

The role of arousal in memory for conversation

BRIAN MacWHINNEY

Carnegie-Mellon University, Pittsburgh, Pennsylvania 15213

and

JANICE M. KEENAN and PETER REINKE
University of Denver, Denver, Colorado 80210

In the first **experimental** study of memory for natural **conversation**, Keenan, MacWhinney, and Mayhew (1977) found that, even after 30 h, subjects had extremely good recognition memory for the exact wording of statements that contained information about a speaker's "intentions, beliefs, and his relations with the listener." Such sentences were said to be high in "interactional content." One possible interpretation of the results of Keenan et al. is that it is the immediate affective response to an utterance, rather than its interactional content, that increases its memorability. In the present study, the strong relationship between interactional content and memory found by Keenan et al. was replicated. Subjects showed excellent recognition memory for high interactional content statements from a conversation, even after a 72-h interval. However, there was very little relationship between arousal, as determined by subjects' electrodermal response (EDR), and subsequent memory. Moreover, involvement had its greatest effect not on memory, but on subjects' arousal as measured by EDR.

Keenan, MacWhinney, and Mayhew (1977) conducted the first experimental study of memory for natural conversation. Their study was motivated by the observation that sentences spoken in natural conversation convey two fundamentally different types of information: propositional information and pragmatic information. Propositional information was characterized as information about the objects and events referred to by the sentence. Pragmatic information was defined as information about the context of the sentence in the social interaction, including information about the speaker and "his intentions, beliefs, and his relations with the listener." Sentences that conveyed a great deal of pragmatic information referencing the speaker-listener interaction were said to be rich in "interactional content." Keenan et al. hypothesized that sentences with high interactional content would be highly memorable because "subsequent interactions with the speaker often depend upon the contents of earlier interactionally significant events" (p. 551). More important, they predicted that listeners would tend to remember the exact surface wording of

This research was partially supported by National Institute of Education Grant NIE-G-78-0173 to J. Keenan. It is based in part on a master's thesis submitted by Peter Reinke to the Psychology Department of the University of Denver. Joseph Campos graciously made available to us the use of his laboratory, facilities, and equipment. He also provided us with valuable technical assistance and advice regarding the psychophysiological measurements. Without his help, the study could not have been conducted. We also wish to thank Suzanne Gendreau for testing the subjects in the control study and Polly Brown for conducting the statistical analyses of the data in the control study. Requests for reprints should be sent to Brian MacWhinney, Psychology Department, Carnegie-Mellon University, Pittsburgh, Pennsylvania 15213.

high-interaction sentences because it is often the particular choice of words made by the speaker that is the carrier of interactional information.

To test these predictions, Keenan et al. (1977) tape-recorded a lecture and conversation at a luncheon research group meeting in the Psychology Department at the University of Denver. In this meeting, Brian MacWhinney discussed a variety of issues in psycholinguistics with a contentious and sometimes boisterous audience made up of both students and faculty. The results of a recognition memory test composed of sentences taken from that meeting showed that sentences with high interactional content were remembered twice as well as sentences with low interactional content. Furthermore, the surface form, that is, the exact wording used by the original speaker, was remembered three times better for high-interaction sentences than for low.

Keenan et al. (1977) argued that subjects remembered the surface forms of high-interaction statements so well because these surface forms carry information that is relevant to the subsequent dynamics of the speaker-listener relationship. Such surface devices as hesitation placement, lexical choice, intonation contour, paralinguistic markers, and constituent ordering all provide a rich data base from which inferences regarding the speaker-listener relationship can be drawn. By attending to the surface form and maintaining this information in memory, at least until the next encounter with the speaker, the listener can more accurately assess his/her sociolinguistic status vis-a-vis the speaker and his/her interactional commitments toward the speaker.

An alternative interpretation of the results of Keenan et al. (1977) is that it is not so much that pragmatic con-

tent makes sentences memorable as that high pragmatic content induces a high arousal level that, in turn, produces good memory. According to this interpretation, high-interaction statements elicit higher levels of attention and arousal at encoding than low-interaction statements do. Because the results of list-learning experiments show that higher levels of arousal at encoding lead to higher levels of memory performance on tests of long-term retention (see Craik & Blankstein, 1975, for a review), it follows that high-interaction statements would be more memorable than low-interaction statements. We refer to this interpretation as the arousal effect.

There are good a priori reasons to expect the arousal effect to be as strong as, or stronger than, the interactional content effect. For example, Zajonc (1980) argued that human beings can often generate direct affective responses to stimuli even before those stimuli have been categorized and evaluated cognitively. Moreover, he held that these immediate affective responses may be important determinants of subsequent cognition. In regard to natural conversation, an analysis like Zajonc's would lead us to believe that affect is the crucial determinant of memory for sentences, because "affect dominates social interaction and it is the major currency in which social intercourse is transacted" (Zajonc, 1980, p. 155).

There is also extensive neurophysiological evidence for a biological substrate connecting arousal and memory. For example, Kety (1970, p. 329) showed that norepinephrine and certain other amines play a central role in the consolidation of memory for experiences that are accompanied by high levels of arousal. Pursuing Kety's analysis, Pettigrew (1978) noted that the locus coeruleus, with its connections to all of the major areas of the brain, might serve as a bridge between the overall arousal registered in the brainstem and the consolidation of particular memory traces in diverse cortical and subcortical areas. In Pettigrew's words, "it is easy to see that it would have made little sense for one of the brains of our predecessors to have gone on endlessly memorizing the details of contours of the ceiling of his lair if he were lying there safe, sound, and satiated; but that it *would* have made evolutionary sense to record even apparently 'trivial' sensory details associated with a life-threatening encounter" (p.73).

