COGNITIVE PSYCHOLOGY 21, 211–232 (1989)

# Understanding by Addressees and Overhearers

MICHAEL F. SCHOBER AND HERBERT H. CLARK

*Stanford University*

In conversation speakers design their utterances to be understood against the common ground they share with their addressees—their common experience, expertise, dialect, and culture. That ordinarily gives addressees an advantage over overhearers in understanding. Addressees have an additional advantage, we propose, because they can actively collaborate with speakers in reaching the mutual belief that they have understood what was said, whereas overhearers cannot. As evidence for the proposal, we looked at triples of people in which one person told another person in conversation how to arrange 12 complex figures while an overhearer tried to arrange them too. All three began as strangers with the same background information. As predicted, addressees were more accurate at arranging the figures than overhearers even when the overhearers heard every word. Other evidence suggests that the very process of understanding is different for addressees and overhearers. 

People understand each other in conversations by gathering evidence about each other's intentions. How do they do that? The traditional view, which we will call the *autonomous view,* is that they listen to the words uttered, decode them, and interpret them against what they take to be the common ground of the participants in the conversation (e.g., Anderson, 1985; Clark & Haviland, 1977; Fodor, 1983; Johnson-Laird, 1983; Sperber & Wilson, 1986; van Dijk & Kintsch, 1983; see also Clark & Schaefer, 1989). An alternative view, the *collaborative view,* is that speakers and their addressees go beyond these autonomous actions and collaborate with each other moment by moment to try to ensure that what is said is also understood. Collaboration takes extra processes and may require extra steps in the conversation (Clark & Wilkes-Gibbs, 1986; Garrod & Anderson, 1987; Goodwin, 1981; Kraut, Lewis, & Swezey, 1982; Schegloff, 1982; Schegloff, Jefferson, & Sacks, 1977).

These two views contrast in the way addressees and overhearers un-

We thank S. E. Brennan, A. Fernald, and M. Lepper for technical expertise and use of laboratory rooms and equipment. We also thank S. E. Brennan, E. V. Clark, F. Edwards, and E. F. Schaefer for valuable counsel on the manuscript. The research was supported by Grant BNS 87-09812 from the National Science Foundation and by a National Science Foundation Graduate Grant. Requests for reprints should be addressed to Michael F. Schober or Herbert H. Clark, Department of Psychology, Jordan Hall, Stanford University, Stanford, CA 94305.



0010-0285/89 $7.50

derstand. Addressees are participants in the conversation at the moment, and overhearers are not (Goffman, 1976; McGregor, 1986). Speakers are responsible for making themselves understood to the other participants, but not to overhearers. Indeed, speakers can take one of several attitudes toward overhearers (Clark & Carlson, 1982; Clark & Schaefer, 1987b). They may try to disclose, conceal, or disguise what they say, or they may be indifferent toward them. We will limit ourselves to the case of indifference. On the autonomous view, overhearers should do as well as addressees in understanding utterances in conversation whenever they have the same background as the addressees. They might even do better because they do not have to worry about what to say next, which can only interfere with the process of understanding. On the collaborative view, however, overhearers should be at a disadvantage even if they have all the right background. Let us see why.

In conversation, the participants accumulate information as part of their common ground—their mutual knowledge, beliefs, and assumptions (Clark, 1985; Gazdar, 1979; Lewis, 1979; Stalnaker, 1978). If Susan and Evan know each other, they begin each conversation with a good deal of common ground, and, as they talk, they add to it. They design their utterances to be understood against their accumulating common ground. When Susan tells Evan "He's here," she intends him to identify who she is referring to by consulting their common ground. Its source may be their conversation so far (Steve was just mentioned), their shared perceptual surroundings (Ed just walked in), previous joint experiences (Susan and Evan had arranged to meet Scott at that time), information universally known or believed in one of the cultural communities to which Susan and Evan know they both belong (the Pope is visiting that day), or some combination of these (Clark & Marshall, 1981).

On both views of understanding, overhearers should be at a disadvantage whenever they are ignorant of critical parts of the participants' common ground. Suppose Liz is overhearing Susan speak to Evan. She should have trouble if she has not caught the first part of the conversation, or if Susan and Evan are old friends, or if Susan and Evan are members of a culture she does not belong to. Her difficulties should occur on just those parts of Susan's meaning that depend on common ground to which Liz is not privy (Clark & Schaefer, 1987b). By the autonomous view, however, Liz should do as well as Evan whenever three conditions hold: (1) Susan, Evan, and Liz are from the same cultures, (2) they do not know each other in advance, and (3) Liz has listened in on the conversation from the start.

On the collaborative view, overhearers should be at a disadvantage even when all three conditions hold. The idea is that the participants in a conversation try to establish the mutual belief that the listeners have

understood what the speaker meant to a criterion sufficient for current purposes. This is a collaborative process, called *grounding,* that requires actions by both speakers and their addressees (Clark & Schaefer, 1987a). A reference or a question, for example, is not considered complete until both speaker and addressees have acknowledged that they have established the mutual belief that it has been understood. Consider this attested example (Svartvik & Quirk, 1980):

A: well wo uh what shall we do about uh this boy then—
B: Duveen?
A: m
B: well I propose to write uh saying . I'm very sorry I cannot- uh teach . at the institute

Although A tries to ask B what should be done about Duveen, A and B don't consider the question complete until A has cleared up B's problem well enough for B to answer "well I propose to write," etc. It takes three turns for A's question to become complete (see Clark & Schaefer, 1989).

Overhearers should be at a disadvantage just because they don't have grounding as a resource. Suppose C overhears the first utterance, and B understands who "this boy" refers to but C does not. B won't have to ask "Duveen?", but C, not being a participant in the conversation, has no way of asking for such a confirmation. A and B will continue in the mutual belief that B has understood A's assertion, whereas C cannot be sure she has understood it. That puts her at a disadvantage.

To make the argument concrete, let us define four time points in the understanding of a speech act such as an assertion or question.

1. *Initiation point.* This is the point at which a speaker initiates the speech act of interest. In our example, the initiation point of A's question is at the word *well.*
2. *Completion point.* This is the point at which the participants deem the grounding of that speech act to be complete. In our example, the completion point of A's question is at the end of the sound *m.*
3. *Recognition point.* This is the point at which the addressee believes that he or she has grasped what the speaker meant. In our example, that presumably coincides with the completion point.
4. *Conjecture point.* This is the point at which the overhearer conjectures that he or she has grasped what the speaker meant.

