Journal of Experimental Psychology:
Learning, Memory, and Cognition
1992, Vol. 18, No. 3, 544-554

0278-7393/92/$3.00

On the Relation Between Feeling of Knowing and Lexical Decision: Persistent Subthreshold Activation or Topic Familiarity?

Lisa T. Connor and David A. Balota
Washington University

James H. Neely
State University of New York at Albany

Experiment 1 replicated Yaniv and Meyer's (1987) finding that lexical decision and episodic recognition performance was better for words previously yielding high-accessibility levels (a combination of feeling-of-knowing and tip-of-the-tongue ratings) in comparison with those yielding low-accessibility levels in a rare word definition task. Experiment 2 yielded the same pattern even though lexical decisions preceded accessibility estimates by a full week. Experiment 3 dismissed the possibility that the Experiment 2 results may have been due to a long-term influence from the lexical decision task to the rare word judgment task. These results support a model in which Ss (a) retrieve topic familiarity information in making accessibility estimates in the rare word definition task and (b) use this information to modulate lexical decision performance.

Incubation effects in problem solving are defined as an improvement in problem-solving performance when subjects are retested on the same problems after a delay interval (Smith & Blankenship, 1989). Although phenomenologically powerful, incubation effects have been difficult to obtain reliably in an experimental setting (Olton, 1979). For example, although Dreistadt (1969), Murray and Denny (1969), and Silveira (1971) provided early evidence for incubation effects, attempted replications of each of these studies were unsuccessful (Dominowski & Jenrick, 1972; Olton, 1979; Olton & Johnson, 1976).

More recently, Yaniv and Meyer (1987, Experiment 2) reported an intriguing study that purportedly provides evidence for a type of incubation effect. They presented subjects with a series of four rare word definitions, such as "The dense, fibrous, opaque, white outer coat of the eyeball," for which subjects tried to retrieve the correct answer (sclera). If subjects were unable to retrieve the word that fit the definition, they responded to questions about their feeling-of-knowing and tip-of-the-tongue states. The feeling-of-knowing rating and the tip-of-the-tongue state were combined to produce a *latent accessibility* rating for those trials in which no response was given for a definition's referent. Accessibility was considered high if the subject was in the tip-of-the-tongue state and rated feeling of knowing as being either a 4 or 5, that is, moderately high or high confidence. Accessibility was considered low if the subject was not in the tip-of-the-tongue state, and feeling of knowing was rated as 1 or 2 (low or moderately low confidence). All other combinations of tip-of-the-tongue state and feeling-of-knowing rating were considered medium latent accessibility. After each set of four rare word definitions, subjects were given a lexical decision task in which the answers to the previous rare word definitions were embedded within a list of new words and nonwords. Finally, after receiving a series of blocks of rare word definition trials and lexical decision trials, subjects were given an unexpected episodic recognition test, again with the answers to the rare word definitions presented along with new distractor items.[1]

Yaniv and Meyer (1987) found that the items that subjects were unable to recall and that they rated as high latent accessibility produced faster reaction times (RTs) in the lexical decision task in comparison with items that were not recalled and that produced a low-accessibility estimate. The episodic recognition results mimicked the lexical decision results. That is, recognition of the items rated high in accessibility produced faster RTs than recognition of items rated low in accessibility.

Yaniv and Meyer (1987) used the subjects' answers to the accessibility questions as an index of the potential retrievability of an item. They proposed that their results were caused by persistent activation (Collins & Loftus, 1975) in the semantic memory network produced by the attempted retrieval of the solution to a rare word definition. Items that are highly retrievable (i.e., that produce a high-accessibility estimate) produce more activation than items that are less retrievable. This increased activation remains in the network long enough

This work was supported by Grant 54297A from the National Institute on Aging awarded to David A. Balota. Portions of this research were reported at the 1990 meeting of the Midwestern Psychological Association in Chicago, Illinois. We thank Thomas Carr, David Meyer, and Ilan Yaniv for their helpful comments on an earlier version of this article. In addition, we thank Anne Poinier and Brian Stephens for their help in data collection and Ken Connor for his software advice.

Correspondence concerning this article should be addressed to David A. Balota, Department of Psychology, Washington University, St. Louis, Missouri 63130. Electronic mail may be sent to C39814DB@WUVMD.

[1] In this study, we focus on Yaniv and Meyer's (1987) second experiment because in their first experiment, subjects were given the lexical decision task after each rare word definition. Hence, there was a minimum delay between the two tasks. With such short delays, it is possible that subjects could have engaged in a postaccess search for a relation between a given letter string and the earlier definition. Finding such a relation could have facilitated the word response in the lexical decision task (see Balota & Lorch, 1986, and Neely & Keefe, 1989, for a discussion of such postlexical checking processes).

to produce facilitation in responses to the definition's referent both in lexical decision and episodic recognition performance.

Although Yaniv and Meyer's (1987) data are consistent with a persistent activation account, several questions remain concerning the feasibility of this explanation. First, one of the major assumptions of the spreading activation model (Collins & Loftus, 1975) is that activation decays rapidly with passing time and intervening activity. The data presented by Yaniv and Meyer challenge this assumption because in their study residual activation would have had to remain in the memory network on the order of 4 to 30 min. Moreover, there was intervening activity during this delay interval. Foss (1982) and Masson (1991) have shown that spreading activation is considerably reduced or eliminated across unrelated intervening items, even at intervals shorter than 30 s (see also Joordens & Besner, 1992; McNamara, 1991). Either a revision of spreading activation theory is necessary to account for Yaniv and Meyer's results or these authors need to more explicitly state why subthreshold activation persists longer than suprathreshold activation.

Second, and more important, there is an alternative interpretation of Yaniv and Meyer's (1987) data that does not appeal to persistent subthreshold activation. This alternative explanation assumes that people make accessibility estimates on the basis of the likelihood that they have knowledge about the general topic of the rare word definition problem (Nelson, Gerler, & Narens, 1984) or that they are familiar with the various elements in the query (Reder & Ritter, 1992). For example, if a person has a considerable knowledge base about the topic of music and the rare word definition that is presented refers to an arcane musical term, the person would rate his or her accessibility level as relatively high. However, if the definition referred to chemistry and the person had little knowledge about chemistry, then the person would be more likely to rate his or her accessibility level as relatively low. Later in the lexical decision task, the person would respond faster, more accurately, or both to the musical term, with which he or she had some familiarity, than to the chemistry term, with which he or she had little or no familiarity. Thus, by this account subjects may be rating their accessibility level on the basis of metacognitive knowledge about a given topic. It should be no surprise that subjects recognize words more quickly from topics for which they have high accessibility in comparison with topics for which they have very little accessibility.