Brown and Kulik (1977) have shown that such arousal-based "flashbulb memories" exist in human beings not only for life-threatening encounters but also for numerous other moving experiences, such as the moment of first hearing of the assassination of President Kennedy in 1963. They argued that arousal is a sine qua non for the creation of such memories; however, they also noted that further consolidation and retrieval of these snapshots may depend not so much on arousal as on the importance of the consequences for subsequent interactions.

One of the major goals of the present study is to compare these two accounts by measuring the relative effects of immediate physiological arousal as opposed to interactional content in determining memory for sentences spoken in natural conversation. The study will examine whether, for a given set of high and low interactional content statements, recognition memory for sentences is proportional to the amount of arousal that occurs when a listener first hears the sentence in a conversational setting. The study will examine the effects of arousal on memory for both surface form and content, although the crucial test will be the effect of arousal on memory for the surface form.

Degree of arousal during listening will be measured by the magnitude of the subject's phasic electrodermal response (EDR), also referred to in the literature as galvanic skin response (GSR). Because EDR bears a monotonic relation with the subjective intensity of the stimulus (Lykken, 1968), this measure is one that has been used in almost all previous studies of the relation between arousal and memory.

The second major goal of the study is to assess the degree to which the effects reported in Keenan et al. (1977) depend on a subject's direct personal involvement with the conversation. If the interactional content effect occurs only for subjects who were themselves participants in the interaction, then future studies of memory for natural conversation would have to confine themselves to the study of actual participants. However, it might be the case that nonparticipants who watch a videotape of a conversation become so fully involved in the conversation, albeit vicariously, that they also show higher memory for the surface form of high-interaction sentences.

If vicarious involvement could be demonstrated, it would mean that it should be possible to use videotapes rather than live conversations in future research on memory for natural conversation. This would then allow us to overcome some of the methodological problems inherent in the study of natural conversation. In particular, studies of live conversation do not permit the investigator any control over the statements that occur. This lack of stimulus control not only hampers attempts to deconfound variables, but it also severely limits the kinds of questions that can be researched.

Studies by Bates, Kintsch, Fletcher, and Giuliani (1980), Bates, Masling, and Kintsch (1978), and Kintsch and Bates (1977) have shown that subjects who do not directly participate in verbal exchanges but simply listen to classroom lectures or watch videotaped soap operas show above-chance recognition memory for certain elements in the surface structures of sentences. However, the levels of memory for surface form found in these studies did not approach those observed in Keenan et al. (1977). The present study examines the extent to which high interactional content can lead to superior memory for the surface form of whole sentences, even when the

interaction is viewed on videotape. If the effect of interactional content holds up, even for the nonparticipants, it may be possible to begin to study memory for natural conversation by using videotape as a stimulus.

In order to test for the effect of participation on memory for natural conversation, we studied three groups of subjects: (1) participants, subjects who knew each other and who directly engaged in natural conversation; (2) familiar nonparticipants, subjects who knew the participants but who did not themselves participate in the conversation and who listened to and watched a videotape of the conversation generated by the participants; and (3) nonfamiliar nonparticipants, subjects who neither knew the participants nor engaged in the conversation and who listened to and watched the discussion on videotape.

These three groups are viewed as varying along a general dimension of involvement: The participants are the most involved and the nonfamiliar nonparticipants are the least involved. Unfortunately, it was not possible to deconfound participation and familiarity in this design, because a conversation between nonfamiliar participants could hardly qualify as an instance of natural conversation. Thus the main comparison of interest is between the participants and the nonfamiliar nonparticipants. Although familiarity is partially confounded with participation, by comparing the performance of the familiar nonparticipants with the nonfamiliar nonparticipants, we can get some idea of the relative importance of familiarity as opposed to participation in memory for natural conversation.

EXPERIMENT 1

Method

Subjects. The participant and familiar nonparticipant groups were composed of graduate students from the clinical area of the department of psychology at the University of Denver. Four conversations, each involving three students plus the experimenter (P.R.), who was also a psychology graduate student, were recorded on videotape. Simultaneously, electrodermal responses were recorded on polygraph tracings. Each conversational group was made up of students from the same entering year. Thus, there was a 4th-year conversational group, a 3rd-year group, a 2nd-year group, and a 1st-year group. Because there were 3 subjects in each group, there was a total of 12 subjects in the familiar participants group.

The familiar nonparticipants group also contained 12 subjects. Each familiar nonparticipant was shown a videotape of the conversation in which the participants were from his own class. In other words, 4th-year familiar nonparticipants watched 4th-year participants, 3rd-year familiar nonparticipants watched 3rd-year participants, and so forth. The familiar nonparticipants had extensive social contact with the participants within the context of the numerous close interactions in the clinical program. This was true because all of the students within a given class spent a great deal of time with each other in supervision meetings, area meetings, and courses.

The nonfamiliar nonparticipants were 12 graduate students in the professional psychology program at the University of Denver. All were complete strangers to those they watched on videotape. However, in terms of clinical interests, academic ability,

age, social class, and so forth, the nonfamiliar nonparticipants were quite comparable to the other two groups of clinical graduate students.

