain activation of the prior turn's discourse content. Discourse-new continuations following specific backchannels, while not predicted by the addressee's response, will not be integrated as novel elaborative information and thus will not interfere with access to prior information. This will be visible as heightened access of prior information, or faster verification times, in comparison to generic backchannels or pauses. However, when the elaborative next turn predicted by the specific backchannel is indeed provided by the storyteller, the rapid integration of the information will be visible as interference with accessing prior information. That is, specific backchannels signal that preceding information should be maintained in memory because the information will soon be elaborated upon. As elaborations are processed, that preceding information is updated and revised – as expected, decreasing access to the original preceding

information. Thus word verification after specific backchannels followed by elaborations should be slower than after generic backchannels or pauses.

### **Summary of Current Studies**

To distinguish between the attention hypothesis and the predictive hypothesis, we compared how specific and generic backchannels influenced overhearers' comprehension as they listened along to spontaneous conversation. We restricted overhearing to situations in which the addressees' responses were clearly secondary to the speaker's, that of storytelling. In narrative dialogues, the roles of speaker and listener are clearly distinct, as the storyteller typically has higher, if not complete, epistemic access to the content of the story. Thus, overhearers could understand the talk completely without the addressee contributions. An effect of backchannels on comprehension would suggest that overhearers are influenced by the interactive role of backchannels, either as responses cueing attentional resources, or as responses shaping the speaker's talk in predictable ways, preparing overhearers for likely next contributions.

### **Experiment 1: Monitoring for Words after Backchannels**

We compared how quickly overhearers monitored for words following specific backchannels, generic backchannels, and pauses, controlling for information before and after the response points. That is, we compared word monitoring in the same stimulus when it was heard either with a specific backchannel, a generic backchannel, or a pause. Target words all occurred in the next turn following the addressee response point.

### **Method**

**Participants.** Eighty-nine students from the University of California Santa Cruz participated in exchange for course credit. Participants were randomly assigned to one of three counter-balanced list conditions.

**Materials.** Stimuli for both experiments were selected from a previously recorded audio corpus of loosely topical spontaneous conversations. In exchange for course credit, undergraduates from the University of California Santa Cruz spoke to each other for 12 minutes. The topic of bad roommate experiences was the starting point, but there was no requirement to stay on this topic. The typical format of the conversations involved taking turns describing experiences, yielding many examples of long speaker turns with a variety of addressee backchannels. From this corpus, 30 audio clips were selected which varied in length from 25 to 85 seconds. Each clip contained either a backchannel that was not spoken in overlap with the speaker's talk or a pause. Audio stimuli were selected in which the voices of both participants in the interaction were heard prior to the target backchannel or pause. Audio stimuli were selected from talk that fit these criteria, such that there was an equivalent number of generic backchannels, specific backchannels, and pauses, and equivalent variation within the general categories of backchannels (see below). This gave 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, 10 contained an authentic specific backchannel, and 10 contained an authentic pause in the speakers' talk in which no backchannel was provided. The discourse level relations across the talk surrounding the target backchannel were categorized as either continuations or elaborations. For continuations, the talk following the backchannel presents a discourse new event, whereas elaborations contained novel or elaborative information on the same discourse event

presented in the turn prior to the target backchannel (see Figure 1). Of the 30 clips, 16 were continuations and 14 were elaborations. All of the authentic generic backchannels were found in discourse continuation contexts, 9 of the 10 authentic specific backchannels were followed by elaborations, and the authentic pauses were evenly split between the two.

(a)

S1: Beth has a similar situation with her roommate right now. Her roommate will not allow any smoking or anything in their room.

S2: [**mhm**/oh/pause]

S1: Beth brought friends to the room once and she freaked out.

(b)

S1: Yeah she totally snored.

S2: That's so funny. Yeah, but what are you gunna say?

S1: Rebecca used to, she used to like totally take the trash bag outta the garbage can and like tie it up and she'd leave it there.

S2: [**mhm/really**/pause]

S1: Like right there in the middle of the floor.

Figure 1: Sample stimuli transcripts of (a) a discourse continuation, in which the speaker presents a discourse new event following the addressee response, and (b) a discourse elaboration, in which the speaker elaborates on the previously presented discourse event. Naturally occurring backchannels in **bold** were replaced with backchannels from the other category as well as pauses.

