JOURNAL OF MEMORY AND LANGUAGE 32, 25–38 (1993)

# On the Course of Answering Questions

VICKI L. SMITH

*Northwestern University*

AND

HERBERT H. CLARK

*Stanford University*

People responding to questions are sometimes uncertain, slow, or unable to answer. They handle these problems of self-presentation, we propose, by the way they respond. Twenty-five respondents were each asked 40 factual questions in a conversational setting. Later, they rated for each question their feeling that they would recognize the correct answer, then took a recognition test on all 40 questions. As found previously, the weaker their feeling of knowing, the slower their answers, the faster their nonanswers ("I don't know"), and the worse their recognition. But further, as proposed, the weaker their feeling of knowing, the more often they answered with rising intonation, used hedges such as "I guess," responded "I don't know" instead of "I can't remember," and added "uh" or "um," self-talk, and other face-saving comments. They reliably used "uh" to signal brief delays and "um" longer ones. 

What is the last name of the first man to run a mile in under four minutes? Answering factual questions like this is often viewed simply as a matter of recall. People search memory for the answer, and when they think they have found it, they produce it, and if they cannot find it quickly enough, they say "I don't know." This is the view, for example, in the traditional research on memory. In most memory experiments, people are instructed to answer such questions as quickly as possible and to limit their responses to a bare answer or "don't know."

In conversational settings, however, responding to such questions often takes a more complicated course. Here are four responses we recorded to the miler question in such a setting.

A: (0.6) Bannister
B: (3.1) hmm (2.0) Coe?
C: (1.6) um (3.0) I don't know
D: (1.6) oh, who was that? (to self) (10.0) oh, I can't remember.

(The numbers in parentheses mark seconds of silence.) These responses vary in several ways. Respondents A and B offered answers, whereas C and D produced what we will call nonanswers. B, C, and D took a long time before their final response and made comments during that interval; A took much less time and added no commentary. In giving her answer, B used a rising intonation (as marked by the question mark); A did not. And although C and D both gave nonanswers, they chose different phrasing—"I don't know" and "I can't remember." These examples suggest that the respondents were trying not only to answer the question, but also to account for their delays, uncertainties, and failures in an-

We thank Edward F. Schaefer for checking and editing the transcripts, and many colleagues for discussions on the issues presented here. The research was supported in part by Grant BNS 83-20284 from the National Science Foundation. Correspondence and reprint requests may be addressed to Vicki L. Smith, Department of Psychology, Northwestern University, Evanston, IL 60208 or Herbert H. Clark, Building 420, Jordan Hall, Stanford University, Stanford, CA 94305.



0749-596X/93 $5.00

swering. In classrooms, courtrooms, job interviews, and police interrogations, these accounts may be especially important. What course do people choose for their responses and how?

*Social Processes*

Responses to questions, we suggest, are guided by two goals that are fundamentally social—exchange of information and self-presentation. The point of most questions and answers is to exchange information. Questioners request information, and respondents, when willing and able, try to honor these requests. So one of the respondents' goals is ordinarily to retrieve and provide the requested information. But if they are slow, uncertain, or unsuccessful in that process, they also have to deal with their self-presentation—how they look to the questioner. For so-called examination questions (the questioner knows the answer and is testing the respondent), as in our study, self-presentation should be especially important. For convenience let us call this the *interactive view* of question answering.

*Confidence.* When people offer an answer to a question, they aren't always certain it is right. According to Grice (1975), speakers in conversation try to adhere to the maxim of quality: "Do not say what you believe to be false" and "Do not say that for which you lack adequate evidence." So when respondents are not certain of their answer, they should say so, or they will imply that they *are* certain. They should also indicate uncertainty in order to save face (Goffman, 1967, 1971). If their answer turns out to be incorrect, they do not want to look foolish. If they indicate little confidence in an uncertain answer, they insure themselves against that embarrassment.

One way to mark one's confidence in an answer is with intonation. In English a rising intonation signals "unfinished business" (Bolinger, 1989). When B answered "Coe?" she was requesting confirmation of whether or not the answer was correct. It was as if she were asking "Is it Coe?" In contrast, when A answered "Bannister" with falling intonation, he was asserting the answer with no special request for confirmation. If so, a rising intonation should signal lack of confidence, and a falling intonation, just the opposite. Another way to mark lack of confidence is with hedges such as "I think" and "I guess."

Respondents should also mark their confidence in nonanswers, and they can do that by their choice of nonanswer. Compare "I can't remember," "I don't know," and "I have no idea." When D said he couldn't remember, he implied he believed he knew the answer but couldn't recall it at the moment. When C said he didn't know, he implied he might not even be able to recognize the answer. If he had said he had no idea, he would have implied he didn't even know where to search. By Grice's maxim of quality, respondents should indicate which of these states they are in, or they might imply they are in another state. They should also do so to save face. If they can't quite recall an answer, they will want to say so to show that they really *do* know the answer—even if they cannot recall it at the moment.

*Delays.* Whenever respondents delay their response, that too may need explaining. Questions project answers; that is, after a question, the relevant next turn is an answer (Schegloff & Sacks, 1973). But the timing of that answer is crucial. According to Sacks, Schegloff, and Jefferson (1974), the respondent "has the right and is obligated to take the next turn to speak" beginning precisely at the completion of the question. That makes any delay in responding open to several interpretations. Respondents may be having difficulty (1) understanding the question, (2) retrieving the requested information, or (3) formulating their response. Or (4) they may even be withdrawing from the interchange altogether. Any misinterpretation could undermine the respondents' self-presentation. If so, respondents should be motivated to account for their delays. For factual questions with one word answers, that generally means accounting for (2).