Instructions for the conversation. The experimenter served as a moderator for each of the four conversations, explaining the use of the polygraph and promoting dialogue. He communicated the following instructions to each of the four groups: "I'm testing a theory I have about communication of affect among individuals in small groups. I'm measuring affect through the use of GSR. I can't really go into any more detail than that without jeopardizing the results I expect to get. Please don't try to do anything you wouldn't do if you were having a semiserious discussion with a group of your friends. Basically, I'd like a sort of representative sample of serious discussion, jokes, and messing around. It is, however, important that there be only one person speaking at a time, and it would be nice if everyone paid attention to what was being said. In the course of this experiment I will be showing these tapes to other graduate students in the psychology department, so, before we begin, I'd like everybody's permission to do that."

Throughout the sessions, the experimenter attempted to elicit a mix of high- and low-interaction statements. At times, he would offer self-revealing statements in hopes that others, too, would be self-revealing. At other times, he would throw in a few preplanned controversial topics to keep the conversation moving. These preplanned topics included such things as the status of professional psychology, the role of women in academics, and the things that people do at graduate student parties.

The nonparticipant subjects were told that this was a study of how people communicate affect in conversations. They were also told the reason for use of the skin electrodes. At the end of the experiment, all subjects were told they would receive a "questionnaire"; no reference was made to the fact that it would be a memory test. The memory test was taken by all subjects between 72 and 96 h after the conversation.

Recording procedure. For the participants, skin conductance was recorded using a Beckman four-channel dc polygraph that permitted simultaneous time-locked recording of all four subjects. For the nonparticipants, who were tested individually, a Lykken box (Lykken & Venables, 1971) and a Grass polygraph were used. For all subjects, two large Beckman skin electrodes, filled with Beckman electrode paste, were placed about 5 cm apart on the heel of the subject's left palm. This placement left the right hand free to gesticulate and hold the refreshments that had been provided by the experimenter. Standard electrode collars were used to hold the electrodes in place.

The participants' conversations took place in the living room of an old fraternity house, now used as a lounge in an infant research laboratory. The participants were seated in a semicircle, and a video camera with a wide-angle lens was positioned so that it could record the facial expressions and the gestures of each of the participants. A second camera was focused on the moving polygraph record. Both cameras were connected to a special effects generator that provided a split-screen image. The participants appeared on the bottom half of the screen and the polygraph record was shown in the upper right corner. The microphone, on a table in the middle of the group, was connected directly into the video recorder. This setup allowed simultaneous video and audio recording of the participants and video recording of the polygraph record. The polygraph, the recorder, the monitor, and the assistant, who monitored the polygraph to adjust basal levels of skin conductance, were all located in another room.

The nonparticipants saw the videotape in the same room used by the participants. They watched the split-screen recording with a card taped over the portion of the screen that contained the polygraph record. A TV camera was used to record the polygraph record, and a microphone next to the TV monitor speaker was connected to the video recorder.

Memory test construction. For this study, a four-alternative forced-choice test was constructed in the following way. The day after the taping of each conversation, the experimenters reviewed the videotape with the polygraph record blocked out. From the pool of high-interaction sentences, 14 were selected that had the highest interactional content. As in Keenan et al. (1977), interactional content was defined in terms of relevance of a statement to the "face" (Goffman, 1972) that both speaker and listener maintained in conversational interaction. Agreement between raters was high ($r = .93$). Next, the 1-h tape was stopped at 14 approximately 4-min intervals and the first utterance that occurred on the tape was selected, as long as it was not 1 of the 14 high interactional content items that had already been selected. Distractors were made up using the procedure described by Keenan et al. (1977). The three distractor types were: (1) a paraphrase of the target, (2) a statement that differs from the target in prepositional content, but whose content was plausible in terms of the discussion, and (3) a paraphrase of Statement 2. This last distractor was a procedural necessity to make Statement 2 a plausible alternative. Examples of both high and low interactional content items are given in Table 1. Table 2 illustrates how test items were constructed.

The memory test was thus a 28-item four-alternative forced-choice test in which each item identified the speaker, for example, "Peter said: (a), (b), (c), and (d)." Thenonfamiliaronparticipants were given the first names of the four participants at the beginning of the viewing session so that they would have no trouble later remembering the names of the participants. Because

Table 1 Illustrations of High and Low Interactional Content Sentences

High Interactional Content Examples:

1. So how come you didn't invite me?
2. I really do feel a little like I'm on the Dick Cavett show.
3. These are my two best friends in Denver.
4. I've come a long way in like the last year.
5. I was going to ask you to take your shirt off, Doug, so that we could see your scars.
6. And then there's this big abdomen that just sort of hangs on and is good for reproduction and that sort of stuff-egg-carrying, inseminating.
7. I had a pretty good idea you didn't have one.
8. Linda has never been known to go anywhere without one one-hundredth of the things that she owns, which is a considerable amount.
9. Take this reality over here and make it look more like the fantasy.
10. Now that we've avoided your initial question.
11. Oh Rachel, my cock beats for you.
12. I mean it pisses me off, but it's much easier to be pissed off at this than to be pissed off at Orin for leaving.

Low Interactional Content Examples:

1. It's made the whole issue surface, that's for sure, and galvanized the gay community and stuff.
2. But you can see an advantage in just increasing offensive maneuvers for their own sake.
3. He's become more disappointed with academia in general.
4. But most, you know, PhDs in clinical, who have a clinical practice, don't do research.
5. And the last thing I did was a math unit, metric measurement.
6. And I don't know how to cut the rest of the chicken.
7. They measure how much you sweat basically.
8. I talked to Mom a little bit later and she said, "Oh gosh, that's one of the 10 best school districts in the country."
9. Well, Aspen is really beautiful.
10. We got a head lifeguard once who was an ex-football coach.
11. We invited 50 people up for Deborah and my's birthday.