For each stimulus item, two additional stimuli were created using digital splicing through the audio editor software Audacity (<http://audacity.sourceforge.net/>). The critical backchannel or pause was replaced with a token from the other two categories. Because the generic and specific backchannel pairs were generally different lengths, with the generic backchannels mostly reduced in length compared to the specific backchannels, the onsets of the backchannels were matched and the audio pairs were then edited by adding additional empty sound after the backchannels so that the onset of the next turn following the

backchannel were matched to within two milliseconds. The average pause added after generic backchannels was 74 ms, ( $SD = 118$ ). Manipulating the audio in this way prevented any confound that might arise from the generally shorter length of generic backchannels. Similarly, pauses were created by taking white noise from elsewhere in the audio recording for the same conversation and replacing the backchannel tokens, matching the length of the pause such that the total time between the end of the speaker's last turn and beginning of the subsequent turn remained the same across all three versions of the stimuli. Differences in loudness between specific and generic backchannels were reduced by adjusting the amplification of the generics to match.

The audio manipulation resulted in 90 total stimuli. Each triad contained the exact same audio except for the critical location, which contained a specific backchannel, a generic backchannel, or a pause. The target generic backchannels included 12 *mhms*, 7 *uh huhs*, and 11 *yeahs*. The manipulated specific backchannels included 10 *ohs*, 11 *reallys*, and 9 from a more varied category of responses including variations of *whoa* and *wow*. Each target backchannel type was used an equivalent number of times across naturally produced and edited stimuli. There were also a number of naturally produced noncritical backchannels present in the stimuli, both generic and specific, discouraging participants from adopting a strategy of responding after hearing backchannels.

Target words were identified for each triad. They consisted of unique content words produced by the speaker in the following turn. 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. For the 15 filler trials, target words did not appear anywhere in the dialogue. This discouraged participants from developing strategies for responding, such as responding more quickly when they felt like they were coming to the end of a stimulus. Like the target stimuli, filler stimuli contained a number of backchannels, further discouraging participants from adopting a strategy of listening for backchannels to predict a word's occurrence.

**Design and Apparatus.** Three lists were created that contained equal numbers of specifics, generics, and pauses, as well as equal numbers of continuative and elaborative turns following the target backchannel, with no stimuli created from the same audio being presented together in the same list. Each list consisted of four training trials, 30 target trials, and 15 filler trials. Naturally produced, unedited stimuli and edited stimuli were balanced across the lists. Critical utterances appeared in fixed order, with the backchannel type presentation order varied across list. The procedure was run using the experimental software SuperLab, along with an RB series response pad (Cedrus, Phoenix, Arizona). Trials were presented in randomized order.

**Procedure.** Instructions were presented on screen. For both the current experiment and Experiment 2, participants were informed that the purpose of the study was to explore how understanding language in conversation might be different from listening to just one person talk. As such, they were asked to “listen along to the conversation.” No additional emphasis was placed on listening to any particular conversational role. 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. Time between target and probe did not differ across conditions. Participants were instructed to press a reaction button as soon as they heard the target word. If they did not hear the target word they were to make no response.

## 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. Average response accuracy for each level of the independent variables was entered into a 2 x 3 repeated measures ANOVA. There was no difference in accuracy across backchannel type and no interaction between backchannel and discourse relation; however, participants did respond accurately to more continuation trials ( $M = 4.40$  trials) than elaboration trials ( $M = 3.68$ ),  $F(1, 78) = 90.28, p < .001$ . In order to test whether editing caused stimuli to sound unnatural, potentially affecting results, we conducted paired-sample  $t$ -tests comparing reaction times for naturally-produced and edited-in backchannels of each type. Across the three tests, no differences in reaction times were found for specific backchannels,  $t(79) = -.82, p = .42$ , generic backchannels,  $t(79) = -.47, p = .63$ , or pauses,  $t(79) = -1.57, p = .12$ .

Response 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,  $F(2,77) = 7.07, p = .002$ ,  $\eta_p^2 = .16$ . Post hoc comparisons with Bonferonni adjusted alpha levels revealed that word monitoring responses following specific backchannels,  $M = 1203\text{ ms}, SD = 442\text{ ms}$ , were slower than when following pauses,  $M = 1012\text{ ms}, SD = 347\text{ ms}$ , mean difference =  $190\text{ ms}, SE = 51\text{ ms}, p = .002, 95\% CI = [62, 319]$ . Similarly, word monitoring responses following specific backchannels were slower than words following generic backchannels,  $M = 1056\text{ ms}, SD = 349\text{ ms}$ , mean difference =  $147\text{ ms}, SE = 55\text{ ms}, p = .016, 95\% CI = [22, 271]$ . Response latencies for words following generic backchannels and pauses did not differ,  $p = 1.0$ . See Figure 2 for a summary of Experiment 1 results.

