Journal of Experimental Psychology:
Learning, Memory, and Cognition
1991, Vol. 17, No. 5, 897–907

0278-7393/91/$3.00

Memory for Things Forgotten

Stephen J. Lupker and Joanne L. Harbluk
University of Western Ontario
London, Ontario, Canada

Andrew S. Patrick
Communications Research Centre
Department of Communications
Government of Canada
Ottawa, Ontario, Canada

As previous research has shown, items not recalled on an initial memory task are not simply forgotten. Often, some can be recalled on a later, second task. Further, subjects can generally predict, in terms of feeling-of-knowing (FOK) ratings, which items will be subsequently recalled. Two experiments were carried out to assess both second-task performance and FOK accuracy for unrecalled items as a function of two factors, encoding manipulations (levels of processing in Experiment 1, study time in Experiment 2) and the nature of the second task (explicit or implicit cued stem completion). Results indicate that although levels of processing affected explicit second-task performance more than implicit second-task performance, it increased FOK accuracy in both types of tasks. Study time, however, affected FOK accuracy only in the explicit second task. Apparently, only when subjects were able to do some elaborative processing on the items did their FOK ratings reflect information relating to factors that drive performance on implicit tasks.

Memory researchers have known for a long time that items are not simply either remembered or forgotten. Items that are unrecallable at one point may be recallable at a subsequent point and vice versa (Belmore, 1981; Roediger, Payne, Gillespie, & Lean, 1982), with factors such as changes in the environmental context (Smith, 1988; Smith, Glenberg, & Bjork, 1978) or in the nature of the task used to assess memory (Tulving & Pearlstone, 1966) accentuating these effects. Further, unrecalled items are typically relearned more easily than unstudied items (MacLeod, 1988; Nelson, 1971, 1978). Finally, although individuals may be unable to recall an item, they may be able to recall partial information about that item (Brown & McNeill, 1966) or to assess accurately their ability to remember the item later with additional cues or in a different context (Blake, 1973; Hart, 1967).

This research was supported by Grant A6333 from the Natural Sciences and Engineering Research Council of Canada to Stephen J. Lupker and by funds from the Information Technology Research Centre to the Centre for Cognitive Science at the University of Western Ontario. Major portions of this article were reported at the 29th annual meeting of the Psychonomic Society, Chicago, Illinois, November 10–12, 1988, and at the 51st annual meeting of the Canadian Psychological Association, Ottawa, Ontario, Canada, May 31–June 2, 1990.

We would like to thank Jim Neely, John Gardiner, Roddy Roediger, Robert Greene, and an anonymous reviewer for their comments on earlier drafts of the manuscript and Silvana Santolupo, Nancy Mann, and Bonnie Williams for their help in the data collection and analysis.

Correspondence concerning this article should be addressed to Stephen J. Lupker, Department of Psychology, University of Western Ontario, London, Ontario, Canada N6A 5C2.

The focus of the present article is those items that cannot be recalled on an initial memory task. The central issues concern subjects' abilities to predict which of those items will be remembered on a second, less demanding memory task. In the present studies, subjects' abilities to predict second-task performance were evaluated using a standard feeling-of-knowing (FOK) design. Subjects first studied item pairs. Subsequently, in the initial task, they were shown the first member of the pair and asked to recall the second member. For unrecalled items, subjects were asked to make FOK ratings based on their perceived ability to remember the second member of the pair on a second memory task. Finally, the second memory task was administered to assess the accuracy of the FOK judgments, measured here in terms of Gamma correlations (Nelson, 1984). The standard result in paradigms of this sort is that the FOK ratings do predict second-task performance, at the level of the individual items, with above-chance accuracy (Blake, 1973; Hart, 1967; Nelson, Leonesio, Shimamura, Landwehr, & Narens, 1982; Schacter, 1983).[1]

Although FOK ratings do predict second-task performance with better-than-chance accuracy, the overall accuracy of these predictions is often not very good. As Nelson, Gerler, and Narens (1984) have argued, FOK ratings are undoubtedly based on a number of sources of information. Some of those sources would support second-task performance and, hence,

[1] Although the second task always involved all the studied items, only performance on the unrecalled items was analyzed. Thus, all subsequent references to second-task performance are meant to refer only to second-task performance for items not recalled on the initial cued-recall task.

account for the significant relationship between FOK ratings and second-task performance. On the other hand, FOK ratings also seem to be based on sources of information that do not support second-task performance and, hence, their effect is to decrease FOK accuracy because they increase the noise in the system. In certain circumstances, some of the sources of information on which FOK ratings are based may even be negatively related to second-task performance (Begg, Duft, Lalonde, Melnick, & Sanvito, 1989; Shaughnessy, 1981; Zechmeister & Shaughnessy, 1980).

In the present article, the specific question was how the accuracy of FOK ratings would vary as a function of two factors: the nature of the encoding instructions and the nature of the second task. Both experiments in the present article involved a two-level encoding manipulation designed to produce a large difference in performance on an initial task. The first question centers on the potential effects of these encoding differences on the unrecalled items. Because the unrecalled items in each encoding condition receive the same type of encoding as the analogous recalled items, one question is, Will the condition with superior initial-task performance show superior second-task performance for unrecalled items? More important, will the generally superior encoding processes provide a useful source of information for deciding which of the items might be recallable on the second task? If so, the result would then be better FOK accuracy for those subjects in the condition showing better performance on the initial task.

Unfortunately, the previous data bearing on this issue do not present a clear pattern. Schacter (1983) showed that although a study–time manipulation (1.5 s vs. 5 s per item) did affect the overall frequency of "yes" FOK judgments, it did not affect either performance on the second task or the accuracy of the FOK ratings. In contrast, Nelson et al. (1982) showed that the amount of "overlearning" (defined as one, two, or four correct recalls) affected not only the overall FOK ratings but also performance on the second task and the accuracy of the FOK ratings. In particular, overall FOK ratings, second-task performance, and FOK accuracy increased monotonically with increased overlearning (although only the four-recalls condition actually produced above-chance accuracy on the FOK ratings). Finally, Carroll and Simington (1986) attempted to replicate Nelson et al.'s results using one- and three-recalls conditions. Although this replication was not completely successful, a similar pattern emerged; better performance on the second task and a nonsignificant trend for higher FOK accuracy.

One possible explanation for these differences would be that second-task effects are somewhat fragile and will only emerge consistently when the initial encoding manipulation is a large one. As will be discussed later, if certain assumptions are made about the recallability of items, performance differences on the second task would be monotonically related to the size of the initial task effect. In Experiment 1, the encoding manipulation was a levels-of-processing (LOP) manipulation (Craik & Lockhart, 1972; Craik & Tulving, 1975). Specifically, subjects either counted the number of vowels in the word pairs or used the pairs to make a sentence. Perhaps more than any other type of encoding manipulation, LOP manipulations have reliably produced large differences in a subsequent memory task. Thus, if encoding manipulations do affect FOK ratings or second-task performance, those effects should be evident here.

