Discourse Processes
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t775653637

Overhearing Single and Multiple Perspectives

Jean E. Fox Tree [a]; Sarah A. Mayer [a]
[a] University of California, Santa Cruz

Online Publication Date: 01 March 2008
To cite this Article: Tree, Jean E. Fox and Mayer, Sarah A. (2008) 'Overhearing Single and Multiple Perspectives', Discourse Processes, 45:2, 160 - 179
To link to this article: DOI: 10.1080/01638530701792867
URL: http://dx.doi.org/10.1080/01638530701792867

Discourse Processes, 45:160–179, 2008

ISSN: 0163–853X print/1532–6950 online
DOI: 10.1080/01638530701792867

Overhearing Single and Multiple Perspectives

Jean E. Fox Tree and Sarah A. Mayer
University of California—Santa Cruz

In 2 spontaneous speech experiments, this study found that multiple perspectives improved overhearers' abilities to select abstract shapes from an array, although single-perspective descriptions were more detailed. Prior findings that overhearers performed better when listening in on dialogues (Fox Tree, 1999) can best be understood as an advantage of the multiple-perspective descriptions that arise out of the entrainment process. Results support the collaborative theory of language use and suggest that vicarious learners would do better listening in on talk containing multiple perspectives than talk containing a single perspective.

Imagine you are trying to follow crucial driving directions on the radio. You might be listening to a single announcer, or you might be listening to a radio pair's dialogue. In both cases, you are an *overhearer* unable to actively participate in the conversation (Schober & Clark, 1989). There is no *a priori* reason to expect that overhearers would have an easier time following a monologue or a dialogue (Fox Tree, 1999). However, the type of spontaneous talk that naturally arises under these different speaking conditions may affect the usefulness of the information conveyed. In particular, language produced in dialogues has been associated with the verbalization of a greater number of perspectives (Fox Tree, 1999). In this study, we tested the role that multiple perspectives have on overhearers' comprehension of spontaneous speech. We present results within the framework of the collaborative theory of language use (Clark, 1996; Clark & Brennan, 1991; Clark & Wilkes-Gibbs, 1986).

Correspondence concerning this article should be addressed to Jean E. Fox Tree, Department of Psychology, Social Sciences II, Room 277, University of California at Santa Cruz, CA 95064. E-mail: foxtree@ucsc.edu

BACKGROUND

Common ground is the mutual knowledge established between speakers during a conversation (Clark, 1996). As a conversation develops, interlocutors continuously update their models of the common ground they share. This process of *grounding* ensures that conversational participants understand what's being said well enough to accomplish the goals of the interaction (Clark & Wilkes-Gibbs, 1986). At the beginning of a conversation, participants establish their mutual knowledge, which they use throughout their dialogue (Isaacs & Clark, 1987). As the conversation develops, participants *lexically entrain* on certain ways of referring to people, places, objects, or topics (Brennan & Clark, 1996); that is, they mutually decide on a certain way to describe something, at least for the purposes of this conversation. So, for example, in describing abstract shapes, interlocutors may begin by labeling a shape with a variety of labels ("the person standing on one leg with the tail" and "an ice skater") but later entrain on one label to refer to the figure ("an ice skater"; Clark & Wilkes-Gibbs, 1986, pp. 13–14). The decision to call an object by a particular label is partner specific; speakers keep track of the labels, or *conceptual pacts,* that they use with different people (Brennan & Clark, 1996; Metzing & Brennan, 2003).

In establishing common ground and entraining on descriptions in a conversation, participants aim for *least collaborative effort* (Clark & Wilkes-Gibbs, 1986); that is, effort at production is balanced by effort at comprehension. One consequence of this is that conversational participants can adjust what they're saying for the knowledge state of their addressees. So, for example, an expert will describe things differently to another expert versus to a novice (Isaacs & Clark, 1987), and informed people will try to describe things in ways that assist the uninformed (Haywood, Pickering, & Branigan, 2005; Schober, 1995). People may also refuse to create conceptual pacts with one another; for example, by choosing to label an aborted fetus as a *baby* or a *fetus* (Brennan & Clark, 1996). In this case, the speakers embrace the extra work of maintaining separate labels to highlight their ideological differences.

Researchers who study grounding often explore the process in the context of active conversational participants. However, in many of our daily interactions with language, we are not participants, but overhearers to spontaneous talk. We can be overhearers to talk that is specifically designed with non-participating addressees in mind, such as when we listen to interviews on TV or the radio, or when we listen to the question–answer session of a lecture or meeting. We can also be overhearers to talk that is not designed for a wider audience, such as when we listen to a conversation between children at the table while preparing dinner at the stove (for a description of the differences between *addressees*, *side participants*, *bystanders,* and *eavesdroppers*, see Schober, 1998b, p. 232). Examining how people glean information from overheard language will help identify ways to improve communication where this setting is the only one possible, such as with broadcasting of emergency information.

Examining how people glean information when they are not direct conversational participants is also of great relevance for education research, especially with respect to *vicarious learning.* Vicarious learning refers to how people learn when they are not an active participant. For example, vicarious learners may learn by observing other's actions, reading transcripts of conversations or instructional monologues, or listening in on other's conversations (Cox, McKendree, Tobin, Lee, & Mayes, 1999; Craig, Driscoll, & Gholson, 2004; McKendree, Stenning, Mayes, Lee, & Cox, 1998). The conversations overheard could be either naturally produced between human speakers or artificially created with computer animated figures (Clark & Krych, 2004; Craig et al, 2004; Driscoll et al., 2003; Fox Tree, 1999; Murfitt & McAllister, 2001). The question of how non-participants best make use of informational resources is increasingly relevant with the widespread availability of computer technologies for distance learning.