Suppose two listeners B and C have the same background knowledge, but B is a participant in the conversation and C is an overhearer. On the average, B's recognition point and C's conjecture point will be the same. Suppose further, however, that these two points will vary from one occasion to the next, and these variances are not completely correlated. On

one occasion, B's recognition point may precede C's conjecture point, and on another occasion, it may follow it.

Because of the grounding process, B's recognition point will ordinarily be identical to A and B's completion point. In our example, the completion point was collaboratively determined by A and B based on B's belief that he understood what A was asserting. Ideally, it should never come *before* B's recognition point, because B should always accept that he has understood. Suppose that B and C have the same mean speed of understanding, but these speeds have partly independent variances. Then C's conjecture point should precede the completion point roughly 50% of the time and follow it the other 50%. Whenever C understands before the completion point, she is in the clear, though she still cannot check on her understanding. But whenever her conjecture comes *after* the completion point, she is at a disadvantage, and for two reasons. First, she is receiving no more information to help her understand. And second, she has to continue processing the last utterance while trying simultaneously to listen to the next one. This should interfere and lead to mistakes on both the last and the current utterances.

The collaborative model leads to several predictions. First, overhearers should have greater difficulties understanding than addressees even when they are equal in background knowledge. The autonomous view predicts no difference. Second, misunderstandings by overhearers should increase dramatically whenever their conjecture point comes after the completion point. About this the autonomous view makes no obvious prediction. And third, overhearers should have even more difficulties understanding when they do not share all the background knowledge of the participants. The autonomous view makes the same prediction. There is also a possible fourth prediction, consistent with both views. Overhearers who have some control over the pacing of the conversation (e.g., they are listening to a tape recording of the conversation and can stop it whenever they want) should have less trouble understanding than overhearers who have no control over the pacing. Those with control will have more time to process the critical information without interference from having to listen simultaneously to the next utterances. Experiment 1 was designed to test these predictions.

## EXPERIMENT 1

### *Method*

The experiment divided into two parts. In the first, 10 pairs of students who were not acquainted with each other carried out a task in which one of them, the *director,* talked with the other, the *matcher,* in order to get the matcher to arrange 12 figures in a particular order. Each pair repeated this task for six trials. In the second part, tape recordings of these conversations were presented to 40 overhearing matchers (from here on, *overhearers*), who

were to arrange the same 12 figures just as the matchers had done. Half of the overhearers listened to the conversations from Trial 1 on, and the other half, only from Trial 3 on. Half of each of these groups were allowed to use the "pause" button to stop the tape whenever they wanted, and the other half were not. The two principal measures of understanding were accuracy and time of placement of the correct figure.

The task in the first part was a version of a communication task originally devised by Krauss and Glucksberg (Krauss & Weinheimer, 1964, 1966, 1967; Krauss & Glucksberg, 1969, 1977; Glucksberg, Krauss, & Higgins, 1975; see also Asher, 1979; Clark & Wilkes-Gibbs, 1986). In our version, the two students were seated at tables on either side of a barrier so that they could not see each other. (Although both sexes took part in all roles in the experiment, we will refer to the director as female, the matcher as male, and the overhearers as female). In front of the matcher on the table was a set of 16 cardboard cards, each displaying a different figure on it (see Fig. 1); the figures were black paper cutouts of Tangram figures (Elffers, 1976) or altered versions of Tangram figures. Also in front of the matcher was a cardboard sorting frame with 12 spaces numbered from 1 to 12. In front of the director was a sheet of paper with the matcher's same 16 figures photocopied onto it in a random order. The first 12 of the 16 figures on the director's sheet were numbered from 1 to 12.

The students were told that the director's job was to get the matcher to place his cards on the sorting frame in the correct order—the director's order—as quickly and as accurately as possible. They could talk with each other as much as they wanted. The director was to go through the positions sequentially, starting with figure number 1 and ending with figure number 12. Only 12 out of 16 figures were used on a trial so that the director and matcher

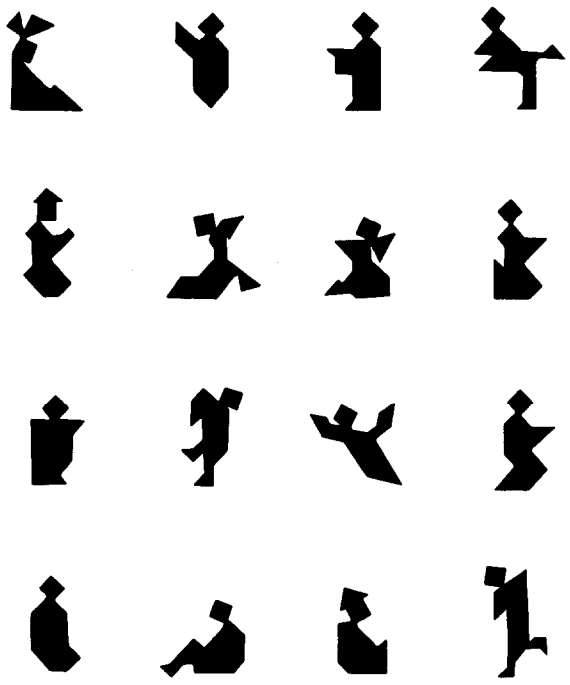

FIG. 1. The Tangram figures.

could not use process of elimination to arrange the last few on any given trial. Each pair played the game six times. The same figures were used in each trial, but the target matrix was in a different random order each time. During each trial, the students were timed and tape recorded. We noted the order in which the matcher put down the cards. The director and matcher were told that other students would listen to the tape recording, but they were not told why. There were 10 male and 10 female participants, with six mixed-sex pairs and two same-sex pairs for each sex.

In the second part of the experiment, 40 overhearers listened to all or part of one of the tape recorded conversations, and performed the matcher's task, sorting the cards, while listening to it. There were four overhearing groups. Half the overhearers heard the entire conversation, and the other half only heard Trials 3 through 6. In each of these groups, half were allowed to press a pause button on the tape recorder, and half were not. Each of the 10 conversations was heard by four different students, one in each of the four overhearing groups. As with the matchers, we noted the order in which the overhearers put down their cards, but we also noted whenever they put down a card noticeably before or after the speakers on the tape had verbally completed their placement. We tape recorded these sessions to be able to study the patterns and lengths of pauses. Of the 40 overhearers, 22 were female and 18 were male.

All participants were Stanford students, native speakers of American English, all but one of whom received Psychology 1 course credit. None of the directors and matchers knew each other before the beginning of the experiment. Only one of the 40 overhearers recognized either of the voices on the tape, so we can assume that the overhearers were not privy to any common ground beyond that which accumulated during the experiment.

