

Feeling-of-Knowing Ratings Distinguish Between Genuine and Simulated Forgetting

Daniel L. Schacter

Unit for Memory Disorders, Department of Psychology, University of Toronto

Three experiments investigated the relation between genuine and simulated forgetting of a specific episode. Subjects who had genuinely forgotten an episode, and subjects who were instructed to simulate forgetting of the same episode, made feeling-of-knowing ratings concerning the likelihood that they could remember the episode on their own or in the presence of cues. They also verbalized their thoughts as they attempted for several minutes to recall the forgotten episode. Although the patterns of feeling-of-knowing ratings made by genuine and simulating subjects were similar in several respects, they also differed systematically: Simulators consistently expressed less confidence that cues would facilitate retrieval than did genuinely forgetful subjects. In contrast, psychologists and psychiatrists who were given verbal protocols of subjects' retrieval attempts could not distinguish between genuine and simulating subjects, even when they expressed certainty that they had. The role of metamnemonic knowledge in attempts to simulate forgetting is discussed.

A widespread assumption underlying psychological studies of memory is that reports of forgetting can be attributed to a failure of mnemonic processes. When a subject in a memory experiment states that he or she is unable to remember a particular item, the experimenter assumes that the subject's report is an accurate reflection of his or her current mental experience. There is no reason to doubt the validity of this assumption in laboratory studies. In everyday life, however, claims of forgetting are not always attributable to memory failure: People sometimes state that they have forgotten a specific event or episode even though they do in fact remember it. In the present article, this phenomenon will be referred to as *simulated forgetting*.

Simulated forgetting occurs in a variety of everyday contexts. For example, when a person commits a violent crime, he or she frequently claims amnesia for the episode (Bradford & Smith, 1979; Guttmacher, 1955; Leitch, 1948; O'Connell, 1960; Schacter, in press-a). Such claims of forgetting can have significant legal consequences (Gibbens & Williams, 1977; Koson & Robey, 1973), but in a large proportion of cases forgetting is simulated (Adatto, 1949; Bradford & Smith, 1979; Hopwood & Snell, 1933;

O'Connell, 1960; Power, 1977; Schacter, in press-a). Other real-life situations in which it is important to distinguish between genuine and simulated forgetting of a specific episode include eyewitness testimony (Loftus, 1979; Neisser, 1981), allegations of plagiarism (Taylor, 1965), compensation claims following injury (Guthkelch, 1980), and civil actions concerning contracts and divorce (Gibbens & Williams, 1977). In view of the potentially far-reaching consequences, there is widespread acknowledgment that distinguishing between genuine and simulated forgetting constitutes a major problem (Bradford & Smith, 1979; Gibbens & Williams, 1977; Koson & Robey, 1973; Power, 1977; Schacter, in press-b). Unfortunately, existing literature provides little or no useful information concerning this issue. Experimental investigators of memory have not yet studied the relation between genuine and simulated forgetting of a specific episode (For research concerning simulated amnesia in hypnosis, see Spanos, Radtke-Bodorik, & Stam, 1980; Williamsen, Johnson, & Eriksen, 1965). Clinicians who have addressed the problem have offered suggestions and speculations that are based on uncontrolled observations (e.g., Bradford & Smith, 1979; Hopwood & Snell, 1933; Koson & Robey, 1973; Power, 1977). Because the appropriate studies have not yet been performed we do not know whether any procedures can distinguish reliably between genuine and simulated forgetting.

The purpose of the present article is to examine the relation between genuine and simulated forgetting of a specific episode and to determine experimentally whether they can be distinguished. Two general approaches can be taken to the problem. The first is to study the phenomenon as it occurs in one of the everyday contexts noted earlier. Although this approach has the advantage of ecological validity, it also has a serious drawback: It is not possible to be certain about the actual status of an alleged case of memory loss unless there is a direct admission of simulation. This uncertainty concerning the status of individual subjects indicates that study of actual cases is not the ideal place to initiate research that attempts to distinguish between genuine and simulated forgetting.

The research reported in this article was supported by the Natural Sciences and Engineering Research Council of Canada Grant U0361 and by a Special Research Program Grant from the Connaught Fund, University of Toronto.

I am indebted to James Worling for invaluable assistance in many phases of this research, thank Elana Joram and Carol A. Macdonald for experimental assistance, and thank Carol A. Macdonald for transcribing the verbal protocols in Experiments 1 and 2 as well as for preparing the manuscript. I am also grateful to Mark Ben-Aron, Gordon Hayman, Steve Hucker, Paul Muter, Karen Raaflaub-Walsh, Susan Rich, Gary Snow, and Michèle Stampf for volunteering to serve as judges. Peter Graf, Michèle Stampf, and Endel Tulving provided helpful comments concerning an earlier version of the manuscript.

Correspondence concerning this article should be addressed to Daniel L. Schacter, Department of Psychology, University of Toronto, Toronto, Ontario, Canada M5S 1A1.

The second approach, the one that will be adopted here, is to attempt to create a laboratory analogue of the phenomenon, that is, to construct a situation in the laboratory that enables one to examine various features of an everyday phenomenon. This approach has two advantages. First, the investigator can control the assignment of subjects to conditions (i.e., genuine vs. simulating) and thus be certain of who has genuine memory failure and who does not. Second, the investigator can manipulate experimental variables and hence examine a wide range of situations, a luxury that is not readily available in everyday life. The main drawback of this approach is that one cannot be certain that results obtained in the artificial laboratory situation will generalize to real-life contexts: A laboratory simulation will necessarily differ from an everyday situation in numerous ways. The present research therefore does not attempt to mimic the exact circumstances that would be encountered in any one of the numerous everyday contexts in which simulated forgetting occurs. Rather, the goal is to determine whether any features of performance reliably distinguish between genuine and simulating subjects in the laboratory. The issue of generalizability could then be addressed empirically once a replicable pattern of results has been established in the laboratory. Because there are no such results at the present time, creation of a laboratory analogue may serve as a useful beginning step toward understanding the relation between genuine and simulated forgetting.

Goals and Logic of the Research

When formulating a laboratory procedure for distinguishing between genuine and simulated forgetting, two key issues must be addressed. First, in real-life cases in which a person claims to have forgotten a particular incident, investigators have no firm knowledge of what actually occurred during the critical episode; they must rely entirely on what the person says about it. Accordingly, a laboratory procedure for distinguishing between genuine and simulated forgetting should rely solely on what a person says when questioned at the time of attempted retrieval; it should not be contingent upon an examiner's knowledge of what actually occurred during the critical episode. Second, it is important to consider what people are likely to believe or know about the characteristics of genuine forgetting of specific episodes. If people possess accurate beliefs or knowledge concerning a feature of genuine forgetting, it is reasonable to suppose that they will be able to simulate forgetting successfully. However, if people do not possess accurate beliefs or knowledge concerning a feature of genuine forgetting, it may be difficult for them to simulate successfully. For example, some investigators have suggested that reports of forgetting in which there is a sudden, sharply defined onset and termination of the forgotten episode are likely to be simulated, whereas reports of a gradual onset and termination are more likely to be genuine (Hopwood & Snell, 1933; Power, 1977). If this is in fact correct, a simulator who "knew" that he or she should report a gradual onset and termination of an allegedly forgotten episode would have a greater chance of being successful than one who did not. Thus, a critical question with respect to the present concerns can be stated as follows: Are there observable aspects of genuine forgetting of a specific episode that would be unknown to a simulator? Existing literature provides little or no guidance concerning this question, because (a) as noted earlier, there are few well-established facts concerning the characteristics of memory loss for individual episodes and

(b) there is no information available concerning people's knowledge and beliefs about the characteristics of such episodes. Accordingly, construction of a laboratory analogue must be guided initially by somewhat speculative hypotheses concerning which features of memory loss are unlikely to be intuitively obvious to naive subjects.

The present article focusses on a phenomenon that satisfies the requirements that have been discussed thus far. The phenomenon is known as the *feeling of knowing*. The feeling of knowing refers to a person's belief that he or she could retrieve or recognize an unrecalled item, event, or fact if he or she were given more powerful hints or cues. A number of studies have documented that (a) people frequently report a feeling of knowing that they will be able to recognize unrecalled items, events, or facts and that (b) these feelings of knowing are in fact correlated with subsequent recognition performance (e.g., Blake, 1973; Eysenck, 1979; Freedman & Landauer, 1966; Hart, 1965, 1967; Koriat & Lieblch, 1974; Nelson, 1984; Nelson, Leonesio, Shimamura, Landwehr, & Narens, 1982; Schacter, 1983; Schacter & Worling, 1985).