According to one theory of how people establish and maintain common ground, direct participation should aid learning more than vicarious participation (Schober & Clark, 1989). Direct participants can negotiate until they have achieved mutual understanding, whereas vicarious participants must do their best with the information they can glean. Because vicarious participants cannot ask questions and are dependent on the information provided by the conversational interactants, they should not understand as well as direct participants. In support of this, direct conversational participants outperform overhearers in identifying abstract shapes (Schober & Clark, 1989). In these *referential card tasks,* a *director* describes a picture, often a *tangram* shape that looks similar to an origami figure, to a *matcher,* who selects that picture out of an array. Similarly, in a tutorial design featuring an electronic tutor, direct tutorial participants outperformed vicarious learners who were overhearing and overviewing the interaction (Craig et al., 2004). Direct participants also outperformed non-participants on a movie retelling task (Kraut, Lewis, & Swezey, 1982) and in a model building task (Clark & Krych, 2004).

However, the theory of common ground does not make clear predictions about what kind of information is most effective for non-participants (Fox Tree, 1999). One hypothesis is that information is best conveyed in a monologue format, free from the messiness of dialogue. For example, listening to a carefully constructed monologue tutorial would be preferable to listening to the messy result of an actual tutorial between an instructor and a learner. However, in contrast to this hypothesis that people learn better when receiving carefully constructed tutelage, people learned how to construct syntactic trees just as efficiently when they read transcripts of a dialogue between a student and a tutor as when they heard an instructional monologue from a tutor without an addressee. The tutorial monologue was intended for use by a future learner, but the unpredictable, naturally developed conversation was just as good (Cox et al., 1999).

There is other evidence that vicarious learners perform as well on information presented in a monologue format as in a dialogue format. People were just as good at identifying tangrams when they heard a description of the tangram jointly produced by two speakers as when they heard a description produced by one speaker (Murfitt & McAllister, 2001). People also recalled the same number of facts about an instructional topic when viewing and listening in on a dialogue between an electronic tutor and tutee and when viewing and listening in on a monologue presented by the electronic tutor (Driscoll et al., 2003). There is even some evidence that overhearers understand information better when they are listening in on dialogues as opposed to monologues; they can identify more tangram shapes correctly in a referential card task (Fox Tree, 1999).

Researchers acknowledge that variation in materials may account for discrepancies across studies, such as the number of words used in dialogue versus monologue stimuli, whether tangram shapes were described as metaphoric figures or in terms of their geometric components, and whether the feedback provided by the addressee in the dialogue conditions did or did not contain quality questions that fostered deeper thinking (Driscoll et al., 2003; Murfitt & McAllister, 2001). Furthermore, in a dialogue there is explicit acknowledgment that each conversational participant's contribution to a conversation has been understood (i.e., there is an explicit *acceptance phase* to the *presentation phase* that, together, make up a conversational *contribution*; Clark, 1996). This means that overhearers have information about the effectiveness of each contribution. Monologues have imagined agreement (the speaker provides information that they imagine is sufficient), but not explicit agreement.

Exploration of some of these factors revealed that they either had no effect on the dialogue advantage (Fox Tree, 1999) or that they "influence novel listener comprehension to a surprisingly small degree" (Murfitt & McAllister, 2001, p. 342). These include the number of words spoken (Fox Tree, 1999; Murfitt & McAllister, 2001), the rate of speech, the number of repetitions of words, the number of restarted ideas (Fox Tree, 1999), the number of repetitions of concepts (Fox Tree, 1999; Driscoll et al., 2003), and the figurativeness of descriptions (Murfitt & McAllister, 2001).

PERSPECTIVES

The label *perspectives* has been used to refer to a variety of phenomena, including "world views," "conversational agendas," "conceptions of how particular phrases are intended," and "physical vantage points" (Schober, 1998a, p. 146), as well as the outcome of combining conflicting concepts to create new metaphorical descriptions (Schön, 1993, p. 138). In this work, we operationally define perspectives as the different descriptions people use to refer to a shape.

A perspective is a set of details that, together, describe one way of viewing an object. People can adopt a perspective that is a metaphorical whole-object label such as "it looks like a canyon," but they can also adopt a geometrical perspective such as "looks like a square and two like symmetrical rocks on each side of it." People can also use multiple perspectives such as "a diamond [. . .] being lifted up by two pillars," "seals [going] after this little square shaped ball," and "the tips of two pencils [. . .] leaning against each other sort of like firewood" (all of these perspectives—the canyon, the square with symmetrical rocks, the diamond and pillars, seals and ball, and pencils like firewood—were suggested for the same tangram shape).

Dialogues contain more perspectives than monologues (Fox Tree, 1999). The process of lexical entrainment can explain this observation. In dialogues, matchers can indicate that they need another perspective to identify a referent, whereas in monologues they cannot. Dialogue participants may foster more perspectives as they negotiate how to label a referent, whereas monologues speakers may be more inclined to stick with the particular perspective they personally favor (Fox Tree, 1999); that is, when people describe abstract shapes under monologue conditions, they are not compelled to increase their ways of viewing a referent and, therefore, would offer, on average, fewer perspectives. In fact, although monologue speakers do sometimes provide multiple descriptions of tangram shapes, they generally do stick with one perspective: In this study, the number of multiple-perspective monologue items was a limiting factor in stimulus selection.

Prior research evidence supports the argument that dialogues encourage multiple perspectives. If people happen to share the same perspective of a referent, they opportunistically adopt that view of the item; if they do not happen to share the same perspective, they negotiate alternatives (Clark & Wilkes-Gibbs, 1986; Schober & Clark, 1989). This is another way to describe overhearers' poorer performances relative to direct addressees' when listening to a referential card task dialogue—they do not share in "exploiting adventitious commonalities" the way direct addressees do (Schober & Clark, 1989, p. 229).