We report three experiments designed to discriminate between these two explanations of Yaniv and Meyer's (1987) results. The first was an attempted replication of Yaniv and Meyer's Experiment 2 using the same materials and procedures. Although the reaction time effect Yaniv and Meyer reported in their second experiment is theoretically intriguing, the size of the effect is relatively small. In fact, because the Item Type × Accessibility Level interaction did not reach significance, Yaniv and Meyer were forced to rely on separate analyses of target items and control items to support effects of latent accessibility level on lexical decision performance. Hence, because of some concern regarding the strength of the effect, we first decided to attempt a replication of the original pattern. The second experiment was a variation of Yaniv and Meyer's Experiment 2, in which their procedure was, in some sense, reversed. In the first session, subjects performed all lexical decision trials. In the second session subjects first attempted to retrieve responses for the rare word definitions, thereby yielding information regarding accessibility level for unretrieved words. After the rare word definition task, subjects received the unexpected recognition task. The rationale for the second experiment is quite simple. If it was the residual activation produced by the prior accessibility level that was modulating lexical decision performance in Experiment 1, then there should be no relation between later accessibility level and earlier lexical decision performance in our Experiment 2. Very simply, if the attempted retrieval during a rare word retrieval task produces residual activation that influences lexical decision performance, then how could there be such an influence when this attempted retrieval occurs 1 week after the lexical decision task? On the other hand, if the topic familiarity hypothesis is correct, subjects should produce faster responses, more accurate responses, or both in the lexical decision task for items that are judged as high in accessibility level, in comparison with items that are judged as low in accessibility level, regardless of whether the lexical decision task follows or precedes the accessibility estimates.

Finally, a third experiment was designed to address the possibility that the results obtained in the second experiment could be caused by an influence from the lexical decision task to the latent accessibility task 1 week later. Although, as discussed later, this possibility is dismissed, it could have considerably compromised the topic familiarity account.

Experiment 1

Method

Subjects

Forty-four undergraduate volunteers recruited from Washington University psychology courses participated in this study.

Apparatus

The experiment was controlled by an Apple IIe microcomputer, equipped with a Mountain Hardware clock that measured reaction times to the nearest millisecond. A software routine synchronized stimulus presentation with the timer.

Materials

The definitions and corresponding referents were the same as those used by Yaniv and Meyer (1987). These consisted of 52 pairs of rare words (matched for length) and their respective definitions. For a given subject, one of the rare words in each pair served as a target for which the definition of that word was presented during the rare word definition task. The other member of the pair served as a control word, and its definition was not seen by the subject. List 1 contained all targets for the odd-numbered subjects, and List 2 contained all targets for the even-numbered subjects. The mean frequency of the targets that were listed in the Kučera and Francis (1967) norms in List 1 was 3 per million (range: 1 to 12 per million), and the mean frequency of the targets in List 2 was 5 per million (range: 1 to 29 per million). Across subjects each word occurred equally often in the target and control word conditions.

For each pair of rare words, two pronounceable nonwords (e.g., *tonkord* and *miximize*) were constructed by replacing one to three letters of a valid English word. In addition, two filler words that were valid English words but were not part of the rare word list were included. Nonwords and filler words were matched for length with the rare word pairs. These six items formed a stimulus block. Each of the items of the stimulus block were presented to all subjects during the lexical decision task with the following constraint on the item's position in the stimulus block: Target and control words were never in the first or second position in the block. Otherwise, items were randomly assigned to a position in the block.

Subjects received a block of four definitions followed by the 24 lexical decision items (4 stimulus blocks of 6 items) that corresponded to the definitions' stimulus blocks. Thus, these 24 lexical decision items included 16 words (4 targets, 4 control words, and 8 fillers) and 8 nonwords. Subjects completed 13 blocks of experimental trials, for a total of 364 trials.

Procedure

The procedure was also based on the Yaniv and Meyer (1987) paradigm. Subjects were assigned alternately to List 1 or to List 2 as they came to participate in the study. Subjects were instructed that on each trial they would receive on the cathode-ray tube (CRT) a definition that corresponded to a rare word. The prompt "The word is . . ." would be presented at the bottom of the screen. The subject's task was to type in the word that corresponded to the definition on each trial and then to press the return key. The definition and prompt remained on the screen until the subject responded.

Subjects were informed that if they were unable to retrieve a word for the definition, they were simply to press the return key on that trial. As in Yaniv and Meyer (1987), the definition and response were erased from the screen when the return key was depressed, and the question "Did you enter a word? (Y/N)" appeared. After the subject's response to this question, one of the following appeared: If the subject responded "Y," a statement appeared, "Rate your confidence in the word you entered." followed by a five-point scale for which the subject was to type in the number that best corresponded to his or her perceived confidence. A 1 corresponded to low confidence, and a 5 corresponded to high confidence. If the subject responded "N", the question "Is the word on the 'tip of your tongue?' (Y/N)" appeared on the screen, and subjects responded by typing in the response that best reflected their state. This was followed by a request for the subject to rate his or her feeling of knowing. The feeling-of-knowing judgment reflected how well the subject thought he or she knew the correct referent but was unable to recall the word. A 1 corresponded to a low feeling of knowing, and a 5 corresponded to a high feeling of knowing.

Subjects responded to each lexical decision item by pressing the 0 key if the item was a word or by pressing the 1 key if the item was a nonword. The message "ERROR" appeared on the screen for 250 ms immediately after an incorrect response was made or if the subject failed to respond within 2 s. Each item remained on the screen until the subject responded or until 2 s elapsed. The interstimulus interval was 1 s. Each subject received a random order of definitions. Subjects were informed that they could take a rest break after each block of trials.

The unexpected recognition task was presented after the last block of definitions and lexical decisions. Subjects were instructed on the computer screen to press the 0 key if the word that appeared had been seen in the experimental session or to press the 1 key if the word had not been seen in the experimental session. The recognition test consisted of 208 randomly presented words, 104 target and control words from the lexical decision task and 104 distractor words. Consistent with Yaniv and Meyer (1987), the distractor words ranged in frequency from 1 to 142 per million, with a mean frequency of 8 per million for those words that were listed in the Kučera and Francis (1967) norms. Stimuli remained on the screen until the subject responded. No error feedback was given during the recognition test. The intertrial interval was 1 s. Rest breaks were given after each set of 52 trials.