The feeling of knowing is well suited to the present concerns for three reasons. First, a feeling of knowing about an unrecalled event can be assessed in the absence of any knowledge of the contents of the event; one can simply ask subjects to rate the strength of their feeling of knowing that they could recall or recognize a forgotten event under specified circumstances. Second, it seems unlikely that simulators would have a great deal of knowledge concerning the kinds of feeling-of-knowing ratings that would be made by subjects who are genuinely unable to remember a particular episode. Although this point admittedly represents an unsubstantiated conjecture, there is no evidence that is contrary to this speculation. Third, the tendency to report a feeling of knowing about an unrecalled event can be influenced experimentally (Nelson et al., 1982; Schacter, 1983). Thus, it is possible that an experimental manipulation would have different effects on feelings of knowing in genuinely forgetful subjects and in simulating subjects. In the present experiments, feeling-of-knowing ratings of genuinely forgetful subjects and simulating subjects were assessed under a wide range of conditions. The experiments are exploratory ones whose main purpose is to sample feeling-of-knowing ratings in a variety of circumstances and to determine whether any reliable differences in the types of ratings made by genuine and simulating subjects can be detected.

Overview of the Experiments

The general strategy invoked in the present experiments was to devise a situation in which one group of subjects is exposed to a target event by a first experimenter (Experimenter A), and then questioned by a second experimenter (Experimenter B) about an aspect of the event that they are genuinely unable to remember. A second group of subjects is exposed to the same event. In this group, however, Experimenter A supplies the correct answer to the question that is subsequently asked by Experimenter B and instructs subjects to try to convince Experimenter B (who is blind concerning their status) that they cannot remember the answer. The former group of subjects represent an analogue of a real-life situation in which a person has genuinely forgotten a specific episode. The latter group represents an analogue of a

situation in which a person claims to have forgotten a particular episode, even though he or she in fact remembers it.

The experiments were comprised of five components. The first was an input phase in which Experimenter A exposed subjects to relatively complex materials—an excerpt from a novel in Experiment 1 and a videotaped story in Experiments 2 and 3. Each set of materials included a critical event or events that subjects were genuinely unable to recall. Forgetting of the critical event was induced by using materials that contained an arousing, highly salient incident, and asking subjects about an event that occurred just prior to the arousing incident. This strategy was suggested by studies that have demonstrated that subjects have difficulty remembering events that occur just prior to a highly salient incident (e.g., Detterman, 1976; Loftus & Burns, 1982; Tulving, 1969).

The second component was an instructional phase that followed exposure to the input materials. At this point, Experimenter A instructed subjects in the genuine condition to try to recall the target event when questioned about it by Experimenter B, and instructed subjects in the simulating condition to try to convince Experimenter B that they had forgotten the event. The third component occurred immediately after subjects were asked about the critical event by Experimenter B and reported that they could not remember it. Experimenter B then asked subjects to make feeling-of-knowing ratings concerning the likelihood that they could gain access to the event under different test conditions. Three successive feeling-of-knowing ratings were made: free-recall, cued-recall, and recognition ratings. For the free-recall rating, subjects were asked to rate the likelihood that they would be able to remember the target event if they were given several more minutes to think of it on their own. For the cued-recall rating, subjects were asked to rate the likelihood that they could remember the target if they were given a hint or cue about it. For the recognition rating, subjects were asked to rate the likelihood that they could recognize the target if it was presented to them along with another item that had not occurred. After making these ratings, subjects were given 2 min to try to remember the target. They then made a second set of feeling-of-knowing ratings, attempted recall for a further 2 min, and subsequently made a third set of feeling-of-knowing ratings. Thus, there were three types of feeling-of-knowing ratings that were made at three different times. Although there is no a priori basis for formulating specific hypotheses concerning any expected differences between genuine and simulating subjects, this design permitted exploration of the possibility that the patterns of feeling-of-knowing ratings made by genuine and simulating subjects would be differentially affected by the type of rating or by the time of rating.

The fourth component of the experiments was the 2-min recall periods that were inserted between the sets of feeling-of-knowing ratings. During each of the two recall periods, subjects were instructed to “think out loud” and to verbalize any thoughts they had as they tried to recover the target episode. The purpose of this procedure was to determine whether the verbal protocols generated by the subjects provided a basis for distinguishing between those who had genuinely forgotten and those who were simulating. To evaluate this hypothesis, the transcribed verbal protocols were given to panels of psychologists and psychiatrists who were blind concerning the actual condition of the subjects, and who were instructed to classify each subject as either genuine or simulating.

The fifth and final component of the experiments was comprised of cued-recall and recognition tests concerning the critical event that were given after the conclusion of the thinking-out-loud procedure and the feeling-of-knowing ratings. The main purpose of these tests was to evaluate the extent to which the genuine subjects had forgotten the critical event. Although the inclusion of recall and recognition tests also makes it possible to examine the accuracy of feeling-of-knowing ratings, this was not done in the present experiments because (a) feeling-of-knowing ratings were made concerning only a single question (Experiment 1) or two questions (Experiments 2 and 3), which does not provide a reliable basis for determining feeling-of-knowing accuracy in individual subjects, (b) recall and recognition performance was at or near chance levels in most experimental conditions, and (c) assessment of feeling-of-knowing accuracy requires knowledge of the contents of the forgotten event and, as discussed earlier, such knowledge is not available in actual cases in which it is necessary to distinguish between genuine and simulated forgetting.

Experiment 1

Method

Subjects. Thirty-two University of Toronto undergraduates, 19 women and 13 men, took part in the experiment. They were paid \$5.00 for their participation.

Materials. The target materials were based on a 900-word passage from a spy novel (Ludlum, 1980, pp. 41–44) that was edited to make it suitable for the experiment. In the passage, a protagonist (Jason) enters a cafe to meet with two men about the possibility of obtaining a forged passport. After the conclusion of their discussion, Jason leaves their table and suddenly becomes involved in a violent altercation with another man in the cafe that is described vividly. The passage concludes with the termination of the fight.

In conformity with the requirements discussed earlier, a target event was chosen from a part of the episode that occurred immediately prior to the violent incident and hence would be difficult for uninstructed subjects to recall. The question is “What was the last thing that was said by the man who was referred to as ‘the connection’?” The correct answer to the question is “I’ll need a photograph.” A pilot study conducted with 16 subjects who listened to the passage and were then asked this question revealed that none of them could remember the correct answer.

Design and procedure. The experiment conformed to a $2 \times 3 \times 3$ mixed design. The between-subjects factor was instructional condition (genuine vs. simulating). Subjects were assigned randomly to either the genuine or simulating group. The within-subjects factors were type of feeling-of-knowing rating, as defined by the test about which the rating was made (free recall, cued recall, or recognition), and time of the feeling-of-knowing rating, as defined by the point in the recall period when the ratings were made (immediate, 2 min, 4 min).

Two experimenters tested the subjects. Experimenter A assigned the subjects to conditions, played them a tape of the critical passage, and provided instructions concerning the next phase of the experiment. All subjects were instructed to rate the passage concerning level of excitement and interest; no mention was made of a memory test. Subjects in the *genuine* condition were then informed that they would be asked a question about the passage by a second experimenter and that they should do their best to remember the answer. Subjects in the *simulating* condition were given similar instructions, except that they were also instructed to convince the second experimenter that they could not remember the answer to the question. It was stressed that it was important for them to do their best to appear as though they genuinely could not remember the critical incident. The experimenter emphasized that they could achieve this goal by simulating in as realistic a manner as possible. Subjects were then told the correct answer to the question to insure that they had access

to the target information during the simulation attempt, just as in real-life situations simulating subjects in some sense have access to the information that they claim they cannot remember. They were further informed that the second experimenter would not know whether they were feigning forgetting or whether they genuinely could not remember, and that they should not mention anything about the instructions that they had been given.