### *Results*

*Collaboration.* Speakers in the first task followed the pattern of collaboration that Clark and Wilkes-Gibbs (1986) observed. The first time that figures appeared on the directors' sheets, the directors *described* them; from then on, they *referred* to them with definite descriptions, which got shorter and shorter as the trials progressed. On Trial 1, the two of them also tended to negotiate for several turns in placing each figure, but by Trial 6, they were down to one turn each. For example, the first time one pair saw one of the figures (second on the second row in Figure 1), the two of them had this exchange[1]:

D: Then number 12 . is (laughs) looks like a, a dancer or something really weird. Um . and, has a square head . and um, there's like, there's uh- the kinda this um .
M: Which way is the head tilted?
D: The head is . eh- towards the left, and then th- an arm could be like up towards the right?
M: Mm-hm.
D: *And . It's- *
M: *an- . a big* fat leg? *You know that one?*

[1] Overlapping speech in adjacent turns is enclosed in asterisks (see Svartvik & Quirk, 1980).

D: *Yeah, a big* fat leg.
M: and a little leg.
D: Right.
M: Okay.
D: Okay?
M: Yeah.

By the last trial, the reference was more compact, and the two of them took only one turn each:

D: Um, 12 . the dancer with the big fat leg?
M: Okay.

This final reference combines the perspectives offered by both the director and the matcher. In general, the perspectives that speakers ended up agreeing on ranged widely. By the last trial, the figure at the bottom left in Fig. 1 was variously referred to as "the rice bag," "the whale," "the complacent one," "the stretched-out stop sign," and "the baby in a straitjacket."

These patterns of collaboration were supported by the statistics. Figure 2 shows that the director and matcher each took fewer and fewer words over trials coming to agreement about each reference. Directors, who in their role spoke more, started at about 73 words per figure and ended at about 13; linear trend: $F(1,45) = 159.77$, $p < .001$. Matchers began with about 39 per figure and ended up at only about 3; linear trend: $F(1,45) = 98.54$, $p < .001$. The average number of turns the director spent discuss-

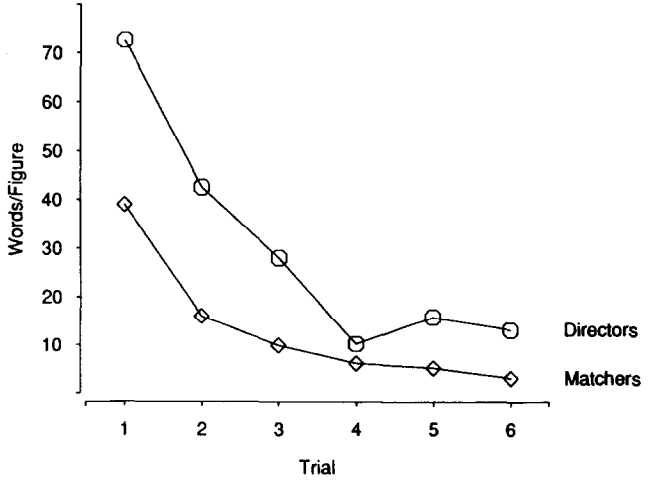


FIG. 2. Number of words which participants use to come to agreement about a figure's identity.

ing each figure decreased from 7.8 to 1.1 over the six trials; linear trend: $F(1,45) = 125.83$, $p < .001$. And the amount of time spent per figure, which is highly correlated with number of words spoken, dropped from about 39 s per position on Trial 1 to about 6 s per position on Trial 6; linear trend: $F(1,45) = 179.15$, $p < .001$. These results corroborate Krauss and Weinheimer's (1964, 1966, 1967) classic observations on repeated references.

*Accuracy of understanding.* Our principal test of the autonomous and collaborative views of understanding was based on accuracy, the percentage of figures placed correctly. By both views, overhearers who entered on Trial 1, whom we will call *early* overhearers, should be more accurate over all trials than those who entered on Trial 3 (*late* overhearers), and they were, 88 to 68%; $F(1,38) = 20.26$, $p < .001$. Even after four trials the late overhearers were not as accurate as early overhearers, $F(1,38) = 15.83$, $p < .001$. By the collaborative model, however, the matchers should do better than the early overhearers even though these overhearers heard every word uttered by the speakers. This is precisely what occurred, 99 to 88%; $F(1,28) = 10.51$, $p < .005$. That is, being witness to the buildup of common ground did not seem to provide all the necessary information for overhearers to understand the references as well as addressees. This is direct evidence against the autonomous view of understanding.

Both matchers and overhearers got more accurate as they went along; linear trend: $F(1,135) = 4.69$, $p < .025$. Figure 3 plots percentage correct

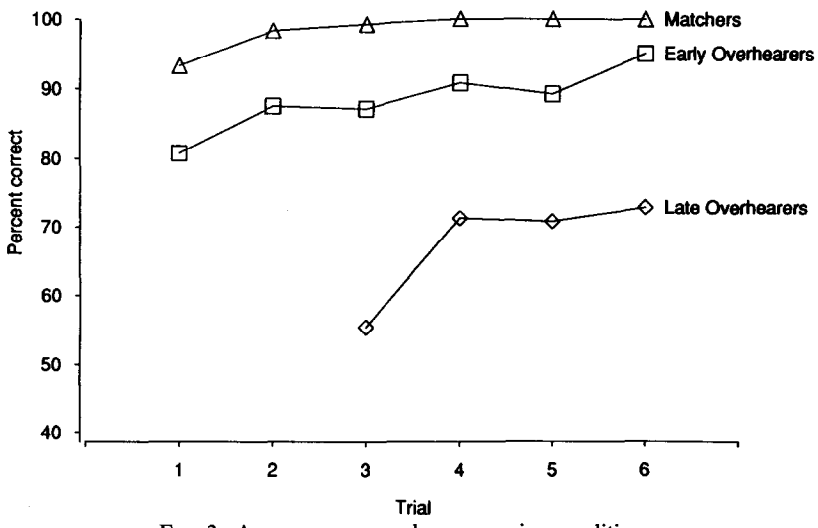


FIG. 3. Accuracy averaged over pausing conditions.

in each condition on each trial. As the figure shows, the matchers were very accurate from the outset. They averaged 93% correct on Trial 1 and were perfect from Trial 4 on; linear trend: $F(1,45) = 20.44$, $p < .001$. In contrast, early overhearers began at 81% correct and by Trial 6 increased to only 95%, linear trend: $F(1,95) = 17.13$, $p < .001$. Late overhearers also improved from 55 to 73%; linear trend: $F(1,57) = 15.41$, $p < .001$. Over the last four trials, they improved at a faster rate than the early overhearers; linear trend: $F(1,114) = 8.58$, $p < .01$. There are several possible reasons why they improved more quickly, but we have no evidence to choose among them.