Subjects participated individually in an experimental session that lasted approximately 75 min. Subjects were seated approximately 40 cm from the computer monitor. Subjects received 3 practice definitions and 40 practice lexical decision items with the experimenter in the room to answer any questions. After the practice trials, the experimenter left the testing room with further instructions provided by the computer.

Results

Targets were categorized into three different accessibility levels. In addition, the control items were also broken down into the same three levels. Of course, the control items did not directly produce latent accessibility scores, because these items, by definition, were never related to the definitions that a given subject received. The latent accessibility measures for the control items were obtained by means of pairing subjects across Lists 1 and 2. For example, if an item fell into the high-accessibility category for Subject 1, List 1, then that same item would be classified in the high-accessibility category for Subject 2, List 2. In this way, the same items were contributing to the target and control conditions for each accessibility level.[2]

We also used a variation of Yaniv and Meyer's (1987) adjustment procedure for taking into account differences in overall item difficulty. First, a grand mean (*GM*) for all subjects' lexical decision responses to word items was calculated. Second, the item mean (*IM*) across all subjects was calculated. Finally, each RT was adjusted by the following formula:

$$\text{Adjusted RT} = \text{Actual RT} + (GM - IM).$$

All means reported were derived from these adjusted RTs. Additionally, error rates (ER) were adjusted using the same procedure, that is,

$$\text{Adjusted ER} = \text{Actual ER} + (GM - IM).$$

Because each subject selected his or her own condition by rating tip-of-the-tongue state and feeling of knowing, there

[2] This method of control analysis is slightly different from Yaniv and Meyer's (1987) method. Yaniv and Meyer calculated the mean RTs for the items that appeared in the alternative list and were in the same stimulus block as the target item. The goal of their analysis was to control for variability across blocks of trials within a given subject because of differences induced by a block of easy or difficult definitions. However, with this procedure, different items are contributing to the means for target and control conditions across the three latent accessibility levels. Our goal was to reduce item effects. Thus, having the same items contributing to target and control conditions avoids this problem.

was a small number of cases in which a subject did not produce responses in all three latent accessibility levels. There were no subjects, however, who failed to produce responses in the correct- and incorrect-recall states or who failed to respond in at least two of the three accessibility levels. To avoid totally eliminating these subjects from the data analyses and to address the problem of missing cells, we used the following procedure.

First, for those subjects who responded in all three latent accessibility levels, the mean and standard deviation of all responses (latency or error rate) were calculated. Second, the mean was calculated for the high-, medium-, and low-accessibility levels separately using these same subjects. Then, for each accessibility level, a z score was calculated that determined how many standard deviations the mean of that accessibility level was from the overall mean of the distribution of all responses.

Next, for a subject who was missing observations in, for example, the high-accessibility level, a mean and standard deviation were computed for all other responses from that subject. Finally, to estimate the missing cell, the subject's mean and standard deviation and the z score for the missing level computed from those subjects who contributed to all three levels were combined in the following manner:

$$\text{Subject's } M + [\text{Subject's } SD \times z \text{ (estimated cell)}].$$

This estimation procedure produces an estimate for the missing cell that is precisely as far from the individual's mean as that condition is from the grand mean of all responses, taking into account the individual subject's relative speed and variability of responding. Only 9 of the 132 possible cells in this experiment were estimated using this procedure.[3]

Lexical Decision Performance

RT outliers were removed in the following manner: The mean RT for correct word responses was calculated for each subject. If a particular RT was three standard deviations above or below that subject's mean, it was considered an error and did not contribute to the mean RT for that cell. The proportion of outliers removed in each condition was as follows: correct, .019; incorrect, .029; high-accessible, .024; medium-accessible, .036; and low-accessible, .025. The total proportion of responses that were considered outliers was .025.

Overall lexical decision performance. Before turning to the critical data concerning whether lexical decision performance is modulated by latent accessibility level, we first present data regarding overall lexical decision performance as a function of target recall state, target versus control word, and lexicality. These data are displayed in Table 1. First, consider the impact of recall state for the target referents to the definitions that a given subject received. As shown in the top three rows in Table 1, if the target word was correctly produced earlier as a referent to a definition, lexical decision RTs were faster and more accurate than in conditions in which subjects either produced an incorrect referent or were unable to produce any referent, that is, were in the no-recall state (all ts > 3.14; all ps $< .003$). Turning to the control words, these items did not differ from either the no-recall state words or the incorrect-recall state words in error rate (both ps $> .15$). RTs to control words were overall faster than RTs to no-recall state words, $t(43) = 2.81$, $p < .01$, but did not differ from RTs to incorrect-recall state words, $t(43) = .41$.

Table 1
Experiment 1—Mean Reaction Time (RT; in Milliseconds) and Percentages of Error for Targets, Controls, and Nonwords as a Function of Recall State

Stimulus type	Mean RT	% Error
Target word		
No recall	905	12
Correct	828	3
Incorrect	876	15
Control word	881	12
Nonword	1,007	22

Table 2
Experiment 1—Mean Reaction Time (RT; in Milliseconds) and Percentages of Error (% Err) for Lexical Decision and Episodic Recognition Tasks as a Function of Latent Accessibility Level and Word Type

Stimulus type	Low		Medium		High	
	RT	% Err	RT	% Err	RT	% Err
Lexical decision						
Target	897	15	914	14	904	7
Control	869	14	898	10	897	10
Episodic recognition						
Target	973	21	929	15	907	12
Control	924	20	987	20	906	21

No-recall state. We shall now turn to the more important data concerning the degree to which latent accessibility scores for targets in no-recall states modulated lexical decision performance. Table 2 shows the mean RTs for the words in the no-recall state as a function of whether the word is a target or control item and its corresponding latent accessibility level. As shown in Table 2, there appears to be very little impact of latent accessibility on RTs for either control items or targets. A 2 (target vs. control) by 3 (latent accessibility level) analysis of variance (ANOVA) on the RT data failed to produce reliable differences (all Fs < 1.42). Thus, our RT data did not yield Yaniv and Meyer's (1987) finding of faster RTs for high-accessibility than for low-accessibility items.

Table 2 also displays the mean percentage of error data as a function of target versus control and latent accessibility level. These data indicate an impact of latent accessibility on errors to the target items but not on errors to the control items. Although the overall ANOVA did not yield a reliable interaction between item type and latent accessibility, $F(2,$

[3] The following are the mean RTs and percentages of error for the accessibility levels for the Experiment 1 lexical decision data without estimates included: high, 912 ms (7%); medium, 913 ms (14%); and low, 897 ms (15%). These means are quite consistent with the means including estimated cells in Table 2. Hence the estimation procedure did not artificially produce this pattern of data.