One way to account for delays is with "uh." According to Levelt (1983), speakers produce "uh" on detecting some unspecified trouble in processing. When C said "um" after 1.6 s he was indicating he was in momentary trouble, and that helped explain the delay in his response. Our proposal is that respondents use "uh" both to signal a delay and to offer a brief account for it. In this view, "uh" is not merely a filled pause—which is an oxymoron anyway. It is a deliberate signal chosen from a range of interjections that include "uh," "um," "hm," "mm," and the tongue click "ts." The commonest of these are "uh" and "um" (see Svartvik & Quirk, 1980), and if they are distinct choices, they should also be used for distinct purposes (see E. Clark, 1983, 1987).

Respondents can also account for delays with self-talk and explicit commentary. As Goffman (1978) argued, there are norms against people talking to themselves in public. Yet in the right circumstances they use self-talk to *display* a current mental attitude to their onlookers. When a man stumbles, for example, he can say "oops" to himself to show onlookers that he realized he stumbled and is not crazy. When D said "oh, who was that?" to himself (self-talk has a delivery that is different from speech directed at addressees), he was showing that he was actively trying to retrieve the answer. So respondents might well use self-talk to account for their delays. They might use other comments as well, such as "Oh I know that one" or "Hang on a minute—this one has potential."

*Metamemory Judgments*

For respondents to mark confidence and account for delays, they need to be able to assess the current state of their memory search. How might they do that?

Answering factual questions, we assume, has three components: memory search, metamemory assessment, and responding. When searching memory, people may try several avenues of inquiry in what Williams and Hollan (1981) have called a retrieval cycle. For the miler questions, they might think of famous athletes, then runners, then milers, and then the first four-minute miler. If that avenue leads to a dead end, they might try another, thinking of when they first heard about the four-minute mile being broken, where it happened, and who did it. In Williams and Hollan's terms, a retrieval cycle consists of finding a context, searching it for a possible answer, and then verifying the answer.

Memory search works in conjunction with metamemory assessments, which in turn influence how people respond. For each potential answer, respondents must judge how confident they are that it is correct. If it passes criterion, they will offer it as an answer. If it does not, they must decide whether or not to continue searching. They should only continue as long as they believe they might retrieve an acceptable answer. This leads to two predictions. First, the more difficult the question, the longer the delay to the answer. And second, the more difficult the question, the sooner respondents will abort the search and give a nonanswer. Both of these predictions hold in memory tasks with constrained responses (Nelson, Leonesio, Landwehr & Narens, 1986; Nelson & Narens, 1980). They should also hold in conversational settings. The interactive view assumes that respondents are also concerned with their self-presentation, and that leads to further predictions. The more difficult the question, the more often respondents should use rising intonation or hedges for their answers, and the less confidence they should mark in their nonanswers.

Metamemory assessments should also help respondents deal with delays. If respondents judge the answer to be easily retrievable, or never retrievable, they know the delay will be brief. But if they judge that the answer *may* be retrievable and believe the delay may be long, they have reason to account for the delay, say, with "uh." We will call this the *anticipation hypothesis:* respondents signal when they anticipate a long delay before their final response. Indeed, they might use one signal (say, "uh") when they expect the delay to be brief, and another (say, "um") when they expect it to be longer. An alternative possibility is that respondents judge how long they have *already* delayed, and if it has been too long, they signal with "uh" or "um." We will call this the *reactive hypothesis:* respondents react to too long a silence (see Jefferson, 1989; Maclay & Osgood, 1959; Sacks, Schegloff, & Jefferson, 1974). Note that metamemory judgments are needed for the anticipation hypothesis but not for the reactive hypothesis. And in the anticipation hypothesis they are needed very early—after only 1.6 seconds for respondents C and D.

Much is known about people's "feeling of knowing" an answer. In a characteristic study of this phenomenon (Hart, 1965), people were asked a series of general knowledge questions (like our miler question). For the questions they couldn't answer, they were asked to predict whether or not they could recognize the correct answer. These predictions were their feeling of knowing judgments. The respondents were then tested on their recognition of the answers. They were more likely to recognize the correct answer the greater their feeling of knowing (e.g., Blake, 1973; Gruneberg & Monks, 1974; Hart, 1965, 1966, 1967). The phenomenon is robust across test materials, recall instructions, and retrieval conditions, and it holds for cued recall, recall on a second attempt, perceptual identification, and savings in relearning (Gruneberg & Sykes, 1978; Nelson, Gerler, & Narens, 1984).

In this study, then, we will examine the course of people's responses to factual questions in a conversational setting. We will take up several issues. First, do the latencies and accuracy of people's responses in our conversational setting parallel the latencies and accuracy in traditional memory experiments? If so, we can move to the issues that inspired our study, namely these: Do people mark the uncertainty of their answers and nonanswers, and if so, how, and how accurately? And do they signal delays in their answers with "uh" and "um" and other such devices? To answer these questions, we will also look at feeling of knowing judgments and recognition accuracy.

## METHOD

### *Overview*

The study had three stages. In the first stage, 25 participants were each asked 40 factual questions in spoken form, and they provided spontaneous spoken answers. The entire session was tape-recorded. In the second stage, without warning, the same participants were asked to rate their feeling of knowing for each of the 40 questions; that is, they judged their confidence that they would recognize the correct answer. In the third stage, again without warning, they were asked to complete a four-alternative, forced-choice recognition test on the 40 questions.