The subjects were then taken to the office of Experimenter B, who was blind concerning the condition to which they had been assigned. Experimenter B asked the critical question and allowed 15 s for a response. When the correct response was not given, subjects were told that they were going to rate the likelihood that they could remember the correct answer under conditions that would be specified by the experimenter. They were instructed in the use of a 7-point scale ranging from *certain that I will not remember the answer* (1), in the specified condition, to *certain that I will remember the answer* (7). Subjects used this scale to provide ratings for each of three consecutive questions that required them to assess the likelihood of remembering the answer under conditions of free recall (Would you remember the answer if you were given several minutes to think of it on your own?), cued recall (Would you remember the answer if you were given a hint?), and recognition (Would you remember the answer if it was shown to you along with an incorrect alternative?).

After completing their ratings, subjects were given 2 min to recall the answer on their own. They were encouraged to "think out loud" during this time and to verbalize any thoughts they had as they tried to remember the answer. Subjects' comments were recorded on cassette tape. After 2 min had passed, subjects made a second set of feeling-of-knowing ratings in response to the same three questions that had been posed earlier. Following these ratings, they engaged in the "think out loud" procedure for two more minutes, and then made a third set of feeling-of-knowing ratings, again in response to the same three questions.

The subjects were then given a cue in an attempt to prompt recall of the answer. The cue was "The connection said 'I'll need _____.'" Subjects were required to guess if they said they still could not remember the answer; they were not told whether their response was correct. They were then given a two-alternative forced-choice recognition test in which the correct response was *photograph* and the lure was *money*.

After completion of feeling-of-knowing ratings with Experimenter B, all subjects were debriefed by Experimenter A concerning the nature and purposes of the experiment. In addition, simulating subjects were queried to determine whether or not they could remember the correct answers and had access to them when they made their feeling-of-knowing ratings.

Judging of transcripts. The recorded protocols of each subject thinking out loud were transcribed into printed form on standard 8½" × 11" sheets of paper. The transcripts were verbatim records of the subjects' verbal output, with the exception that the three sets of feeling-of-knowing ratings and the final cued-recall and recognition tests were deleted from these protocols.

The protocols of the subjects, identified only by numbers assigned arbitrarily to each one, were given to six judges. Two of the judges were experienced forensic psychiatrists who deal regularly with cases of simulated forgetting in clinical and criminological contexts. Two judges were cognitive psychologists with professional interests in memory, and two judges were clinical neuropsychologists with experience assessing both genuine and simulated memory loss.

The judges were provided with a summary description of the experiment and were told that half of the subjects had feigned forgetting of the critical incident and that half genuinely could not remember it. They were instructed to read the transcripts carefully and to classify each subject as either a simulator or a subject who genuinely could not remember. Judges were further instructed to assign each of their classifications a confidence rating, where 1 indicated that they were *uncertain* of their judgment, 2 indicated that they were *fairly certain*, and 3 indicated that they were *certain*.

Results

Recall and recognition. None of the subjects in either condition spontaneously recalled the correct answer at any point during the thinking-out-loud procedure. After the conclusion of this procedure, 60% of the subjects in the genuine condition remembered the correct answer when given a cue (i.e., "I'll need _____."), and 87% chose the correct response on the two-alternative forced-choice recognition test (i.e., *photograph* vs. *money*). By contrast, only 30% of the subjects in the simulating group provided the correct response when given the cue, and 27% chose correctly on the recognition test. (One simulating subject was inadvertently not given a recognition test, so this proportion is based on 15 subjects.) The latter proportion is significantly below the chance expectation of 50% correct on a two-choice recognition test (binomial probability = .042, $p < .05$, for this and all other statistical tests). After the debriefing procedure, however, all simulators gave the correct answers on both tests. When those simulators who had provided the correct responses prior to debriefing on the cued-recall and recognition tests were asked why they had done so, they typically stated that they thought it would be "overdoing it" to continue to simulate forgetting in the presence of recall and recognition cues. Simulators who chose the incorrect response on the recognition test, however, frequently stated that they thought that someone who genuinely could not recall the answer would also have difficulty recognizing it.

Judging of verbal protocols. The recorded protocols of 2 subjects, one in each group, could not be transcribed because of technical problems with the tape. Thus, each judge provided classifications for 30 subjects, 15 in the genuine condition and 15 in the simulating condition.

The protocols produced by individual subjects varied in length, ranging from 72 words for the least verbal subject to 440 words for the most verbal subject. There were various types of statements in the protocols, including attempts to retell the story in sequence up until the critical incident, self-generated strategies for aiding recovery of the correct answer (e.g., "I'm trying to picture them around the table"), guesses concerning the correct answers (e.g., "Did he say he needed a gun?"), and comments concerning the difficulty of attempting to recover the target (e.g., "I've hit a block, this is really hard to do"). Two protocols, one from a simulator and one from a genuine subject, are presented in the Appendix.

Table 1 presents data concerning the accuracy of the judges' classifications. Overall, the judges correctly classified only 53% of the subjects, a level of accuracy that does not differ significantly from chance, ($\chi^2(1, N = 180) = < 1$). The chance level of performance is observed even for those cases in which the judges claimed that they were *certain* of their classification. Judges assigned the *certain* label to 40 of the 180 classifications that were made, but were accurate on just 50% of these *certain* responses. Analysis of individual judge's responses revealed an inability to discriminate between genuine and simulating subjects in each case: No single judge classified more than 18 of the 30 subjects correctly. When the *certain* responses are considered separately, there is a suggestion of accuracy in only a single case: One judge accurately classified 7 of the 9 subjects that he had assigned a *certain* rating.

Table 1
*Proportion of Subjects Classified Correctly
 by Six Judges in Experiment 1*

Type of subject	Confidence rating			Mean
	Guess	Fairly certain	Certain	
Genuine				
Proportion of subjects classified correctly	.67	.43	.50	.50
Raw no. of classifications	17	51	22	90
Simulating				
Proportion of subjects classified correctly	.43	.63	.50	.56
Raw no. of classifications	23	49	18	90
Mean				
Proportion of subjects classified correctly	.55	.53	.50	.53
Raw no. of classifications	40	100	40	180

In spite of their consistently low level of accuracy, the judges claimed to be *fairly certain* or *certain* about 78% of their classifications, and stated that they were *guessing* on only 22% of them. This observation suggests that the judges were responding to misleading or uninformative cues in the verbal protocols that they believed provided a basis for distinguishing between genuine and simulating subjects.

Analysis of the subjects' protocols suggests that the sheer quantity of verbal output—the number of words in a subject's protocol—was one of these misleading cues. The mean number of words per protocol was virtually identical in the genuine group ($M = 208.5$) and the simulating group ($M = 210.4$; $t(30) < 1$). Consider, however, the number of words produced by subjects who judges agreed were either simulators or genuine. An agreement was defined as a case in which either five or six judges assigned the same classification to a subject. By this criterion, there were 8 subjects who the judges agreed were simulators and 7 subjects who they agreed were genuinely forgetful (only 53% of these consensus choices were accurate). A word count revealed that those subjects considered to be simulators produced significantly fewer words ($M = 124.6$) than those subjects considered to be genuine ($M = 224.7$; $t(13) = 2.61$). This finding suggests that judges interpreted an impoverished protocol as a sign of simulation, even though there was in fact no difference between the number of words uttered by subjects in the genuine and simulating conditions.

In summary, the judges failed to discriminate between the two groups on the basis of the protocols, indicating that the simulating subjects were able to feign forgetting of an episode in a convincing manner.

Feeling-of-knowing ratings. The mean feeling-of-knowing ratings for both groups of subjects are displayed in Table 2.¹ For each of the three sets of ratings, both genuine and simulating subjects were least confident that they would remember the target

under conditions of free recall, were more confident that they could remember the target under conditions of cued recall, and were most confident that they could choose the target on a recognition test. This pattern of results was reflected by a highly significant main effect of type of rating, $F(2, 60) = 157.47$, $MS_e = 1.40$. There was also a trend for both groups' feeling-of-knowing ratings to decline across the recall period, with ratings in the free-recall condition dropping more precipitously than ratings in the cued-recall or recognition conditions. Analysis of variance (ANOVA) revealed a significant interaction between type of feeling-of-knowing rating and time of feeling-of-knowing rating, $F(4, 120) = 16.9$, $MS_e = .246$. The three-way interaction of Subject Group \times Type of Rating \times Time of Rating was nonsignificant, $F(4, 120) < 1$, indicating that the pattern reflected by the two-way interaction held for both subject groups.