Table 2 Examples of Test Items

High Interactional Content Targets:

1. Helen said:
 - a. Since I'm a woman, I figured I couldn't be a doctor. (Paraphrase)
 - b. Being a woman, nobody ever told me I could be a doctor. (Target)
 - c. They think they're so much better just cause we're psychologists. (Distractor)
 - d. Because we're psychologists, they think they're so much better than we are. (Distractor)
2. Ellen said:
 - a. Oh, no, not the Geritol commercial. (Distractor)
 - b. I still feel the pain-psychically. (Target)
 - c. The Geritol commercial, again? (Distractor)
 - d. My ego still bears the wounds. (Paraphrase)

Low Interactional Content Targets:

1. Peter said:
 - a. The 50 minutes on the road to and from Evergreen is the main drawback. (Distractor)
 - b. We asked about 40 people and over 50 showed up at our last birthday blow-out. (Paraphrase)
 - c. We invited 50 people up for Deborah and my's birthday. (Target)
 - d. The only thing is, the 50-minute commute is getting to be a real drag. (Distractor)
2. Chris said:
 - a. It seems to me that, you know, that's a pretty natural reaction. (Distractor)
 - b. And yet someone could really argue that it makes you more of a scholar. (Target)
 - c. I think that's a very natural reaction. (Distractor)
 - d. But then one could say it's more scholarly, it builds character. (Paraphrase)

two of the participants in each session were female and two were male, all that the nonfamiliar nonparticipants had to remember was which of the two females or two males had which name. Because participants used each others' names during the interaction, learning their names was a minimal problem for the nonfamiliar nonparticipants. The instructions for the test told the subject that only one sentence had actually been said and that they were to choose that sentence. The tests were scored for both memory for content and memory for surface form. Correct memory for content required choosing either the target or its paraphrase, and correct surface-form memory required choosing the target.

EDR scoring. After the questionnaires were constructed, the polygraph record was examined and scored. This was done by playing back the videotape and marking the position of the end of the each target sentence to the nearest .5 sec. Once the ends of the sentences were marked, EDR amplitude was scored from the polygraph recording using the following procedure: (1) Scan the period from 1 sec before the end of the statement to 3 sec after the end. (2) If there is a 45-deg rise, take its beginning. (3) If there is no 45-deg rise, EDR = 0. (4) If there is a 45-deg rise, find the end of the rise by locating the first 1 sec of negative slope. (5) Response is equal to the change in micromhos from the beginning of the rise to the end of the rise even if the beginning or end lies outside the 4-sec period of Step 2. (6) If there are two such rises in the 4-sec period, record the bigger one.

The criterion of a 45-deg rise was chosen to maximize the detection of significant arousal reactions and minimize the detection of spurious or unrelated arousal reactions. This is a standard cutoff criterion used in EDR research. If the 45-deg

deflection continues for 1 sec at the paper speed of 5 mm/sec, it represents an increase of .1 micromhos in EDR.

All raw EDR scores were transformed into standard scores. This was done by finding each subject's mean EDR and then assigning each of his 28 EDR responses a z score, so that the overall set of scores for that subject conformed to the normal distribution about that mean.

Results and Discussion

Correlational analyses. In general, EDR was not as good a predictor of memory as was interactional content. Table 3 presents, for each of the three subject groups, the correlations of EDR and memory for both surface form and content, the correlations of interactional content with memory for surface form and memory for content, and the intercorrelation of EDR and interactional content. These correlations used each subject's EDR on a given sentence as predictor of his/her memory for that sentence; thus, the analysis employed the standardized EDR scores. As is evident from the table, none of the correlations between EDR and memory was significant, whereas all the correlations between interactional content and memory for surface form were significant. Furthermore, a scatter plot of memory performance as a function of EDR showed that these low correlations were not the result of masking of an underlying quadratic function. Thus, these results provide little support for the arousal hypothesis.

A multiple regression analysis provided further evidence indicating that arousal was not leading to independent increments in memory. When the EDR component was partialled out, there was a significant ($p < .05$) residual correlation ($r = .20$) between interactional content and memory. But, when the interactional content component was removed, there was virtually no residual correlation between EDR and memory ($r = .03$). Results such as these are best understood in terms of a simple model in which high levels of interactional content produce increments in both memory and EDR, but in which EDR does not itself control memory. In other words, these results are best fit by a model in which there are two paths from interactional content, one to memory and one to EDR, but no path from EDR to memory. Such a model clearly fits these data better than the alternative, in which EDR and interactional content make separate contributions to memory.

Group differences in EDR. An analysis of variance with interactional content (high vs. low) and involvement

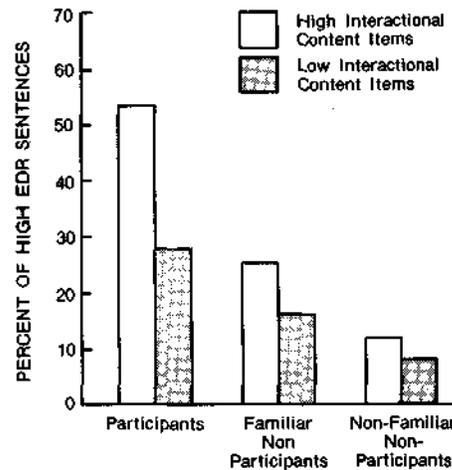


Figure 1. Percent of sentences yielding an EDR greater than 1 micromho.