86) = 1.60, p = .21, MS_e = .016, separate planned one-way ANOVAs for targets and control items indicated that error rates for the target items were being modulated by latent accessibility level in the predicted direction, $F(2, 43) = 5.09$, p = .008, MS_e = .015, whereas the corresponding influence of latent accessibility level for the control words did not reach significance, $F(2, 43) = 1.58$, p = .21, MS_e = .019. Thus, the error data replicate Yaniv and Meyer's (1987) finding of statistically better performance on high-accessibility targets than on low-accessibility targets and statistically equivalent performance on high- and low-accessibility control items.

Episodic Recognition

The overall proportion of outliers in recognition performance was .019. For the correct recall state, the proportion was .010; for the incorrect recall state, .010; for high-accessibility, .012; for medium-accessibility, .004; and for low-accessibility, .014.

In Table 2, the mean episodic recognition RTs are displayed as a function of word type and latent accessibility level. It appears that accessibility is influencing RTs in the predicted direction and primarily for target words. However, the overall ANOVA on the RT data did not yield a reliable Item Type × Accessibility Level interaction, $F(2, 86) = 1.52$, p = .22, MS_e = 41,115. In addition, neither the one-way ANOVAs on the targets nor those on the control words produced a significant main effect of accessibility (both Fs < 2.22). It should be noted, however, that a planned comparison did indicate that RTs for targets rated low in accessibility were reliably slower than RTs for targets rated high in accessibility, $t(43) = 1.78$, $p < .05$, one-tailed. Also, the control words produced an inconsistent pattern, with the medium-accessibility level producing slower RTs than either the high- or low-accessibility levels.

Turning to the error rate data displayed in Table 2, one can see that the recognition error rates decreased across accessibility level from low to high for targets, but for control items the error rates remained constant across accessibility levels. Again, the overall interaction between item type and accessibility level did not reach significance, $F(2, 86) = 1.24$, p = .29, MS_e = .044. However, the one-way ANOVA on the targets did yield a marginal main effect of accessibility level, $F(2, 43) = 3.01$, p = .053, MS_e = .034, and the corresponding analysis for the control words did not approach significance ($F < 1$). It should again be noted that planned comparisons for the targets yielded reliable differences between the low- and medium-accessibility levels, $t(43) = 1.78$, $p < .05$, one-tailed, and between the low- and high-accessibility levels, $t(43) = 2.52$, p = .01. Thus, overall the data from the episodic recognition data conform to the predicted direction. Specifically, both RTs and error rates were lower for nonrecalled words that were rated high in accessibility than for nonrecalled words that were rated low in accessibility.

Discussion

Experiment 1 produced the predicted influence of an item's latent accessibility level on error rates in both lexical decision and in episodic recognition performance. Moreover, there was the predicted impact of latent accessibility level in the RTs in the episodic recognition test.

Although our lexical decision RT data do not replicate the Yaniv and Meyer (1987) RT data, the error rates are consistent with their data. It is possible that subjects in the present study were more likely than subjects in the Yaniv and Meyer (1987) study to stress speed over accuracy in their lexical decision performance. Because Yaniv and Meyer did not report mean error rate as a function of latent accessibility level, we cannot make a direct comparison across studies concerning overall error rate. It should also be noted that although our Item Type × Latent Accessibility Level interaction did not reach significance in the lexical decision error rate, Yaniv and Meyer also failed to produce a significant Type × Latent Accessibility Level interaction in their corresponding lexical decision RT data. As in the present experiment, they relied on separate analyses for targets and control items to provide evidence for differential effects of latent accessibility level on lexical decision performance. Thus, our analyses have the same level of statistical force as Yaniv and Meyer's analyses. The important point for present purposes is that Experiment 1 provides a conceptual replication of Yaniv and Meyer's results. That is, the items that yielded relatively high latent accessibility levels produced better lexical decision performance (in error rates) and episodic recognition performance (in error rates and RTs) than those items that yielded low latent accessibility levels.

We now turn to the more important issue of the underlying mechanism that is producing these effects. As noted earlier, it is possible that the attempted retrieval of the referent to the definition is producing the influence of latent accessibility, as Yaniv and Meyer (1987) have argued, or that subjects' latent accessibility ratings reflect metaknowledge about familiar topics. To discriminate between these two accounts, we reversed the order of tasks in the second experiment. In Experiment 2, subjects first made lexical decisions to the targets and then attempted to provide the referents for the definitions, as well as providing latent accessibility measures. The predictions are straightforward: If Yaniv and Meyer are correct and the influence of latent accessibility on lexical decision is due to the prior attempted retrieval of the definition's referent, then one should not find an impact when the attempted retrieval occurs on average a week after the lexical decision task. On the other hand, if the influence of latent accessibility is simply due to subjects being able to access metaknowledge about familiar topics, then one should still find a relation between latent accessibility ratings and lexical decision performance, even though the latent accessibility measure is obtained 1 week after the lexical decision task.

Experiment 2

Method

Subjects

Forty-two undergraduate volunteers recruited from Washington University psychology courses participated in this study. None had participated in Experiment 1.

Apparatus and Materials

Apparatus and materials were the same as in Experiment 1.

Procedure

Subjects participated in two sessions approximately 7 days apart (M = 8 days; range = 5 to 15 days). The first session lasted approximately 30 min, and the second session lasted approximately 45 min. In Session 1, subjects were presented with the identical lexical decision task that was used in Experiment 1 with no definitions and no breaks between successive stimulus blocks. In the second session, subjects were presented the 52 rare word definitions from List 1 or 2 as described in Experiment 1. Definitions were given in 13 blocks consisting of 4 definitions each. Each block of definitions was followed by a 1-min rest break. As in Experiment 1, subjects were again given an episodic recognition test. However, subjects were instructed to make their recognition judgments on whether they saw a word in Session 1 or 2. Because the task of recognizing words that occurred in the same experimental session or in a session 1 week earlier is a considerably different recognition test than the test used by Yaniv and Meyer (1987) and in our Experiment 1, we focus on the lexical decision data of the second experiment.

Results

Lexical Decision Performance

The overall proportion of outliers in lexical decision performance was .036. The proportions for each of the other conditions was as follows: correct recall state, .020; incorrect recall state, .014; high-accessibility, .014; medium-accessibility, .021; and low-accessibility, .043.