The foregoing findings indicate that subjects in the simulating group were able to mimic a fairly complex pattern of genuine feeling-of-knowing ratings. However, the feeling-of-knowing ratings of genuine and simulating subjects did differ in a systematic way. There was a main effect of subject group on feeling-of-knowing ratings, $F(1, 30) = 4.99$, $MS_e = 9.20$, indicating that the mean ratings of the simulators were lower than the mean ratings of subjects in the genuine condition. More important, there was a significant interaction between subject group and type of feeling-of-knowing rating, $F(2, 60) = 11.58$, $MS_e = 1.40$. The interaction reflects the fact that free-recall ratings of genuine and simulating subjects were similar, whereas cued-recall and recognition ratings of simulating subjects were consistently lower than those of genuine subjects. Post hoc comparisons with a Tukey test revealed no difference between the mean free-recall ratings of subjects in the genuine and simulating groups, and revealed significant ($p < .05$) differences between cued-recall ratings and the recognition ratings of subjects in the two groups. Thus, subjects attempting to simulate forgetting provided lower estimates of their confidence that they could gain access to the inaccessible memory under conditions of cued recall and recognition than did the genuinely forgetful subjects.

These data indicate that feeling-of-knowing ratings can discriminate between groups of genuine and simulating subjects under conditions in which professional psychologists and psychiatrists are unable to do so on the basis of verbal protocols of subjects' retrieval attempts. It must be kept in mind, however, that the differences observed in the two groups' feeling-of-knowing ratings were based on subjects' responses to a single question about a verbally described event, and that the feeling-of-knowing

¹ One cannot assume that an interval scale underlies feeling-of-knowing ratings, hence it would be appropriate to present median ratings and base statistical inferences on the results of nonparametric tests. All of the feeling-of-knowing data in the present experiments have been analyzed with a parametric method (ANOVA) and a nonparametric method (median test). The patterns of data are identical whether means or medians are used, and the statistical conclusions are the same whether parametric or nonparametric tests are used. However, because interactions are of central interest in the present experiments, the results can be described far more clearly in terms of ANOVA than in terms of median tests. Thus, in the interests of expositional clarity, means are presented and ANOVAs are described. See Maki and Berry (1984) for a similar strategy of treating rating scale data.

Table 2
Mean Feeling-of-Knowing Ratings of Genuine and Simulating Subjects in Experiment 1

Time of rating (min)	Type of rating					
	Genuine			Simulating		
	FR	CR	RN	FR	CR	RN
0	3.3	5.1	6.3	3.0	4.0	4.4
2	2.1	4.6	6.2	2.3	3.9	4.6
4	1.3	4.2	5.8	1.5	3.5	4.4
<i>M</i>	2.2	4.6	6.1	2.3	3.8	4.5

Note. FR = free recall; CR = cued recall; RN = recognition.

ratings were made immediately after hearing the passage. It is conceivable that the positive results depend critically on the specific question that was asked, the particular study materials that were used, or the fact that ratings were made immediately after exposure to the target episode. To determine the generality of the results observed in Experiment 1, it is necessary to investigate whether they can be replicated with different input materials and test questions, as well as with a delay interpolated between study and test. These issues were examined in Experiment 2.

Experiment 2

The procedure used in Experiment 2 was generally similar to the one used in Experiment 1, with three differences. First, instead of listening to an excerpt from a novel, subjects were shown a 30-min videotape that depicted a camping excursion marked by a brief violent episode. This tape was used in an attempt to model more closely a situation that might occur in everyday life. The tape included a critical, highly salient incident, so the strategy of inducing subjects to forget the events that occurred just prior to the critical incident could be used again. The second difference was that subjects were asked two questions about different incidents in the tape to reduce the possibility that the findings could be attributed to some idiosyncratic property of a single question. The third difference was that half of the subjects were tested after a 90-min delay. The delay was included to reduce the accessibility of the target event. Sixty percent of the genuine subjects in Experiment 1 remembered the target in the presence of a cue and 87% of them recognized it on the two-choice test. Thus, it could be argued that these subjects had not truly forgotten the target event. It is important to determine whether feeling-of-knowing ratings provide a basis for distinguishing between genuine and simulating subjects under conditions in which cued-recall and recognition performance are lower than in Experiment 1.

Method

Subjects. Forty-eight University of Toronto undergraduates, 26 men and 22 women, participated in the experiment. They were paid \$5.00 for their participation.

Materials. Subjects viewed a 30-min color videotape entitled *Doing It Wrong*. The tape is a professionally acted dramatization that was made by the Ontario Ministry of Health as a warning to teenagers concerning the dangers of drinking and submitting to peer pressure. The tape focusses on a group of four teenaged friends who embark on a camping weekend and become acquainted with another teenaged foursome who are por-

trayed as irresponsible rowdies. After the group becomes thoroughly intoxicated at the campsite, they break into a vacant cottage and vandalize it. The key incident in the tape occurs in the midst of this activity, when one of the rowdies attempts to rape a girl from the other group. She escapes and becomes seriously injured while running through the woods, and the police are ultimately brought in to find her.

The salient incident (the attempted rape) occurs about 20 min into the tape. The two critical questions pertain to events that occur just prior to this incident: Question A—What is the last thing that the bearded man, Sandy, said before the victim was attacked? (Answer—Did you see his face?); Question B—What did the dark-haired woman, Cathy, say after Bruce threw his hat down when they were in the cottage? (Answer—Hey, what are you doing taking my booze?) These questions were selected on the basis of a pilot study conducted with 12 subjects who viewed the tape and were asked several questions immediately afterward. None of them provided the correct answer to the two questions just mentioned.

Design and procedure. The experiment was a $2 \times 2 \times 3 \times 3$ mixed design. The between-subjects factors were instructional condition (genuine vs. simulating) and retention interval (immediate vs. 90-min delay). Subjects were assigned randomly to one of the four groups of subjects that were formed by the orthogonal combination of these two factors. As in Experiment 1, the within-subjects factors were type of feeling-of-knowing rating and time of feeling-of-knowing rating.

Assignment of subjects to conditions and testing of subjects was handled by two different experimenters in the same manner as described in Experiment 1. Subjects first viewed the videotape and were told that they would be asked to rate it concerning the level of excitement and interest. After conclusion of the tape, Experimenter A instructed subjects in the immediate condition in the same manner as was described in Experiment 1. Subjects in the delay condition were told to return to the laboratory in 90 min and were then given the identical instructions by Experimenter A.

When subjects were taken to the office of Experimenter B, they were given the same instructions as were described for Experiment 1 and followed the same procedure. The only change was that the entire procedure was repeated with a second question. For half of the subjects in each condition, Question A was presented first, and for the other half, Question B was presented first. At the conclusion of the three sets of feeling-of-knowing ratings for each question, a cued-recall test was given that contained an added hint concerning the answer, followed by a two-alternative forced-choice recognition test that included the correct answer as well as a similar lure item.

The taped protocols of each subject's thinking out loud were transcribed in the same manner as was described in Experiment 1. They were given to four judges. Two were forensic psychiatrists who had participated in Experiment 1. The other two judges had not taken part in Experiment 1. One of them was a cognitive psychologist with research interests in normal memory, and one was a psychometrist with extensive experience testing patients who have memory difficulties. The judges were provided with a detailed transcript of the videotape as well as a summary of the experimental procedure. The instructions to judges were the same as those in Experiment 1.

Results

Cued recall and recognition. None of the subjects recalled the correct response to either of the questions at any point during the thinking out loud procedure.

The results of the cued-recall and recognition tests that were administered at the conclusion of the thinking out loud period are presented in Table 3. Consider first the results with subjects in the genuine condition. On the cued-recall test, almost all subjects were unable to provide the correct answers to the questions at each of the retention intervals. On the two-choice recognition test, performance did not deviate from chance expectation of

50% correct except on Question A in the immediate condition, which 83% of subjects answered correctly. The fact that delayed recognition performance on both questions was at chance indicates that the interpolation of the 90-min delay had the intended effect of creating conditions in which genuinely forgetful subjects could neither recall nor recognize the correct answers.