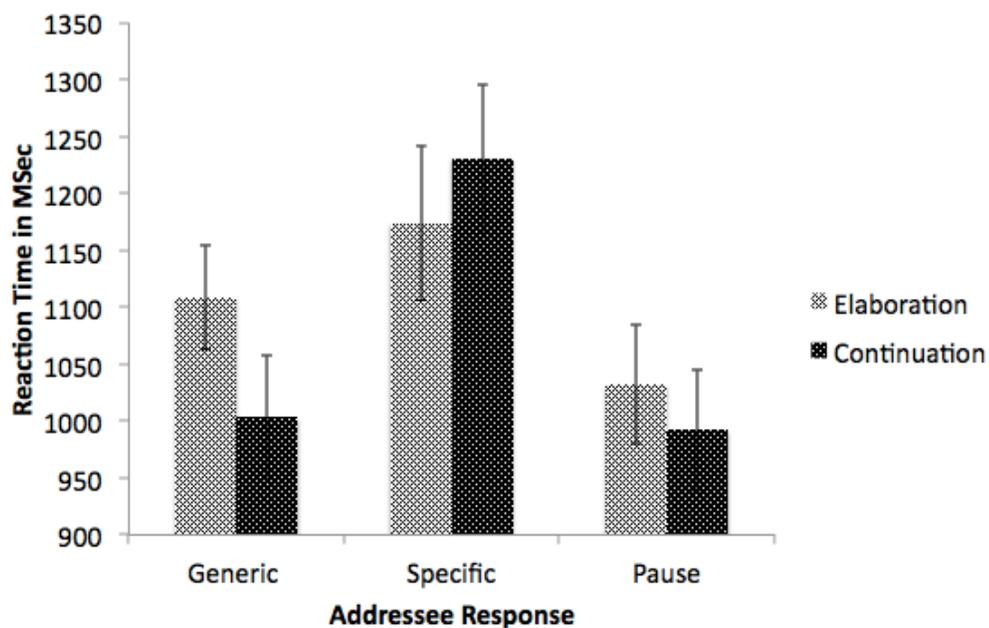


Figure 2: Mean word monitoring latencies after generic backchannels, specific backchannels, and pauses for monitoring within elaborative and continuative next turns. Error bars represent *SE*.

### **Discussion**

Addressees' specific backchannels slowed overhearers' identification of subsequent words in a speaker's talk in comparison to generic backchannels and pauses. Generic backchannels did not lead to better monitoring of the following speaker talk, as predicted by the attention hypothesis in which generic backchannels mark the content of prior talk as uninteresting or previously established within the conversational common ground. However, specific backchannels did cause slower responses than pauses or generics. The attention hypothesis could be revised to accommodate these results: Overhearer attention is influenced, but only by specific backchannels; the drawing of attention back towards prior talk results in reduced monitoring of the speaker's next turn.

Results support the predictive hypothesis, as generic backchannels had the same effect as pauses: Discourse-new information is the default next turn as well as the predicted development following generic responses within the context of conversational story telling. At the same time, specific backchannels slowed word monitoring, supporting the idea that they signaled overhearers to maintain access to the preceding utterance. The effect of specific backchannels was not contingent on whether the following speaker turn was a continuation of the narrative or an elaboration on the previously presented event, but was rather driven by the backchannels themselves. The data suggest that overhearers took the specific backchannels as cues for what might come next in the speaker's talk regardless of whether the prediction was subsequently fulfilled.

The word monitoring experiment demonstrates for the first time that addressee backchannels influence overhearer comprehension. Although the differences in comprehension following generic versus specific backchannels support the predictive hypothesis, they could also support a revised attention hypothesis where specific backchannels are treated as uniquely informative. We further tested how backchannels and pauses affect comprehension with a word verification technique. According to the attention hypothesis, talk preceding specific backchannels should be verified more quickly, regardless of the discourse development of the continuing speaker talk. Contrastively, the predictive hypothesis suggests an interaction between response type and speakers' next turns. Specific backchannels followed by continuations should result in faster word verification in comparison to generic backchannels and pauses followed by continuations. However, specific backchannels followed by elaborations should result in slower verification. Because they are updating prior information, elaborations interfere with recall of the now-overwritten information (to borrow a word processing metaphor).

### **Experiment 2: Verifying Words before Backchannels**

We compared how quickly overhearers verified the presence of words in turns preceding specific backchannels, generic backchannels, or pauses controlling for information before and after the response points. As participants listened to talk following the response point, a word prompt was presented requiring participants to verify whether or not the word had been used previously in the dialogue. This occurred in two discourse relation conditions, one in which the following turn was a continuation, presenting some novel next event, and one in which the following turn was an elaboration, providing further information on the same discourse event described in the prior turn.