Every overhearer given the opportunity to pause *did* pause at least once during the course of the experiment. They paused from 0 to 7 times during a trial, and their pauses ranged from 2 to 79 s in length. Still, all this pausing did not help. The pause and no pause conditions yielded 89 and 88% correct responses for the early overhearers, and 73 and 68% for the late overhearers; these differences were not significant. After the experiment, most overhearers thought the opportunity to pause had helped (even though it had not). In some cases, pausing actually seemed to hurt, because when the conversation became very quick, it was easy for the overhearer to lose her place on the sorting board, and end up missing descriptions. Because the overhearers' performance did not improve even when they had unlimited time to think about the speakers' words, we know that they did not do poorly simply because they had not happened to find a card soon enough whose identity they were already sure about.

*Procedural differences.* In interpreting this task, we assume that the moment at which the matcher or overhearer placed a card corresponds more or less to their recognition or conjecture points—the points at which they believed they had identified the referent. The assumption, of course, is not quite right, because different people surely have different strategies for marking their beliefs and different criteria for making their guesses. The assumption is probably even too strong, because overhearers may put down cards with less confidence than addressees. If anything, however, this works against predictions of the collaborative model.

By the collaborative model, overhearers should put some of their cards down considerably before matchers, and some considerably after, and they did. Table 1 shows that overhearers placed only 63% of their cards down by the speakers' verbal completion points, as compared with the matchers' 99%; $t(94) = 6.76$, $p < .001$. Overhearers made their placements during the director's description of the next figure ("late") or even after that ("very late") 31% of the time, as compared with the matchers' rate of 1%; $t(94) = 5.63$, $p < .001$. And they left blanks in their frames 6% of the time, whereas matchers never did. Early overhearers gave more

TABLE 1
Time and Accuracy of Placements Relative to Speakers' Completion Points

| | Early | On time | Late | Very late | Blank | Total |
|---|---|---|---|---|---|---|
| | Percent placements in each time period | | | | | |
| Matchers | 0 | 99 | 0.4 | 0.6 | 0 | 100 |
| Early overhearers | 15 | 56 | 20 | 7 | 2 | 100 |
| Late overhearers | 6 | 49 | 25 | 10 | 10 | 100 |
| | Probability of error in each time period | | | | | |
| Matchers | — | .02 | .00 | .20 | | |
| Early overhearers | .03 | .05 | .16 | .35 | | |
| Late overhearers | .11 | .20 | .31 | .55 | | |

"early" or "on time" responses than late overhearers, 71 to 55%; $t(94) = 3.36$, $p < .002$.

For the collaborative model, the important prediction is that late placements are more likely to be incorrect. That prediction was confirmed. As Table 1 shows, overhearers were more likely to make an error when they placed a card *after* the speaker's completion point than before it, 28 to 12%; $t(47) = 7.38$, $p < .001$. (The seemingly high "very late" error rate of 20% among the addressees represents one error in only five card placements.)

The matchers amd overhearers also sometimes changed the cards they put down. They would place a card in the sorting frame and later replace it with another, indicating they had changed their choice for the correct referent. On the collaborative view, overhearers should do this more often than matchers, simply because overhearers cannot verbally test out their hypotheses; and they did, 5.1 to 1.6% of the time; $t(38) = 4.26$, $p < .0001$. The matchers' changes were always to correct an error, but only 56% of the overhearers' changes were to the correct referent. Still, among the answers they did get right, both matchers and overhearers overwhelmingly got them right on the first try (as opposed to changing their answers from incorrect ones): 97% for participants, and 96% for overhearers. For the responses that overhearers ultimately got wrong, the rate of changing, 32%, was much higher than their overall rate of 5.1%.

Overhearers distributed their errors fairly evenly among positions on the sorting frame within trials, so they did not have more trouble within a trial as it proceeded, linear trend: $F(1,429) = 0.49$, *n.s.* Getting behind early on did not cause them to do worse on later positions; rather, they tended to try to concentrate on the current reference, only going back to previously missed references when they could. And as expected, some figures were harder for overhearers than others. The percentage of errors ranged from 3.3% on the second Tangram on the bottom row of Figure 1

to 29% for each of the middle two figures on the second row. The differences among figures were significant; $F(15,630) = 5.96$, $p < .001$, but could have arisen for any number of reasons.

*Subjective commentary.* Overhearers often muttered to themselves during trials and made comments to the experimenter between trials. During pauses, several of them repeated the descriptions that the directors had just made, for example, "Triangular shape to the right with rabbit ears . . . triangular shape to the right with rabbit ears . . . ," or "Facing right . . . foot facing right." They demonstrated verbally what we already knew from their changes, that they kept descriptions they had not understood in mind and went back to work on them when they had a chance. One overhearer paused during the description of the 12th position and muttered "What kind of an animal?" repeating the description of the 7th position, which she had left blank. Another overhearer, who had had trouble with "the Number One," exclaimed when she finally figured it out "So *that's* the Number One thing!"

Overhearers clearly realized they did not understand some references. One said "I have trouble with this one" when she heard a rather obscure description for the second time. Another paused the tape after an unusual description, muttering, "Wait, Hoover Tower figure, with the tray on the left . . . Geez, I dunno." When the same figure was described in a later trial, she announced in annoyed tones "I don't *know* the Hoover Tower figure!" Another expressed dismay at not understanding a reference: "I don't know which one-monk, they're talking about." Several overhearers wanted to rewind the tape to remind themselves of descriptions they had forgotten. The point is, if these listeners had been addressees, they would have cleared up these failures before the conversation went on. As overhearers without the opportunity to collaborate, they could not.

One overhearer assumed that over time she would be able to understand the conversation fully, but found that she was not: "It's harder—I thought it'd be easier, 'cuz they just say, y'know, '*this* one'." Another overhearer got right to the point in explaining why he was having trouble: "I think if—it's like if I was more a part of the thing, then it would be . . . They're just talking to each other, and using their definitions."