### *Procedure*

*Question answering.* Participants were told that they were taking part in a memory experiment. They were helping us pretest various general knowledge questions that we hoped to use in future experiments. We were pretesting the questions to determine how difficult each one was to answer. So, they would be asked a series of general knowledge questions with one-word answers. We emphasized that the questions covered a wide range of difficulty; some questions would be easy, others hard, and others they may not be able to answer. They were told that they could take as much time as they needed for each question, and that they could guess. They were told that the session would be tape-recorded, and microphones were attached to their lapels. The participants were 25 Stanford University students fulfilling a course requirement for introductory psychology.

The participant sat at one table and the experimenter sat at another behind an opaque screen. Those who asked about the screen were told that it was to prevent the experimenter from inadvertently giving them clues to the correct answer. The real reason was to force them to use their voice to convey any information to the experimenter that they wished to convey, and this was captured on tape. The experimenter read each question aloud, waited for the participant to finish responding, whether with the correct answer, a guess, or some indication that the answer was not known, and then read the next question. The participants sometimes produced more than one response, but always made it clear through their intonation when they had finished.

We selected the 40 questions from a list of 300 general knowledge questions for which Nelson and Narens (1980) developed norms for the probability of correct recall. The norms provided an independent assessment of the difficulty of each question. We divided the original list of 300 questions into quartiles based on the probability of correct recall and chose 10 questions at random from each quartile. The questions covered a variety of topics: science, literature, sports, geography, art, history, and entertainment. They were presented in one of five random orders.

*Feeling of knowing judgments.* After answering all 40 questions, the participants made feeling of knowing judgments. They were given a questionnaire that listed each question followed by a seven point rating scale that ranged from "absolutely sure I WON'T recognize" to "absolutely sure I WILL recognize." The scales were counterbalanced left to right across the 40 questions. The participants were asked to rate "how likely you think it is that you would be able to recognize the correct answer."

Unlike most studies of feeling of knowing, we asked participants to rate all 40 questions, not just the ones they failed to answer correctly. We did this for two reasons. First, if we asked participants to rate only those questions they had gotten wrong, they could use this information to alter their feeling of knowing judgments on questions they thought they had gotten right, and to rule out alternatives on the recognition test. Second, we wanted to compare the feeling of knowing judgments for the several types of correct answers, those with and without rising intonation, those with and without verbal hedges, and those offered after short versus long intervals.

*Recognition test.* After making feeling of knowing judgments, the participants completed a paper-and-pencil recognition test on the 40 questions. Each question was presented with the correct answer and three plausible alternatives. For example, the question "What is the capital of Chile?" listed Santiago (correct) and three other large Chilean cities, Valparaiso, Valdivia, and San Fernando. The participants were instructed to answer every question, even if they had to guess.

### *Transcriptions and Analyses*

We transcribed the entire response to each question, as shown for these four responses to the question "In which sport is the Stanley Cup awarded?"

a. (0.5) hockey
b. (1.4) um (1.0) hockey
c. (4.3) mm (3.0) pro racing?
d. (4.7) Stanley Cup (to self) (4.0) soccer? (3.0) I don't know.

We transcribed every vocalization we could hear: ordinary words and phrases, like *hockey* and *I don't know;* interjections like *uh, um,* and the tongue click; whistles, sighs, and self-talk; self-corrections; and rising intonation. The transcripts were prepared by Smith and edited by another person experienced in transcribing conversation. For the initial transcripts, we measured the silent intervals in seconds (shown in parentheses in the examples). Later, we slowed the tape to half speed and measured two latencies to the nearest hundredth of a second, namely: time to first utterance; and time to answer (to "hockey" and "pro racing" in *a, b,* and *c* and to "I don't know" in *d*). Responses were classified by what the participant said last. In *d,* the final response was "I don't know," a nonanswer. If it had ended with "soccer?" it would have been an incorrect answer.

In all there were 1000 responses, 40 from each of 25 participants. Because these responses were not independent of each other, analyses across all responses were inappropriate; one person's data could artificially inflate or obscure the relation between any two variables. To avoid this problem, we computed correlation coefficients for each participant individually, transformed the 25 correlations into Fischer's $z_r$ scores, and tested the average $z_r$ score against zero. The average $z_r$ scores were then transformed back into correlations for reporting in this article. Likewise, when comparing means, we computed a mean for each participant, and used these composite scores in our analyses. In any individual analysis, we did not include any participant for whom we could not compute an individual correlation or mean, so some of our statistics are based on a total $n$ of less than 25.

## Results

The results replicate the standard findings on memory, metamemory, and response latency. They also confirm the main predictions of the interactive view of answering questions. We will demonstrate how they replicate the standard findings before turning to the primary issue—how they bear on the interactive view of answering questions.

### *Memory, Metamemory, and Latency*

*Recall and recognition.* The 40 questions varied widely in difficulty. The normative probabilities of recall (to be called *normative recall*) for these questions, as reported by Nelson and Narens (1980) using University of California and University of Washington students, ranged from 0 to .97, with a mean of .49. The probability of recall across our Stanford students (or *obtained recall*) ranged from 0 to 1.00, with a mean of .57, which did not differ significantly from normative recall ($z = 0.76$, n.s.). Obtained recall correlated with normative recall .90 ($p < .001$) across the 40 questions, and an average of .62 ($p < .001$) for each respondent. Recognition varied widely across the 40 questions as well. On average, participants recognized 80% of the correct answers, which is substantially higher than the 25% expected by chance ($z = 6.32$, $p < .001$). Correct recognition correlated .75 ($p < .001$) with normative recall and .80 ($p < .001$) with obtained recall across questions, and an average of .42 ($p < .001$) and .52 ($p < .001$) for each participant. So normative recall, obtained recall, and correct recognition are highly correlated in the orderly way we would expect.