(participants, familiar nonparticipants, and nonfamiliar nonparticipants) as independent variables was performed to examine the effects of these variables on arousal. Because of the highly skewed nature of the distribution of EDR scores, the dependent variable was the number of sentences that elicited an EDR greater than 1 micromho.

The results of this analysis are displayed in Figure 1, which presents the mean percentage of high EDRs within the 14 high and 14 low interactional statements for each of the three subject groups. As can be seen in the figure, there was a highly significant effect of involvement [$F(2,33) = 10.91, p < .01$]. Participants showed high EDRs on an average of 39% of the sentences, whereas familiar nonparticipants showed high EDRs on only 20% of the sentences and nonfamiliar nonparticipants showed high EDRs on only 9% of the sentences. It is interesting to note that, while both familiarity and participation affected arousal, the increase due to familiarity (19%) was almost twice that due to participation (11%).

There was also a significant effect of interactional content on EDR [$F(1,33) = 24.76, p < .001$]. An average of 30% of the high interactional content statements produced high EDRs, compared with only 17% for the low interactional content statements. Finally, there was a significant Involvement by Interactional

Table 3 Correlations of EDR and Interactional Content (IC) with Memory Performance

	EDR Memory		IC and Memory		IC and EDR
	Surface Form	Content	Surface Form	Content	
Participants Familiar	.14	.17	.35*	.44*	.41*
Nonparticipants Nonfamiliar	.05	.13	.30*	.14	.22*
Nonparticipants All Subjects	.10	.04	.34*	.11	.25*
	.09	.11	.33*	.22*	.29*

* $p < .05$.

Content interaction [$F(2,33) = 5.07, p < .025$]. As can be seen in Figure 1, the large difference between high and low interactional content statements observed in the familiar participants (27%) decreased for the familiar nonparticipants (9%) and was virtually eliminated in the nonfamiliar nonparticipant group (3%).

The findings in regard to EDR and its relations to involvement, interactional content, and memory may be easily summarized. Arousal, as measured by EDR, has a close relation to the importance of the stimulus to the listener. If the sentence is high in interactional content, arousal will be high. This is particularly true when the listener is directly involved in the conversational interaction or when he is at least an acquaintance of those who are involved in the interaction. However, this intimate relation of EDR to involvement and interactional content has no obvious consequences for memory.

Overall memory performance. Like the subjects in the study by Keenan et al. (1977), our subjects were clearly able to distinguish between the meanings of statements that did and did not occur, even though the memory test was unexpected and was given after a retention interval of at least 72 h. The average percentage of "old" responses to targets and their paraphrases was 76%. Also, as in the Keenan et al. study, memory for the meanings of high interactional content statements exceeded that of low interactional content statements (82% hits vs. 71%).

Memory for surface form. Of particular interest is the present study's replication of Keenan et al.'s (1977) finding of substantial memory for the surface form of statements. The average percentage of "old" responses to target items was almost twice that to paraphrases of the target (49% vs. 27%). Furthermore, as was found by Keenan et al., memory for surface form was much greater for high interactional content statements than for low interactional content statements. For high interactional content statements, the average percentage of "old" responses was 59% to targets and 23% to paraphrases. For low interactional content statements, the average percentage of "old" responses was 39% to targets and 31% to paraphrases.

Figure 2 shows the mean proportion of correct recognition of surface forms for high and low interactional content statements for each of the three experimental groups, as well as for a control group whose composition will be discussed later. An analysis of variance, involving interactional content (high vs. low), involvement (participants, familiar nonparticipants, and nonfamiliar nonparticipants), and class (1st year, 2nd year, 3rd year, and 4th year) was conducted on these data. As can be seen in Figure 2, the surface forms of high-interaction items were recognized significantly better than the surface forms of low-interaction items [$F(1,6) = 47.31, p < .000001$]. The main effect of involvement also just reached significance [$F(2,24) = 7.83, p < .05$]. Planned comparisons showed that nonfamiliar nonparticipants remembered less about surface form than did either

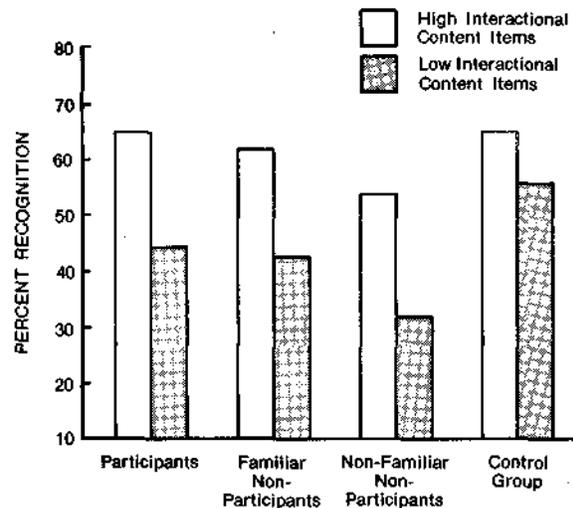


Figure 2. Percent of target sentences correctly recognized, indicating memory for surface form.

participants [$F(1,24) = 7.22, p < .02$] or familiar nonparticipants [$F(1,24) = 4.17, p < .05$]. The main effect of class was significant [$F(3,24) = 8.00, p < .001$]; the 4th-year class remembered more than the 3rd-year class [$F(1,24) = 5.48, p < .03$], the 2nd-year class [$F(1,24) = 7.41, p < .02$], and the 1st-year class [$F(1,24) = 10.05, p < .005$]. This effect of class is difficult to interpret because it reflects not only increasing familiarity over the years but also different conversations. The fact that class did not interact with involvement (i.e., even in the nonfamiliar nonparticipants, 4th-year students remembered more than the other students) suggests that the effect is not due to familiarity over the years, but rather, to differences in the memorability of the conversations. The 4th-year conversation seems to have been particularly memorable.