Simulating subjects, like genuine subjects, revealed a low level of performance on the cued-recall test (Table 3). On the recognition test, however, only 17% of the simulators chose the correct answers on the immediate test; on the delayed test, 17% chose the correct answer to Question A and 8% chose the correct answer to Question B. All of these values are significantly below the chance level (binomial probability $< .05$ in all cases). Thus, as in Experiment 1, the simulators as a group overplayed their role by scoring significantly below chance on a recognition test.

Judging of verbal protocols. The data concerning the judges' classifications of genuine and simulating subjects are presented in Table 4. These data are collapsed across the immediate and delay conditions because a preliminary analysis revealed no differences in the accuracy of judges' classifications of subjects in the two groups. Overall, the judges classified only 53% of the subjects accurately. As in Experiment 1, this value did not exceed chance expectation, $\chi^2(1, N = 192) < 1$, and chance levels of classification were observed for each of the three types of confidence ratings. The judges did, however, make a smaller proportion of *certain* classifications (8%) than did the judges in Experiment 1 (22%). Nevertheless, it should be noted that judges rated 65% of their choices as *fairly certain* or *certain* and that the accuracy of these choices did not exceed chance.

Analysis of the number of words in the protocols indicated that the 6 subjects who judges agreed were simulators (where an agreement is defined as consensus among three or four judges) engaged in less verbal activity than the 6 subjects who judges agreed were genuine ($M = 182.7$ words vs. $M = 216.3$ words). The difference between the two means was in the same direction as was observed in Experiment 1, but it failed to achieve statistical significance, $t(10) = 1.38, p < .10$. The actual number of words produced by simulating and genuine subjects was virtually identical. A 2×2 ANOVA performed on the number of words in the protocols revealed no effects of subject group or retention interval, and no interaction between these variables ($F < 1$ in all cases).

Feeling-of-knowing ratings. The mean feeling-of-knowing ratings are displayed in Table 5. An initial ANOVA was performed in which the individual questions (Question A vs. Question B) were treated as a factor with two levels. There was a nonsignificant main effect of question type on feeling-of-knowing ratings, and there were nonsignificant interactions with all other factors and combinations of factors ($F < 1$ in all cases). These results indicate that the patterns of data to be described hold for both of the critical questions. Thus, the feeling-of-knowing data displayed in Table 5 represent mean ratings to the two questions, and all subsequent analyses were performed on these ratings.

The data in Table 5 replicate the key findings of Experiment 1. As in Experiment 1, simulators successfully mimicked some features of the ratings made by genuinely forgetful subjects. For all subjects, recognition ratings were higher than cued-recall ratings, which were in turn higher than free-recall ratings, as indicated by a highly significant main effect of type of rating, $F(2, 88) = 247.86, MS_e = 2.74$. In addition, there was a significant interaction between time of feeling-of-knowing rating and type

Table 3

Proportion of Subjects Who Provided the Correct Response on the Cued-Recall and Recognition Tests in Experiment 2

Time of test	Genuine		Simulating	
	CR	RN	CR	RN
Question A				
Immediate	.08	.83	.00	.17
Delay	.00	.58	.00	.17
Question B				
Immediate	.00	.58	.00	.17
Delay	.08	.50	.00	.08

of feeling-of-knowing rating, $F(4, 176) = 5.42, MS_e = .338$, reflecting the fact that the decline of feeling-of-knowing ratings across the recall period was most pronounced for free-recall ratings, less pronounced for cued-recall ratings, and least pronounced for recognition ratings. The three-way interaction of Type of Feeling-of-Knowing Rating \times Time of Feeling-of-Knowing Rating \times Retention Interval was nonsignificant, $F(4, 176) = 1.37, MS_e = .338$, indicating that this pattern was observed at both retention intervals. More important, the three-way interaction of Subject Group \times Time of Feeling-of-Knowing Rating \times Type of Feeling-of-Knowing Rating was nonsignificant, $F(4, 176) = 1.08, MS_e = .338$, as was the four-way interaction of Subject Group \times Retention Interval \times Time of Feeling-of-Knowing Rating \times Type of Feeling-of-Knowing Rating, $F(4, 176) < 1, MS_e = .338$. These nonsignificant interactions with subject group indicate that the differential decline of the three types of ratings across the recall period occurred in simulators as well as in genuine subjects.

Although these findings highlight the similarities between genuine and simulating subjects, feeling-of-knowing ratings once again provided a basis for distinguishing between the two groups. There was a significant interaction between subject group and type of feeling-of-knowing rating, $F(2, 88) = 11.28, MS_e = 2.74$, and the three-way interaction of Subject Group \times Type of Feeling-of-Knowing Rating \times Retention Interval was nonsignificant, $F(2, 88) < 1, MS_e = 2.74$. These analyses reflect the fact that at both retention intervals, free-recall ratings of genuine and simulating subjects were virtually identical, whereas cued-recall and recognition ratings of the simulators were lower than those of genuine subjects. Planned comparisons performed on the mean feeling-of-knowing ratings of genuine and simulating subjects revealed no difference between free-recall ratings of the two groups, $t(46) < 1$. In contrast, recognition ratings of simulators were significantly lower than those of genuine subjects, $t(46) = 3.74$, and so were cued-recall ratings, $t(46) = 2.73$.

To summarize, Experiment 2 replicated and extended the main findings of Experiment 1: Feeling-of-knowing ratings distinguished between genuine and simulating subjects at both immediate and delayed tests, even though judges could not discriminate between the two groups of subjects on the basis of verbal protocols. Although simulating subjects, like subjects in the genuine condition, judged that provision of more free-recall time would be of little help, they underestimated the confidence that genuinely forgetful subjects had that they could recover the sought-after memory in the presence of cues. The fact that this

Table 4
Proportion of Subjects Classified Correctly
by Four Judges in Experiment 2

Type of subject	Confidence rating			Mean
	Guess	Fairly certain	Certain	
Genuine				
Proportion of subjects classified correctly	.55	.60	.38	.55
Raw no. of classifications	33	55	8	96
Simulating				
Proportion of subjects classified correctly	.56	.43	.75	.50
Raw no. of classifications	34	54	8	96
Mean				
Proportion of subjects classified correctly	.55	.51	.56	.53
Raw no. of classifications	67	109	16	192

pattern was observed with a different type of input material than was used in Experiment 1, with different test questions, and was observed at a 90-min delay as well as in the immediate condition, suggests that these findings have some generality. It is important to note, however, that the 90-min delay used in Experiment 2 was a relatively short one. In real-life cases, the interval between the critical incident and the time at which memory is probed may be a good deal longer than 90 min. It would thus be desirable to determine whether the observed pattern of results holds over a longer delay, one that is closer to what might be encountered in at least some everyday situations. This issue was examined in Experiment 3 by testing subjects at a 24-hr delay.

A further question concerning the generality of the results reported thus far centers on the 1-7 scale that has been used to assess feeling-of-knowing ratings. We do not know whether the observed outcomes in some way depend on an idiosyncratic property of this scale. Experiment 3 examined whether the same pattern of results is observed when a different scale is used to measure feeling-of-knowing ratings.

Experiment 3

Method

Subjects and materials. Twenty-four University of Toronto undergraduates, 15 women and 9 men, took part in the experiment. They were paid \$5.00 for their participation. The videotape from Experiment 2 was used again, and the same critical questions were also used.

Design and procedure. The design was similar to the one used in Experiment 2, with four changes. First, all subjects were tested at a 24-hr delay. Second, feeling-of-knowing ratings were made only with respect to free-recall and recognition conditions; the cued-recall rating was dropped. This was done both to simplify the procedure and to determine whether successful discrimination between genuine and simulating subjects depends on including a cued-recall rating. Because the cued-recall rating was eliminated, the subsequent cued-recall test was also dropped; only the two-choice recognition test was given after feeling-of-knowing ratings were made. Third, subjects made their feeling-of-knowing ratings on a different scale than the one used in Experiments 1 and 2. Instead of the

1-7 scale, a 1-21 scale was used, ranging from *certain that I will not remember the answer* (1), in the specified condition, to *certain that I will remember the answer* (21). The fourth and final difference between Experiment 3 and the two preceding ones is that judging of verbal protocols was not done. In view of the fact that judges performed at the chance level in Experiments 1 and 2, and that the present focus is on delineating the range of conditions under which feeling-of-knowing ratings discriminate genuine from feigned forgetting, the judging procedure was deemed unnecessary. All other aspects of the design and procedure were identical to Experiment 2. Thus, the design of this experiment can be summarized as a $2 \times 2 \times 3$ (Subject Group \times Type of Feeling-of-Knowing Rating \times Time of Feeling-of-Knowing Rating) design.