The second major factor in this study was the nature of the second task. For all subjects, this task was a cued stem-completion task. Subjects were shown the first member of each pair and the first three letters of the second member (e.g., lobster-sho ______). Their task was to complete the second word. Half of the subjects were asked to perform this task under explicit instructions (complete the stem with the second member of the pair from the study trials), and the other half were asked to perform this task under implicit instructions (complete the stem with the first word that comes to mind). One advantage that this particular second task has is that it allows the creation of explicit and implicit tasks that are quite similar. As such, results from the two tasks should be as comparable as possible.

The previous discussion about the effect of LOP on subject performance has been based on the assumption that the subjects were engaged in explicit memory tasks. The implications for an implicit second task are slightly different. That is, although some researchers (i.e., Graf & Schacter, 1985; Srinivas & Roediger, 1990) have shown that it is possible to get LOP effects in implicit tasks like that used here, typically, LOP effects tend to be either muted or nonexistent under implicit retrieval instructions (Graf, Squire, & Mandler, 1984; Jacoby & Dallas, 1981; Schacter & Graf, 1986). A similar result was reported by Gardiner (1988). In this study, subjects were asked in a recognition task whether they "remembered" the item from the study list (i.e., could explicitly retrieve the event) or simply "knew" that it must have been there. Gardiner reported an effect of LOP for the items that subjects "remembered" but not for the items that subjects simply "knew." The implication of these results is that the factors that drive performance on an explicit memory task tend to be different from those that drive performance on an implicit memory task.

The question, then, is how FOK accuracy will vary as a function of the nature of the second task. Presumably, when subjects are making FOK ratings, they are making assumptions about the nature of the second task. Seemingly, the accuracy of those ratings would vary as a function of the nature of the second task. Because FOK ratings in the present study were made following a failure to recall in an explicit task, one possibility is that none of the factors that support performance on implicit tasks were considered. A second possibility is that the factors that drive performance on an implicit task simply may not be available to consciousness and thus cannot be considered when making FOK ratings. In either case, the result may be that the ratings would not be predictive of implicit second-task performance at all.

As Nelson, Gerler, and Narens (1984) have argued, however, the sources of information that subjects use when making FOK judgments may be quite varied and may include many that are of minimal importance to explicit task performance. Thus, it is possible that some of these sources may be ones that support performance on implicit tasks. For example, as Roediger, Jacoby, and colleagues (Jacoby, 1983; Jacoby &

Dallas, 1981; Roediger & Blaxton, 1987; Roediger & Weldon, 1987; Roediger, Weldon, & Challis, 1989) have argued, explicit memory tasks may involve more conceptually driven processing, whereas implicit memory tasks may involve more data-driven processing. Although this distinction may not capture all of the differences between explicit and implicit tasks, it accounts for many of the performance differences found in the two types of tasks. If Nelson et al.'s suggestion that FOK ratings involving multiple sources of information is correct, then under some circumstances, subjects may integrate information about data-driven as well as conceptually driven factors into FOK ratings. If so, these ratings should be predictive of performance on both explicit and implicit tasks.

Previous research does, in fact, suggest that FOK ratings are related to implicit task performance. Nelson et al. (1984) examined FOK accuracy in both what appears to be an implicit second-task (perceptual identification) and an explicit second-task (recognition). The Gamma correlation in the perceptual identification task (.16) was not only significant, but it was also not significantly different from the Gamma correlation in the recognition task (.29). Unfortunately, because the authors' focus was not to evaluate FOK accuracy in an implicit task, they made no attempt to isolate their perceptual identification task from the effects of explicit retrieval operations. The stimuli in their initial task had been general information questions from the Nelson and Narens (1980) norms. During the perceptual identification task, each question was displayed on the computer screen above the to-be-identified item. More important, subjects were told that the item that they were trying to identify was the answer to the question. Thus, during the perceptual identification task, subjects may have been explicitly searching their memories in an effort to retrieve these items. As such, the contributions of implicit versus explicit operations to the perceptual identification results are unclear.

In a second study bearing on the question of FOK accuracy in an implicit task, Yaniv and Meyer (1987) calculated Gamma correlations between FOK ratings to rare word definition questions and subjects' lexical decision latencies to these words. A significant Gamma correlation of .32 was obtained in Experiment 1 and a marginally significant Gamma correlation of .16 was obtained in Experiment 2, supporting the suggestion that FOK ratings are related to performance in implicit tasks.

Yaniv and Meyer suggest that both the FOK ratings and lexical decision results are reflective of subthreshold activation of the target that resulted from trying to retrieve it from memory. If so, the significant correlation would be a clear demonstration that the type of information that drives implicit task performance is not only available to consciousness but is used in making FOK ratings. What is not clear, however, is to what extent these lexical decision reaction times (RTs) are actually reflective of subthreshold activation rather than postretrieval decision processes, for example, some sort of question–target integration process (Balota & Chumbley, 1984; Forster, 1981; Lupker, 1984; Norris, 1986; Seidenberg, Waters, Sanders, & Langer, 1984; West & Stanovich, 1982).

A second interpretation of Yaniv and Meyer's data that has been provided recently by Connor, Balota, and Neely (1990) is that the significant correlations are at least partially due to differential subjective familiarity with the various topics used. That is, when the word and the definition are from a general topic area that a subject is familiar with, both a shorter lexical decision latency and a higher FOK rating will result. In fact, as Conner et al.'s data show, a similar relationship between lexical decision latency and FOK ratings exists when the lexical decision task is performed a week before the FOK ratings are made.

If this interpretation is correct (and a similar interpretation could also be applied to the Nelson et al., 1984, data) it would indicate that subjective familiarity could be a means by which FOK ratings would reflect performance in an implicit task. What these results would not suggest, however, is that FOK ratings incorporate any information related to the results of the data-driven processing (or any other relevant processing), which presumably takes place during study. That is, in neither instance was there any study of the items before making the FOK ratings. As such, there was no opportunity to integrate information from this processing into the ratings. Thus, what this analysis suggests is that, for the present article, the question of whether there is a relationship between FOK ratings and implicit task performance is too broad. A more appropriate way of posing the question is to ask whether there are any factors relating to the study event and subsequent attempted retrieval that are integrated into FOK ratings and that at the same time relate to implicit task performance.

Experiment 1

Method

Subjects and design. One hundred and twenty University of Western Ontario undergraduates participated to fulfill a course requirement or for monetary compensation ($6.00). Data from 19 other subjects were eliminated because the subjects either (a) made too many incorrect guesses (10 or more) on the cued-recall task (6 subjects) or (b) failed to follow instructions on the FOK ratings task (13 subjects).

Two between-subjects factors were combined to produce four different conditions. These factors were LOP at study (vowel counting or sentence generation) and type of instructions for a cued stem-completion task (explicit or implicit). Subjects were assigned to conditions at their arrival for testing, and they were tested individually.

Materials. Forty critical word pairs were formed by selecting 80 words (4–10 letters in length) of moderate frequency (mean frequency = 44.6 occurrences per million) from the Kucera and Francis (1967) norms. These words were randomly paired to form the 40 stimulus–response pairs (e.g., lobster-shorts). A further constraint was that the initial three letters of each response word (the stem) had to be unique in the set of 80 words yet have at least 10 possible completions, according to a pocket dictionary. Sixty additional word pairs having the same characteristics were also selected to be used as filler pairs in the cued stem-completion task.