When people say they cannot recall an answer, they are still able to predict whether they can recognize it. The standard finding is that the higher the feeling of knowing, the more accurate the recognition. Our data replicate this finding. When participants gave a nonanswer (e.g., "I don't know"), feeling of knowing and recognition correlated an average of .46 ($p < .001$) for each participant. Our data, however, show that the same relation holds for answers, whether correct or incorrect. When participants gave an incorrect answer (e.g., "Coe?" to the miler question), the average correlation between feeling of knowing and recognition was .28 ($p <$

.001). Even for questions answered correctly (e.g., "Bannister"), the average correlation was .61 ($p < .01$). So, the more confident people are that they will recognize the correct answer, the more likely they are to recognize it, whether they give a correct answer, incorrect answer, or nonanswer.

*Metamemory.* For our participants, feeling of knowing was a better index of the difficulty of a question than was normative recall. Table 1 shows the average correlations among feeling of knowing, normative recall, obtained recall, and recognition for each participant for all questions together, and Table 2 shows the same correlations for correct answers, incorrect answers, and nonanswers considered separately. Feeling of knowing was better than normative recall at predicting recognition overall (.55 to .42, $z = 2.83$, $p < .004$). It was better for correct answers alone (.61 to .07, $z = 2.33$, $p < .02$) and nonanswers alone (.46 to .28, $z = 3.68$, $p < .001$), although only marginally better for incorrect answers alone (.28 to .18, $z = 1.64$, $p < .10$). When Nelson et al. (1986) looked at incorrect answers and nonanswers together, they found feeling of knowing *less* effective than normative recall at predicting recognition. But in their study, the correlation between recognition and normative recall was about .45, whereas in ours it was only .26. This is not surprising. The two measures were taken on students from the identical cohort in their study, but on students from different cohorts in ours.

TABLE 1
Average Correlations across Participants for All Questions

| | 1 | 2 | 3 |
|---|---|---|---|
| 1. Normative recall | | | |
| 2. Obtained recall | .62*** | | |
| 3. Feeling of knowing ratings | .53*** | .68*** | |
| 4. Probability of recognition | .42*** | .52*** | .55*** |

*Note.* ***$p < .001$.

TABLE 2
Average Correlations across Participants of Correct Recognition Computed Separately for Correct Answers, Incorrect Answers, and Nonanswers

| Correlation of recognition with | Normative recall | Feeling of knowing |
|---|---|---|
| Correct answers alone | .07 | .61** |
| Incorrect answers alone | .18*** | .28*** |
| Nonanswers alone | .28*** | .46*** |

*Note.* **$p < .01$; ***$p < .001$.

In our data, feeling of knowing and normative recall each contributed unique information to predicting recognition for individual participants. Feeling of knowing had an average correlation of .43 ($p < .001$) with recognition when normative recall was partialed out. At the same time, normative recall had an average correlation of .19 ($p <$ .001) with recognition when feeling of knowing was partialed out. So feeling of knowing and normative recall are not redundant in predicting recognition. Each contributes unique information, although feeling of knowing contributes more.

*Response latency.* What does response latency tell us about memory search and metamemory judgments? First, according to previous studies, the harder the question, the longer it should take to produce the correct answer. As Table 3 shows, latency to correct answer was predicted by normative recall with an average correlation of $-.23$ ($p < .001$) and by feeling of knowing with an average correlation of $-.57$ ($p < .001$). Second, the harder the question, the longer it should take to produce an incorrect answer as well. This held for feeling of knowing ($r = -.51$, $p < .001$), though not for normative recall ($r = -.03$, n.s.). (Compare Nelson et al., 1984, who found nonsignificant correlations of latency with feeling of knowing for incorrect answers.) Third, the harder the question, the

TABLE 3

AVERAGE CORRELATIONS ACROSS PARTICIPANTS OF LATENCY OF FINAL RESPONSE WITH NORMATIVE RECALL, OBTAINED RECALL, FEELING OF KNOWING, AND PROBABILITY OF RECOGNITION

| Correlation of latencies with | Normative recall | Obtained recall | Feeling of knowing | Recognition |
|---|---|---|---|---|
| All responses together | −.31*** | −.53*** | −.25*** | −.15*** |
| Correct answers alone | −.23*** | — | −.57*** | −.17 |
| Incorrect answers alone | −.03 | — | −.51*** | −.01 |
| Nonanswers alone | .36*** | — | .80*** | .37*** |

*Note.* Correlations cannot be computed for the three cells marked with dashes; ***$p < .001$.

sooner people should realize they cannot answer it and should say so. For nonanswers, latency was predicted by normative recall with an average *positive* correlation of .36 ($p < .001$) and by feeling of knowing with an average positive correlation of .80 ($p < .001$). So just as feeling of knowing was remarkably accurate in predicting how well respondents could recognize the correct answer, it was also remarkably accurate in predicting how long they would search for that answer before giving up.

Latencies should also correlate with correct recognition. People often know quickly what they do not know (Glucksberg & McCloskey, 1981; Kolers & Palef, 1976; McCloskey & Bigler, 1980), and there is no point in searching long for information judged not to be in memory. So the longer respondents search for an answer and yet give a nonanswer, the more confident they should be that they actually do know the correct answer, and the more likely they should be to recognize it. These two predictions were confirmed. As noted earlier, the longer the latency, the higher the feeling of knowing for each participant ($r = .80$, $p < .001$). Also, the longer the latency, the better the recognition for each participant ($r = .37$, $p < .001$). In contrast, the longer respondents have to search for an answer they are willing to offer, the *less* confident they should be that they could recognize it and the lower their recognition should be. Both of these correlations were negative, though not reliably so, perhaps because of the more restricted range of the latencies.