Once again, each subject was matched with a subject that received definitions from the other list. In this experiment, only 7 of the 126 cells were estimated using the procedure described earlier.[4]

Overall lexical decision performance. First, we review the overall performance of subjects in the lexical decision task as a function of recall state, word type, and lexicality. Table 3 presents these data. As shown, responses to targets that produced correct responses were faster and more accurate than responses to targets that produced the no-recall state (both ts > 2.10; both ps $< .04$). The correct-recall state did not differ from the incorrect-recall state in RTs, $t(41) = .75$, but did reliably differ in error rate, $t(41) = 4.14$, $p < .001$. Turning to the control items, we see that responses to controls were faster than responses to no-recall items, $t(41) = 4.47$, $p < .001$, but were no more accurate, $t(41) = .95$. Responses to control items, however, did not reliably differ from responses to the incorrect-recall items in either RTs or error rates (both ts $< .20$).

No-recall state. Mean RTs as a function of word type and latent accessibility level are shown in Table 4. There appears to be some tendency in these data for latent accessibility to modulate RTs in the predicted direction for target words but not control items. Specifically, both the medium- and high-accessible items produced faster RTs than low-accessible items for targets, but for controls the medium-accessible items produced slower RTs than both the low- and high-accessible items. This observation was supported by a reliable Word Type × Accessibility Level interaction, $F(2, 82) = 5.29$, $p < .01$, $MS_e = 5546$. The results of planned comparisons indicated that for the target items the low-accessibility level produced, as predicted, slower RTs than the medium-accessibility level, $t(41) = 2.91$, $p < .01$, whereas for the control items, the medium-accessibility level actually produced slower RTs than both the high- and low-accessibility levels (both ts > 2.57; both ps $< .013$).

Table 4 also displays the mean percentage of errors as a function of word type and latent accessibility level. Here, errors strongly conform to the predicted pattern. The overall ANOVA yielded a marginal main effect of latent accessibility level, $F(2, 41) = 2.95$, $p = .056$, $MS_e = .013$, and, more important, a reliable interaction, $F(2, 82) = 7.21$, $p = .001$, $MS_e = .018$. Planned one-way ANOVAs produced a highly reliable main effect of latent accessibility level for targets, $F(2, 41) = 9.58$, $p < .001$, $MS_e = .013$. The low-accessibility condition produced higher error rates than the medium-accessibility condition, which in turn produced higher error rates than the high-accessibility condition (both ts > 2.03; both ps $< .05$). For control items, the one-way ANOVA did not yield a reliable effect of accessibility level, $F(2, 41) = 2.24$, $p = .11$, $MS_e = .018$.

Table 3
Experiment 2—Mean Reaction Time (RT; in Milliseconds) and Percentages of Error for Targets, Controls, and Nonwords as a Function of Recall State

Stimulus type	Mean RT	% Error
Target word		
No recall	815	13
Correct	787	6
Incorrect	780	12
Control word	782	12
Nonword	936	22

Table 4
Experiment 2—Mean Reaction Time (RT; in Milliseconds) and Percentages of Error (% Err) for the Lexical Decision Task and Episodic Recognition Task as a Function of Latent Accessibility Level and Word Type

Stimlus type	Low		Medium		High	
	RT	% Err	RT	% Err	RT	% Err
Lexical decision						
Target	834	18	799	13	811	7
Control	784	12	822	10	781	16
Episodic recognition						
Target	1,382	39	1,312	36	1,462	35
Control	1,331	35	1,145	30	1,323	41

[4] The following are the mean RTs and percentages of error for the accessibility levels for the Experiment 2 lexical decision data without estimates included: high, 813 ms (7%); medium, 800 ms (13%); and low, 838 ms (18%). These means are quite comparable to the means including estimated cells in Table 4. Hence the estimation procedure did not artificially produce this pattern of data.

Episodic Recognition

Because the recognition task in Experiment 2 demanded that subjects discriminate new words from words that were seen either immediately (in the definition task) or, on average, 1 week earlier (in the lexical decision task), it is not comparable to the recognition task of Experiment 1. The recognition task in Experiment 2 is relatively more difficult, as is reflected by subjects' relatively slow RTs and high error rates, in comparison with those of Experiment 1. The overall RTs and percentage of error were 1,385 ms (37%) for targets and 1,266 ms (35%) for controls. The mean RTs as a function of accessibility level and item type are displayed in the bottom half of Table 4. No comparisons attained statistical significance.

Discussion

Once again, the pattern of data for the error rates conforms to the original pattern reported by Yaniv and Meyer (1987). For targets, the low-accessibility level items yielded a higher error rate than either the medium- or high-level items. For control items, there was no effect of latent accessibility level.

The importance of these findings is that this pattern was obtained under conditions in which the rare word definition followed the presentation of the definition's referent by 1 week. Hence, there was no opportunity for the lexical decision to that referent to be influenced by the activation of the referent's node in memory by prior exposure to the definition. Therefore, the persistent subthreshold activation account Yaniv and Meyer (1987) provided for their data is unable to accommodate the present findings. The topic familiarity account, however, correctly predicts better performance for the high-accessibility items when the lexical decisions both precede and follow the rare word definition task.

There is, however, an alternative account of the results of Experiment 2. Specifically, the obtained pattern may not be because of topic familiarity but rather could be because of an influence from the earlier lexical decision task to performance in the rare word definition task a week later. Consider the plausible notion that attention fluctuates across trials within the lexical decision task. On some trials, subjects are relatively alert (high-attention trials), whereas on other trials, subjects are relatively less alert (low-attention trials). On high-attention lexical decision trials, one might expect subjects to be relatively faster and more accurate than on low-attention trials. Moreover, on high-attention trials, one might expect that subjects would be more likely to encode the target word into memory more fully. In fact, this more fully developed encoding could influence performance on the rare word definition task given a week later. Thus, it is possible that high-attention lexical decision trials might produce better memory traces, and hence, one might observe a relation between lexical decision performance and rare word definition performance, even though the lexical decision task occurred a week earlier.[5]

Interestingly, a comparison of the results from the first and second experiments provides some support for this possibility. In particular, the mean number of observations that fell within either the correct-recall state or the high-accessibility category was slightly higher in Experiment 2 (24.5), when the lexical decision task preceded the rare word definition task, than in Experiment 1 (20.2), when the rare word definition task preceded the lexical decision task, $t(84) = 2.41$, $p = .02$. Unfortunately, however, this comparison is compromised by the fact that the subject sample in Experiment 2 (mean age = 20.6 years) was more likely to involve older and more advanced undergraduates than in Experiment 1 (mean age = 19.5 years), $t(84) = 2.14$, $p = .03$. Thus, it is possible that this difference may not be attributed to attentional fluctuation but rather to subject differences.