An important and potentially useful finding of the study was the absence of an interaction between interactional content and involvement. Whereas in the EDR data the difference between high and low interactional content statements was quite large for the participants and increasingly smaller for the familiar nonparticipants and nonfamiliar nonparticipants, in the memory data the difference between high and low interactional content statements was virtually the same for all three experimental groups. As stated earlier, one might expect that what counts as high interactional content for a given listener would depend both on the listener's knowledge of the speaker and his role in the interaction. Thus, when the listener neither knows the speaker nor is required to respond to the speaker's utterances, one might expect that the speaker's statements would carry less interactional content for the listener than when the listener is actively engaged in interacting with a friend.

Of course, the absence of an interaction between

involvement and interactional content also leaves open the possibility that factors other than interactional content may be contributing to the difference in memorability between high- and low-interaction statements. For example, it may be that high-interaction sentences contain more memorable lexical items or have more memorable syntactic structures than low-interaction sentences have. This possibility was tested in the control study that follows. Alternatively, it may be the case that the conversations were sufficiently high in human interest value that the familiar nonparticipants and the non-familiar nonparticipants became so deeply involved in the interaction that high interactional content statements, although less arousing than for the participants, carried almost as much interactional content as they did for the participants. We will attempt to determine which of these explanations is correct after reporting the data on memory for content.

Memory for content. Figure 3 shows the mean proportion of correct content recognition for each group. A three-way analysis of variance, parallel to that of the surface-form analysis, was performed. The main effect of interactional content was again highly significant [$F(1,6) = 32.00, p < .00001$]. The main effect of involvement was also significant [$F(2,24) = 13.45, p < .0005$]. Planned comparisons showed that the nonfamiliar nonparticipants remembered less than both the familiar participants [$F(1,24) = 13.01, p < .002$] and the familiar nonparticipants [$F(1,24) = 5.53, p < .03$]. Finally, the effect of class was significant [$F(3,24) = 4.76, p < .01$], indicating once again that the 4th-year class remembered more than any of the other classes. As in the surface-form analysis, the interaction of interactional content with involvement again failed to reach significance [$F(2,24) = 2.83$]. Thus, unlike EDR, memory for surface form and memory for

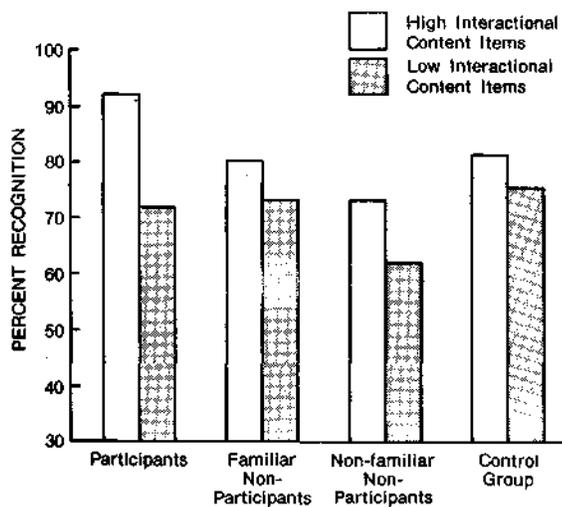


Figure 3. Percent of target or paraphrase sentences correctly recognized, indicating memory for meaning.

EXPERIMENT 2: CONTROL STUDY

In order to test the possibility that factors other than interactional content may have been responsible for the difference in memory between high- and low-interaction statements, a control study was performed to determine whether the high-interaction statements used in this experiment were simply more inherently memorable than the low-interaction statements. The control study, like that reported by Keenan et al. (1977), examined memory for the experimental materials presented in a standard list-learning situation.

According to the definition of interactional content, removing the statements from the context of meaningful, purposeful communication between real people should drastically reduce, if not totally eliminate, the interactional content of the statements. Thus, if it is the interactional content of the statements that is responsible for the difference in memorability, no difference between high and low interactional content statements should be obtained in a list-learning situation. If we still find a difference in memorability in a list-learning situation, then there is reason to believe that the effect of interactional content in the main experiment was due to factors other than interactional content. Keenan et al. (1977) found that presenting the statements used in their study in a list-learning task eliminated the large difference in memorability between high- and low-interaction statements that they had observed in the context of a conversation. A similar result was expected for this control study.

Method

Subjects. Thirty-six students from the University of Denver were recruited by an advertisement in the school newspaper. Each subject was paid \$5 for participation.

Procedure and Materials. Subjects listened to a tape consisting of 150 conversational-type sentences. The list of sentences included the 56 high and 56 low interactional content sentences from the four conversations, along with 38 filler sentences. Fillers were included to prevent ceiling effects on the memory test. All of the sentences were recorded by a single speaker who spoke in a conversational tone. The order of the sentence presentation was randomized, with the constraint that the first 10 and last 10 sentences of the list be filler items to control for primacy and recency effects. Subjects were instructed to listen carefully to the sentences in order to do well on a subsequent memory test; the nature of the test was not specified. As part of the informed consent procedure, it was necessary to warn the subjects that a few of the sentences contained profanity, highly suggestive sexual references, or both and that, if they did not want to listen to such language, they did not have to participate in the experiment. All subjects, however, agreed to participate.

