

To cite, please refer to: Fox Tree, J. E. (1999). Listening in on monologues and dialogues. *Discourse Processes*, 27, 35-53.

Listening in on Monologues and Dialogues

Jean E. Fox Tree

University of California, Santa Cruz

Running Head: LISTENING IN ON MONOLOGUES AND DIALOGUES

mailing address:

Jean E. Fox Tree

Psychology Department

Social Sciences II, room 277

University of California, Santa Cruz

Santa Cruz, CA 95064

Abstract

A study compared the communicative effectiveness of spontaneous monologues and dialogues on nonparticipating addressees overhearing talk. Overhearers were more accurate at following instructions in a referential communication task when listening in on dialogues than when listening in on monologues. Several extraneous variables could not account for the result. Two explanations for overhearers' better performances on the dialogues are that the greater number of discourse markers helps overhearers follow the speech and that the perspectives of multiple interlocutors are more informative than one perspective. Extensions of the collaborative model of language use are discussed.

Listening in on Monologues and Dialogues

Although face-to-face conversation is our most natural form of communication, we often find ourselves understanding talk in settings where we do not participate actively. There are at least two sorts of nonparticipation. One is when nonparticipants are the intended addressees, such as members of an audience attending a lecture. Another is when nonparticipants are not the intended addressees, such as people listening to a recording of a lecture that took place several years earlier. In this case, the nonparticipants are overhearers. The two groups are not always distinct; if a professor gives a lecture to a group of students registered for a course, but some students' parents attend it, those parents are in some sense addressees because visual feedback of their presence may affect the professor's lecture. But in another sense they are overhearers, because the professor has not designed the lecture for them.

Information can be communicated in a variety of formats. Two of these formats are monologues and dialogues. Each of these formats can in turn be presented with different levels of spontaneity. On the one hand, the speech can be prepared in advance and carefully delivered, such as with an inaugural address (a monologue) or a discussion between two soap opera characters (a dialogue). On the other hand, the speech can be produced on the fly without any advance planning, such as with an answering machine message (a monologue) or an informal interview (a dialogue).

The research presented here focusses on the ability of overhearers to extract information from spontaneously produced monologues and dialogues. There are at least two predictions that can be made about which presentation format would be easier for an overhearer to understand. Both predictions are compatible with the collaborative theory of language use, which is taken as a starting point. A referential communication task is used to test the communicative effectiveness of spontaneous monologues and dialogues, and extensions to the collaborative theory of language use are discussed.

A Collaborative Theory of Communication

The collaborative theory holds that conversational participants collaborate on what the meaning of an utterance will be taken to be, working together to achieve mutual understanding (Clark & Brennan, 1991; Clark & Wilkes-Gibbs, 1986; Isaacs & Clark, 1987; Krauss & Weinheimer, 1966; Schober, 1993; Schober & Clark, 1989; Wilkes-Gibbs & Clark, 1992).

A conversation does not just involve participants expressing their thoughts out loud to one another in an orderly fashion. Instead, speakers say things they believe their participants will understand, tailoring their utterances to their addressees, and monitoring their addressees' comprehension. Addressees, in turn, acknowledge the speakers' contributions and provide evidence of their understanding, either explicitly or by providing replies that demonstrate understanding to the speakers (Clark & Schaefer, 1987, 1989).

In the collaborative theory, conversational participants keep working to ensure understanding until they have reached their grounding criteria, or until the participants "mutually believe that the addressees have understood what the speaker meant to a criterion sufficient for current purposes" (Clark & Schaefer, 1989, p. 20). Participants make more or less effort to ensure comprehension depending on what they need to accomplish. When pointing out a passerby to discuss something about that person, participants will talk until satisfied that each knows who is indicated. But when pointing out a passerby as a preamble to a story concerning someone the passerby reminds them of, participants may choose not to go to the effort of exact identification.

The collaborative theory predicts that participating addressees have an advantage over overhearers in understanding because the talk is designed for the addressees and because addressees can work together with speakers until understanding is reached. In

contrast, overhearers have to take what they can get out a conversation without the opportunity to ask for clarification or display their current understanding. Schober and Clark (1989) confirmed this prediction by means of a careful analysis of addressees' and overhearers' performances on a referential communication task. In their task, matchers try to put a randomly arranged set of abstract shapes into the order described by a director. The more shapes they placed correctly, the more successful the communication. Matchers either conversed with directors or listened in on conversations between directors and different matchers. According to the collaborative model, addressees' privileged abilities to participate in the grounding process should lead them to outperform overhearers, and this is just what Schober and Clark found.

More precisely, when participating addressees reach their grounding criteria with the speakers after overhearers have been able to identify the shape a speaker is describing, then overhearers should perform as well as addressees. But when participating addressees reach their grounding criteria before overhearers identify the shape, overhearers will never be able to make up for their lack of understanding. In reaching a grounding criterion, speakers and participating addressees can also exploit any common perceptions they happen to share about the abstract shapes, such as that a particular figure looks like "the knight on a chessboard," as two participants in the current experiment called one figure. Overhearers who do not share this perspective of the shape will be confused. Because they are monitoring only participating addressees' understanding, speakers will move to the next shape before overhearers know which shape was placed, leaving overhearers at a disadvantage.