These data suggest that multiple perspectives should offer more commonalities for overhearers. The presence of multiple perspectives increases the chances that one of the perspectives will match the overhearer's perspective (Fox Tree, 1999). However, at the same time, multiple perspectives may be more confusing to overhearers than single perspectives. Because overhearers cannot ask for clarification, a single detailed perspective may be clearer than multiple perspectives.

Fox Tree (1999) hinted that multiple perspectives were the driving force behind overhearers' enhanced performances listening in on dialogues; that is, it was not the setting (monologue or dialogue) that drove performance, but the type of information typically conveyed in that setting (one or multiple perspectives). However, although this was proposed as an explanation, it was not tested directly; and it is possible that something else drove the dialogue advantage in this study, such as the greater number of discourse markers in the dialogues (Fox Tree, 1999). Alternatively, the information provided in the dialogue may be more likely to be considered useful by overhearers because it was vetted by at least one addressee in the acceptance phase of the conversational contribution; no such acceptance is present with monologue items.

Dialogues yield more perspectives as a byproduct of the grounding process. However, without a direct test, we cannot firmly conclude that the number of perspectives advantage overhearer comprehension. It could be that listening in on dialogues is simply more pleasurable than listening in on monologues, prompting overhearers to pay more attention. Even portions of dialogues that can be made to appear like monologues, by snipping a section that includes only one voice, may still elicit a dialogue advantage if they are produced in a more attention-eliciting style because of the presence of an interactive addressee. There are a variety of other ways dialogues and monologues differ (even across one-voice snippets), any of which could affect overhearer performance including how many subparts a description has, the duration of words, and the vantage point of descriptions (Murfitt & McAllister, 2001; Schober & Brennan, 2003).

EXPERIMENTS

In these experiments, we tease apart the contribution of perspective number (one or multiple) and production setting (dialogue or monologue). Participants performed a referential card task. They listened to a description of an abstract shape and attempted to select the shape being described from an array presented on a computer screen. Descriptions were divided into multiple perspectives that were originally from dialogues, multiple perspectives that were originally from monologues, single perspectives that were originally from dialogues, and single perspectives that were originally from monologues.

The single perspectives we compared were presented as single perspectives by directors to matchers in the monologue productions or negotiated as single perspectives in the dialogue productions; that is, our single perspectives were not the final conceptual pact after multiple iterations of the task. In dialogues, end-product labels can be idiosyncratic to the couple engaged in the negotiation; the entrainment process can lead to obtuse conceptual pacts (Clark & Wilkes-Gibbs, 1986; Schober & Clark, 1989). Using conceptual pacts as the single perspectives would have artificially increased the difficulty of interpreting single perspectives as opposed to multiple perspectives.

ADDITIONAL FACTORS

Unscripted, unrehearsed talk can vary on numerous dimensions, even when the task is as restricted as describing tangrams. Any of these dimensions could covary with accuracy and, in turn, underlie any potential relation discovered between accuracy and perspective number or production setting.

In addition to our primary exploration of the use of differing numbers of perspectives in monologues and dialogues, we also evaluated the role of eight additional factors on people's abilities to select an abstract shape out of an array. These were factors that could most obviously have a bearing on people's success at the referential card task. They are listed as follows, along with a description of their potential impact on the results:

1. *The number of words used in the descriptions:* The number of words could influence accuracy under the simple principle that the more words heard, the more information provided. Multiple-perspective items, or dialogues, may be longer and therefore provide more information than single-perspective items, or monologues. Indeed, more words have led to higher accuracy in selecting referents in dialogue settings (but not in monologue settings; Murfitt & McAllister, 2001).

2. *The number of details in the descriptions:* It may not be the number of words that is most helpful, but the number of details those words provide. People could select referents more accurately based on the number of details in the descriptions rather than the number of perspectives or production setting. Therefore, for example, the following description has two details, *dog* and *walking dog* (asterisks indicate overlap):

(a) Director: it's like a dog that's walking
Matcher: a walking dog *let me see*
Director: *a walking dog*
Matcher: oh ok
Director: think there's only one dog
Matcher: nkay

However, the following description, which is similar in length, has five details (arrow, block beneath it, someone sitting down, hunchback, and has a hat on):

(b) Director: an arrow with a block beneath it tk or, I know, what is that? or maybe it's someone sitting down with a hunchback? and has a hat on?

Details were evaluated by breaking up the tangram descriptions into the smallest informational chunks.

3. *The number of discourse markers*: Discourse markers can potentially organize spontaneous speech in a way that makes talk easier for overhearers to follow; for example, they might indicate which ideas are more important, when an upcoming piece of information will be disjointed from prior information, or when to expect a revision of something just said (Fox Tree, 2000; Fox Tree & Schrock, 1999, 2002). People might identify referents better when they have followed the talk better because of the discourse markers used, not the number of perspectives or production setting. The discourse markers present in these materials were *I dunno, I mean, oh, like, ok* and *kay,* and *let's see*.

4. *The number of disfluencies*: Disfluencies have been shown to affect the comprehension of spontaneous speech (Brennan & Williams, 1995; Brennan & Schober, 2001; Fox Tree, 1995, 2001, 2002). An item with many restarts may be harder to follow than one with no restarted ideas, as is suggested in reading the following transcript of a relatively disfluent item:

(a) looks like, what does it look like? looks like um a leaf or leaf-like or it may be a bird, may, it looks like a bird maybe, um

As with prior factors, the number of disfluencies may vary systematically across experimental conditions, potentially driving any effect found. The disfluencies evaluated included repetitions, false starts, and *um*s and *uh*s.