Procedure. There were three main components to the procedure: instructions and study, cued-recall and FOK ratings, and cued stem completions. The presentation and timing of stimuli were controlled by an Apple II microcomputer, and subjects wrote their responses in booklets. The instructions were read aloud by the experimenter.

Instructions and study. In this phase, the subjects performed a task that required either vowel counting (shallow LOP) or sentence generation (deep LOP). Vowel counting involved reading the word pair aloud, counting the number of vowels in the words, and saying that number aloud. Sentence generation involved saying aloud a sentence that related the words in a meaningful way. Subjects in both tasks were informed that our interest was in how well they could remember the word pairs using the specified instructions. In an effort to have subjects adhere to the study instructions, they were told that their responses (number of vowels or generated sentences) were being recorded.

The subjects received 5 practice word pairs to familiarize themselves with the procedure. The 40 critical word pairs followed at a rate of 7 s per pair. Finally, a 5-min distractor task was given in which the subjects were required to generate and write down the names of the states in the United States of America.

Cued recall and FOK ratings. For the cued-recall task, subjects were presented with the stimulus member of each pair and asked to recall the response member. If they could recall a given item, they recorded it in the response booklet. If they could not recall an item, they were asked not to guess but to give a rating for their feeling-of-knowing (FOK) for that item. A scale of 1 (*low FOK*) to 5 (*high FOK*) was used.[2] The items were presented in a random order that differed from the study session. Fifteen s were allowed for each answer. Finally, a 5-min distractor task was given in which subjects were presented with a first name and the first letter of a last name (e.g., Pierre T______), and their task was to generate a complete last name and record it in the booklet.

Cued stem completions. For the cued stem-completion tasks, all subjects received the same materials but performed the task under different instructions. The stimuli were in the form of stimulus–stem pairs (e.g., lobster-sho____), and the task was to complete the stem to form a word (7 s were allowed for each pair). The 40 critical word pairs were randomly combined with 60 stimulus–stem fillers. Thirty-six subjects in each study conditions (vowel or sentence LOP) received explicit instructions. That is, they were told that some of the cues and stems were derived from word pairs that they had studied previously and that for those stimuli, they should complete the stem with the second word from the studied pair. All other stems could be completed with the first word that came to mind. The remaining 24 subjects in each condition were told simply to complete the stems with the first word that came to mind in a free-association fashion (implicit instructions). For these subjects, no reference was made to the original study session. Finally, at the end of the experiment, all of the implicit task subjects were asked whether they had made any attempts to recall the target items explicitly. All 48 subjects indicated that they had not.

Results

Initial cued recall. On the initial cued-recall task, the subjects who generated sentences recalled a significantly larger proportion of the items (.48) than the subjects who counted vowels (.02), $F(1, 118) = 346.00$, $MS_e = 0.0185$, $p < .001$, demonstrating the usual LOP effect. In the analyses that follow, only the items not recalled on this initial task were included.

FOK ratings. The mean FOK ratings for the items not recalled in the first task were: vowel condition = 2.41, sentence condition = 2.91. This difference in FOK ratings was significant, $F(1, 118) = 18.88$, $MS_e = 0.4117$, $p < .001$, indicating that subjects felt that they were more likely to remember the items that they could not recall after generating sentences than after counting vowels.

Cued stem completions. The mean proportion of stems completed with the studied word for the two LOPs and the two types of instructions are presented in Figure 1. A large LOP effect is apparent for the explicit instructions group, whereas there is a much smaller effect for the implicit instructions group. This impression was confirmed by the analysis of variance (ANOVA), which showed a main effect of LOP $F(1, 116) = 23.71$, $MS_e = 0.0172$, $p < .001$, a main effect of task instructions, $F(1, 116) = 16.27$, $MS_e = 0.0172$, $p < .001$, and an interaction of LOP and task instructions, $F(1, 116) = 9.48$, $MS_e = 0.172$, $p < .005$. A simple main effects analysis revealed that the LOP effect was significant for the explicit group, $t(70) = 5.85$, $p < .001$, but not for the implicit group, $t(46) = 1.32$, $p > .05$ (one-tailed).

The results in Figure 1 also show that subjects completed the stems with a higher frequency than they "correctly" completed the new filler items in all conditions, vowel–explicit $t(35) = 11.06$, vowel–implicit $t(23) = 11.70$, sentence–explicit $t(35) = 15.19$, sentence–implicit $t(23) = 7.48$, all $ps < .001$. Finally, in all conditions, almost all of the initially recalled items were also produced in the cued stem-completion task (93% in the vowel–explicit condition, 95% in the vowel–implicit condition, 96% in the sentence–explicit condition, 85% in the sentence–implicit condition).

FOK accuracy. Following the procedure recommended by Nelson (1984), Gamma correlations were calculated between FOK ratings and cued stem-completion scores for each subject. These Gamma correlations represent how accurate the FOK ratings are in predicting the cued stem-completion performance. The mean Gamma correlations for the four groups are presented in Figure 2. As Figure 2 shows, there appears to be a strong LOP effect for both the implicit and explicit instructions conditions. This impression was confirmed by the ANOVA, which showed a significant main effect of LOP, $F(1, 116) = 4.99$, $MS_e = 0.1418$, $p < .05$, but no main effect for task instructions and no interaction, both $Fs < 1.0$. (The median Gamma correlations were .23, .10, .40, and .10 in the sentence–explicit, vowel–explicit, sentence–implicit, and vowel–implicit conditions, respectively. Thus, the medians showed the same pattern as the means.)

The mean Gamma correlations were also analyzed to determine which correlations differed from zero. The Gamma correlations for the sentence conditions were significantly greater than zero, explicit $t(35) = 2.44$, $p < .01$ (one-tailed), implicit $t(23) = 3.47$, $p < .001$ (one-tailed), but the correlations for the vowel conditions were not, explicit $t(35) = 1.17$, implicit $t(23) = 1.61$, both $ps > .05$ (one-tailed).

Discussion

Explicit instructions conditions. The results for the subjects who received explicit recall instructions provide a nice

[2] Because FOK ratings were not given when subjects generated incorrect answers (i.e., what Krinsky and Nelson (1985) refer to as *commission errors*), those trials were not analyzed.

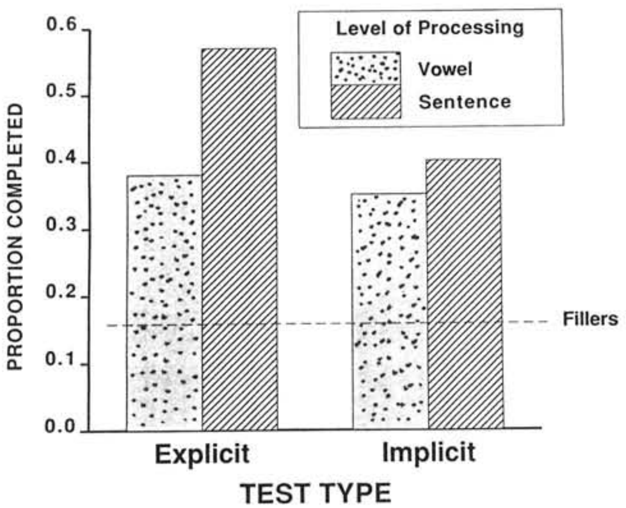

Figure 1. Mean proportion of stems completed with the target word for unrecalled items as a function of type of instructions and level of processing in Experiment 1.

replication of Nelson et al. (1982). That is, deeper LOPs acted just like Nelson et al.'s additional overlearning trials in terms of producing (a) better initial task performance, (b) better second-task performance and higher overall FOK ratings, and (c) a larger Gamma correlation.