The collaborative model predicts that conversational participants will understand talk better than nonparticipants, and the evidence supports this prediction. But what about nonparticipants listening in on different kinds of talk, such as talk with differing amounts of collaboration and feedback? Although collaboration puts a nonparticipant at a disadvantage compared to a participant, does overhearing the collaboration of others

provide any advantage or disadvantage compared to overhearing monologues?

Kraut, Lewis, and Swezey (1982) made headway at an answer to this question in an experiment where speakers summarized movies to listeners with either full feedback between the speakers and listeners, partial feedback (backchannels and one-word questions only), or no feedback. In the full feedback and partial feedback conditions, there were always two listeners, and they were of two types: participating addressees and nonparticipating addressees. In the no feedback condition, there were only nonparticipating addressees. Like Schober and Clark (1989), Kraut, Lewis, and Swezey (1982) found that participating addressees followed the summaries better, giving higher quality and more complete retellings of the summaries than the nonparticipating addressees. But they also found that both participating and nonparticipating addressees gleaned more from the descriptions the more feedback the speaker had. They argued that speakers can better design their speech for all addressees when they have information about any addressee's comprehension.

There is reason to believe, however, that the speakers' talk in Kraut, Lewis, and Swezey's partial- and no-feedback conditions is different from naturally occurring partial- and no-feedback situations. In the movie experiment, speakers believed that they were getting full feedback at all times. This was accomplished by disclosing the partial feedback instructions only to the addressees (not to the speakers), and by secretly disconnecting the nonparticipating addressees' microphones after 30 seconds in the partial- and no-feedback conditions. When feedback is believed possible speakers might interpret its absence differently from when feedback is not believed possible. When feedback is believed possible, speakers might choose to interpret the lack of addressee response as implying lack of understanding. But they might also choose to interpret the lack of addressee response as implying presence of understanding. After all, if addressees were not following, they could say something. In contrast, when feedback is not believed possible, speakers know that the lack of addressee response is unrelated to addressee

understanding.

Kraut, Lewis, and Swezey's (1982) results might be an artifact of their design. It might not be that speakers could not summarize the movies well without feedback, but that they stopped short of elaborating because they did not have any objection from addressees they believed to be fully participating. Perhaps with a true monologue, where speakers know they would not receive feedback, speakers would make a greater effort to communicate their messages. The current study is designed to test the difference between two non-deceptive versions of the full- and no-feedback conditions.

Monologues versus Dialogues

At first blush, the talk of a single person seems worlds apart from a conversation. But monologues and dialogues actually have a lot in common. Although monologue speakers do not negotiate what they say with individual members of an audience, they do take audiences into account. Orators use feedback in the form of clapping, cheering, or booing to carefully time what is said and how speech is delivered (Atkinson, 1984). For example, some politicians continue to speak after people have started to clap to convey a sense of modesty and sincerity, as if praise were secondary to conveying the weighty content of their speech, even though in reality the clapped-over speech is usually of little importance (Atkinson, 1984).

Even in situations without audience feedback, speakers keep a particular audience in mind. Radio announcers often interject their own comments for the benefit of an imagined audience, even when they are reading aloud pre-written text. Sometimes announcers cannot help putting themselves into their readings (adapted from Goffman, 1981, p. 302):

(1) Try this wonderful new bra... It's lightly padded and I'm sure you'll love it. I do!... I mean I like the looks of it... Well...what I am trying to say is that I don't need one myself naturally, as a man... but if you do, I recommend it... How do I

know? I really don't... I'm just reading the commercial for Mary Patterson who is

ill at home with a cold!

Even when they do not have a live audience in front of them, speakers keep the audience in mind as they talk.

Dialogues differ from monologues in that speakers can use specific verbal feedback from their addressees in order to design their utterances. In the following example, a director and matcher check each other's understanding continually in order to coordinate an appropriate referent

for an abstract shape (the asterix indicates overlapping speech; examples from here on are from the speech used in the current experiment):

(2) director: ok the next one is like a very tall guy, kay, he's very tall,

he has he looks like he has way broad shoulders

matcher: is he kicking one leg back?

director: yes uh huh *cool*

matcher: *now wait* now wait this isn't a christmas tree guy is it?

director: a christmas tree guy, no

matcher: he's got he looks like a number seven body?

director: exactly

matcher: ok num*ber* is that number six?

director: *exactly* yes that was number six

Like in monologues, the dialogue director adjusts what he says for his addressee. But unlike in monologues, what the dialogue director says more closely matches what the addressee needs or wants to know.

Extensions of the Collaborative Theory

One view that is consistent with the collaborative theory predicts that overhearers would understand more when listening in on monologues because monologues are designed for nonparticipating addressees and therefore would not be as tailored to particular people as are dialogues. Instead of using participating addressee's feedback to limit the information conveyed

to what that addressee needs, monologue speakers might provide more detailed and thorough information to cover all potential problems they imagine a nonparticipating addressee might have. But even if a dialogue were to contain the same level of detail and thoroughness as a monologue, the dialogue might still be harder for an overhearer to follow because the information provided is tailored to a particular addressee. If the overhearer's needs do not match those of the participating addressee's at a particular moment, then the overhearer will be at a disadvantage.