Results

Recall and recognition. None of the subjects recalled the correct answer to either question during the thinking-out-loud procedure.

In the genuine condition, 54% of subjects chose the correct response to Question A on the recognition test, and 50% chose the correct response to Question B. Thus, recognition performance was once again at or near the chance level. For the simulators, recognition performance was again significantly below chance: 17% of simulating subjects chose the correct answer for Question A ($p = .016$), and 8% chose correctly for Question B ($p = .003$).

Feeling-of-knowing ratings. Mean feeling-of-knowing ratings are displayed in Table 6. A preliminary ANOVA that included question type as a factor revealed a nonsignificant main effect and nonsignificant interactions with other factors ($F < 1$ in all cases). Thus, the data displayed in Table 6 represent mean ratings to the two questions, and all subsequent analyses were performed on these ratings.

The data in Table 6 differ from those in the previous two experiments in that recognition ratings are numerically a good deal higher than the ones displayed earlier, reflecting the wider scale that was used in the present experiment. Free-recall ratings, however, were numerically similar to the previous ratings.

The most important feature of the data in Table 6 is that they

Table 5
Mean Feeling-of-Knowing Ratings of Genuine and Simulating
Subjects in Experiment 2

Time of rating (min)	Type of rating					
	Genuine			Simulating		
	FR	CR	RN	FR	CR	RN
Immediate test						
0	2.5	4.2	5.5	2.1	2.9	3.8
2	1.6	3.7	5.5	1.5	2.4	3.8
4	1.3	3.6	5.4	1.3	2.1	3.5
<i>M</i>	1.8	3.8	5.5	1.6	2.5	3.7
Delayed test						
0	1.9	3.7	5.5	1.8	3.0	4.5
2	1.5	3.3	5.3	1.7	3.0	4.5
4	1.3	3.1	5.3	1.6	3.0	4.4
<i>M</i>	1.6	3.4	5.3	1.7	3.0	4.5
Grand <i>M</i>	1.7	3.6	5.4	1.7	2.8	4.1

Note. FR = free recall; CR = cued recall; RN = recognition.

replicate the main findings of the previous experiments. Recognition ratings were a good deal higher than free-recall ratings in both groups, as indicated by a significant main effect of type of rating, $F(1, 20) = 113.98$, $MS_e = 60.31$. There was also a significant Type of Rating \times Time of Rating interaction, $F(2, 40) = 4.25$, $MS_e = 4.26$, indicating that free-recall ratings dropped more precipitously across the recall interval than did recognition ratings. The three-way interaction of Subject Group \times Type of Rating \times Time of Rating was nonsignificant, $F(2, 40) < 1$, $MS_e = 4.26$, indicating that the differential decline of free-recall versus recognition ratings was characteristic of both subject groups. Simulators in this experiment, like those in the previous two experiments, were in this respect able to mimic the pattern of feeling-of-knowing ratings made by genuinely forgetful subjects.

In spite of these similarities, there was once again a significant Subject Group \times Type of Feeling-of-Knowing Rating interaction, $F(1, 20) = 5.43$, $MS_e = 60.31$. The interaction indicates that free-recall ratings of the two groups did not differ, whereas recognition ratings of the simulators were lower than those of controls. Thus, even at a 24-hr delay, simulators underestimated the confidence that a genuinely forgetful subject would have that he or she could recognize the correct answer.

General Discussion

The results of the present research indicate that feeling-of-knowing ratings provide a basis for distinguishing between groups of genuine and simulating subjects under conditions in which psychiatrists and psychologists did not discriminate between the two types of subjects on the basis of verbal protocols of retrieval attempts. The phenomenon that provided a basis for distinguishing the two groups of subjects—simulating subjects provided lower cued-recall and recognition ratings, relative to free-recall ratings, than did genuine subjects—was observed in three separate experiments that differed with respect to subjects, materials, questions, test procedures, retention intervals, and measurement scales. The occurrence of the critical phenomenon across all of these changes indicates that it is a robust one. Let us now consider some of the practical and theoretical implications of the present results.

The results discussed thus far have been based on differences between groups of subjects. However, in actual cases that are encountered in everyday life, it is necessary to make accurate discriminations concerning individual subjects. To make such discriminations, it would be desirable and perhaps necessary to possess a measure on which genuine and simulating subjects are characterized by nonoverlapping distributions of scores. Although the present research indicates that differences in patterns of feeling-of-knowing ratings provided a basis for distinguishing between groups of subjects, there was a considerable amount of overlap in the ratings made by individual subjects in the two groups. Thus, it is not yet possible to use feeling-of-knowing ratings to determine definitively whether or not an individual subject is simulating or is genuinely forgetful. An important task for future research will be to determine how the feeling-of-knowing procedure used in the present experiments can be refined so that it yields nonoverlapping distributions for genuine and simulating subjects.

It should also be kept in mind that only a single index was used to distinguish between the genuine and simulating groups in this experiment. In everyday life, one would not want to make

Table 6
Mean Feeling-of-Knowing Ratings of Genuine and Simulating Subjects in Experiment 3

Time of rating (min)	Type of rating			
	Genuine		Simulating	
	FR	RN	FR	RN
0	4.2	14.7	3.4	10.4
2	1.6	14.3	2.2	10.1
4	1.3	13.8	1.8	9.7
<i>M</i>	2.4	14.3	2.5	10.1

Note. FR = free recall; RN = recognition.

discriminations about an individual solely on the outcome of a single measure, however accurate it might be; one would want to have multiple measures that provide converging information concerning the individual's status. What measures might be used to supplement feeling-of-knowing ratings? There is a good deal of evidence in the deception literature that visual and auditory cues provide useful information for detecting deception (e.g., Ekman & Friesen, 1969, 1974; Streeter, Kraus, Geller, Olson, & Apple, 1977; Zuckerman, Amidon, Bishop, & Pomerantz, 1982; see Zuckerman, DePaulo, & Rosenthal, 1981, for review). Perhaps these cues could be combined with feeling-of-knowing ratings to yield a classification of individual subjects that has potential for application in everyday life.

There are a number of other uncertainties concerning the generalizability of the present results that should also be acknowledged. For instance, it is possible that the findings are specific to the population of college students that was used. It would thus be important to determine whether these data can be replicated with subject populations that more closely resemble those that are typically encountered in real-life contexts. Another difficulty is that in cases of simulated forgetting, the critical episode is frequently an emotional one, whereas in the present experiments, the critical episode consisted of conversational details and exchanges that had no emotional significance for the subjects. Perhaps the present results would not be obtained when emotional events are involved. Although this issue remains to be explored, the more general hypothesis at stake here is that the pattern of feeling-of-knowing judgments made about a forgotten event depends on the quality or contents of that event. The fact that the results of Experiment 1 were virtually identical to those of Experiments 2 and 3, even though the contents of critical incidents differed, speaks against this hypothesis. Future studies could test this notion further by varying the contents and quality of the critical incidents and examining whether feeling-of-knowing ratings are sensitive to these changes.

The foregoing discussion highlights the fact that it is not yet possible to generalize the data concerning the feeling-of-knowing beyond the bounds of the present research. The same caution also applies to the inability of judges in Experiments 1 and 2 to discriminate between genuine and simulating subjects. It is conceivable that the same judges could have discriminated between the two groups of subjects had they seen and listened to tapes of subjects during the thinking-out-loud procedure, or had they been able to interview the subjects themselves. As noted earlier, the deception literature indicates that visual and auditory cues

are important for detecting deception; indeed, they may be more useful than verbal cues alone (DePaulo, Lanier, & Davis, 1983).