## Method

**Participants.** Eighty-eight students from the University of California Santa Cruz participated in exchange for course credit. Participants were randomly assigned to one of three counter-balanced lists.

**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. They 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 and Apparatus.** The design and apparatus was the same as in Experiment 1.

**Procedure.** 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 target backchannel or pause. No other addressee 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 with an average time of 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 (6 total). As before, latencies three standard deviations above the mean were also removed (32 data points). As in Experiment 1, there was no difference in accuracy across backchannel type and no interaction between backchannel and discourse relation; however, participants did respond accurately to more continuation trials ( $M = 4.41$ ) than elaboration trials ( $M = 3.81$ ),  $F(1, 81) = 50.38, p < .001$ . As in Experiment 1, paired-sample  $t$ -tests comparing reaction times for natural and inserted backchannels of each type revealed no differences for specifics,  $t(81) = -1.45, p = .15$ , generics  $t(81) = .47, p = .64$ , or pauses,  $t(81) = 1.41, p = .16$ .

Remaining latencies were analyzed with a 2 (discourse relation: continuation or elaboration) x 3 (backchannel type: specific, generic, pause) repeated measures ANOVA. There was a significant interaction between backchannel and discourse relation,  $F(2, 80) = 8.02, p = .001, \eta_p^2 = .17$ . See Figure 3 for a visual display of the Experiment 2 interaction. Separate one-way repeated measures ANOVAs were run for the different discourse relations.

For continuative next turns, there was a main effect of backchannel type,  $F(2, 80) = 4.0, p = .02, \eta_p^2 = .09$ . Adjusted post hoc comparisons of backchannel type revealed that words prior to specific backchannels,  $M = 1243\text{ ms}, SD = 345\text{ ms}$ , were verified more quickly than words prior to pauses,  $M = 1359\text{ ms}, SD = 484\text{ ms}$ , mean difference =  $-116\text{ ms}, p = .02$ , 95% CI =  $[-219, -12]$ . However, words prior to specific backchannels were not verified more quickly than words prior to generic backchannels,  $M = 1295\text{ ms}, SD = 350\text{ ms}$ , nor was there a difference in verification latencies for words prior to generic backchannels compared to pauses,  $p = .53$  and  $p = .63$  respectively.

For elaborative discourse relations across speaker turns, there was also a main effect of backchannel type,  $F(2, 80) = 4.6, p = .013, \eta_p^2 = .10$ . Adjusted post hoc comparisons revealed that for next turn elaborations, words prior to specific backchannels were verified more slowly than words prior to pauses,  $M = 1507\text{ ms}, SD = 461\text{ ms}$ , for specific backchannels and  $M = 1350\text{ ms}, SD = 374\text{ ms}$ , for pauses, mean difference =  $157\text{ ms}, p = .035, 95\% CI = [7, 275]$ . Words prior to specific backchannels were also verified more slowly than words prior to generic backchannels,  $M = 1366\text{ ms}, SD = 424\text{ ms}$ , for generic, mean difference =  $156\text{ ms}, p = .018, 95\% CI = [21, 293]$ . There was no difference in verification latencies for words prior to generic backchannels compared to words prior to pauses,  $p = 1.0$ .

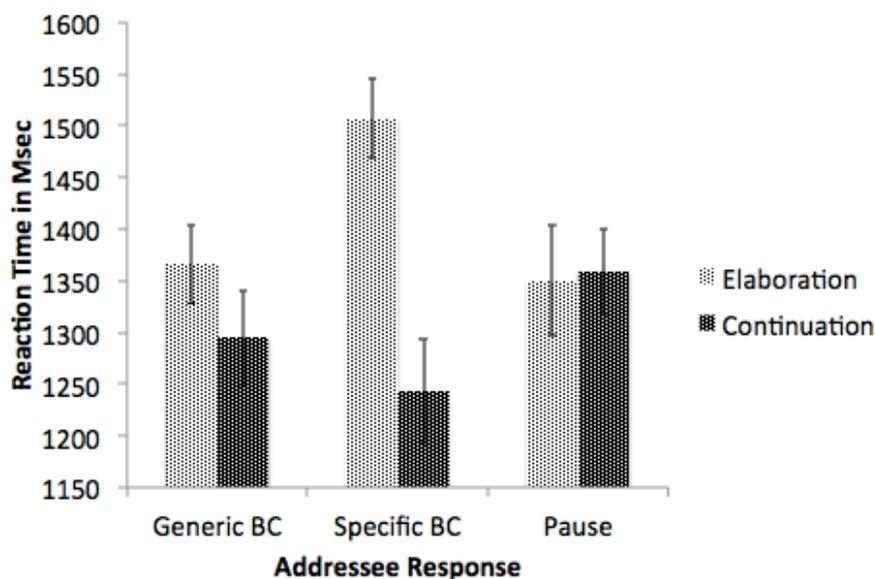


Figure 3: Mean word verification latencies prior to generic backchannels, specific backchannels, and pauses by whether or not the target word was verified during an elaboration or a continuation. Error bars represent *SE*.