5. *The liveliness of the descriptions*: In student–teacher interactions, liveliness is known to have a positive impact on student learning (cf. *expressiveness* and *energy*; Solomon, Rosenberg, & Bezdek, 1964). Dialogues may encourage more lively descriptions, and so a dialogue advantage may result from variation in this one factor rather than anything gained by a multiple party exchange over a monologue. Liveliness was operationalized as the average liveliness rating provided by ten listeners from the same community as the participants. Each of the 40 items was rated by each of the 10 listeners on a 7-point scale from 1 (*boring*), 4 (*neither*), to 7 (*lively*). The most lively description was the following:

(a) Director: kind of like a not very distinguished looking one but it it's like it's looking off to the left and you can make out two legs it's sort of hard to tell what animal it is but it's kind of walking up stairs to the left sort of
Matcher: kind of looks like a polar bear
Director: yeah, it does look like a polar bear, yeah, totally

The least lively was the following:

(b) Director: is tk a hand? Or I guess maybe a hand reaching out? Or a mm I dunno um maybe a bear maybe looks like a bear or a very odd looking thing ha ha um I think it looks like plier hands? Like reaching out to you or something like that

Although it may not be apparent from reading these items what makes one livelier than the other, listeners were able to draw distinctions when hearing the items. Liveliness ratings may reflect differences in vocal intonation.

6. *The number of explicit acceptance parts to conversation contributions, in the form of backchannels*: Dialogues may be further advantaged by providing information to overhearers about effective communication; that is, when the matcher says “ok” to “Director: that’s a- it’s like the cat that’s lying down,” she is indicating that the label “cat that’s lying down” was sufficient for her to understand the director’s instruction. This may in turn lead overhearers to process materials differently; they may select the object most likely to be a lying-down cat instead of continuing to search for competing objects, as they might do if the label “cat that’s lying down” were insufficient.

Backchannels included *yes, yeah, I see, mhm, mmm, ok,* and *right*. Although backchannels are not the only way to indicate acceptance of a presentation (Clark, 1996), they do provide more acceptance information than an utterance with no backchannels (and no facial or gestural information, as in these materials); that is, more information about addressee acceptance is provided in the dialogues than in the monologues through the presence of backchannels.

7. *The number of quality questions asked by the matcher in the dialogues*: Yet another way dialogues may advantage listeners is in the quality of the contributions made by the matcher to the director’s presentations. For example, in the following, the matcher suggests a novel perspective:

(a) Director: with the guy kicking like way back kind of like a a woman like a ballerina or something?
Matcher: yeah, looks like a Christmas tree?
Director: yeah

In contrast, in the following, the matcher merely repeats a perspective the director already offered (example repeated here):

(b) Director:it’s like a dog that’s walking
Matcher: a walking dog *let me see*
Director: *a walking dog*
Matcher: oh ok
Director: think there’s only one dog
Matcher: nkay

Matcher questions or replies that contributed something new to the discourse may improve comprehension. Indeed, more thoughtful questions led to greater learning in a vicarious learning setting (Driscoll et al., 2003). Quality questions were operationalized as matcher contributions that enhanced the information supplied by the director.

8. *The number of metaphorical versus geometrical perspectives*: Descriptions can vary in how metaphorical or geometrical they are, and this factor may influence reference selection. For example, multiple-perspective items may tend to present geometrical perspectives, and these may be easier for overhearers to follow because they rely less on idiosyncratic labels. For example, "it has a piece on top of a square on top of a triangle" and "an arrow with a block beneath it" seem more concrete than "a chess piece" and "someone sitting down with a hunchback."

EXPERIMENT 1

People attempted to identify abstract shapes that were produced either in a monologue or dialogue setting and that contained either multiple perspectives or single perspectives.

Method

Participants. Eighty-two University of California at Santa Cruz undergraduates participated in partial fulfillment of course requirements. All were native English speakers.

Materials. Forty spoken verbal item descriptions were selected from a corpus of tangram descriptions produced under either a monologue or a dialogue condition. The speech tested was produced by 14 different people; all were native English speakers.

Tangrams are abstract figures composed of a fixed set of geometric shapes (5 triangles, 1 square, and 1 parallelogram). Our corpus contained descriptions of tangrams that either resembled people or resembled animals. The monologues consisted of directors describing tangram figures to matchers in an adjoining room who could hear but not speak. The dialogues consisted of directors describing figures to matchers in an adjoining room who could converse without restriction. Because matchers could freely ask questions and elaborate on descriptions supplied by the directors, descriptions in the dialogue condition could be produced jointly. For example, the director could describe "a weird man hunched over," and the matcher could contribute "like a bear?"

Ten stimuli were multiple-perspective auditory descriptions originally produced in monologues, as in the following:

> There's one that has a square well a diamond and it looks like it's being lifted up um by two pillars? um I dunno the pillars could easily be seels and they're after this little square shaped ball or you could see it as like a canyon? where it goes up and then um it goes down it has a rock sitting on top of it, like those rock formations in Arizona?

and um uh I dunno they look like the tips of two pencils, two dull pencils laid against leaning against each other sort of like firewood with a diamond resting on it.

Ten stimuli were multiple-perspective auditory descriptions originally produced in dialogues, as in the following:

Director: um there's one it looks like a fish
Matcher: a fish
Director: yeah
Matcher: ok is it kind of a real big *solid shape*
Director: *there's a big blob* and then a little triangle on the bottom
Matcher: ok I got it

Nine stimuli were one-perspective auditory descriptions originally produced in monologues, as in the following (ellipses indicate card number information, such as "number five," which was removed from all stimuli):

… a graduate, or um somebody who's graduating? and has a hat looking down so you can see the hat the top of the hat and then um the robe or whatever

Eleven stimuli were one-perspective auditory descriptions originally produced in dialogues, as in the following:

Director: … is a dog I guess but he's got he's facing the right he has a little tail, an- but his ears are a lot bigger, it's about half the size of his head?
Matcher: ok and his legs come to points?
Director: yeah
Matcher: bottom
Director: but I think that his forelegs come down further than his back legs
Matcher: right just a little bit
Director: yeah
Matcher: ok
Director: and I guess his tail is a triangle not the parallelogram
Matcher: yeah, ok got it

The descriptions were complete; that is, they were not manipulated to contain one perspective only, or to contain multiple perspectives.