Let us now turn to the theoretical implications of the present research. It was noted in the introduction that successful simulation may depend on subjects' beliefs and knowledge concerning forgetting of individual episodes. The term *metamemory* has been used to describe the knowledge, beliefs, and assumptions that people have about the characteristics of memory function (e.g., Flavell & Wellman, 1977). Experimental studies have revealed that both children and adults possess a good deal of knowledge concerning the conditions under which memory is likely to succeed and to fail (see Cavanaugh & Perlmutter, 1982, and Flavell & Wellman, 1977, for review). In addition, there is evidence that people possess a variety of beliefs concerning numerous attributes of their own memory function (Schulster, 1981). The task for simulators in the present experiments can be conceptualized as one that draws heavily on metamnemonic knowledge: Subjects must make use of their beliefs and knowledge concerning characteristics of forgetting in order to construct a pattern of feeling-of-knowing ratings that they think is characteristic of a genuinely forgetful person. Viewed from this perspective, the finding that simulators gave lower cued-recall and recognition ratings than did genuine subjects implies a failure of metamemory: Simulators did not know, and apparently did not infer, just how confident a genuinely forgetful person would remain that he or she could gain access to the target in the presence of cues. Simulators' feeling-of-knowing ratings were, however, generally higher in the recognition condition than in the cued-recall condition, and generally higher in the cued-recall condition than in the free-recall condition. These results suggest that simulators had a "general idea" about the kinds of feeling-of-knowing ratings that would be made by a genuinely forgetful subject, but that their metamnemonic knowledge was not precise enough to simulate in an entirely convincing manner. By contrast, other features of the data suggest that the simulators' metamemory was quite accurate in at least two respects. First, the free-recall ratings of the two groups did not differ. Second, free-recall ratings declined more across the recall interval than did either cued-recall or recognition ratings in both subject groups. This pattern of results raises a number of questions: How did the simulators "know" what kind of free-recall ratings would be provided by a genuinely forgetful subject? How did they "know" that free-recall, cued-recall, and recognition ratings should decline differentially across the recall period? And, in view of these metamnemonic successes, why did metamemory fail with respect to the level of cued-recall and recognition ratings?

The present research does not provide a basis for answering these questions, but some suggestions can be made regarding tactics for investigating them. One possibility is that the simulators had some pre-established, explicitly formulated beliefs about the characteristics of genuine forgetting that they consulted when making their feeling-of-knowing ratings. However, it seems unlikely that prior to the experiment, subjects would have formulated specific ideas concerning the attributes of feelings of knowing that accompany forgetting of an episode. A second, more likely possibility is that simulators attempted to infer how a genuinely forgetful subject would perform by calling to mind their own experiences and observations of forgetting in everyday life. For example, after being debriefed about the experiment many simulators reported that they tried to recall actual situa-

tions in which they had forgotten an event, or imagined that they had been assigned to the genuine condition of the experiment and actually could not remember the answer, and then tried to infer what kinds of ratings they would have made in those circumstances. Others reported that they tried to construct some sort of prototype of a person who is unable to remember a designated event, and then infer what such a person would do if asked to make feeling-of-knowing ratings. The puzzle posed by the present results concerns why these kinds of metamnemonic strategies led to accurate simulation of certain aspects of feeling-of-knowing ratings, but not of others. Perhaps future studies could focus specifically on the kinds of strategies reported by simulators, with a view toward distinguishing between those that lead to successful simulation and those that do not. More generally, these considerations suggest that studies in which subjects attempt to simulate an aspect of memory failure may provide a useful tool for determining the conditions under which metamemory is accurate or inaccurate, and may also provide clues concerning the sources of metamnemonic knowledge.

The poor performance of the judges in the present experiments can also be conceptualized in terms of metamnemonic function. Both Experiments 1 and 2 indicated that judges tended to call subjects simulators when the verbal protocol was relatively impoverished and tended to call subjects genuine when the protocol contained a lot of verbal activity. Informal inspection of the protocols indicates that the verbal subjects were those who engaged in a variety of recall strategies, generated possible retrieval cues, and generally took an active approach to recovering the inaccessible memory. This observation suggests that the judges believed that an active approach is characteristic of someone who is genuinely unable to remember an episode, whereas a passive approach is characteristic of simulators, and that they made their judgments at least partly on that basis. What the judges apparently did not take into account, however, is that (a) many genuine subjects do not use active strategies and (b) many simulating subjects may have had the same beliefs about genuine forgetting as they did, and hence attempted to appear "active" during the recall period. Thus, one possible reason why the judges failed to discriminate accurately is that they drew upon the same metamnemonic beliefs about characteristics of memory loss that many of the simulators did. As noted in the introduction, there are no well-established facts concerning the features of genuine versus simulated forgetting of a single episode. Accordingly, the judges probably did not have access to any specialized information concerning genuine forgetting that would be unknown to simulators and could thus provide a reliable basis for detecting them. Though speculative, the hypothesis that judges relied on the same set of metamnemonic beliefs as did simulators could be investigated by systematically manipulating the contents of artificially created verbal protocols and giving them to judges with instructions similar to the ones used in the present study. More generally, there are two reasons why the study of classificatory responses made by judges to genuine and simulating subjects is an issue worthy of investigation in its own right. First, such studies provide another method for investigating the nature, accuracy, and origin of people's metamnemonic beliefs and assumptions. Second, expert testimony concerning simulated forgetting can have a significant influence on the outcome of a legal case. However, there is no evidence that experts can reliably detect simulated psychological or psychiatric symptoms (Alpert, Fox, & Kahn, 1980;

Heaton, Smith, Lehman, & Vogt, 1978; Resnick, 1984). The results of Experiments 1 and 2 indicate that when psychologists and psychiatrists judge protocols of genuine and simulating subjects, there can be a significant disparity between the certainty and accuracy of their choices: The judges performed at the chance level even when they claimed to be certain that they were correct. We do not know, of course, whether a similar disparity between subjective certainty and accuracy takes place when experts give testimony in actual cases. However, disparities between accuracy and certainty are known to occur in many different situations (e.g., Fischhoff, Slovic, & Lichtenstein, 1977; Tulving, 1981; Wells & Murray, 1984), and the present results at least raise the possibility that in real-life situations experts may sometimes express certainty in their judgments concerning the genuineness of forgetting even though they are inaccurate.

In conclusion, the present research has revealed that even in a relatively simple laboratory situation, it is difficult to distinguish between genuine and simulated forgetting. Although feeling-of-knowing ratings did provide a basis for discriminating between the two groups, the successes of the simulators were just as striking as their failures. These findings suggest that metamemory can be surprisingly precise. It is not entirely accurate, however, and the subtle imperfections of metamemory can be exploited to distinguish between genuine and simulated forgetting.