## Discussion

Addressees' 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, but sped verification when the subsequent talk was continuative. Responses to generic backchannels and pauses were similar across discourse relations. Results do not support the revised version of the attention hypothesis, in which solely specific backchannels influenced attention, highlighting prior talk to which the addressees respond as either discourse new or worthy of evaluation (see discussion of Experiment 1). For the revised hypothesis to work, responses should be similar across discourse relations for specific backchannels, but they were not.

Results support the predictive hypothesis, which proposes that overhearers make use of the predictive relationship between addressee backchannels and the speaker's following talk. As predicted, there was no difference in effects between generic backchannels and pauses for either continuative or elaborative next turns. Also as predicted, there were differences in effects of specific backchannels depending on whether the next turn was continuative or elaborative. If specific backchannels signal that prior information should be held in memory, discourse continuations on the part of the speaker should not interfere with recall, and they did not. At the same time, because they provide novel information to replace or overwrite the prior talk, elaborations should interfere with access to original prior information, and they did.

One effect was not predicted however. In continuations, words prior to specific backchannels were predicted to be verified more quickly than words prior to generic backchannels and pauses, because the specific backchannel signals that prior information

should be held in memory. While words prior to specific backchannels were verified more quickly than words prior to pauses when presented in continuations, they were not verified more quickly than words prior to generic backchannels. One possibility is that the positive effect of the retention of information on integrative processes later in the speech stream is harder to observe than the negative effect of overwriting information on the accessibility of the prior information. However, it is also possible that generic backchannels act here as a particular type of predictive cue as well. Previously, generic backchannels such as *mhm* and *uh huh* have been analyzed as markers specialized for horizontal transitions within a joint activity (Bangerter & Clark, 2003). Under this analysis, generic backchannels act as signals of the addressee's disinclination to take over the conversational floor and commits to the continuation of the speaker's current action, which in the case of our stimuli would be the production of narrative. Thus, while we have used discourse continuations as the default next contribution, generic backchannels and pauses may differ in the degree to which they confirm the likelihood of this. As such, hearing a generic backchannel as compared to no response at all, a pause, may allow the overhearer to predict that the next speaker turn will be somehow relevant to the story, a continuation rather than say a switch to a new topic or the end of the interaction. This may then lead to some smaller degree of maintained activation of the prior turn.

### **General Discussion**

Previous researchers demonstrated that speakers take up addressee backchannels (Norrick 2010; 2012) and that different types of backchannels result in different patterns of subsequent narrative development in the following turn (Tolins & Fox Tree, 2014). In two experiments we show that addressee backchannels influence overhearers' on-line processing

of spontaneously produced dialogues, such that overhearers capitalize on the predictive relationship across conversational partners' contributions. By hypothesis, when overhearers hear a specific backchannel, they maintain activation of the prior turn's discourse content because of the expected response to the specific backchannel by the speaker. 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 and pauses suggest an upcoming continuation, the default next discourse event. After hearing generic backchannels or pauses, overhearers process the next turn without maintaining heightened access to previous talk. Specific backchannels, in contrast, lead overhearers to maintain strengthened access to the previous turn, slowing recognition of upcoming words.

Second, specific backchannels' activation of prior talk increases access to prior talk when subsequent talk is a continuation. As a context in which the typical pattern of collaborative production of discourse is not met, discourse-new continuations after specific backchannels provide insight into how overhearers make use of the predictive relation across interlocutors. The specific backchannels' activation of prior talk heightened access to the prior talk in comparison to pauses. Generic backchannels' role in highlighting prior talk in the context of continuations fell somewhere between specific backchannels' and pauses', and was not significantly different from either.

Third, specific backchannels' activation of prior talk decreases access to prior talk when subsequent talk is an elaboration. As a context in which the typical pattern of collaborative production of discourse is met, discourse elaborations after specific backchannels also provide insight into how overhearers make use of the predictive relation

across interlocutors. Activating prior talk in the context of an elaboration results in the prior talk being updated during the next turn, decreasing access to the pre-updated, original information.