Immediately after listening to the 150 sentences, the subjects were given a composite of the four memory tests used in the conversation experiment. The items were the same four-alternative forced-choice questions; however, in the test given to the control group, the speaker was not identified by name. The test consisted only of the 112 high- and low-interaction statements from the original conversations.

Results

Figure 2 compares the control group with the experimental groups in terms of memory for correct surface

content do not clearly vary with the listener's objective

form. An analysis of variance was performed in which the 36 experimental subjects were compared with the 36 control subjects. Overall, the control group had better memory for surface form than did the experimental group [$F(1,70) = 12.99, p < .001$]. However, this superiority held only for low interactional content statements. In terms of memory for the surface form of high interactional statements, the two groups did not differ. Thus, the fact that the control group knew they were going to take a memory test, together with the fact that the test followed immediately, rather than after a 3- or 4-day delay, allowed the control subjects to score higher overall. However, this advantage was confined to low-interaction items.

Of major interest is the fact that, contrary to our prediction, the control group remembered the surface forms of high-interaction sentences better (65%) than those of low-interaction sentences (56%) [$t(35) = 8.83, p < .0001$]. However, note that the size of this difference in memorability is much smaller (9%) than that observed for any of the experimental groups (21%). This is reflected in the highly significant interaction between group and interactional content [$F(1,70) = 14.31, p < .0005$]. Similar results were obtained for the analysis of memory for content. As can be seen in Figure 3, the control group remembered more of the content of the high-interaction sentences (82%) than of low-interaction sentences (76%) [$t(35) = 5.26, p < .0001$]. Again, this difference (6%) is much smaller than the average difference observed in the experimental groups (13%), as reflected in the significant interaction between group and interactional content [$F(1,70) = 5.28, p < .03$].

In order to clarify the role of interactional content in determining these differences between control and experimental groups, the responses of the control group subjects were utilized as covariates in an analysis of variance contrasting high- and low-interaction items. To our surprise, once the responses of the control group had been covaried out, the only effect that was significant was the three-way interaction of Interactional Content by Items by Involvement [$F(26,77) = 1.71, p < .03$]. Although the data presented in Figure 2 seem to indicate quite different patterns for the control and experimental groups, the main effects of interactional content [$F(1,77) = 5.96, p < .11$] and involvement [$F(2,77) = 4.24, p < .08$] were only marginally significant.

The presence of a strong interaction involving item differences made us suspicious regarding the effects of the instructions to the control group. In particular, we were worried that, because subjects knew that they would receive a memory test and because they had been specifically warned of the presence of sexually suggestive and profane language, they may have actually devoted some attention to trying to remember the more off-color sentences.

To test this possibility, we asked three independent raters to go through the sentences, noting those in which

there was either "profanity or overt mention of sexual activities." There was complete agreement on the six sentences involving profanity. The words involved were "tight-assed," "cock," "shit," "pisser," and two cases of "ass." The raters also agreed, completely or with only one dissenter, about the mention of sexual activities in sentences discussing (1) women sitting close to a speaker, (2) the use of "big body movements," (3) insemination through the abdomen, (4) "two women making it," (5) jealousy about a wife's activities, and (6) someone "hacking himself up."

A new subset of sentences was formed by excluding from the original set of 112 test sentences the 12 profane and suggestive sentences, which were all high interactional content sentences, and 12 low interactional content sentences chosen at random. This new subset had a total of 88 sentences. As before, this "expurgated" set of sentences was then submitted to analysis of covariance. The results this time were exceptionally clear. The only effect that reached either significance or marginal significance was the effect of interactional content [$F(1,59) = 68.43, p < .01$].

Discussion

The purpose of this control study was to determine whether the difference in memorability between high- and low-interaction sentences observed in the experimental groups was in fact due to the interactional content of the sentences or whether it was due to some other property of the test sentences. The latter possibility was raised by the finding that the difference in memorability between high- and low-interaction statements did not vary as a function of subjects' involvement in the conversations.

The results showed that those 12 high interactional content sentences with profane and sexually suggestive language elicited responses quite different from those elicited by the remaining 44 high interactional content sentences. Because the participants in the conversation studied by Keenan et al. used no profanity, we had not been alerted to what appears, in retrospect, to be a quite obvious fact: Sentences with off-color language possess a memorability that is quite independent of their role in a conversation. In future studies of this type, memory for such sentences should be analyzed separately from the beginning.

Having separated out the contribution of profanity and sexually suggestive language to the data, the control study showed that the effect of interactional content on memory for surface form remains even after the memorability of the sentences in the list-learning control is covaried out. Thus, the basic finding of Keenan et al. (1977) is repeated and supported.

GENERAL DISCUSSION

This study yielded four major results. The first is that arousal, as measured by EDR, is not a good predictor of

long-term memory for natural conversation. Although EDR seems to be a useful indicator of people's immediate reactions to conversation, it does not appear to be a major determinant of the processing and storage of sentences for long-term retention. Thus, like Brown and Kulik (1977), we found that arousal may be important in the initiation of rich memories for perceptual detail. However, in accord with Keenan et al. (1977), we found that the further retention of these memories depends on their consequences for subsequent interactions. Furthermore, unlike the Brown and Kulik study, the present study provided a direct physiological measure of arousal. This allowed us to make a much more direct assessment of the role of arousal in memory for conversation. The second major set of results relates to the impact of involvement on memory. Involvement varied from high involvement among the actual participants to low involvement in the nonfamiliar nonparticipants. Unfortunately, the design of the study did not allow us to fully deconfound participation and familiarity as separate dimensions of involvement. However, to the degree that we were able to measure familiarity and participation as components of involvement, each of these factors was shown to lead to increases in arousal as measured by EDR.