Two of the eight additional factors to be explored were tested in the stimulus preparation stage. These two factors, the number of words (Murfitt & McAllister, 2001) and the number of discourse markers (Fox Tree, 1999), had been shown in prior work to be related to accuracy in a tangram referential card task.

The materials did not differ on the number of words in the single- versus multiple-perspective conditions; average 62.8, $SD = 43.1$, in single perspectives and 78.5, $SD = 42.4$, in multiple perspectives, $t(38) = 1.17$, $p = .25$. However, monologues were longer than dialogues; average 90.8, $SD = 43.8$, in monologues and 52.4, $SD = 33.7$, in dialogues, $t(38) = 3.13$, $p = .003$. The range across all items was

11 to 167 words. Because prior research predicts better performance listening in on dialogues, the longer monologues work against the hypothesis.

The materials did not differ in the number of discourse markers in the single- versus multiple-perspective conditions; average 0.8, $SD = 1$, in single perspectives and 1.3, $SD = 1.3$, in multiple perspectives, $t(38) = 1.23$, $p = .23$. They also did not differ in the rate of discourse markers in the single- versus multiple-perspective conditions, about 2 per 100 words in both the single and multiple, average 2, $SD = 3.7$, and average 1.7, $SD = 1.7$, $t(38) = 0.27$, $p = .79$. The range across all items was 0 to 5 discourse markers.

Across monologues and dialogues, materials did not differ in the number of discourse markers, but they did differ in the rate—with a higher rate of markers in the dialogues. By number, there were on average 0.9, $SD = .9$, discourse markers in the monologues and 1.1, $SD = 1.4$, in the dialogues, $t(38) = 0.67$, $p = .51$. By rate, there were about 1 per 100 words in the monologues, $SD = .8$, and 2.7 per 100 words in the dialogues, $SD = 3.7$, $t(22) = 2.33$, $p = .03$. Both the number and rate of discourse markers were correlated with accuracy to test if the two were related (see section on Analysis of Additional Stimuli Materials).

Two visual stimuli were prepared, one corresponding to the people set of tangrams and the other the animal set. Each stimulus contained 16 different tangram figures. Under each figure, a unique letter was written to identify it.

The experiment was prepared using PsyScope (Cohen, MacWhinney, Flatt, & Provost, 1993) and run on Macintosh computers. The auditory stimuli were synchronized with the visual stimuli such that when a people description was heard, the people visual stimulus was shown; and when an animal description was heard, the animal stimulus was shown.

Design. The design was 2×2 within-subjects and between-items, with the factors setting (monologue or dialogue) and type of description (one perspective or multiple perspectives). The dependent variable was the number of items accurately identified. Because the items were not matched across conditions, effects were assessed with $minF'$ (Raiijmakers, 2003; Raaijmakers, Schrijnemakers, & Gremmen, 1999).

Procedure. Participants were individually seated in front of a computer. They were instructed to listen to a series of item descriptions and to select the item that matched the description from the set of choices visible on the computer. A beep indicated that the description had ended. For each item, participants were told to wait until the beep before making their selection. After pressing the keyboard letter corresponding to their tangram selection, the next trial began. Each participant heard every item. Utterances were presented in a different random order for each participant. At the end of the experiment, participants completed a post-experiment questionnaire soliciting any comments they had about the experiment. American Psychological Association standards for the treatment of human participants were followed.

Results

There was neither a ceiling nor a floor effect in this experiment; the overall range of accuracy was 8 to 37 out of 40. No participant got all the tangrams right, and no tangram was accurately chosen by all the participants.

Overhearers tended to select tangram figures more accurately when they heard multiple-perspective descriptions than when they heard single-perspective descriptions, 79% (SD = 12%) correct as compared to 67% (SD = 11%), $minF'(1, 43) = 3.37$, $p = .07$.

Accuracy was not affected by whether the description was originally produced in a monologue or a dialogue, 70% (SD = 12%) correct as compared to 77% (SD = 11%), $minF'(1, 42) = 1.16$, $p = .29$. Although this result is inconsistent with earlier tests comparing performance when listening to a string of descriptions (i.e., hearing a dozen or more descriptions in a row and ordering the shapes from a jumble of cards on a table; Fox Tree, 1999), it accords with findings using the similar methodology of hearing one description and selecting the shape from an array (Murfitt & McAllister, 2001), including that the direction of the effect favors listening in on dialogues.

There was no interaction between perspective number and production setting, $\min F'(1, 41) = 0.06$, $p = .8$.

Analyses of Additional Stimuli Variables

We tested whether accuracy correlated with seven of the eight stimuli variables: (a) the number of words, (b) the number and rate of details, (c) the number and rate of discourse markers, (d) the number and rate of disfluencies, (e) the liveliness of the descriptions, (f) the number and rate of backchannels, and (g) the number and rate of quality questions asked. Because the hypotheses go in one direction, one-tailed correlational analyses were used for all but the disfluency analyses; that is, the hypothesis is that the greater the number of words, details, discourse markers, liveliness, backchannels, and quality questions, or the greater the rate, the higher the accuracy. Because disfluencies may both hinder or help comprehension (Clark & Wasow, 1998; Fox Tree, 1995, 2001), a two-tailed test was used for this analysis. For the eighth variable, we tested whether (h) there was a difference in accuracy across geometrical versus metaphorical descriptions.