The better second-task performance in the sentence condition suggests that the unrecalled items were also affected by the encoding manipulation and, thus, were more likely to support performance in a somewhat less demanding second task. This type of result is actually predicted from a classic memory strength approach to an item's recallability. That is, if one assumes that the items in each condition represent a distribution of memory strengths and that those distributions satisfy the typical assumptions (e.g., approximately normal, equal variances), the size of the second-task effect should be a monotonic function of the size of the initial-task effect. What this result argues against is any suggestion that the unrecalled items in the sentence condition contained a large percentage of low-strength and, hence, difficult-to-retrieve items (i.e., the strength distributions were bimodal or severely negatively skewed). These arguments are also supported by the fact that the overall FOK ratings mirrored second-task performance. Apparently, at some level, subjects are aware of the effect of encoding on later performance for the unrecalled items.

The larger Gamma correlation in the sentence condition indicates that the subjects in this condition were better able to assess the likelihood of recall for the individual items. (It is important to realize that because the Gamma correlation is a relative measure, it is quite possible to have higher levels of FOK ratings and higher second-task performance and yet find no difference in Gamma correlations.) What was a bit disappointing was the size of Gamma correlations observed here. That is, the correlation in the sentence condition was no better than typically reported in the FOK literature, whereas in the vowel condition, the correlation was not even significantly different from zero. Nonetheless, the present results do suggest that subjects' abilities to predict later retrieval are not independent of the nature of the encoding done at study.

Implicit instructions conditions. For the subjects who got implicit instructions, the results were only slightly different in that there was no significant effect of encoding instructions for the unrecalled items on the second task. (Note, however, that overall performance was higher in the sentence condition because almost all of the recalled target items were produced again in the cued stem-completion task.) This diminished LOP effect essentially replicates the typical finding in tasks of this type (Jacoby & Dallas, 1981; Schacter & Graf, 1986). Furthermore, it suggests that our subjects were following the implicit instructions and were not making an explicit effort to recall the items. Nonetheless, performance in both of these conditions was much higher than in the filler conditions, indicating that the earlier study of these pairs was influencing performance. What did not differ between this condition and the explicit instructions condition was the pattern of results for the Gamma correlations. Specifically, although FOK accuracy in the vowel condition was close to zero, FOK accuracy in the sentence condition was, if anything, even higher than in the analogous explicit condition.

As discussed previously, both Yaniv and Meyer (1987) and Nelson et al. (1984) reported a significant relationship between FOK ratings and performance in an implicit second task. Both of these results suggest that under certain circumstances FOK ratings are based on information from sources that support performance in implicit memory tasks. In both instances, however, two issues were unclear. The first was the extent to which second-task performance actually reflected implicit retrieval operations. In Nelson et al., instructions in the implicit task actually encouraged explicit retrieval operations. Results in the primed lexical decision task in Yaniv and Meyer could have been due to decision rather than to retrieval operations. The second issue was, even if the effects did reflect implicit retrieval operations, whether the relationship between

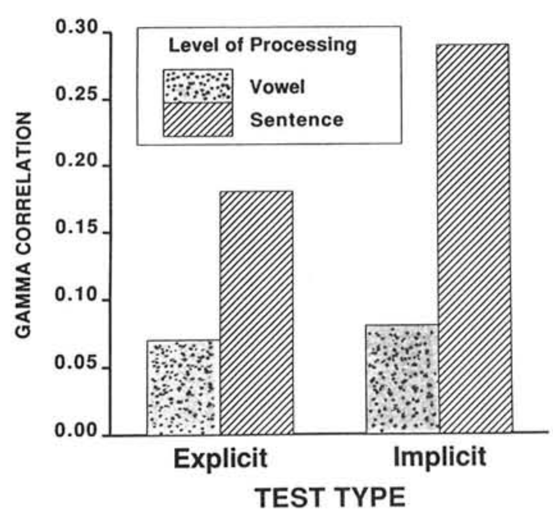

Figure 2. Mean Gamma correlation as a function of type of instructions and level of processing in Experiment 1.

FOK ratings and second-task performance was due solely to subjective familiarity with the general topics and thus had little to say about recallability of individual items.

The task used in Experiment 1 is more closely tied to retrieval operations and, because it is not based on answering general information questions, should not be affected by preexperimental subjective familiarity. Thus, the present results provide a clearer answer to the question of whether FOK ratings are based on information derived during study and subsequent retrieval failure.

One question that might be raised is whether the subjects were really treating the present task as an implicit retrieval task. It is always possible, of course, that subjects in tasks such as these do not follow instructions and instead attempt to complete the stems using explicit retrieval operations. In the present situation, this concern may be heightened by the recent results of Bowers and Schacter (1990). Bowers and Schacter reported that the only subjects who showed a performance difference between the studied pairs and the fillers were those who were aware that the cued stem-completion stimuli contained studied pairs. Given that our subjects showed a large difference between the studied pairs and the fillers, it seems likely that they were all "aware" subjects and hence could have chosen to complete some stems using explicit retrieval operations. Although this possibility cannot be ruled out, two observations suggest that such was not the case. First, all implicit conditions subjects indicated when they left the experiment that they had followed the implicit retrieval instructions. Second, the diminished LOP effect for the implicit completion conditions supports the subjects' claims that they were following the implicit instructions rather than explicitly trying to retrieve the second member of the pair. Thus, the present data provide a clearer demonstration that FOK ratings can be, at least to some extent, based on the type of information that supports performance in an implicit memory task. Further, the fact that these same ratings also predicted performance in an explicit task provides support for Nelson et al.'s basic argument that the FOK ratings draw on information from many sources.

Experiment 2

Although the results from the explicit instructions conditions in Experiment 1 provided a nice replication of Nelson et al.'s (1982) results using an overlearning manipulation, they are somewhat contradictory to Schacter's (1983) results using a study–time manipulation. Although Schacter reported a study–time effect on overall FOK ratings, he found no effect on either second-task performance or FOK accuracy. There would seem to be a number of possible reasons for this result. One possibility might be that performance on specific pairs in Schacter's second task (recognition with semantically related foils) was strongly influenced by the discriminability between the correct items and the foils, a factor that could not be evaluated by subjects while making FOK ratings. Another possibility, as discussed earlier, is that these second-task effects are somewhat fragile unless the encoding manipulation is strong enough to produce a large initial-task effect, like that observed in Experiment 1.