The lack of participating addressees has other potential advantages in addition to the avoidance of tailored messages. Monologues do not contain overlapped speech or interruptions from participating addressees that could make speakers' instructions hard to hear or follow. Monologues also allow speakers to develop and organize information in the absence of participating addressees' contributions. This might lead to more orderly directions than those produced in dialogues. Finally, monologues might avoid some potentially distracting elements of dyadic talk, such as expressions of friendship or politeness.

An alternative view that is also consistent with the collaborative theory predicts that overhearers would understand more when listening in on dialogues because dialogues contain potentially useful information that monologues lack. One type of information dialogues contain is participating addressees' perspectives. When overhearers listen to dialogues they have information from two sources, not just one. In any overheard talk, there is a chance that overhearers will not share the same perspectives as speakers. A lack of a shared perspectives can impede overhearers' understanding. If the disparate perspectives arise in monologues, overhearers will not be able to recoup their losses. But if they arise in dialogues, overhearers have another opportunity, in the form of participating addressees' contributions to the discourse, to catch the drift of the conversation.

A second type of information dialogues contain is participating addressees' feedback about their understanding. Participating addressee feedback might prompt speakers to elaborate on vague points, clarify confusing thoughts, or streamline wordy explanations. Although

monologue speakers might anticipate some problems by imagining how nonparticipating addressees would interpret their talk and adjusting what they say accordingly, they would still be likely to miss other problems. There might be no substitute for participating addressees.

Both the monologue-superiority and dialogue-superiority views are compatible with the collaborative theory. In one view, collaboration between two or more people is always to the detriment of an overhearer, who would be better served by listening in to a monologue addressed to a nonparticipant. In another view, collaboration increases the amount of useful information overhearers have to interpret a discourse, and will always be an advantage in comparison to collaboration-free monologues.

So, what is better for overhearers: listening in on a monologue with no feedback, or listening in on a dialogue with feedback, even if that feedback was not intended for them? Schober and Clark (1989) compared original addressees and the feedback they supplied with overhearers who heard that feedback but gave none of their own. The current study extends that comparison to overhearers listening in on a description containing other people's feedback to speakers (Schober and Clark's condition) and overhearers listening in on a description containing no feedback. Each of these circumstances of language use has real world analogues. For example, there is a difference between (a) a university lecture where a particular student does not participate but overhears other students engaging the professor in discussion and (b) a university lecture where all students listen to the professor's monologue with no audience feedback.

To investigate how listening in on talk with varying amounts of feedback from addressees affects comprehension, I compared overhearers' abilities to do referential communication tasks when listening in on monologues versus listening in on dialogues.

Method

Overview

To create the stimuli, pairs of people were divided into directors and matchers. The director's goal was to describe a set of ordered abstract shapes to the matcher so that the matcher could put an equivalent set of randomly scattered shapes into the same order. Directors either

conversed freely with matchers (dialogue condition) or gave instructions without matcher feedback (monologue condition). Each session was taped. In the main experiment, eight dialogues and eight monologues on which the matchers ordered all shapes correctly were played to new matchers, and these overhearers' abilities to correctly place the figures were compared across conditions.

Subjects

Subjects were 167 University of California students who participated to fulfill a course requirement. All subjects were native speakers of English.

Materials

Thirty-two abstract shapes were selected from a collection of tangram figures and divided into two sets. One set could be broadly characterized as human forms, the people tangrams, and the other as animal forms, the animal tangrams. The tangrams are shown in Figure 1.

Insert Figure 1 about here.

Of the 16 tangrams per set, 12 were part of the correct order, and 4 were extra. Each figure was affixed to a 5" by 6" piece of cardboard. A response board was made by dividing a 22" by 28" foam board into 12 sections, three rows of four.

Spoken monologues and dialogues were created by 16 pairs of Stanford University, San Jose State University, and University of California undergraduates who participated either to fulfill a course requirement or for \$5 pay. All pairs were native speakers of English. Thirteen of the 16 pairs did not know each other (rating of 1 on a 1 to 7 scale of how well they knew each other with 1 being not at all), 2 pairs knew each other somewhat (ratings from 2 to 4), and 1 pair knew each other very well (rating of 7). Three judges who listened to all 16 monologues and dialogues were unable to identify the friends; their averaged rating of 2.33 for the critical pair (using the same 1 to 7 scale) was no different from their averaged rating overall, 2.52 ($z = .15$).

One member of each pair was randomly designated the director, and the other the matcher. Eight directors described tangram figures to the matcher with matcher's feedback (dialogue condition), and eight described the figures without matcher's feedback (monologue condition).