So far, then, the results follow the standard findings on normative recall, obtained recall, feeling of knowing, recognition, and response latency. They show especially clearly how accurate people's feeling of knowing is in accounting for their latencies and later recognition.

### *Form of Responses*

By the interactive view, respondents should also design their responses to reflect their current feeling of knowing. We will take up two dimensions of these responses—confidence and delays.

*Confidence.* When our participants offered an incorrect answer—and 20% of their answers were incorrect—they should have been less confident for several reasons. First, they might have realized that the question addressed an unfamiliar topic. Indeed, they gave more answers that were incorrect to questions that were more difficult as indexed by normative recall (.36 vs. .66; $t(24) = 12.29$, $p < .001$). Second, participants might have noted how long it took them to find an answer that met criterion. Indeed, incorrect answers had a much longer latency than correct answers (6.16 vs. 1.08 s; $t(24) = 6.21$, $p < .001$). Or third, they might have directly assessed their feeling of knowing before producing the answer. Consistent with this, they judged feeling of knowing to be lower after incorrect than correct answers (5.18 vs. 6.65; $t(24) = 10.11$, $p < .001$); of course, they also recognized fewer correct answers after incorrect than correct answers (.64 vs.

.98; $t(24) = 6.61$, $p < .001$). Whatever the basis, participants might have guessed that the incorrect answer they were offering was in fact incorrect.

If so, according to the interactive view, participants should mark these answers for lack of confidence, and they did. Although they used rising intonation on 25% of their answers, they did so more often on incorrect than correct answers (44% to 21%, $t(24) = 4.26$, $p < .001$). Also, feeling of knowing was lower after answers with a rising than a falling intonation (5.61 vs. 6.04; $t(24) = 3.03$, $p < .01$), even though recognition of the right answer was not reliably lower (.80 to .82; $F < 1$).

Respondents should also use more hedges with answers that were incorrect. Hedges were rare, the most frequent being "I guess" (used in 11 responses), "I think" (4 responses), and "something" (4 responses), as in these examples:

(4.5) terrible l- (to self) (2.0) uhh (4.0) um, uh, dinosaurs I guess
(9.7) Brazil, I think
(1.0) hh (1.0) Mayfield or something.

Despite their paucity, they were strongly associated with lower recall (.19 vs. .89, $t(9) = 7.23$), $p < .001$), lower feeling of knowing (5.64 vs. 6.66, $t(9) = 4.31$, $p < .001$), and lower recognition (.63 vs. .95, $t(9) = 3.36$, $p < .01$).

Participants should also indicate level of confidence for their nonanswers. The three most frequent forms of nonanswer were: "I don't know," used by 24 participants on 76% of the nonanswers; "I have no idea/clue," used by 12 participants on 11% of the nonanswers; and "I don't/can't remember," used by 11 participants on 9% of the nonanswers. Other ways included "I'm not sure," used by two participants, and "next" (a call for the next question), used by three participants. Of the three main choices, participants should choose "I can't/don't remember" to indicate the highest feeling of knowing, and "I have no idea/clue" the lowest.

The three types of nonanswers were ordered as predicted, as shown in Table 4. Feeling of knowing ranged from a high of 5.59 for "I don't remember," through 3.27 for "I don't know," to a low of 2.04 for "I have no idea" ($F(2,44) = 24.72$, $p < .001$). This ordering was reflected in later recognition (.73, .52, and .29; $F(2,44) = 6.54$, $p < .01$). So respondents might have based their choice of nonanswer directly on their feeling of knowing. Or they might have based it on how long they had searched, because the three responses were also ordered from highest to lowest in latency to final response (21.98 s, 5.57 s, 3.12 s; $F(2,44) = 21.63$, $p < .001$). They could not have based it entirely on the difficulty of the question; although the three responses differed in normative recall (.40, .22, and .22; $F(2,44) = 6.48$, $p < .003$), there was no difference between "I don't know" and "I have no idea." The difference between "I don't remember" and "I have no idea" was enormous in absolute terms. In latencies, it was more than 7 to 1; in feeling of knowing, it was more than 3.5 units on a 7-point scale; and in recognition, it was 44%. These differences suggest that respondents are very sensitive to their feeling of knowing or its behavioral manifestations.

Were the three nonanswers accurate? When respondents said "I have no idea," they were right. Their recognition of .29 was not significantly above guessing of .25 ($z = 0.32$, n.s.). When they said "I don't

TABLE 4
MEANS OF COMPOSITE SCORES BY PARTICIPANT FOR THE THREE MOST FREQUENT NONANSWERS

| Variable | Form of nonanswer: "I don't remember" | "I don't know" | "I have no idea" |
|---|---|---|---|
| Frequency | 25 | 212 | 31 |
| Latency to final response | $21.98_a$ | $5.57_b$ | $3.12_b$ |
| Feeling of knowing | $5.59_a$ | $3.27_b$ | $2.04_c$ |
| Recognition | $.73_a$ | $.52_a$ | $.29_b$ |
| Normative recall | $.40_a$ | $.22_b$ | $.22_b$ |

*Note*. On each line, means with the same subscript are not significantly different.

remember," they were also right to imply that they knew the answer but couldn't recall it. Their recognition was .73, substantially greater than chance of .25 ($z$ = 3.66, $p < .001$). But when they said "I don't know," their recognition of .52 was also significantly greater than guessing ($z$ = 3.07, $p < .001$). So with "I don't know," respondents either underestimated their knowledge or understated their true feeling of knowing.