Experiment 2 was an attempt to provide an evaluation of the study–time manipulation, using a second task in which target–foil discriminability was not an issue. Subjects were presented word pairs and told to repeat the pairs aloud at the rate of one repetition per second (in time with a flashing asterisk). Half of the subjects did this for 2 s, and half did it for 7 s. The purpose of requiring the repetitions was to maintain the attention of the subjects in the 7-s condition (i.e., to prevent them from using the study time to engage in deeper processing or from stopping study before the entire time had elapsed). The tasks, both initial and second, were the same as in Experiment 1.

Included in Experiment 2 were also implicit instructions conditions. As with LOP effects, the general finding has been that study–time effects tend to be smaller in implicit tasks than in explicit tasks (Greene, 1986; Jacoby & Dallas, 1981; Richardson & Bjork, 1982). Nonetheless, subjects in the 7-s condition are spending a longer time viewing the item pairs and hence may be engaged in more perceptually based processing. The result may be slightly higher implicit task performance in the 7-s condition. Further, it may also be the case that the study–time effect in the explicit instructions conditions will be somewhat smaller than the analogous LOP effect in Experiment 1. Thus, even if subjects do follow the implicit instructions faithfully, it may not be possible to observe a significantly smaller study–time effect in the implicit conditions than in the explicit conditions.

The final issues concern the FOK accuracy. On the basis of the results of Experiment 1, it appears that low levels of encoding produce rather low levels of FOK accuracy in both implicit and explicit tasks. Thus, the same result would be expected here in the 2-s conditions. The more central question in Experiment 2 is whether longer study times will, like deeper LOPs, allow subjects better access to sources of information related to second-task performance. If so, depending on what type of performance those sources support, we could observe increased accuracy on the explicit task, the implicit task, or possibly both.

Method

Subjects and design. Ninety-six University of Western Ontario undergraduates participated to fulfill a course requirement. Data from 3 other subjects were eliminated because (a) the subjects indicated that they did not follow instructions in the implicit second task (2 subjects); or (b) they made too many incorrect guesses on the cued-recall task (1 subject).

There were two between-subjects factors: study time (2 or 7 s) and type of instructions for the cued stem-completion task (explicit or implicit). These factors were combined to produce four different conditions, with 24 subjects in each condition.

Materials. The same materials were used as in Experiment 1.

Procedure. The procedure was identical to that of Experiment 1 with the following exceptions during the instructions and study phase. Subjects viewed the word pairs for either 2 or 7 s. They were instructed to repeat the word pair aloud at the rate of once per second. An asterisk on the computer screen flashed once per second to assist with pacing the repetitions. Subjects' repetitions were recorded. The rest of the procedure followed that of Experiment 1.

Results

Initial cued recall. Subjects who studied the items for 7 s recalled significantly more items (0.22) than subjects who studied the items for 2 s (0.07), $F(1, 94) = 27.66$, $MS_e = 0.0192$, $p < .001$. As in the first experiment, the items that were recalled on this initial cued-recall test were eliminated from the subsequent analyses.

FOK ratings. Subjects' ratings of the unrecalled items were greater for items studied for 7 s (2.74) than for items studied for 2 s (2.43), $F(1, 94) = 11.50$, $MS_e = 0.2048$, $p < .005$, indicating that subjects were more confident of recalling items studied for longer rather than shorter durations.

Cued stem completions. The mean proportion of stems completed with the studied word for the two study durations for both the implicit and explicit tasks are presented in Figure 3. The main effects for study time, $F(1, 92) = 9.93$, $MS_e = 0.0106$, $p < .005$, and test type, $F(1, 92) = 14.16$, $MS_e = 0.0106$, $p < .001$, were significant. The interaction of those two factors was not, $F < 1.0$. These results indicate that (a) subjects were more likely to complete the stems with the originally studied items if the items had been studied for 7 rather than 2 s; and (b) stems were more likely to be completed with the originally studied items when explicit rather than implicit instructions were given.

The results in Figure 3 also show that subjects completed the items with a higher frequency than they correctly completed the new filler items in all conditions, 2 s–explicit $t(23) = 12.30$, 2 s–implicit $t(23) = 7.84$, 7 s–explicit $t(23) = 9.63$, 7 s–implicit $t(23) = 12.63$, all $ps < .001$. Finally, as in Experiment 1, almost all of the recalled items were also produced in the cued stem-completion task (96% in the 2 s–explicit condition, 91% in the 2 s–implicit condition , 97% in the 7 s–explicit condition, 88% in the 7 s–implicit condition).

FOK accuracy. The mean Gamma correlations for the four conditions are presented in Figure 4. The main effect for

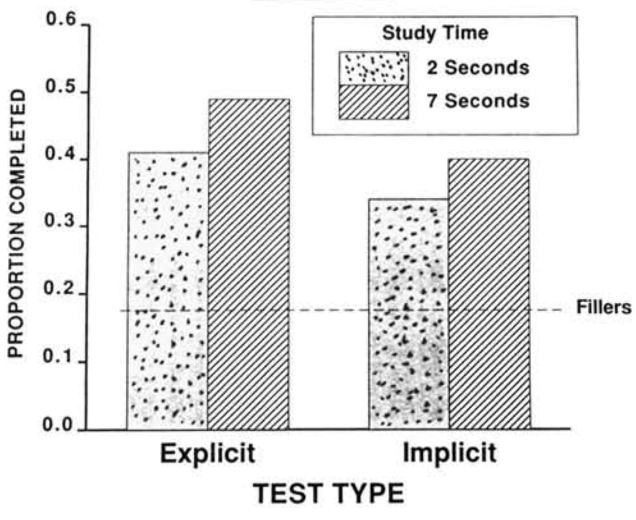

Figure 3. Mean proportion of stems completed with the target word for unrecalled items as a function of type of instructions and study time in Experiment 2.

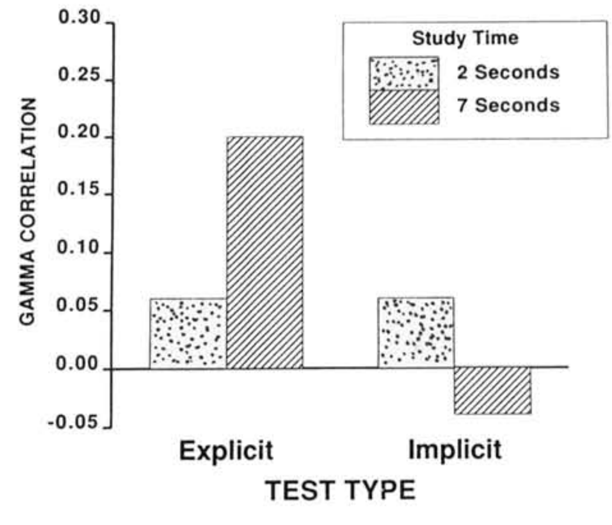

Figure 4. Mean Gamma correlation as a function of type of instructions and study time in Experiment 2.

study time did not reach significance, $F < 1.0$. The main effect for task instructions was significant, $F(1, 92) = 4.74$, $MS_e = 0.0785$, $p < .05$, however, it was qualified by a significant Task Instructions × Study Time interaction, $F(1, 92) = 4.37$, $MS_e = 0.0785$, $p < .05$. (The median Gamma correlations were .18, .09, −.07, and .08 in the 7 s–explicit, 2 s–explicit, 7 s–implicit, and 2 s–implicit conditions, respectively. Thus, the medians showed the same pattern as the means.)