Although overhearers do not actively participate in dialogues, they do make use of addressees' active participation when listening to conversations. Overhearers may miss out on key aspects of communication, such as the ability to request the information they need, which can leave them at a disadvantage in comparison to addressees (Schober & Clark, 1989). However, the relationship between active participation and overhearing is not necessarily dichotomous (Fox Tree & Clark, 2013; Wilkes-Gibbs & Clark, 1992). For example, one addressee might produce frequent comments on a speaker's talk but another might provide only a few backchannels; the backchannel-providing addressee may have more in common with an overhearer than the more active conversational participant. Even when minimally-interactive addressees and overhearers are primarily focused on a speaker's talk, however, they may still be gleaning something from addressees in addition to the speaker. Prior researchers showed that the feedback of any addressee helps the comprehension of all conversational participants (Fox Tree & Clark, 2013). We add to this knowledge that even feedback in the form of backchannels affects overhearers' comprehension.

Specifically, we provide evidence that overhearers make predictions about upcoming speech not only based on the primary speaker's discourse (Federmeier, 2007; Hagoort & van Berkum, 2007; Pickering & Garrod, 2013), but also based on how addressees respond to the primary speaker's talk. Responses from active listeners affect the relationship between the prior and subsequent discourse. People reading transcripts of others' talk use responses when

asked to predict what a speaker will say next (Tolins & Fox Tree, 2014). The current experiments show that overhearers use responses in on-line processing of dialogue as well, as cues to expected next discourse-level developments.

In order for overhearers to make predictions across interactants' contributions within a dialogue, the activity in which overhearers are engaged must be reconceptualized. Emphasizing the interactional basis of language use and comprehension, we demonstrate that even in strictly non-participatory roles within the participation frame of the interaction (Goffman, 1981; Goodwin, 2000), overhearer comprehension makes use of interactive processing strategies (De Jaegher & Di Paolo, 2007; Gallagher, 2001). Given that the predictive relation used by overhearers is across interlocutors, the current experiments demonstrate dialogue comprehension as involving the perception of the mutually constructed and maintained joint activity (Brennan et al., 2010; Shockley, Richardson, & Dale, 2009). This perspective opens a number of possible avenues for exploring third party sensitivity to synergistic social interaction.

The present experiments explored the influence of a particular subset of addressee responses on overhearer comprehension: generic and specific verbal responses. The current findings invite further exploration of the comprehension effects of different kinds of backchannels. One direction might be to explore the role of backchannels at a scope higher than single word monitoring or verification. For example, if specific backchannels assist overhearers in properly integrating speaker's talk into discourse events, their presence or absence may influence how overhearers remember or retell narratives as a whole. Another direction may be to explore nonverbal backchannels, which may or may not overlap with speaker's turns. Affective responses such as nodding or frowning influence the production of

narratives (Beukeboom, 2009), so they may also be used by observers of the interaction. A third direction might be to explore different interactive contexts. The focus of the current experiments was narrative production, which is common in the unstructured dialogue of chit-chat. Different backchannels are more prevalent in other interactions. For example, task-oriented dialogues have few specific backchannels, but many markers of transitions among projects and sub-projects (Bangerter & Clark, 2003). Hearing an *mhm* suggesting a horizontal transition between sub-projects instead of an *alright* suggesting a vertical transition between projects could cause overhearers to hold talk in memory differently, or focus on different information in upcoming talk, or join sections of talk before and after the backchannels differently based on these backchannels' role as signals of upcoming transitions.

Pauses are another topic of future exploration. In the current experiments, the effects of pauses were generally similar to those of generic backchannels, with both anticipating the default next contributions. However, pauses can do other things in conversation, such as indicating disagreement with assessments (Pomerantz, 1984), or an unwillingness to conform to a request (Davidson, 1984; Roberst, Margutti, & Takano, 2011), or that a speaker was planning a lie, was uncomfortable with the topic, or was having trouble speaking (Fox Tree, 2002). Thus, even the withholding of a response by pausing may be informative to overhearers of collaborative talk.

### **Conclusion**

Talk is produced collaboratively, with active and overt participation both from both speakers and listeners shaping the unfolding discourse (Clark & Krych, 2004; Clark & Schaefer, 1989; Goodwin, Goodwin, & Olsher, 2002). The relationship between listener responses and the production of talk such as narrative is predictive, with addressee

expressions steering speakers to develop their stories along particular trajectories (Bavelas, et al., 2000; Beukeboom, 2009; Tolins & Fox Tree, 2014). While the collaborative nature of language use in conversation has steadily gained traction within psycholinguistics, observational paradigms relying on overhearers has for the most part focused on the relation between current speakers and third party listeners. However, given the interactivity of conversation, we argue that dialogue presents a unique format of comprehension for both active participants and listeners with reduced participation status. In finding overhearer sensitivity to predictive patterns across conversational participants, we demonstrate that the collaborative nature in which dialogue is constructed influences non-participatory listening as well.