The major goal of Experiment 3 was to further explore the attentional fluctuation account of the results from the second experiment. In pursuit of this goal, as in Experiment 2, the lexical decision trials preceded the rare word definition task by 1 week. However, in Experiment 3, subjects only received half (52) rather than all of the definition referents embedded in the lexical decision task in Session 1. A week later, subjects received all 104 rare word definitions, consisting of the 52 definition referents that were presented in Session 1 and 52 definition referents that were never presented. Immediately after the rare word definition task, all of the definition referents were presented embedded in a final lexical decision task.

If the attentional fluctuation model is correct, and lexical decision performance actually influences later rare word definition performance across a week interval, then one should find higher recall, higher accessibility levels, or both for definitions that correspond to target words that were presented in the previous week, in comparison with items that were not presented. On the other hand, the topic familiarity model predicts little if any difference across these conditions because prior exposure of the word should not substantially modulate the subject's familiarity with the topic.

In addition, Experiment 3 affords a replication of the basic pattern observed in both Experiments 1 and 2. Specifically, one should find that the high-accessibility items should yield fewer errors than the low-accessibility items for both the lexical decision session that precedes the rare word definition task by 1 week and the lexical decision session that immediately follows the rare word definition task. More important, according to the topic familiarity model, the size of this effect should be similar across the two sessions. This pattern would provide additional support for the notion that the act of retrieving the rare word definition is not producing subthreshold activation that later modulates lexical decision performance.

Experiment 3

Method

Subjects

Thirty-eight undergraduate volunteers recruited from Washington University psychology courses participated in this study. None had participated in Experiments 1 or 2.

[5] We thank David Meyer and Ilan Yaniv for providing this alternative explanation of the results from Experiment 1.

Apparatus and Materials

Apparatus and materials were the same as in Experiment 1.

Procedure

Subjects participated in two sessions approximately 7 days apart ($M = 6.9$ days, range = 5 to 10 days). The first and second sessions lasted approximately 25 and 75 min, respectively. In Session 1, subjects were presented with the lexical decision items as in the previous two experiments with one change. The control words were not included, because these words would serve as the nonrepeated targets in the second session. Subjects received 260 lexical decision items (5-item stimulus set × 52 sets) with no rest breaks between successive blocks. The two lists that served as target items in Session 1 were counterbalanced across subjects with the odd- and even-numbered subjects receiving Lists 1 and 2, respectively.

In the second session, subjects were presented with all 104 rare word definitions for which half of the referents had been seen in the previous session and half had not been seen. Definitions were given in 26 blocks consisting of 4 definitions each. Instructions for responding to the rare word definition task were identical to those of the previous experiments. Subjects did not participate in an episodic recognition task.

Results

Overall lexical decision performance. First, we present the lexical decision data as a function of recall state and lexicality broken down by session and whether the item was repeated or not. Table 5 displays these data. For Session 1, responses to correct-recall state targets were faster and more accurate than responses to no-recall state targets (both ts > 3.64; both ps < .002). Responses to the correct-recall state items were not faster than the responses to the incorrect-recall state items, $t(37) = 1.17$, but were more accurate, $t(37) = 3.02$, $p < .005$.

For Session 2 repeated targets, responses to the correct-recall state items differed reliably from responses to the no-recall state and the incorrect-recall state items in RTs and in error rate (all ts > 2.32; all ps < .03). For Session 2 nonrepeated targets, the same pattern of data was obtained. Responses to correct-recall state items were reliably faster and more accurate than responses either to no-recall state items or to incorrect-recall state items (all ts > 6.69; all ps < .001).

In addition, as shown in Table 5, there is clear evidence of a long-term repetition effect. That is, subjects were both faster and more accurate in Session 2 for targets that appeared in Session 1 than for targets that were not presented during Session 1. This long-term repetition effect is somewhat attenuated in the correct-recall condition. This reduction most likely indicates that the recent exposure to the definition referent by means of correct recall overshadowed any long-term repetition priming.

A 2 (repetition) by 3 (recall state) ANOVA supported these observations. There were reliable main effects of repetition, $F(1, 37) = 20.04$, $p < .001$, $MS_e = 7{,}891$, $F(1, 37) = 91.47$, $p < .001$, $MS_e = .003$, and recall state, $F(2, 37) = 39.00$, $p < .001$, $MS_e = 3{,}134$, $F(2, 37) = 29.30$, $p < .001$, $MS_e = .006$, along with reliable interactions between these two factors, $F(2, 74) = 3.58$, $p = .03$, $MS_e = 2{,}127$, $F(2, 74) = 11.27$, $p < .001$, $MS_e = .004$, in both RTs and error rates, respectively. A series of planned comparisons indicated that the only condition that did not yield a reliable repetition effect was the error rate for the correctly recalled words condition, $t(37) = 1.80$, $p = .08$.

No-recall state. To address a potential criticism of our previous data analysis techniques, and to decrease the number of estimated cells, a procedural change was made in the data analyses performed. As shown in Table 6, the distribution of responses in the three latent accessibility levels for all three experiments reveals that subjects' responses were skewed toward the low-accessibility level. Yaniv and Meyer (1987), however, reported a roughly equal number of responses in each of the three latent accessibility levels. Possibly, the results of our previous experiments may have been somehow compromised by the unequal frequencies of responses in the three accessibility levels. To better equalize the number of observations in the latent accessibility categories, we collapsed across the high- and medium-accessibility levels into a single high/medium-accessibility level.[6] As can be seen in Table 6, the distribution of responses is then approximately equal in

Table 5
Experiment 3—Mean Reaction Time (RT; in Milliseconds) and Percentages of Error for Targets and Nonwords as a Function of Session and Recall State

Stimulus type	Mean RT	% Error
Session 1: Target word		
No recall	878	14
Correct	845	9
Incorrect	857	14
Session 1: Nonword	975	15
Session 2, referents previously seen:		
Target word		
No recall	803	8
Correct	743	4
Incorrect	794	8
Session 2, referents not previously seen:		
Target word		
No recall	853	20
Correct	777	6
Incorrect	868	17
Session 2: Nonword	961	17

[6] If one also collapses across the high and medium categories in Experiments 1 and 2, the results do not substantially change. Specifically, the recognition data of Experiment 1 and the lexical decision data of Experiment 2 clearly indicate that there is a larger effect of latent accessibility (in the predicted direction) in both RTs and error rates for the target items than for the control items. The only noticeable change is in the error rates for the lexical decision data of Experiment 1. That is, the effect of latent accessibility in error rates was quite similar for the control and target items—4% and 5%, respectively. There are two points to note here. First, there is some tendency for the high/medium-accessibility control items to actually produce slower (22 ms) RTs than low-accessibility control items. Hence, the 4% influence of accessibility for the control items in error rates may reflect some trade-off of speed for accuracy. Second, and more important, Experiment 1 is merely a replication of Yaniv and Meyer's (1987) second experiment, and the results from our third experiment provide a successful conceptual replication of Yaniv and Meyer's results in terms of the relation between latent accessibility and lexical decision error rates.