The mean Gamma correlations were analyzed to determine if the correlations differed from zero. Although the Gamma correlation in the 7 s–explicit condition was significantly greater than zero, $t(23) = 3.16$, $p < .005$ (one-tailed), none of the others approached significance, 7 s–implicit $t(23) = -.80$, 2 s–explicit $t(23) = 1.12$, 2 s–implicit $t(23) = 1.04$, all $ps > .10$ (one-tailed).

Discussion

Explicit instructions conditions. The results in the explicit instructions conditions of Experiment 2 provide a replication of Experiment 1, rather than a replication of Schacter's (1983) results. Specifically, the longer study–time condition produced both better initial task performance and higher overall FOK ratings, as it did in Schacter's study. Unlike in Schacter's study, however, this condition also produced better second-task performance and higher FOK accuracy.

One hypothesis advanced for why Schacter observed no second-task effect was that the initial-task effect and, hence, the strength of the encoding manipulation, was not large. Although this hypothesis is undoubtedly not the entire explanation, the results of Experiment 2 suggest that it has some validity. In particular, both initial- and second-task effects in Experiment 2 were smaller than the analogous effects in Experiment 1. That is, the initial-task difference (46% in Experiment 1, 15% in Experiment 2), the FOK ratings difference (0.50, in Experiment 1, 0.32 in Experiment 2) and the second-task difference (17% in Experiment 1, 8% in Experiment 2) were all larger in Experiment 1. It should also be noted, however, that the initial-task effect in Experiment 2 was essentially the same as the difference between the two-recalls and four-recalls conditions in Nelson et al. (1982) (10%) and the initial-task effect in Schacter's (1983) study (16%). Thus, Schacter's failure to find a second-task effect cannot be due solely to the strength of his initial-task manipulation.

A second hypothesis was that the foil selection procedure in Schacter's second (recognition) task may have created sufficient noise in those data to mask any encoding effects. In Schacter's recognition task, the first item of the pair was presented together with the second item and five semantically similar foils. The subjects, of course, had no way of knowing what the foils would be, and hence how difficult the discrimination would be, when they were making their FOK ratings. Thus, in some instances, a positive FOK rating may have been made because subjects felt that they had some knowledge of the meaning of the target item, whereas in other instances a positive FOK rating may have been made because subjects felt that they had some knowledge of the target's spelling. In general, the type of foils used would tend to lower second-task performance in the former instance and to increase it in the latter instance. Given that Schacter's study–time effect in the initial task was not large in the first place, this type of second task may have made second-task effects even harder to observe.

The other difference between the present study and Schacter's is that FOK accuracy was significantly higher with longer study time. This effect was, in fact, virtually identical to the LOP effect in Experiment 1. Furthermore, also analogous to Experiment 1, the effect seems to be due mainly to the fact that only the Gamma in the more complete encoding condition was significantly different from zero.

Implicit instructions conditions. For the subjects who got the implicit instructions, the effects of study time were essentially the opposite of those of LOP in Experiment 1. First of all, the study–time effect on the second task was virtually the same for the explicit (8%) and implicit (6%) groups. We would like to suggest, however, that this lack of an interaction is not evidence that the subjects failed to follow the instructions in the implicit conditions. Rather, as discussed, and as indicated by the results in the initial recall task, the study–time manipulation in Experiment 2 was simply a weaker manipulation than the LOP manipulation in Experiment 1. The result was that the second-task effect in the explicit condition was substantially smaller than (essentially half the size of) the LOP effect in Experiment 1. On the other hand, the study–time effect in the implicit task was almost the same size as the analogous LOP effect for Experiment 1 (5%). As we have argued previously, subjects in the implicit conditions in Experiment 1 seemed to be following the task instructions. Thus, there seems to be no reason to take the lack of a Study Time × Task Instructions interaction here as evidence that subjects were not following the implicit task instructions.

This argument is further supported by the existence of a second difference (FOK accuracy) between results in the implicit conditions in the two experiments. In Experiment 2, there was a significant interaction between study time and task instructions, owing to the fact that there was no study-time effect in the implicit condition. (In fact, if anything, the Gamma correlation in the 7-s–implicit condition is lower than the Gamma correlation in the 2-s–implicit condition.) The apparent implication is that our study–time manipulation was much less effective than a LOP manipulation in giving subjects access to sources of information about unrecalled items. That is, as indicated in Experiment 1, a deeper LOP gives the subjects enhanced access to sources of information that support performance in both explicit and implicit tasks. As suggested by these data, longer study time gives the subjects enhanced access to the sources of information that support performance only in explicit tasks. Thus, the question of whether FOK ratings reflect factors relating to study and retrieval that also relate to performance on implicit tasks is not a straightforward one. In fact, a better question might be, Under what circumstances will factors relating to implicit task performance be incorporated into FOK ratings? Although the present results do not provide a complete answer to this question, they do indicate that these factors are not simply incorporated automatically. Rather, it appears that some sort of integrative processing of the stimulus pair is necessary, a point that will be returned to in the General Discussion section.

General Discussion

The purpose of the present article was to gain additional insight into subjects' ability to predict future recallability of initially unrecalled items. Within this context, the specific question that was asked was how the accuracy of prediction would vary as a function of two factors, the nature of the encoding process and the nature of the second task, specifically an implicit versus an explicit task. The results suggest that although encoding differences do affect FOK accuracy, the reason is that less complete encoding leads to very low accuracy rather than more full encoding leading to high accuracy. The results further suggest that not only does the nature of the second task affect FOK accuracy but also that these effects interact with the nature of the encoding. That is, although certain types of encoding may lead to parallel effects in explicit and implicit tasks, under other circumstances these effects may be limited to explicit tasks.

FOK Accuracy on Explicit Secondary Tasks

The basic question is why and how LOP and study time affected FOK accuracy in the observed fashion. FOK ratings will be accurate to the extent that (a) subjects have knowledge about the sources of information that will ultimately drive second-task performance, (b) subjects use that knowledge in making the ratings, and (c) subjects do not use information that is either irrelevant or negatively related to task performance. If one assumes that subjects will use whatever knowledge they think is relevant, one question becomes, How do they obtain accurate information about relevant sources? It seems most likely that this knowledge is garnered at two points, during study and during the initial recall task.

During study, of course, an item's recallability is increased. At the same time, it is quite possible that there would be a parallel buildup of information about the ease or difficulty of this processing. For example, a subject might gain some impression of how easily the two words fit together on either a semantic or a phonological level. These types of impressions may be the type of information that the subjects will later use when making FOK ratings. They may also turn out to be reasonably predictive of second-task performance (Begg et al., 1989). In both the vowel-counting condition in Experiment 1 and the 2-s study condition in Experiment 2, a subject's exposure to all of the stimulus pairs was quite limited compared with the exposure gained in the sentence construction and 7-s conditions. Thus, with less opportunity to gain information about future recallability during study, it would follow that FOK accuracy in the vowel-counting and 2-s conditions would be somewhat poorer.