References

- Adatto, C. P. (1949). Observations on criminal patients during narcoanalysis. *Archives of Neurology and Psychiatry*, *62*, 82-92.
- Alpert, S., Fox, H. M., & Kahn, M. W. (1980). Faking psychosis on the Rorschach: Can expert judges detect malingering? *Journal of Personality Assessment*, *44*, 115-119.
- 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.
- Bradford, J. W., & Smith, S. M. (1979). Amnesia and homicide: The Padola case and a study of thirty cases. *Bulletin of the American Academy of Psychiatry and Law*, *7*, 219-231.
- Cavanaugh, J. C., & Perlmutter, M. (1982). Metamemory: A critical examination. *Child Development*, *53*, 11-28.
- DePaulo, B. M., Lamer, K., & Davis, T. (1983). Detecting the deceit of the motivated liar. *Journal of Personality and Social Psychology*, *45*, 1096-1103.
- Detterman, D. K. (1976). The retrieval hypothesis as an explanation of induced retrograde amnesia. *Quarterly Journal of Experimental Psychology*, *28*, 623-632.
- Ekman, P., & Friesen, W. V. (1969). Nonverbal leakage and clues to deception. *Psychiatry*, *32*, 88-105.
- Ekman, P., & Friesen, W. V. (1974). Detecting deception from the body or face. *Journal of Personality and Social Psychology*, *29*, 288-298.
- Eysenck, M. W. (1979). The feeling of knowing a word's meaning. *British Journal of Psychology*, *70*, 243-251.
- Fischhoff, B., Slovic, P., & Lichtenstein, S. (1977). Knowing with certainty: The appropriateness of extreme confidence. *Journal of Experimental Psychology: Human Perception and Performance*, *3*, 552-564.
- Flavell, J., & Wellman, H. (1977). Metamemory. In R. V. Kail & J. W. Hagen (Eds.), *Perspectives on the development of memory and cognition*. Hillsdale, NJ: Erlbaum.
- Freedman, J. L., & Landauer, T. K. (1966). Retrieval of long-term memory: "Tip-of-the-tongue" phenomenon. *Psychonomic Science*, *4*, 309-310.
- Gibbins, T. C. N., & Williams, J. E. H. (1977). Medicolegal aspects of amnesia. In C. W. M. Whitty & O. L. Zangwill (Eds.), *Amnesia*. London: Butterworths.
- Guthkelch, A. N. (1980). Posttraumatic amnesia, post-concussional symptoms and accident neurosis. *European Neurology*, *19*, 91-102.
- Guttmacher, M. S. (1955). *Psychiatry and the law*. New York: Grune & Stratton.
- 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.
- Heaton, R. K., Smith, H. H., Lehman, R. A. W., & Vogt, A. T. (1978). Prospects for faking believable deficits on neuropsychological testing. *Journal of Consulting & Clinical Psychology*, *46*, 892-900.
- Hopwood, J. S., & Snell, H. K. (1933). Amnesia in relation to crime. *Journal of Mental Science*, *79*, 27-41.
- Koriat, A., & Lieblich, I. (1974). What does a person in a "TOT" state know that a person in a "don't know" state doesn't know? *Memory & Cognition*, *2*, 647-655.
- Koson, D., & Robey, A. (1973). Amnesia and competency to stand trial. *American Journal of Psychiatry*, *130*, 588-592.
- Leitch, A. (1948). Notes on amnesia in crime for the general practitioner. *Medical Press*, *219*, 459-463.
- Loftus, E. F. (1979). *Eyewitness testimony*. Boston, MA: Harvard University Press.
- Loftus, E. F., & Burns, T. E. (1982). Mental shock can produce retrograde amnesia. *Memory & Cognition*, *10*, 318-323.
- Ludlum, R. (1980). *The Bourne identity*. New York: Richard Marek Publishers, Inc.
- Maki, R. H., & Berry, S. L. (1984). Metacomprehension of text material. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*, 663-679.
- Neisser, U. (1981). John Dean's memory: A case study. *Cognition*, *1*-22.
- Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin*, *95*, 109-133.
- Nelson, T. O., Leonesio, R. J., Shimamura, A. P., Landwehr, R. F., & Narens, L. (1982). Overlearning and the feeling of knowing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *8*, 279-288.
- O'Connell, B. A. (1960). Amnesia and homicide. *British Journal of Delinquency*, *10*, 262-276.
- Power, D. J. (1977). Memory, identification and crime. *Medicine, Science, and the Law*, *17*, 132-139.
- Resnick, P. J. (1984). The detection of malingered mental illness. *Behavioral Sciences and the Law*, *2*, 21-37.
- Schacter, D. L. (1983). Feeling of knowing in episodic memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *9*, 39-54.
- Schacter, D. L. (in press-a). Amnesia and crime: How much do we really know? *American Psychologist*.
- Schacter, D. L. (in press-b). On the relation between genuine and simulated amnesia. *Behavioral Sciences & the Law*.
- Schacter, D. L., & Worling, J. R. (1985). Attribute information and the feeling of knowing. *Canadian Journal of Psychology*, *39*, 467-475.
- Sehulster, J. R. (1981). Structure and pragmatics of a self-theory of memory. *Memory & Cognition*, *9*, 263-276.
- Spanos, N. P., Radtke-Bodorik, L., & Stam, H. J. (1980). Disorganized recall during suggested amnesia: Fact not artifact. *Journal of Abnormal Psychology*, *89*, 1-19.
- Streeter, L. A., Kraus, R. M., Geller, V., Olson, C., & Apple, W. (1977). Pitch changes during attempted deception. *Journal of Personality and Social Psychology*, *35*, 345-350.
- Taylor, F. K. (1965). Cryptomnesia and plagiarism. *British Journal of Psychiatry*, *111*, 1111-1118.
- Tulving, E. (1969). Retrograde amnesia in free recall. *Science*, *164*, 88-90.
- Tulving, E. (1981). Similarity relations in recognition. *Journal of Verbal Learning and Verbal Behavior*, *20*, 479-496.

Wells, G. L., & Murray, D. M. (1984). Eyewitness confidence. In G. L. Wells & E. F. Loftus (Eds.), *Eyewitness testimony*. Cambridge, England: Cambridge University Press.

Williamsen, J. A., Johnson, H. J., & Eriksen, C. W. (1965). Some characteristics of posthypnotic amnesia. *Journal of Abnormal Psychology*, 70, 123-131.

Zuckerman, M., Amidon, M. D., Bishop, S. E., & Pomerantz, S. D. (1982). Face and tone of voice in the communication of deception. *Journal of Personality and Social Psychology*, 43, 347-357.

Zuckerman, M., DePaulo, B. M., & Rosenthal, R. (1981). Verbal and nonverbal communication of deception. *Advances in Experimental Social Psychology*, 14, 2-59.

Appendix

Transcripts of a Simulating and a Genuinely Forgetful Subject

Printed below are two verbatim transcripts of subjects who participated in Experiment 1. Each transcript includes two 2-min retrieval attempts that were separated by a set of feeling-of-knowing ratings. These feeling-of-knowing ratings were edited from the transcript, as were the ratings made before and after the retrieval attempts. One of the transcripts represents a simulating subject; the other represents a genuinely forgetful subject. The reader is invited to inspect the transcripts and come to his or her own decision regarding the identity of the subjects before reading the paragraph that follows the two transcripts. (E denotes experimenter and S denotes subject.)

Subject A

- E. Well let's try then. Try to get it on your own now. I'll give you a couple of minutes to think.
- S. What did you have on board, diamonds? Okay, diamonds.
- E. Okay. That wasn't the last thing he said.
- S. That wasn't the last thing.
- E. No. Just keep trying.
- S. Umm . . . so what about the thousand dollars. I don't . . . I can't remember anything after that. Umm . . . he'd look after things or he'd . . .
- E. Keep thinking out loud. Try to remember what the last thing the connection said was. . . . Okay, I'll ask you for these ratings again.
-
- E. Okay, let's go after it again for a couple of more minutes. What was the last thing the connection said? . . . Think out loud if you think of anything. What was the last thing the connection said?
- S. I seem to have a mental block for that part of the story.
- E. Well, just keep trying to think of what it is.
- S. Something well . . . the other guy said it was talent, so maybe he said something about talent. I don't . . .
- E. No, that's not the . . . What's the last thing he said. I'll give you a bit more time to try to remember.
- S. I'll have it ready in 24 hours.
- E. That wasn't the last thing he said. . . . Okay, well let's take the ratings again.

Subject B

- E. Try to remember the answer on your own. I'll give you a couple of minutes to think about it.
- S. Okay. Well if the last thing he did was give him the photograph then he either said something before that or in response to that. And if I recall, the last thing he said before that was that he would take the abuse for a thousand for making it in one day. Then after that . . . umm . . . I don't remember.
- E. Well, just keep trying. Think out loud if you can.
- S. I keep going over how he came down to the table but . . . so they talk about how long it'll take, he offers him another thousand. They talked about how risky a business it was before, so it wasn't that. And it wasn't the artistry.
- E. Let's try these ratings again.
-
- E. Okay why don't you try for another couple of minutes. Think out loud. Try to remember what was the last thing the connection said.
- S. Okay . . . Probably something that wasn't anything to do with the passport. I remember that either the captain or him said at one point something that was just like social in passing that had nothing to do with the story. I don't remember which one it was. And he didn't shake hands or anything. I don't think they gave out any money. . . .
- E. I'm going to give you a little more time. Think out loud.
- S. Umm . . . okay, he got the photograph from an arcade . . . a small black and white photograph. And I remember thinking that's what I had to do for residence. I had to get them a photograph. What did he say? Umm . . . I don't remember. He hands him the photograph.
- E. Let's try these ratings again.

The foregoing transcripts were selected to illustrate one of the findings of the experiment. Most people think that Subject A is simulating, and think that Subject B genuinely cannot remember. Subject B is "active" and tries out different strategies and possibilities; Subject A is relatively passive and uses stereotyped phrases such as "I seem to have a mental block." However, Subject A genuinely cannot remember, whereas Subject B is simulating forgetting.

Received March 25, 1985

Revision received May 13, 1985 ■