Three of the stimuli variables correlated with accuracy. The livelier the description, the more overhearers identified it correctly, $r(38) = 0.3$, $p = .03$. Therefore, overhearers' attention is modulated by how engaging the talk is that they are listening to. Students may either pay more attention to lively descriptions, or less attention to boring descriptions, or both.

In addition, the more disfluencies the item contained, the fewer overhears identified it correctly, $r(38)= -0.31$, $p = .05$. However, this effect did not carry over to rate, $r(38) = -0.13$, $p = .41$. Therefore, for these materials, the number of disfluencies negatively affected comprehension.

Finally, the higher the rate of discourse markers in the item, the fewer overhearers correctly identified it, $r(38) = -0.3$, $p = .03$. This effect did not carry over to number, $r(38) = 0.1$, $p = .27$. More important, it is in the opposite direction of the one predicted by prior work, where more discourse markers were related to greater accuracy (Fox Tree, 1999).

None of the other variables affected accuracy: $r(38) = -0.12$, $p = .23$, for the number of words, $r(38) = -0.18$, $p = .13$, for the number of details, and $r(38) = -0.14$, $p = .2$, for the rate; $r(18) = 0.12$, $p = .3$, for the number of backchannels, and $r(18) = 0.05$, $p = .43$, for the rate (only tested in dialogue stimuli); and $r(18) = -0.12$, $p = .31$, for the number of quality questions asked, and $r(18) = -0.22$, $p = .17$, for the rate (only tested in dialogue stimuli).

Accuracy was similar across one-perspective metaphorical and geometrical perspectives, means of 68% to 65%, $t(18) = 0.19$, $p = .85$; and across multiple-perspective metaphorical-only versus metaphorical plus geometrical perspectives, means of 85% to 76%, $t(18) = 1.5$, $p = .15$.

Discussion

Numerically, overhearers performed better on the dialogues and on the multiple perspectives, but only the multiple-perspective factor approached significance. The two factors did not interact.

Five of the eight additional stimuli variables were unrelated to accuracy including the number of words, the number of details, the number of backchannels, the number of quality questions asked, and whether the descriptions contained metaphorical or geometrical perspectives.

The number of disfluencies, the rate of discourse markers, and liveliness were related to accuracy and may have influenced the main effects. However, disfluencies had an effect only by number, and discourse markers only by rate, and that in the opposite of the predicted direction.

To control for the liveliness of the descriptions, in Experiment 2 we created new single-perspective item descriptions by shortening the multiple-perspective descriptions down to one perspective. This procedure also controlled for other spurious factors, such as the quality of the speaker's voice or the speaker's natural tempo.

EXPERIMENT 2

Experiment 2 is similar to Experiment 1, with the change that single-perspective items were shortened versions of the multiple-perspective items.

Method

Participants. Seventy University of California at Santa Cruz undergraduates participated in partial fulfillment of course requirements. All were native English speakers.

Materials. Materials were the same as Experiment 1 except that the single-perspective items from Experiment 1 were replaced by shortened versions of the Experiment 1 multiple-perspective items, as in the following:

> Looks like a very tall neck tie, ha ha it's basically vertical straight up and down I'll start describing from the bottom, there's a triangle which is wider on the bottom and then there's a square a perfect square, and then on the top is kind of this like big I dunno sort of has a rectangle on its left side but I think probably from if you look at [sic] from the bottom it's a triangle and then a square on top of the triangle and then another structure

This two-perspective item ("very tall neck tie" and geometric description) was edited to contain only one perspective by removing "looks like a very tall neck tie." Sometimes the single perspective retained from the multiple perspectives was geometric, as in this example; sometimes it was metaphoric. An even distribution was necessary to avoid confounding single perspectives with type of perspective (geometric or metaphoric). Single- and multiple-perspective items were no longer matched for length, average 37.3 words, $SD = 27.1$, in single perspectives and 78.5 words, $SD = 42.4$, in multiple perspectives, $t(19) = 6.2$, $p < .001$.

Monologues were still longer than dialogues, average 73.4, $SD = 42.4$, in monologues and 42.4, $SD = 33.5$, in dialogues, $t(38) = 2.57$, $p = .01$; range across all items was 7 to 144 words.

In addition, materials did now differ in the number of discourse markers in the single versus multiple-perspective conditions with, on average, about one more discourse marker in the multiple perspectives, average 0.4, $SD = .7$, in single perspectives and 1.2, $SD = 1.3$, in multiple perspectives, $t(19) = 3.8$, $p = .001$. However, they still did not differ in the rate of discourse markers, 1.4, $SD = 2.7$, in single perspectives and 1.6, $SD = 1.7$, in multiple perspectives, $t(19) = 0.37$, $p = .72$.

Materials also now tended to differ in the number of discourse markers in the monologues and dialogues with more in the dialogues, average 0.5, $SD = 0.8$, in monologues and 1.1, $SD = 1.3$, in dialogues, $t(38) = 1.75$, $p = .09$. They continued to differ in the rate of discourse markers in the monologues and dialogues, average 0.6, $SD = 0.9$, in monologues and 2.5, $SD = 2.6$, in dialogues, $t(24.5) = 3.12$, $p = .005$.

Finally, the materials differed in the number of disfluencies in the single and multiple perspectives with more in the multiple perspectives, average 1.3, $SD = 1.5$, in the single perspectives and 3, $SD = 2.2$, in the multiple perspectives, $t(19) =$

5.16, $p < .001$. However, they did not differ in the rate of disfluencies, average 4.3 per 100 words, $SD = 5.2$, in the single perspectives and 4.1, $SD = 3.1$, in the multiple perspectives, $t(19) = 0.16$, $p = .87$.