During the initial cued-recall task, subjects would also have an opportunity to gain information to be used in making FOK judgments. Specifically, they may gain information about (a) the current potential for recall for each item (assuming that this potential does decay over time) and (b) what levels of recallability are necessary to support successful performance. With respect to (b), the vowel-counting and 2-s conditions would not provide much information. That is, initial-task performance in these conditions was somewhat low and, thus, subjects may have been less able to evaluate how the knowledge they had about the pairs would ultimately relate to second-task performance. (For other discussions about the usefulness of an opportunity to receive feedback on performance by taking an initial test, see King, Zechmeister, & Shaughnessy, 1980; Lovelace, 1984; and Glenberg, Sanocki, Epstein, & Morris, 1987.)

Although the argument that FOK ratings are based on information from both study and test periods is reasonable, recent results suggest that information gained during study may not contribute much to FOK ratings. In particular, Leonesio and Nelson (1990) reported that ratings made at study (both ease of learning and judgments of knowing) were not particularly strongly correlated with FOK ratings made at test. In their task, however, there was a 4-week interval between study and test. Thus, memory of the study experience may have been much more limited and, hence, less influential in their studies than in the present experiments.

Interestingly, both judgment of learning ratings and FOK ratings did significantly predict performance on unrecalled items in Leonesio and Nelson's second task (recognition). Furthermore, both types of ratings predicted performance to approximately the same extent as our FOK ratings did (i.e., Gamma correlations were approximately .20 in all cases). If our subjects actually were drawing on information from both study and test, one might have expected that our Gamma correlations would have been slightly larger than Leonesio and Nelson's. As noted, however, subjects simply do not seem to be very good at predicting second-task performance. Gamma correlations rarely exceed .30 in the reported literature, leading to the conclusion that subjects are unable to evaluate relevant sources very accurately (Jameson, Narens, Goldfarb, & Nelson, 1990) or that FOK ratings are inevitably influenced by a number of factors that are irrelevant or negatively related to task performance. In fact, Begg et al. have recently argued that a number of these irrelevant factors are evaluated when subjects predict performance right after learning. It is not clear to what extent a similar case could be made for FOK ratings made after recall failure, however, Krinsky and Nelson (1985) have documented at least one way in which a problem of this sort could arise.

FOK Accuracy on Secondary Implicit Tasks

The final question is why LOP produces similar FOK accuracy effects in both explicit and implicit tasks, whereas study time affects FOK accuracy only in explicit tasks. In essence, this result indicates that only in the sentence construction condition was there much evidence that FOK ratings were based on the types of factors that drive implicit task performance. Working on the assumption that FOK ratings are based on whatever type of information subjects have available and deem relevant, the implication is that subjects did not have this type of information in the other conditions, even in the 7-s condition of Experiment 2. The further implication is that this information became available in the sentence condition as a result of the particular encoding operation.

One possible way to think about this result would be based on the idea that performance in implicit tasks is based primarily on data-driven processing (Roediger, Weldon, & Challis, 1989). During study, some word pairs will be processed in a way that will lead to their production on a data-driven task, whereas others will not. Subjects will be able to predict performance on data-driven tasks to the extent that they can discriminate among items on this dimension. Thus, the argument would be that by constructing sentences, subjects are somehow better able to obtain the information necessary to make this discrimination than by doing any of the other study tasks in the present experiments.

Although this argument cannot be rejected out of hand, it does seem somewhat bizarre. Presumably, the type of processing during study that would lead to production of an item on a data-driven task would be quite perceptually based. Yet, in this particular condition, more so than in any other, the subject's job is to engage in extensive elaborative processing, almost at the expense of perceptual processing. Furthermore, the degree of elaborative processing will probably be quite varied from item pair to item pair, which means that this factor should be a predominant influence in creating FOK ratings. Thus, if anything, the FOK ratings in this condition should be the least predictive of implicit task performance.

An alternative explanation for this result could be derived from Graf and Schacter's (1985) demonstration that, under certain circumstances, it is possible to get a LOP effect in an implicit task. Using encoding tasks virtually identical to those in Experiment 1, Graf and Schacter found that performance in a sentence construction condition was significantly higher than in a vowel comparison condition. Furthermore, although performance in the vowel comparison condition was above that in the filler condition, it was equal to performance in a condition in which the target stems were presented with different cues.

The conclusion that Graf and Schacter drew was that the sentence construction condition provides an additional benefit by allowing the creation of a unitized representation of the two words. This type of representation, which is not a data-driven representation in the normal sense, could then drive performance in an implicit task if a strong enough cue was given (i.e., the cue word alone may not be sufficient in a free-association task, but the cue with a stem may be sufficient). Pairs in the vowel comparison condition would not be integrated and, thus, performance would only be a function of having seen the target previously. As such, performance in this condition would be independent of the nature of the cue, as was reported. Finally, words presented alone, as they tend to be in other investigations of LOP effects in implicit tasks, would presumably already have unitized representations. Thus, no construction of unitized representations would be undertaken, and no real LOP effects should be observed in those implicit tasks.

If this hypothesis is correct, it would explain why only the sentence construction condition showed any significant FOK accuracy. Only in this condition was there any information about the existence of a unitized representation that could be incorporated into the FOK ratings. Furthermore, it might also explain why Yaniv and Meyer (1987) found a significant relationship between FOK ratings and RT in their lexical decision task in spite of the fact that the subjects did no data-driven processing on the items before making FOK ratings. (The assumption is being made, of course, that Yaniv and Meyer's effect was due neither to postretrieval, decision processes nor simply to subjective familiarity.) In their task, subjects gave FOK ratings to word definitions if they could not retrieve the word being defined. Words with higher FOK ratings tended to produce shorter lexical decision RTs. Yaniv and Meyer argued that when the question was presented, the right answer was activated to a subthreshold level, depending on the strength of the links between the definition and the word. Potentially, there would be two sources of information available here to aid in making FOK ratings, the level of subthreshold activation reached, and the number and strengths of the links extending from the definition to the word's node. Both of these would presumably contribute to the item's future recallability and hence could affect performance on explicit tasks. The second of these, however, could be thought of as an index of unitization and may affect performance on implicit tasks. If so, one would expect that the FOK ratings would predict performance on both types of tasks, as Yaniv and Meyer reported. (One could make a similar argument to explain Nelson et al.'s results, using a perceptual identification task.)

Clearly, much of the above explanation is speculative. To some extent, this is due to the fact that at this point, many of the central concepts discussed here are still being developed. That is, although the literature on explicit versus implicit memory has burgeoned in recent years, researchers are only beginning to understand why explicit and implicit tasks seem to be driven by different factors. In comparison, although the phenomenon is older, research concerning the other main concept, FOK (specifically, how subjects make FOK ratings), is still in its infancy. Nonetheless, both areas of research would seem to have much to tell us about how memory works, especially memory for those things that are not immediately recallable.