In terms of memory, significant effects of both interactional content and involvement were demonstrated. However, no interaction of interactional content with involvement was observed. We believe that this interaction failed to appear because the videotapes were sufficiently high in their human interest value to insure full vicarious involvement even for the nonfamiliar nonparticipants. It should be remembered that the nonfamiliar nonparticipant group was composed of other students of clinical psychology enrolled in a separate, but partially competing, program at the same university. Even though these students did not know the participants, they had a real interest in many of the topics being discussed. Despite the fact that these students showed little arousal as measured by EDR, they seemed to get involved in the interaction, much as the viewers of a soap opera get involved with the dialogue of their favorite characters (cf. Bates et al., 1980). As a result, the effect of interactional content held across all three groups. However, the overall memory of the nonfamiliar subjects was poorer than that of the familiar subjects. This result can be explained in terms of the relative richness of the conceptual structures available to the subjects in encoding the statements (Keenan & Baillet, 1980). Familiarity with a speaker provides the listener with a rich cognitive framework that serves to elaborate new inputs and increase their retrievability. Nonfamiliar subjects lack these rich cognitive structures, and thus, their encodings of the statements are less elaborated and less memorable.

The third major result is a consequence of the second. Because subjects were able to become vicariously involved

with videotaped conversations and because they showed superior memory for the surface form of high interactional content items, it should be possible now to begin to study memory for natural conversation using videotape as a stimulus. As stated earlier, this will immensely simplify the execution of these studies, as well as give us more control over the statements, thus allowing for a broader range of questions to be investigated. From a practical point of view, this is perhaps the most important result of the study.

The fourth major result of this study is the demonstration of the replicability and generalizability of the findings of Keenan et al. (1977). All of the major findings of that study were replicated here on four new conversational interactions, with new data collection procedures and with two types of nonparticipant groups. It was shown that recognition memory for both the form and content of high-interaction statements maintains itself at a very high level even after 3 days.

A final comment is in order regarding the categorization of statements according to interactional content. In both the present study and the one reported by Keenan et al. (1977), this categorization has been highly reliable. However, the decision procedure has not been operationalized. In the Keenan et al. paper, we claimed that this was a major limitation of this research and proposed that future research be directed toward specifying the criteria that render statements high in interactional content. We now realize, however, that this is an impossible task. In those cases in which sentences have profanity or sexually suggestive language, there is no reason to resort to the concept of interactional content to explain memorability. In cases in which sentences are not memorable in themselves, it is the history of shared personal experiences that makes one statement interactionally significant and another not. Thus, although it is relatively easy to get reliable categorizations of interactional content from people who have observed a natural conversation, it is certainly difficult and probably impossible to objectify the criteria that people use in making these judgments.

REFERENCES

- BATES, E., KINTSCH, W., FLETCHER, C. R., & GIULIANI, V. The role of pronominalization and ellipsis in texts: Some memory experiments. *Journal of Experimental Psychology: Human Learning and Memory*, 1980, 6, 676-691.
- BATES, E., MASLING, M., & KINTSCH, W. Recognition memory for aspects of dialogue. *Journal of Experimental Psychology: Human Learning and Memory*, 1978, 4, 187-197.
- BROWN, R., & KULIK, J. Flashbulb memories. *Cognition*, 1977, 5, 73-99.
- CRAIK, F. I. M., & BLANKSTEIN, K. R. Psychophysiology and human memory. In P. H. Venables & M. J. Christie (Eds.), *Research in psychophysiology*. London: Wiley, 1975.
- KEENAN, J. M., & BAILLET, S. Memory for personally and socially significant events. In R. S. Nickerson (Ed.), *Attention and performance*. Hillsdale, N. J: Erlbaum, 1980.

- KEENAN, J. M., MACWHINNEY, B., & MAYHEW, D. Pragmatics in memory: A study in natural conversation. *Journal of Verbal Learning and Verbal Behavior*, 1977, **16**, 549-560.
- KETY, S. D. The biogenic amines in the central nervous system: Their possible roles in arousal, emotion, and learning. In F. Schmitt (Ed.), *The neurosciences: Second study program*. New York: Rockefeller University Press, 1970.
- KINTSCH, W., & BATES, E. Recognition memory for statements from a classroom lecture. *Journal of Experimental Psychology: Human Learning and Memory*, 1977, **3**, 150-159.
- LYKKEN, D. T. Neuropsychology and psychophysiology in personality research. In E. F. Borgatta & W. W. Lambert (Eds.), *Handbook of personality theory and research*. New York: Rand McNally, 1968.
- LYKKEN, D. T., & VENABLES, P. H. Direct measurement of skin conductance: A proposal for standardization. *Psychophysiology*, 1971, **8**, 656-672.
- PETTIOKEW, J. D. The locus coeruleus cortical plasticity. *Trends in the Neurosciences*, 1978, **1**, 73-74.
- ZAJONC, R. B. Feeling and thinking: Preferences need no inferences. *American Psychologist*, 1980, **35**, 151-175.

(Received for publication May 21, 1981;
revision accepted February 22, 1982.)