To produce the materials, an extra response board and set of tangram cards were created. Directors' tangrams were arranged in a predesigned order before them on their response board, with the four extra cards visible. Matchers' identical 16 tangrams were randomly laid out on a desk in an adjoining room, in the same orientation as the directors'. Matchers were seated before a blank response board. The task was for the directors to describe to the matchers the order of their tangrams such that the matchers could place their tangrams in the same order. Headphones and microphones linked the two participants to each other and to the recording equipment, with the connecting wires going underneath the closed doors of the two rooms. In this way, only verbal information could be transmitted; neither gestures, eye contact, nor the rustling of cards could provide information regarding the success or failure of a placement instruction. In the monologue condition, the director described the tangrams to the matcher without any matcher feedback; in the dialogue condition the director and matcher conversed freely to complete the task. Participants were told that the experiment involved no deception, that directors would get no feedback from matchers (if the two were in the monologue condition) or that they could converse freely (if the two were in the dialogue condition), and that both directors' and matchers' goals were for matchers to lay their cards out in the same order as the director's cards were laid out.

Directors and matchers were introduced to each other before engaging in the task. The presence of real matchers emphasized the directors' need to effectively communicate the order of the tangram figures. Directors were not describing tangrams for themselves, for the experimenter, or for some unknown future matcher or overhearer. Fellow subjects, whom directors might recognize from their subject pool classes, sat in the next room relying on the directors' descriptions.

If the matcher ordered all the tangrams correctly, the recording of the session was

retained for the experiment. This way, any errors an overhearer makes listening to the materials is not likely to be due to something in the director's talk that made correct ordering impossible, such as a missing or inaccurate description. At least one person, the original matcher, heard the director's descriptions and got all the tangrams in the correct order. Director/matcher pairs were run until the monologues and dialogues retained consisted of four monologues on the animal tangrams, four monologues on the people tangrams, four dialogues on the animal tangrams, and four dialogues on the people tangrams.

As can be expected from a speech production situation where people can say whatever they have to to get the job done, speakers' productions differed greatly from one another. Though directors generally described the tangrams one by one in order, they sometimes skipped forwards or backwards to similar-looking tangrams, or returned to shapes they had already covered for added clarification. Some chose to describe the extra four shapes, others did not. There was also a great difference in the style of description, with some people preferring geometrical breakdowns of the components of the tangram and others preferring figurative labels for what the tangram looked like. Most people used both these styles. The length of the sessions also varied greatly. The shortest was 5 minutes, 25 seconds; the longest 19 minutes, 32 seconds. This difference was not systematic across conditions, however. The average monologue length was 11 minutes, 38 seconds and the average dialogue length was 10 minutes, 52 seconds, $t(14) = .35$.

Another source of variance was the tangram set used. It was much more difficult to get a flawless performance on the people tangrams than on the animal tangrams. One reason for this difference is that the people tangrams were more similar to one another. Another possible reason is that the people tangrams were less amenable to general labels for the whole figure.

To control for some of this variance, description sessions were matched on length and tangram set. A 5 min 25 sec animal monologue, for example, was matched with a 5 min 39 sec animal dialogue. A 9 min 5 sec people monologue was matched with a 11 min 30 sec people dialogue. Time alone could be a factor in overhearer's performance regardless of whether they

listened to monologues or dialogues. Having more time would allow overhearers to evaluate the different shapes independently, perhaps allowing them to identify a priori shapes that were similar to each other and easily confusable, or allowing them time to imagine the possible perceptions of different tangrams before the tangrams were even described on the tape. Other potentially systematic sources of variance are discussed in the next section.

From these eight monologues and eight dialogues, 16 tapes were made, each containing one monologue and one dialogue, one describing people tangrams, and the other animal tangrams. The 16 tapes consisted of eight pairs. Each pair of tapes had the same monologues and dialogues, but in different orders: one tape from a pair had the monologue first and the other had the dialogue first. In this way each overhearer, listening to a single tape, would hear one tangram set described with two voices (the dialogue) and the other tangram set described with one voice (the monologue).

Counterbalanced for order, tangram type, and feedback condition, the 16 tapes can be described as four sets with the following pattern: tape 1 contains item 1 (monologue, animals) followed by item 2 (dialogue, people); tape 2 contains the reverse order, item 2 (dialogue, people) followed by item 1 (monologue, animals); tape 3 contains monologues and dialogues on alternate tangram sets from tapes 1 and 2, such as item 3 (monologue, people) followed by item 4 (dialogue, animals); and tape 4 contains the reverse order of tape 3, item 4 (dialogue, animals) followed by item 3 (monologue, people). That is, this pattern describes the four tapes made from items 1 - 4. Similarly, four tapes were made for items 5 - 6, four tapes for items 7 - 10, and four tapes for items 11 - 16. A ten second silence separated the two items on each tape. Subjects were instructed to pause the tape when they heard this silence.

Procedure

Subjects were tested individually. They were seated at a desk with a tape player, headphones, an answer board, and 16 randomly arranged tangrams in front of them. They were instructed to listen to the tape and to do their best to put the tangrams into the order described on the tape. After they had listened to one description session, subjects paused the tape and left the

experiment room. The experimenter noted which tangrams were placed correctly, but did not report the results to the subjects. After recording the data, the experimenter cleared away the cards and laid out the new set of cards in preparation for the second description. Subjects then returned to the experiment room and listened to the second description session. Each subject heard one tape (one monologue and one dialogue), using the animals tangrams during one session and the people tangrams during the other. This way, subjects' natural abilities or lack of abilities at making sense of abstract shapes and following instructions is controlled across conditions. The experiment lasted between 25 and 45 minutes, depending on the tape the subject heard.