*Delays.* When participants were unable to respond quickly, what did they do? All they were required to do was produce the final response, but most did more. They often used interjections, such as "uh" and "oh," and sometimes sighed, whistled, or talked to themselves. We will call these items *fillers*. Of the 1000 responses, 24% had at least one filler before the final response.

When were fillers used? They occurred more often with both incorrect answers (39%) and nonanswers (32%) than with correct answers (17%; both $ps < .001$). As shown in Table 5, fillers were an indication of lower confidence when participants gave substantive answers. For correct answers, feeling of knowing was lower with fillers than without them ($t(20) = 3.02$, $p < .01$). The same held for incorrect answers ($t(18) = 2.05$, $p < .06$). For nonanswers, in contrast, fillers indicated *higher* confidence: feeling of knowing was higher with fillers than without them ($t(20) = 4.68$, $p < .001$).

TABLE 5
FREQUENCY, MEAN FEELING OF KNOWING RATINGS, AND MEAN LATENCIES FOR THREE TYPES OF RESPONSES WITH AND WITHOUT FILLERS

| Measure | Type of answer | Without filler | With filler | Total |
|---|---|---|---|---|
| Frequency | Correct | 479 | 95 | 574 |
| | Incorrect | 91 | 57 | 148 |
| | Nonanswer | 188 | 90 | 278 |
| Feeling of knowing | Correct | 6.74 | 6.13 | 6.44 |
| | Incorrect | 5.33 | 4.71 | 5.02 |
| | Nonanswer | 3.09 | 4.22 | 3.66 |
| Latency to final response | Correct | 0.97 | 5.00 | 2.98 |
| | Incorrect | 5.34 | 9.49 | 7.41 |
| | Nonanswer | 3.39 | 10.30 | 6.85 |

What this suggests is that fillers are associated with delay: the longer the delay, the more likely respondents are to use fillers. The pattern is confirmed by the latencies listed in Table 5. The time to final answer averaged 5.06 s longer with fillers than without. The increase was 4.03 s for correct answers ($t(20) = 3.43$, $p < .01$), 4.15 s for incorrect answers ($t(18) = 3.12$, $p < .01$), and 6.91 s for nonanswers ($t(18) = 4.28$, $p < .001$), so it was also consistent across types of response. But what is the cause of what? Do respondents use fillers to account for past delays, as in the reactive hypothesis? Or do they use fillers to warn of anticipated delays, as in the anticipatory hypothesis? Or do they use them for both purposes? The answer may depend on the filler examined.

The two most common interjections in our data were "uh" and "um." The choice between them might be crucial, because it would enable participants to distinguish long from short delays—whether in reaction or in anticipation. To examine the issue, we looked at the 39 "uh"s and 38 "um"s that were produced as the very first item in a response. As Table 6 shows, there was a slightly longer pause immediately before "um" than before "uh," (3.70 s to 2.54 s, $t(75) = 2.27$, $p < .05$). This suggests that "um" is reactive to longer past delays than "uh." There was also a much longer pause immediately after "um" than after "uh" (4.12 s to 1.00 s, $z = 3.21$, $p < .001$), and a very much longer delay to the final answer (8.83 s to 2.65 s, $t(75) = 2.69$, $p < .01$). So the choice of "um" over "uh" was informative, and more so about the delay it anticipated than about the delay it reacted to.

Both the reactive and anticipatory hypotheses are further supported by an analysis on "uh" and "um" in the London–Lund corpus of English conversation (Svartvik & Quirk, 1980). The analysis (Clark & Fox Tree, unpublished) was carried out on detailed transcripts of sponta-

TABLE 6
AVERAGE LATENCIES BY PARTICIPANT (IN SECONDS) FOR INITIAL "UH" AND "UM" AND PERCENT PAUSES IN SPONTANEOUS CONVERSATION (LONDON–LUND CORPUS)

| | Interjection | | |
|---|---|---|---|
| | "uh" | "um" | Difference |
| *Present data* | | | |
| Pause immediately before | 2.54 | 3.70 | +1.16* |
| Pause immediately after | 1.00 | 4.12 | +3.12*** |
| Time to final response | 2.65 | 8.83 | +6.18*** |
| *Spontaneous conversation* | | | |
| Percent pauses immediately before | 32 | 47 | +15*** |
| Percent pauses immediately after | 23 | 62 | +39*** |

*Note.* $^*p < .05$; $^{***}p < .001$.

neous face-to-face conversations of groups of two to four speakers of British English, mostly academics. The transcripts used contained about 170,000 words, all recorded surreptitiously, covering a total of 78 different speakers (texts S.1.1 through S.3.6 in Svartvik & Quirk, 1980). With overlapping speech excluded, there were 2249 "uh"s and 1738 "um"s. The corpus marks both "brief pauses (of one light syllable)" and "unit pauses (of one stress unit or 'foot')," so we tallied how many times "uh" and "um" were preceded or followed by either a light or a unit pause. The percentages of each are shown in Table 6. "Um" and "uh" were preceded by pauses 47% and 32% of the time. Both were used to break off pauses—"um" more often than "uh." But the main difference between "um" and "uh" was in the pauses that followed them. "Um" was followed by a pause 62% of the time, and "uh" only 23% of the time. (Both differences are highly significant.) What this suggests is that speakers use "uh" to signal a minor problem and "um" a major one. "Uh" therefore anticipates a small break in the utterance and "um" a larger one.