They did not differ in either the number or rate of disfluencies across monologues and dialogues, average 2.6, $SD = 2.2$, for number in monologues and 1.6, $SD = 1.8$, for number in dialogues, $t(38) = 1.5$, $p = .14$; and average 4.6, $SD = 4.5$, for rate in monologues and 3.7, $SD = 3.9$, for rate in dialogues, $t(38) = 0.64$, $p = .53$.

Design. The design was 2×2 within-subjects and mixed-items, with the factors setting (monologue or dialogue, between-items) and type of description (one perspective or multiple perspectives, within-items). Each participant heard five single-perspective stimuli originally produced in a monologue, five multiple-perspective stimuli originally produced in a monologue, five single-perspective stimuli originally produced in a dialogue, and five multiple-perspective stimuli originally produced in a dialogue. The dependent variable was the number of items accurately identified. Because items were matched for the type of description, the effect of type of description was assessed by a simple subject analysis (Raaijmakers, 2003; Raaijmakers et al., 1999). Because items were not matched for setting, the effect of setting was assessed with $minF'$ (Raiijmakers, 2003; Raaijmakers et al., 1999).

Procedure. The procedure was the same as in Experiment 1.

Results

There was neither a ceiling nor a floor effect in this experiment; the overall range of accuracy was 4 to 19 out of 20. No participant got all the tangrams right, and no tangram was accurately chosen by all the participants.

Overhearers selected tangram figures more accurately when they heard multiple-perspective descriptions than when they heard single-perspective descriptions, 74% ($SD = 14\%$) correct as compared to 69% ($SD = 14\%$), $F(1, 43) = 7.68$, $p = .007$.

Accuracy was not affected by whether the description was originally produced in a monologue or a dialogue, 68% ($SD = 13\%$) correct as compared to 75% ($SD = 15\%$), $minF'(1, 22) = 0.8$, $p = .38$.

There was no interaction between perspective number and production setting, $minF'(1, 27) = 0.31$, $p = .58$.

The number of words used in an item description did not correlate with accuracy: $r(38) = 0.08$, $p = .63$. The number and rate of discourse markers used in descriptions also did not correlate with accuracy: $r(38) = 0.21$, $p = .1$, for number and $r(38) = 0.001$, $p = .5$, for rate; nor did the number or rate of disfluencies: $r(38) = -0.16$, $p = .16$, for number and $r(38) = -0.2$, $p = .11$, for rate. Finally, the liveliness of the descriptions did not correlate with accuracy: $r(38) = 0.1$, $p = .27$. Because

Experiment 1 provided predicted directions for the disfluency, discourse marker, and liveliness effects, these tests are one-tailed.

Discussion

Experiment 2 replicated the effects found in Experiment 1. Overhearers selected multiple-perspective items more accurately than single-perspective items, regardless of whether the items were originally produced in a monologue or dialogue setting.

Although monologues were longer than dialogues in both Experiments 1 and 2, accuracy was the same across these settings. There was also no correlation between accuracy and the number of words used in a description in either experiment. This supports the argument that it was number of perspectives, and not length of materials, that drove the perspective effect.

In Experiment 2, there were more discourse markers in the dialogues (by number and rate) and more discourse markers in the multiple perspectives (by number). Earlier research suggested the number of discourse markers may influence comprehensibility of descriptions. However, the number of discourse markers did not positively correlate with accuracy in either Experiment 1 or 2; in Experiment 1, a counter-hypothesis negative correlation between the rate of discourse markers and accuracy was observed.

Finally, although the liveliness of the stimuli was positively correlated with accuracy in Experiment 1, and the number of disfluencies were negatively correlated with accuracy in Experiment 1, neither factor was related to accuracy in Experiment 2. In Experiment 1, the livelier descriptions may have heightened the accuracy of the multiple-perspective items. However, in Experiment 2, the single- and multiple-perspective items were the same in liveliness, as they were matched on this factor. This means that liveliness was not responsible for the superior performance of the multiple-perspective items in Experiment 2.

GENERAL DISCUSSION

Multiple perspectives do, indeed, enhance overhearer performance. When engaged in a referential card task, overhearers were more accurate at selecting the card described when they were presented with more than one way of looking at the abstract figure. The greater number of discourse markers in dialogues did not contribute to the multiple-perspective advantage. Finally, the multiple-perspective advantage held regardless of whether the description was originally produced in a monologue or a dialogue. We conclude that the dialogue advantage documented in earlier work (Fox Tree, 1999) was a result of the greater number of perspectives available to the overhearers to identify the shapes.

In these experiments, we also found evidence that the amount of detail provided was less important than the number of perspectives. In Experiment 1, the single- and multiple-perspective items did not differ on the number of words. What this meant, in effect, was that the single-perspective items contained more detailed descriptions of one way to view the shape, and multiple-perspective items contained concise listings of alternative ways to view the shape. A priori, one could argue that depth would be more important to overhearers because it avoids the potential confusion of rapidly presented, short alternative descriptions, none of which overhearers may happen to share with the original conversational participants. However, one could also argue that depth only provides more detail about a perspective an overhearer may already be struggling with, and thus cuts off other avenues for comprehension that multiple perspectives provide. In this referential card task, breadth was favored over depth.

A number of alternative explanations for the results could be ruled out including that the effects were driven by the number or rate of words used in the descriptions, the number or rate of details, the number or rate of discourse markers, the number or rate of disfluencies, the liveliness of the descriptions, the number or rate of explicit acceptance parts to the conversational contributions (backchannels), the number or rate of quality contributions made by the matcher, and the use of metaphorical versus geometrical perspectives.