Results

Seven subjects' data were removed from the analyses for failure to follow instructions, leaving a planned design of 160 subjects. These useable subjects were evenly divided across the 16 tapes, 10 people per tape. Therefore, each item was heard by 20 subjects, 10 times as the first tangram ordering a subject performed, and 10 times as the second tangram ordering.

Overhearers were more successful at placing tangrams in the correct position when they listened to a dialogue than a monologue. When they listened to a dialogue, they made on average 1.57 errors, but when they listened to a monologue, they made on average 2.27 errors ($t_1(159) = 2.82$, $p < .01$ in the subject analysis; $t_2(7) = 2.36$, $p = .05$ in the item analysis). The performance of overhearers on the tangrams description produced by the pair of friends did not differ from the mean performance in that condition, the monologue condition (average of 2.20 wrong on that item compared to an average of 2.27 wrong over all; $z = .14$).

In debriefing, participants were asked whether listening to monologues or dialogues made a difference in their abilities to order the tangrams. Many found the dialogues more difficult, saying that they were desultory, overly repetitive, and had too many interruptions and distracting chatter. Nonetheless, almost 70% of the people who said they found listening in on dialogues harder actually performed equally well on the monologues and dialogues or, in fact, better on the dialogues. People might expect to perform better listening in on monologues because

monologues might be a more familiar format for receiving instructions. When getting roadway directions, for example, people might be used to hearing a more-or-less monologic stream of information; it is less likely that they would be overhearers to a dialogue between the directions-giver and someone else. Also, in an academic environment, people might be used to hearing information provided in the form of a monologue rather than a dialogue. The experimental situation might bring to mind the academic environment and lead people to expect monologue instructions, making them preferable.

Although the manipulated variable in this experiment, the presence or absence of feedback, was related to overhearers' performance, it is possible that one of the many other uncontrolled variables in the materials was systematically distributed across conditions and contributing to the effect. A quantitative comparison was made of several potentially important differences across conditions to evaluate the contribution of extraneous variables.

Analyses of Experimental Materials

Although there are an infinite number of possible discrepancies between monologues and dialogues, a few stand out as potentially confounding variables. To check whether these variables did in fact vary systematically across conditions, a quantitative comparison was made.

One possible confounding variable that may have influenced the results is the number of words spoken. Although the length of time people spoke did not vary across conditions, as reported above, the number of words might have. If people spoke more in dialogues, then overhearers may have had more information to use. Although prior research has not shown a consistent relationship between the amount of feedback and the amount of information speakers provide (compare Chantraine & Hupet, 1994; Krauss & Weinheimer, 1966; Kraut, Lewis, & Swezey, 1982), it is still possible that this factor might play a role in the current results. But it turns out that the mean number of words used in the monologues, 1355, did not differ from the mean number of words used in the dialogues, 1627, $t(14) = -.82$, so this factor could not have been driving the effect. Overhearers did not have more information, as measured by number of words, in the dialogues than in the monologues.

Interestingly, if matchers' talk is excluded from the dialogues, the number of words spoken by the directors alone in the monologues, 1355, still does not differ from the mean number of words used in the dialogues, 1062, $t(14) = 1.02$. So in both conditions, overhearers heard the same number of words from the person with the main job of providing it, the director. This lack of difference for number of words across conditions is perhaps unsurprising given the wide variation in the number of words necessary for effective tangram descriptions. In one stimulus, 2304 words were used to describe the tangram ordering, but in another stimulus only 478 words were used.

Another potentially confounding variable is rate of speech. If speakers spoke more quickly in the monologues, then listeners may have had more difficulty understanding monologue talk than dialogue talk. Faster speech may lead to more coarticulation and less clear enunciation, which might slow comprehension. Although the length of talk and number of words did not vary across the conditions, the rate of speech did, but not in the direction that the overhearers' performances would predict. People spoke more quickly in the dialogue than the monologue, 115 versus 153 words per minute, $t(14) = -2.81$, $p < .05$. So a faster rate of speech in the monologues could not have been causing overhearer's poorer performance in that condition.

A third potentially confounding variable is the amount of repetitions of concepts. In the following example taken from a monologue, the italicized concept looking upward occurs three times:

(3) it's a guy praying and his head is a square and he's looking up this time and when I mean well, is he look? yeah he's definitely looking up, like if you could look at the squares, the one the square for the head is tilted back more then it kind of appears like he's looking up.

Concepts are also repeated in dialogues, as in the following example where the same concept also occurs three times:

(4) director: nine looks kinda like number one, but the guy is looking up a little bit towards the right?

matcher: ok.

director: kinda like he he he's this time he's like praying up to the sky

matcher: right his head's tilted a little more upward maybe?

director: uh huh.

matcher: ok ok I got him.

The amount of conceptual repetition in each monologue and dialogue was evaluated by underlining each concept that was repeated, and then counting the underlines in two ways. The instances of conceptual repetition were evaluated by counting the number of concepts which were repeated at least once in each stimulus. So in (5) above, the three repetitions of looking upward would count once towards the total instances of conceptual repetition for that stimulus. The total number of conceptual repetitions was evaluated by counting the number of times a concept was repeated. So in (5) above, the two repetitions of looking upward would count twice towards the total number of conceptual repetitions for that stimulus.