the two new categories. That is, the low-accessibility level for the repeated targets has 425 observations, and the high/medium-accessibility level for the repeated targets has 380 observations. For the nonrepeated targets, the low-accessibility level has 444 observations, and the high/medium-accessibility level has 416 observations. The number of estimated cells in this analysis is 3.9% of the total number of cells in this experiment.[7]

The mean RTs for the two lexical decision tasks as a function of latent accessibility level and target type (repeated vs. nonrepeated) are shown in Table 7. It appears that there are no RT differences as a function of accessibility level. This observation was confirmed by an overall 3 (target type) × 2 (accessibility level) repeated measures ANOVA. There was a main effect of target type, $F(2, 74) = 10.01$, $p < .001$, $MS_e = 7{,}746$, no effect of accessibility level, $F(1, 37) = .362$, $p > .5$, $MS_e = 3{,}809$, and no interaction, $F(2, 74) = 1.28$, $p = .28$, $MS_e = 6{,}642$.

The mean percentage of errors as a function of accessibility level and target type are also displayed in Table 7. These data suggest a strong relation between percentage of error and accessibility level. More important, the size of this relation did not change when the lexical decision task was presented an average of 7 days before the rare word definition task in relation to when the lexical decision task immediately followed the rare word definition task. These observations were confirmed by an overall 3 (target type) × 2 (accessibility level) repeated measures ANOVA. The ANOVA revealed a main effect of target type, $F(2, 74) = 20.92$, $p < .001$, $MS_e = .02$, and a main effect of accessibility level, $F(1, 37) = 13.66$, $p = .001$, $MS_e = .03$, but no evidence of an interaction, $F(2, 74) = .577$, $p > .5$, $MS_e = .02$. Planned comparisons revealed that the high/medium-accessibility level produced a lower percentage of error than its low-accessibility level counterpart across all three target type conditions (all ts > 1.94; all ps < .029, one-tailed).

Repeated versus nonrepeated targets. To address the attentional fluctuation hypothesis, a test was conducted to determine whether a difference existed in the mean number of observations per subject for the high-accessible and the correct-recall state for repeated and nonrepeated targets. As noted earlier, the attentional fluctuation model predicts that recall and accessibility levels should be higher for repeated targets than for nonrepeated targets. This comparison demonstrated that there was no difference between the mean number of observations for these categories, $t(37) = 1.32$, $p = .191$. The repeated targets produced a mean of 21.4 high-accessible nonrecalled and recalled items, and the nonrepeated targets produced a mean of 20.2. A similar comparison of the high/medium category yielded the same pattern, $t(37) = 1.13$, $p = .27$. In fact, as shown in Table 6, there were slightly more items in the high-accessibility category for the new items than for the old items. Clearly, this does not support the notion that the appearance of items in the prior lexical decision task was somehow influencing performance 1 week later on the rare word definition task.

Table 6
Distribution of Observations in Latent Accessibility Levels for all Three Experiments

Experiment	Low	Medium	High
Experiment 1	447	192	165
Experiment 2	472	273	219
Experiment 3			
Repeated targets	425	245	135
Nonrepeated targets	444	260	156

Table 7
Experiment 3—Mean Reaction Time (RT; in Milliseconds) and Percentages of Error (% Err) for Lexical Decision Tasks as a Function of Session and Latent Accessibility Level

	Low		High/Medium	
Session	RT	% Err	RT	% Err
Session 1	878	17	863	10
Session 2				
Nonrepeated targets	852	26	854	15
Session 2				
Repeated targets	795	10	822	4

General Discussion

The results of Experiment 1 conceptually replicated Yaniv and Meyer's (1987) results. Specifically, when a subject failed to retrieve a word that corresponded to its definition, subsequent lexical decision performance and episodic recognition performance on that word was better when the accessibility level for that unretrieved item was high and the subject reported a tip-of-the-tongue state. Yaniv and Meyer interpreted their findings as follows: (a) Activation levels in the node corresponding to the definition's referent are higher for items receiving higher accessibility estimates even when these activation levels are not sufficient for surpassing the threshold for recall, and (b) the better subsequent lexical decision performance for that item is the result of the subthreshold activation remaining in the semantic memory network long enough to facilitate performance. However, the results of Experiment 2 challenge Yaniv and Meyer's position by showing that the positive relation between accessibility level and lexical decision performance exists even when lexical decisions precede estimates of accessibility by a week. Under these conditions, it is impossible that persistent subthreshold activation produced by the accessibility estimates could have mediated lexical decision performance.

Although Yaniv and Meyer's (1987) persistent subthreshold activation account is not adequate to explain the results of Experiment 2, the topic familiarity account provides an acceptable account of the data. According to this latter account, accessibility level and lexical decision performance are both

[7] The following are the mean RTs and percentages of error for the accessibility levels for the Experiment 3 lexical decision data without the estimates included: high/medium, Session 1, 869 (12%); low, Session 1, 867 (18%); high/medium, Session 2, old, 810 (7%); low, Session 2, old, 795 (10%); high/medium, Session 2, new, 851 (15%); and low, Session 2, new, 842 (23%). As seen in Table 7, these means are quite comparable to the means that include the estimated cells.

influenced by the familiarity that a subject has with a particular topic. The accessibility estimate reflects the subject's current assessment of the level of expertise he or she has in a topic, whereas the speeded lexical decision performance simply reflects speeded recognition of an item that falls into a category with which the subject is highly familiar (Balota & Chumbley, 1984). Thus, topic familiarity modulates performance on both accessibility estimates and lexical decision, and the ordering of these two tasks is not critical for the relation to be evident because both effects reflect a third, mediating factor.

The results of Experiment 3 demonstrate that the results of the second experiment were not simply caused by some long-term influence of item-specific attentional fluctuation during the lexical decision task on performance in the subsequent rare word definition task. Specifically, the earlier presentation of the definition referent had no effect on the subsequent probability of correctly recalling that rare word definition referent or on the subsequent probability of that referent being in a high-accessibility state. In addition, the results of Experiment 3 provide a replication of the results of the previous two experiments. Specifically, the size of the influence of accessibility was the same whether the lexical decision task preceded or followed the rare word definition task. These results are again more consistent with the topic familiarity account than Yaniv and Meyer's (1987) subthreshold activation account.