*Discussion*

Our results so far are clear. Overhearers who did not witness the buildup of common ground between conversational participants understood fewer references than the participants themselves. But so did overhearers who *did* witness the buildup (for similar findings, see Kraut et al., 1982). Nor did it help overhearers to be able to control the pacing of the conversation by pausing the tape. Overhearers appeared to use their time differently from participants. They made more guesses, and they made

these over a broader distribution of time. Although the first of our results is consistent with the autonomous view of understanding, the remainder are not. All these results are consistent with the collaborative model.

Still, we were not entirely satisfied with Experiment 1. First, there was an alternative explanation for why the overhearers did so badly in Experiment 1. Listening to a tape recording of a conversation just is not as vivid, as engaging, as easy as listening to the conversation live. That might have been the reason overhearers were worse off. Second, we had only a crude test of the timing predictions of the collaborative model. In Experiment 1, we were able to classify the matchers' and overhearers' responses only as early, on time, late, very late, or missing. That did not allow a very sensitive test of these predictions. Experiment 2 was designed to overcome both of these problems. We had overhearers listen to live conversations, and we videotaped and later timed the matcher and overhearer as they placed their cards in the sorting frames.

## EXPERIMENT 2

### *Methods*

The task was the same as in Experiment 1. Fourteen pairs of students played the Tangram-matching game six times with the figures in a different random order each time. Once again, only 12 out of 16 figures were used on each trial. This time an overhearer was present in the room along with the director and matcher, all separated by visual barriers. A single hidden video camera was trained on the sorting boards in front of the matcher and overhearer and recorded the movements of their hands and the cards. And once again, the director, matcher, and overhearer began as strangers.

Running the experiment with all three people in the room caused a logistics problem. We had to ensure that the director and matcher, who knew there was another person in the room, would be indifferent to her in the design of their utterances. The excuse we fashioned was that she was a coder present to reduce experimental bias. We explained this to the director and matcher when we read them the instructions. The overhearers, therefore, had to come early to hear their instructions. This meant that they also listened to all the instructions given to the director and matcher. This gave them more time to think about the task, but any possible advantage this gave them would work against our hypothesis. The matcher and overhearer sat at exactly the same distance from the director, and their cards, apparatus, and surroundings were otherwise identical.

After the experiment, all three students filled out questionnaires. They were asked whether they had previously known any of the other students in the room, how difficult they had found the task and why, and what they thought the experiment was testing. The last question was designed to determine whether the director and matcher had been aware of the overhearer's role, and none had. The students were members of the Stanford community, all native speakers of English, who were either paid or received experimental credit for participating. One triple of students was eliminated because they had known each other before the experiment, another because they failed to follow directions, and 2 more because the director and matcher were still making two or more errors out of 12 on the sixth trial. These last 2 triples are highly unusual for experiments of this kind, where matchers typically make no errors by the sixth trial (see Experiment 1; Clark & Wilkes-Gibbs, 1986; Isaacs & Clark,

1987; Clark & Schaefer, 1987b). Removing these 2 triples, however, does not alter the main results. The main results, then, are based on 10 triples.

### *Results*

*Accuracy.* By the collaborative model, the matchers should be more accurate than the overhearers, and they were, 98 to 85% correct; $F(1,18) = 10.83$, $p < .005$. (When we included the two error-prone triples, the matchers were still more accurate, 95 to 85% correct; $F(1,22) = 7.00$, $p < .015$.) Matchers started out with 95% correct on Trial 1, and, by Trial 6, they all matched every reference correctly. In contrast, overhearers started out with only 78% correct and only improved to 89% by the last trial. (With the two error-prone triples included, overhearers started out with 80% correct and ended up at 91% by the last trial). Both matchers and overhearers improved over the course of the trials; linear trend: $F(1,90) = 10.89$, $p < .01$, but overhearers improved a bit more than the participants; linear trend: $F(1,90) = 7.01$, $p < .01$, perhaps because the participants started out so well. Figure 4 summarizes these results. As in Experiment 1, overhearers changed their minds about the cards they put down more often than matchers, 4.2 to 1.3% of the time; $t(9) = 3.99$, $p < .005$. In short, it did not help overhearers to listen to the conversations live. They still did not do as well as the matcher. This is decisive evidence against the autonomous view of understanding in conversation.

*Placement times.* The director and matcher engaged in the same collaborative strategies as in Experiment 1, at first describing figures and then referring to them with increasingly abbreviated references. Because we videotaped the matchers and overhearers, we were able to time the initiation points, completion points, and card placements to the nearest

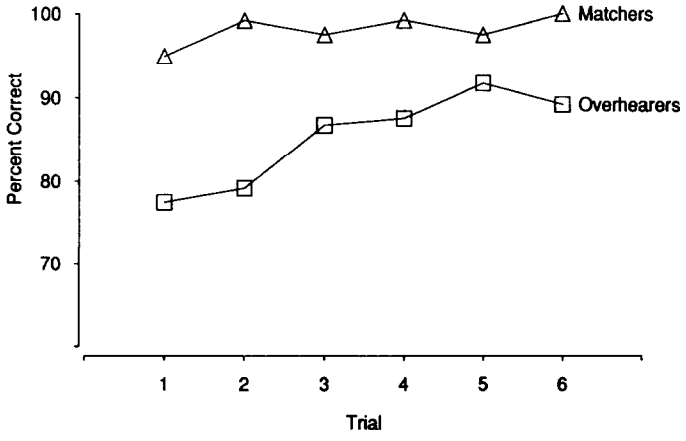


FIG. 4. Accuracy on the matching task.

10th of a second. The first measure we calculated was what we will call *placement times*. The placement time for a matcher or overhearer on a figure was the time duration from the initiation point for that figure (the moment the director began talking about it) to the placement point (the moment the matcher or overhearer put his or her final card choice down). As in Experiment 1, the average placement time per figure got shorter and shorter over the six trials; linear trend: $F(1,45) = 97.38$, $p < .001$. Nevertheless, the placement times were almost the same for overhearers as for matchers; $F(1,9) = 0.16$, *n.s.* On Trial 1, the mean placement time per figure was 31.1 s for matchers and 30.2 s for overhearers. By Trial 6, these times were down to 6.3 and 8.3 s (see Fig. 5). Nor were there reliable differences between matchers and overhearers in the *median* placement times per figure.

Next we compared each placement point (the moment a matcher or overhearer made his or her final card placement) relative to the completion point (the point at which the director and matcher made it clear verbally that they were ready to go on to the next figure). By the collaborative model, the matcher's placement points should ordinarily come just before the completion points, and on the average they did, by 0.5 s. The matcher and director should not be willing to go on until the matcher had signaled he was able to put down his card. But on the average the overhearer's placement points also came just before the completion points, by 1.2 s. These two differences are not reliably different.