If dialogues had more conceptual repetitions, either in the number of different concepts repeated or the number of iterations of any particular concept, then this could explain why overhearers performed better listening in to these items. If they missed an important piece of information, either because they did not hear it correctly or because their attention was focussed elsewhere, they would have a second chance to catch it later in the description. The number of iterations could also provide information about the prominence of a particular feature. If the feature was mentioned several times, this could cause overhearers to assign more weight to it in searching their tangrams for that feature, and also to de-emphasize other elements of the description. But it turned out that there was no difference in the amount of conceptual repetitions across conditions. There were on average 25 instances of conceptual repetitions in both monologues and dialogues. There were on average 35 iterations of a concept in the monologues and 42 iterations in the dialogues, $t(14) = -.75$. Repetition of concepts cannot be causing the dialogue advantage.

A fourth potentially confounding variable is the amount of restarted ideas. People

sometimes abandon a line of thought and start over, as in “it’s got- it’s all one big connected shape.” These restarts can slow listeners’ comprehension (Fox Tree, 1995). If there are more instances of restarts in the monologues, then this may contribute to overhearers’ poorer performance in that condition. But the number of restarts is the same across groups, with an average of 19 restarts per monologue and 20 restarts per dialogue, $t(14) = -.06$. Restarting ideas cannot be causing the monologue disadvantage.

The last potentially confounding variable considered was the number of discourse markers in a stimulus. Discourse markers are words like well, oh, I mean, and you know that occur frequently in spontaneous talk. They are sometimes thought to be disruptive or unnecessary, so it is possible that having many of them in a stimulus may have disrupted overhearers’ abilities to follow the talk. On the other hand, discourse analysts have suggested that these words might serve a beneficial function by forewarning a suspension in speech (James, 1973), forewarning upcoming repairs (Erman, 1987; Levelt, 1989; Schiffrin, 1987), indicating what kind of a repair to expect (Erman, 1987; Levelt, 1989; Schiffrin, 1987), indicating how far back an error occurred (Levelt, 1989), providing cues to discourse structure (Erman, 1987; Flowerdew & Tauroza, 1995; Redeker, 1991, Schourup, 1985; Schiffrin, 1987), providing information about what’s going on in the speaker’s mind (Schourup, 1985), or creating a congenial atmosphere (Stubbe & Holmes, 1995). So if they occur more frequently in monologues or more frequently in dialogues, they may be linked to overhearers’ performances, either for the better or the worse.

Although discourse markers themselves have not frequently been studied systematically across different speaking conditions, a similar speech element, ums and uhs, have. These speech fillers are sometimes considered speech errors or epiphenomena of speech production trouble (Levelt, 1989; Reynolds & Paivio, 1968), but they are also sometimes thought to be intentionally used words (Clark & Fox Tree, in preparation). If they serve a pragmatic function that is useful to overhearers, a greater amount in dialogues may have contributed to improved performances in this condition. But there was no difference in filler rate across conditions. There were on average

39 fillers in the monologues and 46 in the dialogues, $t(14) = -.41$. There were 34 monologue ums versus 36 dialogue ums, $t(14) = -.11$, and 5 monologue uhs versus 10 dialogue uhs, $t(14) = -1.20$.

Unlike fillers, the number of standard discourse markers used did vary across conditions. Discourse markers come in a wide range of forms, including so anyway and or whatever, and connectives and conjunctions like then, and, and but. To get an estimate of the disparity in discourse markers across conditions, we counted the occurrences of five common markers that are relatively easy to tease apart from their nonpragmatic uses: like, oh, well, I mean, and you know. Nonpragmatic uses were defined as those that could carry normal nominal, verbal, or adverbial functions, as in do you know which one I mean? and it looks like a seal. Pragmatic uses did not fit these criteria, as in the other foot is supporting him and like he's leaning against the wall and it's just you know pushed towards the left. There were on average 6 expressions per monologue, but 23 per dialogue, $t(14) = -4.58$, $p = .001$. Most of these showed the same disparity across conditions when considered individually, although only oh remained significant after a Bonferroni correction of the p value to .01 (like: monologue = 4, dialogue = 11, $t(14) = -2.28$, $p < .05$; oh: monologue = 1, dialogue = 9, $t(14) = -4.58$, $p < .01$; well: monologue = 1, dialogue = 3, $t(14) = -2.13$, $p = .05$; I mean: monologue = .1, dialogue = .3, $t(14) = -.45$; you know: monologue = 0, dialogue = .5, $t(14) = -2.65$, $p < .05$).

There are more discourse markers in dialogues, and overhearers perform better in dialogues. This could be evidence for the beneficial impact of discourse markers, but it could also be a red herring. There may be more discourse markers in dialogues simply because they help speakers and addressees manage a conversation. To test this, overhearers' differences in performance across the monologue and dialogue pairs was correlated with the differences in number of discourse markers in the pair. That is, the average number wrong for the 20 people who heard the monologue of the pair minus the average number wrong for the 20 different people who heard the dialogue of the pair (yielding eight data points for the eight pairs) was correlated with the number of discourse markers in the dialogue minus number of discourse markers in the monologue of the pair (likewise, eight data points for the eight pairs). There was a

strong correlation ($r = .73$, $p < .05$). When there were about the same number of discourse markers in the paired items, the differences between conditions was small. As the disparity grew, with more discourse markers in the dialogue than the monologue, so did the difference in conditions, with more wrong orderings in the monologues compared to the matched dialogues. Discourse markers are not only more prevalent in dialogues, they are correlated with overhearers' improved performances.