One reason for using fillers, in the interactive view, is to tell or show questioners that the respondents are still actively engaged in answering the question. When they use "uh" or "um," for example, they also imply that they are still actively engaged in answering the question. "Uh," "um," "mm," and "hm" appeared in 124 responses, often more than once per response, apparently with this function.

Respondents also indicated active engagement by talking to themselves. When people talk to themselves, they *show* others what they are thinking without directly *telling* them (Goffman, 1978). Self-talk was used in 43 responses. These included 25 interjections (which are distinct in their delivery from interjections addressed to the questioner), 10 repeats of part of the question (as in "capital of Chile" and "terrible l-"), and 8 preliminary answers (as in "um Fowler, hh I don't know"). Respondents also showed their frustration using tongue clicks ($n$ = 56), whistling ($n$ = 2), and sighing ($n$ = 15). All of these tactics displayed active engagement in the task.

Respondents also made direct comments about the progress of their memory search. Four times they explicitly asked the questioner to wait while they worked on the answer, as in "(0.7) the Nile (1.0) wait, no, South American, the Amazon." Three times they asked for clarification of the question. For "What is the name of the extinct reptiles known as 'terrible lizards'?" one person responded: "(16.0) extinct reptiles? [Experimenter: mmhmm] (6.0) a dinosaur." Presumably, this respondent had been searching for an answering during the first 16 s, and the request for clarification was an attempt to redirect the search. Three times respondents asked the experimenter to repeat the question.

Another reason for using fillers is to save face. When respondents could not come up with an answer quickly, they often displayed embarrassment and tried to save face. We have already noted four ways of saving face. One was by showing active engagement in memory search. That way respondents implied they understood the question and thought they might be able to answer it. Another way was by indicating lack of confidence in an answer with rising intonation or hedges. If the answer turned out to be wrong, they would have indicated that they were not so foolish as to be certain it was correct. Still another way was by the choice of nonanswer. "I can't remember" implied they knew the answer but couldn't quite recall it.

The most direct moves to save face came in explicit comments. Sometimes respondents said they really did know the answer, even if they couldn't retrieve it at the moment, as in: "shoot hang on a minute (1.0) this one has potential," "I should know that one hh," "this is the tip of the tongue phenomenon," and "I can't think of that right now hh." Other times they showed how they knew something about the answer, as with "it's one of the Andes, isn't it?" On 13 occasions, they commented that it was the particular subject matter of the question that caused the trouble, not their general ignorance: "geography is really embarrassing," "don't know South America, sorry," or "I don't know that one" (suggesting that they *did* know other ones).

In short, the auxiliary interjections, self-talk, comments, and noises that respondents produced when they got into trouble seemed to serve two main purposes. One was to let the questioner in on how they were progressing. This helped explain why they were taking so long. Another was to save face by showing that there were good reasons for their trouble.

## Discussion

Language use is a social process. The common view is that it is primarily the exchange of information. So when speakers formulate utterances, they try to package the information they have to exchange as neatly as they can, and their addressees interpret them as having tried to do this (Grice, 1975). We have argued that language use is also guided by self-presentation. In formulating and producing utterances, speakers attend to their audience's interpretation of what they are doing moment by moment. Our evidence comes from the process of answering questions.

In the ideal answer to a question, respondents (1) provide the requested information (2) confident in its accuracy and (3) without undue delay. Traditional models of answering questions assume as much when they expect respondents to answer as accurately and as quickly as possible. Problems arise when respondents cannot meet that ideal. If they cannot provide the requested information, or are not confident in its accuracy, or are unduly delayed, their audience may judge them to be uncooperative, ignorant, poor in judgment, or slow-witted. These are inferences they may want to forestall. If so, they need to explain their nonanswers, signal their level of confidence, and account for any undue delays. How do they do that?

Dealing with problems of self-presentation takes two steps. The first is to recognize that there is a problem—a current or imminent departure from the ideal answer. The second is to do something about it. In the first step, we have argued, respondents should monitor the state of their memory—assessing their metamemory—just as they do in memory search, and the evidence suggests that they do.

In the standard view of memory search, respondents do not just retrieve the right answer. They evaluate candidate answers until they find one that passes criterion, and then they utter it. If they cannot find a passable answer soon enough, they say "don't know." Such a view accounts for many features of our data. The harder the question, the longer it should take respondents to retrieve the answer, and the sooner they should give up and produce a nonanswer. This we found. Also, the respondents' subjective impression of question difficulty, their feeling of knowing, should determine how long they search for an answer. Indeed, the stronger the feeling of knowing, the quicker were the answers and the slower the nonanswers. Moreover, people's feeling of knowing was highly veridical; it strongly predicted their ability to recognize the right answer later.

In the interactive view, respondents should also signal their level of confidence and account for their delays, and they have many devices for doing this. They can use rising intonation and hedges to indicate lack of confidence in answers. They can select their nonanswer to signal their level of subjective knowledge. They can choose between "uh" and "um" to signal the depth of their ongoing retrieval problem—to account for past delays and project future delays. And they can talk to themselves and make explicit comments to indicate what trouble they are having and why.