We present the first evidence that overhearers' on-line comprehension of spontaneously produced dialogue is modulated by the presence and type of backchannels the addressee provides the speaker. Generic and specific backchannels steer upcoming narrative in different ways, and language users are aware of these differences, making use of this predictive relation even in contexts where they are not active participants. The presence of backchannels in dialogue leads overhearers to make predictions about what discourse development is likely to happen next, depending on whether an addressee produces a generic or specific backchannel. A generic backchannel anticipates the default next turn, which is a continuation of the story. A specific backchannel anticipates an elaboration of the prior turn. This influences how the surrounding speaker talk is processed, causing overhearers to hold the prior turn in memory while listening to the subsequent turn, in order to better integrate the predicted elaborative information. The slower access to prior information demonstrated in Experiment 2 is a result of the successful, fast inclusion of the elaborative information in the

discourse model. More generally, these findings suggest that the comprehension of talk in dialogue is best conceptualized as comprehending a joint activity or synergistic whole (Dale, Fusaroli, Duran, & Richardson, 2014), in which information across the interactants is integrated and used to constrain predictions of likely next actions. The present findings suggest that third party comprehension, which has often focused on overhearers' comprehension of speakers' contributions, is better conceptualized as overhearers' comprehension of speaker-addressee collaborative wholes, where overhearers apply their previous experience engaging in interactions when listening to others talk.

### **Acknowledgements**

This research was supported by faculty research funds granted by the University of California Santa Cruz. Parts of this manuscript were presented at the 36<sup>th</sup> Annual Conference of the Cognitive Science Society and appeared in the proceedings. We thank our many research assistants who aided in data collection and coding, with a special thanks to Jasmine Norman, Gabriel Taylor, Akosua Busia, and Nicolas Russell.

### References

- Akhtar, N. (2005). The robustness of learning through overhearing. *Developmental Science*, 8, 199-209.
- Akhtar, N., Jipson, J., & Callanan, M. A. (2001). Learning words through overhearing. *Child Development*, 72, 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.
- Bavelas, J. B., & Gerwing, J. (2011). The listener as addressee in face-to-face dialogue. *International Journal of Listening*, 255, 178-198.
- Beukeboom, C. J. (2009). When words feel right: How affect expressions of listeners change a speaker's language use.
- Bly, B. (1993). *Uncooperative Language and the Negotiation of Meaning* (Unpublished doctoral dissertation). Stanford University, Stanford, CA.
- 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 Experimental Psychology*, 22, 1482-1493.
- Brennan, S. E., Galati, A., & Kuhlen, A. K. (2010). Two minds, one dialog: Coordinating speaking and understanding

- Brown-Schmidt, S. (2009). Partner-specific interpretation of maintained referential precedents during interactive dialog. *Journal of Memory and Language*, *61*, 171-190.
- Brunner, L. (1979). Smiles can be back channels. *Journal of Personality and Social Psychology*, *37*(5), 728-734.
- Clancy, P. M., Thompson, S. A., Suzuki, R., & Tao, H. (1996). The conversational use of reactive tokens in English, Japanese, and Mandarin. *Journal of Pragmatics*, *26*, 355-387.
- 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.
- Dale, R., Fusaroli, R., Duran, N. D., & Richardson, D. C. (2014). The self-organization of human interaction. *Psychology of Learning and Motivation*, *59*, 43-95.
- De Jaegher, H., & Di Paolo, E. (2007). Participatory sense-making: An enactive approach to social cognition. *Phenomenology and the Cognitive Sciences*, *6*, 485-507.
- Dynel, M. (2011). Revisiting Goffman's postulates on participant statuses in verbal interaction. *Language & Linguistics Compass*, *5*, 454-465.