This last potential explanation highlights the generalizability of these studies. A priori, it may seem that geometrical descriptions would be more helpful to a non-participating audience than metaphorical ones, because everyone can identify triangles and parallelograms, but not everyone will see a shape as a seal or a canyon. Furthermore, tangrams are more abstract than most things people convey in everyday life such as descriptions of people, buildings, and objects. Therefore, perhaps tangrams are described in a specialized way that does not have much bearing on other types of referential communication. The fact that there was no superiority of metaphorical versus geometrical perspectives on accuracy suggests that even if tangrams are described geometrically, that is not the reason people follow some descriptions better than others. In addition, it turns out that speakers were much more likely to describe the tangrams in metaphorical terms. Even when a tangram was too abstract to describe as an animal or person, people still chose metaphorical descriptions (such as “a tie”) over a description of the type and orientation of shapes.

These findings suggest that in a vicarious learning situation, learning can be enhanced by maximizing different viewpoints. Only the communication of visual descriptions of static images was tested here, but extrapolation would suggest more successful learning when overhearing multiple approaches to solving a problem. Also, only overhearing talk not explicitly intended for a wider audience was tested here, but extrapolation would suggest more successful communication when overhearing talk with an intended audience of non-participating listeners as well. Therefore, for example, following driving directions on the radio would be enhanced by describing the route in multiple ways. Of course, further studies are needed to verify these hypotheses.

As the original conversational participants grounded their descriptions, they verbalized more ways to look at the abstract shapes than the monologue directors invented while doing the task alone. It was these additional perspectives that aided overhearers in selecting the abstract shapes. The multiple-perspective descriptions that arise out of the grounding process are useful not only for original addressees, but for overhearers as well.

ACKNOWLEDGMENT

This research was supported by a Faculty Research Grant awarded by the Committee on Research at the University of California, Santa Cruz.

REFERENCES

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 1482–1493.

Brennan, S. E., & Schober, M. F. (2001). How listeners compensate for disfluencies in spontaneous speech. *Journal of Memory and Language, 44*(2), 274–296.

Brennan, S. E., & Williams, W. (1995). The feeling of another's knowing: Prosody and filled pauses as cues to listeners about the metacognitive states of speakers. *Journal of Memory and Language, 44*(2), 274–276.

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

Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In L. B. Resnick, J. M. Levine, & S. D. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 127–149). Washington, DC: American Psychological Association.

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

Clark, H. H., & Wasow, T. (1998). Repeating words in spontaneous speech. *Cognitive Psychology, 37,* 201–242.

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

Cohen, J. D., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments, and Computers, 25*(2), 257–271.

Cox, R., McKendree, J., Tobin, R., Lee, J., & Mayes, T. (1999). Vicarious learning from dialogue and discourse. *Instructional Science, 27,* 431–458.

Craig, S. D., Driscoll, D. M., & Gholson, B. (2004). Constructing knowledge from dialog in an intelligent tutoring system: Interactive learning, vicarious learning, and pedagogical agents. *Journal of Educational Multimedia and Hypermedia, 13,* 163–183.

Driscoll, D. M., Craig, S. D., Gholson, B., Ventura, M., Hu, X., & Graesser, A. C. (2003). Vicarious learning: Effects of overhearing dialog and monologue-like discourse in a virtual tutoring session. *Journal of Educational Computing Research, 29,* 431–450.

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. (2000). Coordinating spontaneous talk. In L. R. Wheeldon (Ed.), *Aspects of language production* (pp. 375–406). Philadelphia: Psychology Press.

Fox Tree, J. E. (2001). Listeners' uses of um and uh in speech comprehension. *Memory and Cognition, 29,* 320–326.

Fox Tree, J. E. (2002). Interpreting pauses and ums at turn-exchanges. *Discourse Processes, 34,* 37–55.

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.

Fox Tree, J. E., & Schrock, J. C. (2002). Basic meanings of you know and I mean. *Journal of Pragmatics, 34,* 727–747.

Haywood, S. L., Pickering, M. J., & Branigan, H. P. (2005). Do speakers avoid ambiguities during dialogue? *Psychological Science, 16,* 362–366.

Isaacs, E. A., & Clark, H. H.. (1987). References in conversations between experts and novices. *Journal of Experimental Psychology, 116,* 26–37.

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

McKendree, J., Stenning, K., Mayes, T., Lee, J., & Cox, R. (1998). Why observing a dialogue may benefit learning. *Journal of Computer Assisted Learning, 14,* 110–119.

Metzing, C., & Brennan, S. E. (2003). When conceptual pacts are broken: Partner-specific effects on the comprehension of referring expressions. *Journal of Memory and Language, 49,* 201–213.

Murfitt, T., & McAllister, J. (2001). The effect of production variables in monolog and dialog on comprehension by novel listeners. *Language and Speech, 44,* 325–350.

Raaijmakers, J. G. W. (2003). A further look at the "language-as-fixed-effect fallacy." *Canadian Journal of Experimental Psychology, 57,* 141–151.

Raaijmakers, J. G. W., Schrijnemakers, J. M. C., & Gremmen, F. (1999). How to deal with "the language-as-fixed-effect fallacy": Common misconceptions and alternative solutions. *Journal of Memory and Language, 41,* 416–426.

Schober, M. F. (1995). Speakers, addressees, and frames of reference: Whose effort is minimized in conversations about locations? *Discourse Processes, 20,* 219–247.

Schober, M. F. (1998a). Different kinds of conversational perspective taking. In S. R. Fussell & R. J. Kreuz (Eds.), *Social and cognitive approaches to interpersonal communication* (pp. 145–174). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

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

Schober, M. F., & 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 Associates, Inc.

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

Schön, D. A. (1993). Generative metaphor: A perspective on problem-setting in social policy. In A. Ortony (Ed.), *Metaphor and thought* (pp. 137–163) Cambridge, England: Cambridge University Press.

Solomon, D., Rosenberg, L., & Bezdek, W. E. (1964). Teacher behavior and student learning. *Journal of Educational Psychology, 55,* 23–30.