Respondents should exploit these devices based on metamemory assessments, and the evidence suggests they did. The longer it took them to answer, and the lower their later feeling of knowing, the more often they marked their answers with rising intonation or hedges, and the more often they expressed their nonanswers as "I can't remember," then "I don't know," and finally "I have no idea." Also, the longer the delay to the final response, the more often they used interjections, self-talk, or explicit comments. They didn't use these fillers just to fill time. Why do that? They used them to indicate that they were still actively engaged in memory search and to help explain why they hadn't answered yet. Both of these helped them maintain self-esteem—save face. What is remarkable is how early in the process they exploited their metamemory assessments. They often used "uh" or "um" after only 1 s to indicate the depth of their retrieval problem. And the deeper the trouble, the more often they used "um" over "uh," accounting both for a longer initial delay and projecting a longer delay to their response.

Answering questions, then, is more complicated than generally supposed. Most models assume that it requires (1) memory search or activation and (2) monitoring that search (see, e.g., Anderson, 1985). The first process generates candidate answers, and the second evaluates them, deciding whether to continue or abort the search. In conversational settings, however, answering questions also requires a third process: (3) assessing how the response will be viewed. That process is needed to decide whether or not to introduce interjections, self-talk, and other commentary, what these fillers should imply, and how to express the final answer or nonanswer. Speakers keep remarkably close track of all three types of information from the moment they begin formulating their responses. How they do so must be accounted for in any adequate model of speech production.

## References


Anderson, J. R. (1985). *Cognitive psychology and its implications* (2nd ed.). San Francisco: Freeman.

Blake, M. (1973). Prediction of recognition when recall fails: Exploring the feeling-of-knowing phenomenon. *Journal of Verbal Learning and Verbal Behavior*, 12, 311–319.

Bolinger, D. L. (1989). *Intonation and its uses*. Stanford, CA: Stanford University Press.

Clark, E. V. (1983). Meanings and concepts. In J. H. Flavell, & E. M. Markman (Eds.), *Handbook of child psychology, Vol 3: Cognitive development* (pp. 787–840). New York: John Wiley & Sons.

Clark, E. V. (1987). The principle of contrast: A constraint on language acquisition. In B. MacWhinney (Ed.), *Mechanisms of language acquisition* (pp. 1–33). Hillsdale, NJ: Erlbaum.

Glucksberg, S., & McCloskey, M. (1981). Decisions about ignorance: Knowing that you don't know. *Journal of Experimental Psychology: Human Learning and Memory*, 7, 311–325.

Goffman, E. (1967). *Interaction ritual: Essays on face to face behavior*. Garden City, NY: Doubleday.

Goffman, E. (1971). *Relations in public*. New York: Harper and Row.

Goffman, E. (1978). Response cries. *Language*, 54, 787–815.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics, Vol. 3: Speech acts* (pp. 113–128). New York: Seminar Press.

GRUNEBERG, M. M., & MONKS, J. (1974). "Feeling of knowing" and cued recall. *Acta Psychologica*, **38**, 257–265.

GRUNEBERG, M. M., & SYKES, R. N. (1978). Knowledge and retention: The feeling of knowing and reminiscence. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory* (pp. 189–196). New York: Academic Press.

HART, J. T. (1965). Memory and the feeling-of-knowing experience. *Journal of Educational Psychology*, **56**, 208–216.

HART, J. T. (1967). Memory and the memory-monitoring process. *Journal of Verbal Learning and Verbal Behavior*, **6**, 685–691.

HART, J. T. (1966). Methodological note on feeling-of-knowing experiments. *Journal of Educational Psychology*, **57**, 347–349.

JEFFERSON, G. (1973). A case of precision timing in ordinary conversation: Overlapped tag-positioned address terms in closing sequences. *Semiotica*, **9**, 47–96.

JEFFERSON, G. (1989). Preliminary notes on a possible metric which provides for a "standard maximum" silence of approximately one second in conversation. In D. Roger & P. Bull (Eds.), *Conversation*. Philadelphia: Multilingual Matters Ltd.

KOLERS, P. A., & PALEF, S. R. (1976). Knowing not. *Memory & Cognition*, **4**, 553–558.

LEVELT, W. J. M. (1983). Monitoring and self-repair in speech. *Cognition*, **14**, 41–104.

MACLAY, H., & OSGOOD, C. E. (1959). Hesitation phenomena in spontaneous English speech. *Word*, **15**, 19–44.

MCCLOSKEY, M., & BIGLER, K. (1980). Focused memory search in fact retrieval. *Memory & Cognition*, **8**, 253–264.

NELSON, T. O., GERLER, D., & NARENS, L. (1984). Accuracy of feeling-of-knowing judgments for predicting perceptual identification and relearning. *Journal of Experimental Psychology: General*, **113**, 282–300.

NELSON, T. O., LEONESIO, R. J., LANDWEHR, R. F., & NARENS, L. (1986). A comparison of three predictors of an individual's memory performance: The individual's feeling of knowing versus the normative feeling of knowing versus base-rate item difficulty. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **12**, 279–287.

NELSON, T. O., & NARENS, L. (1980). Norms of 300 general-information questions: Accuracy of recall, latency of recall, and feeling-of-knowing ratings. *Journal of Verbal Learning and Verbal Behavior*, **19**, 338–368.

SACKS, H., SCHEGLOFF, E., & JEFFERSON G. (1974). A simplest systematics for the organization of turn-taking in conversation. *Language*, **50**, 696–735.

SCHEGLOFF, E., & SACKS, H. (1973). Opening up closings. *Semiotica*, **8**, 289–327.

SVARTVIK, J., & QUIRK, R. (Eds.). (1980). *A corpus of English conversation*. Lund, Sweden: Gleerup.

WILLIAMS, M. D., & HOLLAN. H. D. (1981). The process of retrieval from very long-term memory. *Cognitive Science*, **5**, 87–119.

(Received December 20, 1989)
(Revision received April 8, 1992)