The story was very different, however, for the standard deviation of these placement-completion intervals. As predicted, overhearers placed their cards down both much earlier *and* much later than matchers. So

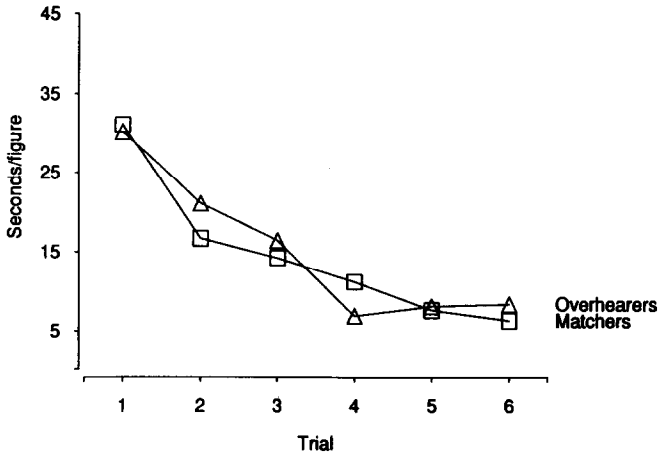


FIG. 5. Time until card placement.

while the mean placement-completion intervals were the same for matchers and overhearers, the standard deviation was about five times larger for overhearers than for matchers, $F(1,18) = 12.64, p < .005$. The average standard deviations of these intervals for matchers and overhearers are plotted over trials in Fig. 6. The difference between the two standard deviations held up even when we removed null responses (to which we had assigned a time corresponding to the end of the entire trial); $F(1,18) = 9.26, p < .007$. Another way of looking at the placement-completion interval is to correlate the placement time for each figure on each trial with the corresponding completion time. On the average, this correlation was .96 for matchers but only .68 for overhearers; $t(11.5) = 3.21, p < .01$. This means that matchers tended to put their cards down at about the same time as they finished establishing each reference with their partners, while overhearers did not track the participants' completion points so closely.

*Pacing and accuracy.* By the collaborative model, pacing should have direct effects on the overhearers' accuracy. Overhearers who have understood a reference before the completion point will be prepared for the next reference, whereas those who have not will have to contend with the next reference while trying to complete understanding of the last one. One prediction, then, is this: Overhearers are more likely to be incorrect on those placements that follow the completion point than on those that precede it. And this was the case. They made errors 9.6% of the time on placements before the completion point, but 21% of the time on placements after the completion point; $F(1,9) = 15.06, p < .005$.

For more detailed evidence for this prediction, we ranked all 120 over-

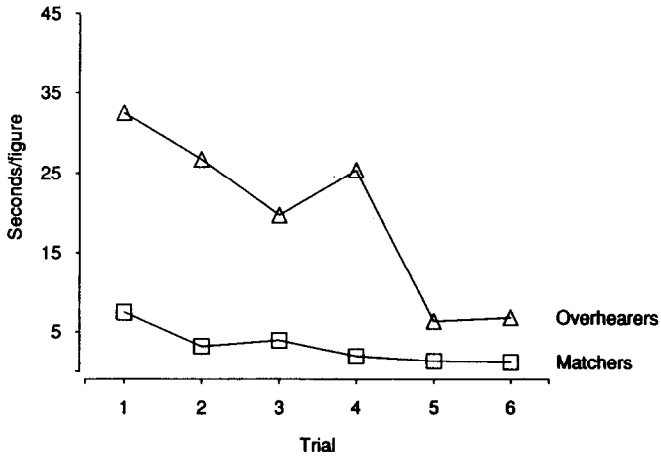


FIG. 6. Standard deviations of card placements.

hearer placements within Trial 1 from earliest to latest relative to the completion points. We have plotted these by deciles in Fig. 7. As the figure shows, overhearers were very accurate (100% correct) when their placements were very early relative to the completion points. They became a bit less accurate as their placements approached the completion points, but their accuracy dropped precipitously once their placements came after the completion points. They were correct on only 42% of their placements in the latest decile. This, then, is further striking evidence for the collaborative model.

Another possible consequence of the collaborative view is that a late placement will compete specifically with the very next placement. We looked to see whether overhearers would do worse for placements that directly followed late placements than for those following early or on-time placements, and we found no difference in accuracy at all, 86 to 85%, *n.s.* As we noted in the first study, overhearers tended to concentrate on the reference currently under discussion and went back to a previous figure only when they had understood the current one.

Directors, matchers, and overhearers, of course, should differ in their effectiveness in communicating. Some pairs of directors and matchers should be more effective as conversational partners than others, and their choice of perspectives and their pacing should be right for some overhearers and confusing for others. By the collaborative model, these vari-

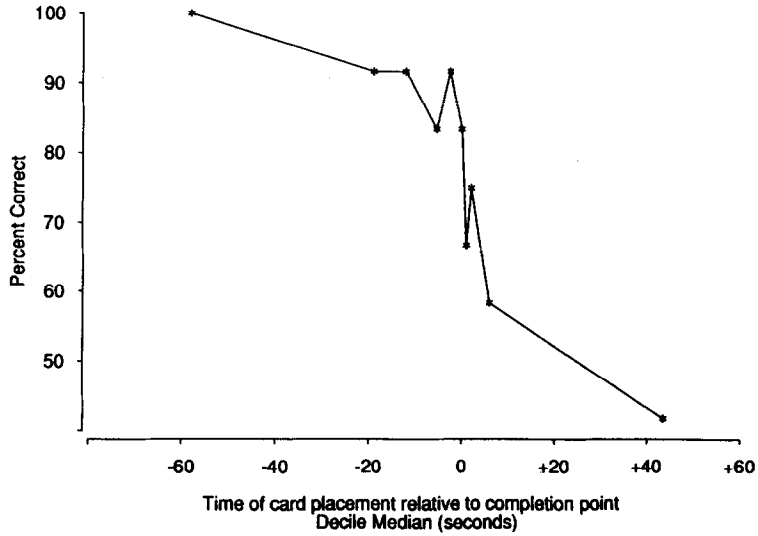


FIG. 7. Overhearers' accuracy on Trial 1, relative to conversational partners' completion points.

ations should affect the overhearers' accuracy. The speedier a director and matcher are compared to the average director and matcher, the less time they give an overhearer to grasp each reference, the more likely she is to make her placement late, and the more likely she is to make an error. This was in fact borne out. On Trial 1, the matcher's mean time of card placement correlated -.78 with the median difference between the overhearer's and matcher's placement times; $t(8) = 3.54$, $p < .005$, one-tailed. The overhearer's number of errors on Trial 1 also correlated .58 with the median difference between the overhearer's and matcher's placement times; $t(8) = 2.01$, $p < .05$, one-tailed.