Discussion

Dialogue is the most fundamental form of human communication. It is the means by which we learn language and the process by which we communicate to each other every day. Monologues, on the other hand, are not particularly natural. Few people find themselves in the position to speak without feedback on a daily basis. Some people who do are radio announcers, news broadcasters, public speakers, and professors. But often the speech produced is not a feedback-free monologue; instead, what is said is modulated by audience reaction. Students start to doze and professors pick up tempo and volume. Nonetheless, monologues have more potential to be odd than dialogues. One director-matcher pair who were not included in the current study provides a good example of this oddness: when the matcher had selected a tangram, she knocked on the wall separating herself and the director to let him know he could move on to the next description. It was her only way of supplying feedback, and for her, the task could not be completed without it.

Although an interactive conversation is more natural than a monologue from the addressee's perspective, as demonstrated by the wall-knocking matcher, it is not clear whether the same is true for overhearers. If a person has no ability to supply feedback, such as when watching a videotaped lecture, does the person learn more when the lecturer delivers a straight monologue or when the lecturer engages in discussions with students? Of course, there are many variables involved, such as the relevance of a student's question, the ability of the lecturer to respond on the fly, how distracted from the study plan a lecturer becomes, and whether the lecturer was reading notes aloud or delivering a spontaneous monologue. But above and beyond

these differences, is

there any way in which one form of communication might be more informative to overhearers than the other?

The collaborative theory of language use could accommodate either potential finding, that monologues are more effective than dialogues or that dialogues are more effective than monologues. Monologues might be more effective than dialogues because they lack a participating addressee, thereby avoiding the problems for overhearers that Schober and Clark (1989) identified. That is, there would be no addressee to find a shape before an overhearer, thereby cutting short (from the overhearer's perspective) the description the speaker provided. There would also be no addressee to share the same happenstance perspective of an object as the speaker, thereby allowing an ambiguous description (from the overhearer's perspective) to stand as sufficient. Otherwise said, monologues might contain more thorough and unambiguous descriptions than dialogues.

Monologues might also be preferable to dialogues for overhearers for reasons not related to participating addressee's abilities to place their cards faster. Monologues do not contain overlapped speech that could render talk uninterpretable to an overhearer, either because the overhearer cannot make out the words or because the overlapping talk is confusing. Likewise, they do not contain addressee interruptions or remarks that might either confuse an overhearer or throw off a directors' orderly descriptions. And they do not contain potentially distracting expressions of friendship or politeness that often arise in conversational talk.

Despite the plausibility of overhearers' performing better when listening in on monologues rather than dialogues, in fact overhearers performed better when listening in on dialogues. The monologues and dialogues studied represented a wide range of personal styles, manners of description, lexical choice, facility with abstract shapes, and amount of detail in descriptions, from a few brief words to lengthy listings of distinguishing features. Despite these variations, there was still a clear advantage to overhearers in listening in on dialogues over monologues.

With the overlapping speech, interruptions, and potentially distracting addressee comments working against them, why are dialogues so much more effective at conveying information to overhearers than monologues? Analyses of the monologues and dialogues showed that there were a number of variables that could not be driving overhearers' differential performances, including the number of words spoken, the rate of speech, the number of conceptual repetitions, the number of restarts, and the number of fillers. But one variable, the amount of discourse markers, did vary across conditions. Dialogues had more discourse markers than monologues and the presence of discourse markers correlated with improved performances. But it is unclear what this correlation means. The absence of markers might be hurting performance in the monologues, or the presence of markers might be aiding performance in the dialogues. Markers might also be correlated with some other factor that is the true cause of the dialogue advantage. For example, if markers help speakers organize their thoughts, then stimuli with more of them might also be better organized and easier to follow. Direct tests of overhearers' performances when listening in on descriptions with discourse markers and descriptions with the expressions excised would yield a clearer picture.

There is another explanation for the greater effectiveness of dialogues which addresses the other way the collaborative model could be extended. Dialogues might be more effective than monologues because they contain a variety of perspectives, increasing the chances that an overhearer's perspective will match what's said. With a monologue, only one person's point of view and descriptive abilities come into play. When two people collaborate on describing the shapes, diverse interpretations can arise:

(5) director: yeah I've got two turkeys *come to* think of it

matcher: *oh*

director: and they're both looking to the left, but one of them kind of has a leg
and a foot

matcher: that looks like an ostrich

director: yeah well this is the one without a leg or a foot

Unlike the earlier suggestion that conceptual repetition might be an advantage, it may be conceptual disparity that is an advantage. In a dialogue, if an overhearer does not see the shape the same way as the director, the overhearer still has another chance to identify the shape by the matcher's contribution. So, in the last example, the overhearer could use the matcher's information that the competing shape could be seen as an ostrich, not just a turkey. The overhearers in the monologue condition have to rely on what the directors think is sufficient, but in the dialogue condition they can use the extra input from the participating addressees.