- Emberson, L. L., Lupyan, G., Goldstein, M. H., & Spivey, M. J. (2010). Overheard cell-phone conversations: When less speech is more distracting. *Psychological Science*, *10*, 1383-1388.
- Federmeier, K. (2007). Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology*, *44*, 491-505.
- Fox Tree, J. E. (1995). The effects of false starts and repetitions on the processing of subsequent words in spontaneous speech. *Journal of Memory and Language*, *34*, 709-738.
- Fox Tree, J. E. (1999). Listening in on monologues and dialogues. *Discourse Processes*, *27*, 35-53.
- Fox Tree, J. E. (2001). Listeners' uses of *um* and *uh* in speech comprehension. *Memory & Cognition*, *29*(2), 320-326.
- Fox Tree, J. E. (2002). Interpreting pauses and *ums* at turn exchanges. *Discourse Processes*, *34*(1), 37-55.
- Fox Tree, J. E. & Clark, N. B. (2013). Communicative effectiveness of written versus spoken feedback. *Discourse Processes*, *50*, 339-359.
- Fox Tree, J. E. & Mayer, S. A. (2008). Overhearing single and multiple perspectives. *Discourse Processes*, *45*, 160-179.
- Fox Tree, J. E. & Schrock, J. C. (1999). Discourse markers in spontaneous speech: Oh what a difference an oh makes. *Journal of Memory and Language*, *40*, 280-295.
- Fusaroli, R., Raczaszek-Leonardi, J., & Tylén, K. (2014). Dialog as interpersonal synergy. *New Ideas in Psychology*, *32*, 147-157.

Gallagher, S. (2001). The practice of mind: Theory, simulation or primary interaction?

*Journal of Consciousness Studies*, 8, 83-108.

Gardner, R. (2001). *When Listeners Talk*. Amsterdam: Benjamins.

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. *Philosophical Transactions of the Royal Society B*, 362, 801-811.

Heritage, J. (1984). A change-of-state token and aspects of its sequential placement. In: J. M. Atkinson & J. Heritage (Eds.) *Structures of Social Action* (pp. 299-347). Cambridge: Cambridge University Press.

Kendon, A. (1967). Some functions of gaze direction in social interaction. *Acta Psychologica*, 26, 22-63.

Kraut, R., Lewis, S., & Swezey, L. (1982). Listener responsiveness and the coordination of conversation. *Journal of Personality and Social Psychology*, 43(4), 718-731.

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

Martin, A., & Vouloumanos, A. (2013). *Sixteen-month-olds recognize the roles of common ground and intentions in third party communication*. Paper presented at the biennial meeting of the Society for Research in Child Development, Seattle, WA.

- Norrick, N. (2010). Incorporating listener evaluation into stories. *Narrative Inquiry*, 20, 183-204.
- Norrick, N. (2012). Listening practices in English conversation: The responses responses elicit. *Journal of Pragmatics*, 44, 56-576.
- Oreström, B. (1983). *Turn-taking in English conversation*. Lund, Sweden: Gleerup.
- Pickering, M. J. & Garrod, S. (2013) Forward models and their implications for production, comprehension, and dialogue. *Behavioral and Brain Sciences*, 36, 377-392.
- 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* (pp. 71-93). Washington, D. C.: Georgetown University Press.
- 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). Mahwah, NJ: Lawrence Erlbaum.
- Schober, M. & Clark, H. H. (1989). Understanding by addressees and overhearers. *Cognitive Psychology*, 21, 211-232.
- Schober, M. & Brennan, S. E. (2003). Processes of interactive spoken discourse: The role of the partner. In A. C. Graesser, M. A. Gernsbacher, & S. R. Goldman (Eds.), *Handbook of discourse processes* (pp. 123-164). Mahwah, NJ: Lawrence Erlbaum.
- Schockley, K., Richardson, D. C., & Dale, R. (2009). Conversation and Coordinative Structures. *Topics in Cognitive Science*, 1, 305-319.
- Stivers, T. (2008). Stance, alignment and affiliation during storytelling: When nodding is a token of affiliation. *Research on Language and Social Interaction*, 41, 31-57.

Tolins, J. & Fox Tree, J. E. (2014). Addressee Backchannels Steer Narrative Development.

*Journal of Pragmatics*, 70, 152-164.

Tolins, J., Zeamer, C. A., and Fox Tree, J. E. (under review). Community Commonality in

Referring Expressions across Monologues and Dialogue.

Tomlinson, J. M., Jr., & Fox Tree, J. E. (2011). Listeners' comprehension of uptalk in

spontaneous speech, *Cognition* 119, 58–69.

Wilkes-Gibbs, D., & Clark, H. H. (1992). Coordinating beliefs in conversation. *Journal of*

*Memory and Language*, 31, 183-194.

Wilkinson, S., Kitzinger, C. (2006). Surprise as an interactional achievement: Reaction tokens

in conversation. *Social Psychology Quarterly*, 69, 150-182.

Yngve, V. H. (1970). On getting a word in edgewise. In: M. A. Campbell (Ed.), *Papers from*

*the Sixth Regional Meeting, Chicago Linguistics Society* (pp. 567-578). Chicago:

Department of Linguistics, University of Chicago.