*Subjective comments.* On the questionnaires, directors, matchers, and overhearers alike commented on the collaborative nature of the task. Many directors and matchers noted that the task had been difficult until they had established "common names for referring to the figures," "a 'vocabulary' for the figures," "familiar ideas about what they [the shapes] could represent," or " 'names' coding the figures," or until they "were on the same wavelength, spoke common descriptor-language." It was precisely this vocabulary that overhearers often complained about. They noted that the director and matcher sometimes "had a name for a figure which I didn't remember or which described it in a way not clear for me," or "agreed on terms that I didn't quite catch," or "developed their own terms and it took a while to catch on to their terminology." They realized they were at a disadvantage because they "couldn't communicate with the players," "couldn't ask questions to clarify some of the shapes," and "disagreed with a few of the interpretations given for the symbols and I kept wanting to add my input!"

Overhearers were also aware that pacing was a problem. One was "forced to rush on to the next figure whether or not I'd gotten the previous one." Another found it "difficult to keep up with the other two . . . occasionally I would fall behind and forget what needed to be filled in." For another, "sometimes the matcher found her shape before I found mine and I missed some of the next description." Another overhearer claimed that "at the end I knew the terms but they went very quickly." One described it succinctly: "it went too fast."

Many directors, matchers, and overhearers, then, recognized the very processes predicted by the collaborative model. Most overhearers realized that they were at a disadvantage because they were not part of the grounding process or in control of the pacing. But not all of them. One argued that "on the whole . . . it was good that I just *listened* instead of trying to describe *and* discern at the same time." She seemed to be echoing the autonomous view of understanding, suggesting that overhearers should understand at least as well as the participants. The data do not, of course, bear her out.

## CONCLUSIONS

Our results suggest that the *social* process of interacting in conversation plays a central role in the *cognitive* process of understanding. Listeners who participate in a conversational interaction go about understanding very differently from those who are excluded from it. It is because of these differences that addressees understand faster and more accurately than overhearers. If understanding in conversation were an autonomous process, there should be no such differences. The conclusion, then, is that understanding is part of a collaborative process (see also Kraut et al., 1982).

Our findings show that the process of understanding differs for addressees and overhearers in several ways. Consider speaker A, addressee B, and overhearer C.

*1. Collaboration.* B ordinarily collaborates with A as he tries to understand her, whereas C does not. One thing it means for B to collaborate with A is for B to monitor his understanding and keep A informed of the state of his understanding. If all goes well, B need only say "Okay" or "Yeah"—he has understood. But when he gets into trouble, he has to identify the trouble, describe it to A, and resolve it. C monitors *her* understanding as well, but she does not have to describe her troubles to anyone nor does she have the opportunity to enlist anyone's help. The only way she can resolve her problems is via conjectures based on her beliefs about A and B's common ground and what A said. So B reaches his final state of understanding in collaboration with A, whereas C has to do it all by conjecture.

*2. Criterion for understanding.* B's criterion for understanding is the belief that he and A mutually believe he has understood her well enough for current purposes. He can work until he has understood as well as he wants. C's criterion is a different matter. She can only reach the belief that she has understood as well as she *could*. She has no way of working until she has understood as well as she wants.

*3. Perspective.* Part of what B does in collaboration with A is search for a common perspective, a perspective shared with A on what A is trying to say. In our task that meant searching for a shared way of conceiving each figure—for example, as a rice bag, a whale, a stretched-out stop sign, a baby in a straitjacket, or a complacent one. B can even introduce his own perspective, as long as he gets A to agree to it. C, in contrast, is forced to accept whatever perspective A and B throw her way. If it is not a perspective she can grasp, that is her tough luck.

The two types of listeners in our experiments reflected these differences. As for collaboration, the matchers did not sit idly by as the directors described the figures. They actively collaborated with them from the

very start. The process was lengthy at first, but soon became very efficient. As for the criterion of understanding, the matchers almost always committed themselves by placing their cards right at the completion points of the directors' assertions. They worked until they were satisfied they understood, and they were almost always right. The points at which the overhearers placed their cards were not so closely tied to the completion points, and their criteria for placing cards were lower. They often changed their minds, and they were often wrong. They were forced to accept a lower criterion simply because the crucial resource—the grounding process—was not available to them. Finally, the matchers worked hard to find perspectives they could share with the directors. Ironically, these were often the same perspectives that the overhearers complained that they could not grasp and that kept them from identifying the right figures.

Why does collaboration leave overhearers at a disadvantage? It is known that speakers accommodate to their particular interlocutors in everything from loudness and speed to dialect and pronunciation (Bell, 1984; Giles, Mulac, Bradac, & Johnson, 1987; Street & Giles, 1982; Thakerar, Giles, & Cheshire, 1982). But factors like these do not seem able to account for our findings. Speakers also adjust to the expertise of their interlocutors, often supplying them with or acquiring from them the needed expertise as they talk (Isaacs & Clark, 1987). Discrepancies in expertise per se do not offer a plausible account of our findings either since our matchers and overhearers began on a par in expertise. We must look instead at the heart of collaboration, the process of grounding.

Grounding is really an opportunistic process. It succeeds in part by exploiting adventitious commonalities between speakers and addressees. In our task, A offers one way of viewing a figure—say, as a whale—and if B happens to be able to see it that way, he accepts it, and they go on. If he cannot see it that way, the two of them try another perspective. The process is opportunistic in that it takes advantage of the first perspective A and B find they can agree on. If it is a perspective C can grasp, she is in luck, but if it is not, she is likely to fail. Her state of understanding plays no role in A and B's decision to stop or go on. With an opportunistic process like this, C is at a disadvantage, and the damage may accumulate.

Do these findings apply to other types of conversation? In their essential features, the answer must be yes. Grounding has been documented to be a central process in ordinary English conversation (Clark & Schaefer, 1989), telephone calls to directory enquiries (Clark & Schaefer, 1987b), and a variety of task-oriented conversations (Clark & Wilkes-Gibbs, 1986; Garrod & Anderson, 1987; Isaacs & Clark, 1987). What we have demonstrated is that listeners who participate in grounding have an advantage over those who do not. So wherever grounding occurs, addressees should

have an advantage over overhearers, all else being equal. In real conversations, particular speakers always bring particular perspectives to bear upon particular topics. Understanding can only be guaranteed for listeners who actively participate in establishing these perspectives.