Of course, the number of discourse markers and the number of perspectives in a dialogue may be closely related. When two people are revising each other's talk, as happens with negotiating different perspectives, they may be more likely to use discourse markers (e.g., the use of oh and well in Example 5). Alternatively, when people use discourse markers they may build a more friendly or creative atmosphere that fosters multiple perspectives.

Conclusion

Nonparticipating addressees understand more when listening in on spontaneous instructions produced by two people engaged in a conversation than spontaneous instructions produced by a single person delivering a monologue. Although future research can reveal the relative effectiveness of prepared monologues and scripted dialogues, such as those found in television commercials, movies, and news anchors' deliveries, the current research has implications for the interpretation of a wide array of other talk frequently encountered, such as town hall meetings, impromptu press conferences, open-microphone rallies, and even messages left on answering machines. The research described here suggests that comprehension will be better when overheard talk contains feedback from anyone than when overheard talk contains no feedback. More research along these lines may also be particularly useful in the deployment of teleconferencing technology. If people learn better when listening to collaborative dialogue, then a teleconferencing classroom where a lecturer has even a small interactive audience would be preferable to one where students watch a lecture on television without the ability to either give their own feedback or hear other people's feedback.

References

- Atkinson, M. (1984). Our Master's Voices: The Language and Body Language of Politics. New York: Methuen.
- Chantraine, Y., & Hupet, M. (1994). Efficiency of the addressee's contribution to the establishment of references: Comparing monologues with dialogues. Current Psychology of Cognition, 13(6), 777-796.
- Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In L. B. Resnick, J. M. Levine, & S. D. Teasley (Eds.), Perspectives on Socially Shared Cognition. Washington, D. C.: American Psychological Association.
- Clark, H. H., & Fox Tree, J. E. (in preparation). A Portrait of "Uh" and "Um".
- Clark, H. H., & Schaefer, E. F. (1987). Collaborating on contributions to conversations. Language and Cognitive Processes, 2(1), 19-41.
- Clark, H. H., & Schaefer, E. F. (1989). Contributing to discourse. Cognitive Science, 13, 259-294.
- Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. Cognition, 22(1), 1-39.
- Erman, B. (1987). Pragmatic Expressions in English: A Study of You Know, You See, and I Mean in Face-to-Face Conversation. Stockholm, Sweden: Almqvist & Wiksell International.
- Flowerdew, J., & Tauroza, S. (1995). The effect of discourse markers on second language lecture comprehension. Studies in Second Language Acquisition, 17, 455-458.
- 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(6), 709-738.
- Goffman, E. (1981). Forms of Talk. Philadelphia: University of Pennsylvania Press.
- Isaacs, E. A., & Clark, H. H. (1987). References in conversation between experts and novices. Journal of Experimental Psychology: General, 116(1), 26-37.

James, D. (1973). Another look at, say, some grammatical constraints on, ok, interjections and hesitations. Papers from the Ninth Regional Meeting. Chicago Linguistic Society.

Krauss, R. M., & Weinheimer, S. (1966). Concurrent feedback, confirmation, and the encoding of referents in verbal communication. Journal of Personality and Social Psychology, 4(3), 343-346.

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

Levelt, W. J. M. (1989). Speaking: From Intention to Articulation. Cambridge, MA: The MIT Press.

Redeker, G. (1991). Linguistic markers of discourse structure. Linguistics, 29, 1139-1172.

Reynolds, A., & Paivio, A. (1968). Cognitive and emotional determinants of speech. Canadian Journal of Psychology, 22(3), 164-175.

Schiffrin, D. (1987). Discourse Markers. New York : Cambridge University Press.

Schober, M. F. (1993). Spatial perspective-taking in conversation. Cognition, 47(1), 1-24.

Schober, M. F., & Clark, H. H. (1989). Understanding by addressees and overhearers. Cognitive Psychology, 4(21), 211-232.

Schourup, L. C. (1985). Common Discourse Particles in English Conversation. New York: Garland.

Stubbe, M., & Holmes, J. (1995). You know, eh and other "exasperating expressions": An analysis of social and stylistic variation in the use of pragmatic devices in a sample of New Zealand English. Language and Communication, 15(1), 63-88.

Wilkes-Gibbs, D., & Clark, H. H. (1992). Coordinating beliefs in conversation. Journal of Memory and Language, 31(2), 183-194.

Author Note

This research was supported by faculty research funds granted by the University of California, Santa Cruz and a National Science Foundation Graduate Fellowship for that portion of the research done at Stanford University. I thank Millicent Alvarado, Carolina Botero, Denise Cardoso, Renee Ferigo, Vicki Olthof, and Heidi Smoot for help with data collection and transcript creation and coding. I also thank four anonymous reviewers for their helpful comments on an earlier version of this manuscript. Correspondence concerning this paper should be addressed to Jean E. Fox Tree, Psychology Department, Social Sciences II, University of California, Santa Cruz, Santa Cruz, CA 95064.

Figure Caption

Figure 1. Correctly Ordered Tangrams