References

Balota, D. A., & Chumbley, J. I. (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human Perception and Performance, 10,* 340–357.

Begg, I., Duft, S., Lalonde, P., Melnick, R., & Sanvito, J. (1989). Memory predictions are based on ease of processing. *Journal of Memory and Language, 28,* 610–632.

Belmore, S. M. (1981). Imagery and semantic elaboration in hypermnesia for words. *Journal of Experimental Psychology: Human Learning and Memory, 7,* 191–203.

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.

Bowers, J. S., & Schacter, D. L. (1990). Implicit memory and test awareness. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 404–416.

Brown, R., & McNeill, D. (1966). The tip-of-the-tongue phenomenon. *Journal of Verbal Learning and Verbal Behavior, 5,* 325–337.

Carroll, M., & Simington, A. (1986). The effects of degree of learning, meaning, and individual differences on the feeling-of-knowing. *Acta Psychologica, 61,* 3–16.

Connor, L. T., Balota, D. A., & Neely, J. H. (1990, May). *Activation and metacognition of inaccessible stored information: A further examination.* Paper presented at the annual meeting of the Midwestern Psychological Association, Chicago, IL.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11,* 671–684.

Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General, 104,* 268–294.

Forster, K. I. (1981). Priming and the effects of sentence and lexical contexts on naming time: Evidence for autonomous lexical processing. *Quarterly Journal of Experimental Psychology, 33A,* 465–495.

Gardiner, J. M. (1988). Functional aspects of recollective experience. *Memory & Cognition, 16,* 309–313.

Glenberg, A. M., Sanocki, T., Epstein, W., & Morris, C. (1987). Enhancing calibration of comprehension. *Journal of Experimental Psychology: General, 116,* 119–136.

Graf, P., & Schacter, D. L. (1985). Implicit and explicit memory for new associations in normal and amnesic subjects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11,* 501–518.

Graf, P., Squire, L. R., & Mandler, G. (1984). The information that amnesic patients do not forget. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 164–178.

Greene, R. L. (1986). Word stems as cues in recall and completion tasks. *The Quarterly Journal of Experimental Psychology, 38A,* 663–673.

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

Jacoby, L. L. (1983). Remembering the data: Analyzing interactive processes in reading. *Journal of Verbal Learning and Verbal Behavior, 22,* 485–508.

Jacoby, L. L., & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology: General, 110,* 306–340.

Jameson, K. A., Narens, L., Goldfarb, K., & Nelson, T. O. (1990). The influence of near-threshold priming on metamemory and recall. *Acta Psychologica, 73,* 55–68.

King, J. F., Zechmeister, E. B., & Shaughnessy, J. J. (1980). Judgments of knowing: The influence of retrieval practice. *American Journal of Psychology, 93,* 329–343.

Krinsky, R., & Nelson, T. O. (1985). The feeling of knowing for different types of retrieval failure. *Acta Psychologica, 58,* 141–158.

Kucera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English.* Providence, RI: Brown University Press.

Leonesio, R. J., & Nelson, T. O. (1990). Do different metamemory judgments tap the same underlying aspects of memory? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 464–470.

Lovelace, E. A. (1984). Metamemory: Monitoring future recallability during study. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 756–766.

Lupker, S. J. (1984). Semantic priming without association: A second look. *Journal of Verbal Learning and Verbal Behavior, 23,* 709–733.

MacLeod, C. M. (1988). Forgotten but not gone: Savings for pictures and words in long-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 195–212.

Nelson, T. O. (1971). Savings and forgetting from long-term memory. *Journal of Verbal Learning and Verbal Behavior, 10,* 568–576.

Nelson, T. O. (1978). Detecting small amounts of information in memory: Savings for nonrecognized items. *Journal of Experimental Psychology: Human Learning and Memory, 4,* 453–468.

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., 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., 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.

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

Norris, D. (1986). Word recognition: Context effects without priming. *Cognition, 22,* 93–136.

Richardson, A., & Bjork, R. A. (1982, November). *Recognition vs. perceptual identification: Effects of rehearsal type and duration.* Paper presented at the annual meeting of the Psychonomic Society, Minneapolis, MN.

Roediger, H. L., III, & Blaxton, T. A. (1987). Retrieval modes produce dissociation in memory for surface information. In D. Gorfein & R. R. Hoffman (Eds.), *Memory and cognitive processes: The Ebbinghaus Centennial Conference* (pp. 349–379). Hillsdale, NJ: Erlbaum.

Roediger, H. L., III, Payne, D. G., Gillespie, G. L., & Lean, D. S. (1982). Hypermnesia as determined by level of recall. *Journal of Verbal Learning and Verbal Behavior, 21,* 635–655.

Roediger, H. L., III, & Weldon, M. S. (1987). Reversing the picture superiority effect. In M. A. McDaniel & M. Pressley (Eds.), *Imagery and related mnemonic processes: Theories, individual differences, and applications* (pp. 151–174). New York: Springer-Verlag.

Roediger, H. L., III, Weldon, M. S., & Challis, B. H. (1989). Explaining dissociations between implicit and explicit measures of retention: A processing account. In H. L. Roediger III & F. I. M. Craik (Eds.), *Varieties of memory and consciousness: Essays in honour of Endel Tulving* (pp. 3–41). Hillsdale, NJ: Erlbaum.

Schacter, D. L. (1983). Feeling of knowing in episodic memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9,* 39–54.

Schacter, D. L., & Graf, P. (1986). Effects of elaborative processing on implicit and explicit memory for new associations. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12,* 432–444.

Seidenberg, M. S., Waters, G. S., Sanders, M., & Langer, P. (1984). Pre- and post-lexical loci of contextual effects on word recognition. *Memory & Cognition, 12,* 315–328.

Shaughnessy, J. J. (1981). Memory monitoring accuracy and modification of rehearsal strategies. *Journal of Verbal Learning and Verbal Behavior, 20,* 216–230.

Smith, S. M. (1988). Environmental context-dependent memory. In G. M. Davies & D. M. Thompson (Eds.), *Memory in context: Context in memory* (pp. 13–34). New York: Wiley.

Smith, S. M., Glenberg, A. M., & Bjork, R. A. (1978). Environmental context and human memory. *Memory & Cognition, 6,* 342–353.

Srinivas, K., & Roediger, H. L., III. (1990). Classifying implicit memory tests: Category association and anagram solution. *Journal of Memory and Language, 29,* 389–412.

Tulving, E., & Pearlstone, Z. (1966). Availability versus accessibility of information in memory for words. *Journal of Verbal Learning and Verbal Behavior, 9,* 381–391.

West, R. F., & Stanovich, K. E. (1982). Source of inhibition in experiments on the effect of sentence context on word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8,* 385–399.

Yaniv, I., & Meyer, D. E. (1987). Activation and metacognition of inaccessible stored information: Potential bases for incubation effects in problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13,* 187–205.

Zechmeister, F. B., & Shaughnessy, J. J. (1980). When you know that you know and when you think that you know but don't. *Bulletin of the Psychonomic Society, 15,* 41–44.

Received June 18, 1990
Revision received December 10, 1990
Accepted January 2, 1991 ■