Understanding by addressees is rarely studied in experiments on comprehension, because in most of them the participants are treated as if they were overhearers. First, the participants are made to listen to tape-recorded speech in isolation from the speaker. They have no way of influencing the pace or form of the speaker's utterances, as addressees normally do. They do not have to prepare to speak while listening, or identify and make their misunderstandings known to the speaker. Second, the speech they hear is sanitized. Rarely does it have repeats, *uhs,* interruptions, mistimings, or self-corrections, all of which change the very nature of parsing. At the same time, it does not make these features available for the participants to exploit in the grounding process, as addressees normally do (see Clark & Schaefer, 1989). And third, because there is no interaction between speaker and listener, the grounding criterion is not even definable. The participants have to be satisfied with a lower, and different, standard of understanding.

Understanding, in short, can never be fully captured in traditional theories of understanding. These theories, like the experiments they depend on, assume that listeners do what they do autonomously—that is, without direct collaboration with the speaker. If we are right, these theories will have to be revised to deal with the grounding process. Just how radical the revisions must be remains to be seen.

## REFERENCES

Anderson, J. R. (1985). *Cognitive psychology and its implications* (2nd ed.). New York: W. H. Freeman.

Asher, S. R. (1979). Referential communication. In G. J. Whitehurst & B. J. Zimmerman (Eds.), *The functions of language and cognition* (pp. 175–197). New York: Academic Press.

Bell, A. (1984). Language style as audience design. *Language in Society,* **13,** 145–204.

Clark, H. H. (1985). Language use and language users. In G. Lindzey & E. Aronson (Eds.), *Handbook of social psychology* (3rd ed., pp. 179–231). New York: Harper and Row.

Clark, H. H., & Carlson, T. B. (1982). Hearers and speech acts. *Language,* **58**(2), 332–373.

Clark, H. H., & Haviland, S. E. (1977). Comprehension and the given-new contract. In R. O. Freedle (Ed.), *Discourse production and comprehension* (pp. 1–40). Hillsdale, NJ: Erlbaum.

Clark, H. H., & Marshall, C. R. (1981). Definite reference and mutual knowledge. In A. K. Joshi, B. Webber, & I. A. Sag (Eds.), *Elements of discourse understanding* (pp. 10–63). Cambridge: Cambridge Univ. Press.

Clark, H. H., & Schaefer, E. F. (1987a). Collaborating on contributions to conversation. *Language and Cognitive Processes,* **2,** 19–41.

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

Clark, H. H., & Schaefer, E. F. (1989). Contributing to discourse. *Cognitive Science,* in press.

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

Elffers, J. (1976). *Tangram. The ancient Chinese shapes game.* New York: McGraw–Hill.

Fodor, J. A. (1983). *Modularity of mind.* Cambridge, MA: MIT press.

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

Gazdar, G. (1979). *Pragmatics: Implicature, presupposition, and logical form.* New York: Academic Press.

Giles, H., Mulac, A., Bradac, J. J., & Johnson, P. (1987). Speech accommodation theory: The first decade and beyond. In M. L. McLaughlin (Ed.), *Communication Yearbook 10* (pp. 13–48). Beverly Hills: Sage.

Glucksberg, S., Krauss, R. M., & Higgins, E. T. (1975). The development of referential communication skills. In F. E. Horowitz (Ed.), *Review of child development research* (Vol. 4, pp. 305–345). Chicago: Univ. of Chicago Press.

Goffman, E. (1976). Replies and responses. *Language in Society,* **5,** 257–313.

Goodwin, C. (1981). *Conversational organization: Interaction between speakers and hearers.* New York: Academic Press.

Isaacs, E. A., & Clark, H. H. (1987). References in conversation between experts and novices. *Journal of Experimental Psychology: General,* **116**(1), 26–37.

Johnson-Laird, P. N. (1983). *Mental models.* Cambridge, MA: Harvard Univ. Press.

Krauss, R. M., & Glucksberg, S. (1969). The development of communication: Competence as a function of age. *Child Development,* **40,** 255–256.

Krauss, R. M., & Glucksberg, S. (1977). Social and nonsocial speech. *Scientific American,* **236,** 100–105.

Krauss, R. M., & Weinheimer, S. (1964). Changes in reference phrases as a function of frequency of usage in social interaction: A preliminary study. *Psychonomic Science,* **1,** 113–114.

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

Krauss, R. M., & Weinheimer, S. (1967). Effect of referent similarity and communication mode on verbal encoding. *Journal of Verbal Learning and Verbal Behavior,* **6,** 359–363.

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.

Lewis, D. K. (1979). Scorekeeping in a language game. *Journal of Philosophical Logic,* **8,** 339–359.

McGregor, G. (1986). Listening outside the participation framework. In G. McGregor & R. S. White (Eds.), *The art of listening* (pp. 55–72). London: Croom Helm.

Schegloff, E. A. (1982). Discourse as an interactional achievement: Some uses of uh-huh and other things that come between sentences. In D. Tannen (Ed.), *Analyzing discourse: Text and talk. Georgetown University Roundtable on Languages and Linguistics 1981* (pp. 71–93). Washington, D.C.: Georgetown Univ. Press.

Schegloff, E. A., Jefferson, G., & Sacks, H. (1977). The preference for self-correction in the organization of repair in conversation. *Language,* **53,** 361–382.

Sperber, D., & Wilson, D. (1986). *Relevance.* Oxford: Basil Blackwell.

Stalnaker, R. C. (1978). Assertion. In P. Cole (Ed.), *Syntax and semantics: Pragmatics* (Vol. 9, pp. 315–332). New York: Academic Press.

Street, R. L., Jr., & Giles, H. (1982). Speech accommodation theory: A social cognitive

approach to language and speech behavior. In M. E. Roloff & C. R. Berger (Eds.), *Social cognition and communication* (pp. 193–226). Beverly Hills: Sage Publications.

Svartvik, J., & Quirk, R. (1980). *A corpus of English conversation*. Lund, Sweden: Gleerup.

Thakerar, J. N., Giles, H., & Cheshire, J. (1982). Psychological and linguistic parameters of speech accommodation theory. In C. Fraser & K. R. Scherer (Eds.), *Advances in the social psychology of language*. New York: Cambridge Univ. Press.

van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York: Academic Press.

(Accepted October 24, 1988)