Although Yaniv and Meyer's (1987) results and our findings are all easily accommodated by the topic familiarity account, this does not mean that Yaniv and Meyer's interpretation of their own results is necessarily wrong. That is, it could be that persistent subthreshold activation works in conjunction with topic familiarity to produce better lexical decision performance on words that have been previously judged to be highly accessible. However, we question this possibility for the following reasons: First, if the persistent subthreshold activation accompanying high-accessibility estimates for particular words does indeed prime subsequent lexical decisions to those words, the difference in error rates to high- and low-accessibility items should have been greater in Experiment 1 (in which such priming could have occurred) than in Experiment 2 (in which it could not have occurred). If anything, the opposite was true; that is, the 9% fewer errors for high- than for low-accessibility targets that occurred in Experiment 1 actually increased slightly to an 11% difference in Experiment 2. Second, the largest difference in RTs in Yaniv and Meyer's Experiment 2 was 38 ms (estimated from their Figure 4), which is quite similar to the 35 ms difference found between the medium- and low-accessibility targets in our Experiment 2. Once again, if at least part of Yaniv and Meyer's results were due to persistent activation, then one would predict that their effects would have been substantially larger than the effects found in our Experiment 2. Third, the results of our Experiment 3 indicated that there was little change in the size of the accessibility effect for lexical decision trials given before or after the rare word definition task. Thus, although it is possible that persistent subthreshold activation works in conjunction with topic familiarity to mediate lexical decision performance in this paradigm, we do not believe that there is sufficient evidence available at the present time to support the notion of persistent subthreshold activation as an explanatory mechanism.

A potential criticism of the topic familiarity account is that it is contravened by data from the episodic recognition literature. That is, because high-frequency words are more poorly recognized than low-frequency words (e.g., Balota & Neely, 1980; Duchek & Neely, 1989; Glanzer & Adams, 1985), one might expect that unretrieved words that receive a high-accessibility estimate (which by the topic familiarity account would be high-frequency words for that subject) would actually produce slower episodic recognition RTs. Yet the recognition data from Experiment 1 and Yaniv and Meyer's (1987) second experiment produced the opposite pattern. There are two replies to this criticism. First, normative word frequency, which has been manipulated in the previous recognition memory studies, may not have the same effect on recognition as idiosyncratic familiarity (cf. Gernsbacher, 1984, for clear demonstrations of such dissociations). Second, and more important, Schulman (1976) has shown that recognition memory for low-frequency words is high only if subjects know they are words. Thus, on the basis of Schulman's (1976) data, the topic familiarity account actually predicts better performance on high-accessibility words than on low-accessibility words, as was observed in Yaniv and Meyer and in our first experiment.

It is noteworthy that there is a consistent pattern of data across Yaniv and Meyer's (1987) two experiments and our first two experiments that is consistent with the topic familiarity account but inconsistent with the subthreshold activation account. Specifically, low-accessible targets yield longer RTs and higher error rates than low-accessible controls in all of the nine possible comparisons across the four experiments. The topic familiarity account predicts this pattern in the following manner: Words that are rated as low in accessibility will likely have lower familiarity values than the mean familiarity rating of control words that subjects never received. Specifically, for the targets, there is information indicating a lower level of familiarity (i.e., the subject's rating), whereas for the controls, there is no such information available, and hence the control items should be closer to the overall higher mean familiarity value. This would produce slower RTs and higher error rates for the low-accessibility targets than for the low-accessibility controls. However, the persistent subthreshold activation account does not predict a consistent difference across targets and controls in the low-accessibility condition. Specifically, there is no reason to expect less activation in the memory network for targets that produced an attempted retrieval in comparison with control words that subjects never received.

Finally, we have been arguing that our results are caused by topic familiarity. However, one could just as easily argue that the relation between latent accessibility and lexical decision performance is mediated by item familiarity, that is, familiarity with a given lexical representation. Fortunately, the results of Experiment 3 provide informative data in this regard. The results of Experiment 3 yielded a clear long-term repetition effect on lexical decision performance. Such a long-term repetition effect most likely involves changes in item familiarity because of the earlier presentation of the same word. However, the results of Experiment 3 also yielded no evidence of an influence of the earlier presentation of the item on the rare word definition task. Thus, although item familiarity does appear to modulate lexical decision performance, it does not modulate the rare word definition task. Hence, it is more likely that preexisting familiarity with a given topic is the mediator in both the rare word retrieval task and lexical decision performance. These results are intriguing in light of recent arguments that meaning-level representations (i.e., familiarity with a given topic) can modulate lexical decision performance and possibly early processes in isolated word recognition (see Balota, 1990; Balota, Ferraro, & Connor, 1991).

In summary, although the persistent subthreshold activation account of Yaniv and Meyer's (1987) data is intriguing, we favor a topic familiarity account because it provides an explanation of why the same relation exists between accessibility level and lexical decision performance, regardless of whether the lexical decision task precedes or follows the accessibility estimates. At the very least, our work demonstrates the importance of producing an effect of latent accessibility above and beyond that of familiarity in the Yaniv and Meyer paradigm. In this light, we are forced to await future demonstrations of such an effect and the development of a more sensitive paradigm to document the long-lasting impact of attempted retrievals on activation levels of nonretrieved words.

References

Balota, D. A. (1990). The role of meaning in word recognition. In D. A. Balota, G. B. Flores D'Arcais, & K. Rayner (Eds.), *Comprehension processes in reading* (pp. 9–32). Hillsdale, NJ: Erlbaum.

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.

Balota, D. A., Ferraro, R. F., & Connor, L. T. (1991). On the early influence of meaning in word recognition: A review of the literature. In P. Schwanenflugel (Ed.), *The psychology of word meanings* (pp. 187–222). Hillsdale, NJ: Erlbaum.

Balota, D. A., & Lorch, R. F. (1986). Depth of automatic spreading activation: Mediated priming effects in pronunciation but not in lexical decision. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12,* 336–345.

Balota, D. A., & Neely, J. H. (1980). Test-expectancy and word-frequency effects in recall and recognition. *Journal of Experimental Psychology: Human Learning and Memory, 6,* 576–587.

Collins, A. M., & Loftus, E. (1975). A spreading-activation theory of semantic processing. *Psychological Review, 82,* 407–428.

Dominowski, R. L., & Jenrick, R. (1972). Effects of hints and interpolated activity on solution of an insight problem. *Psychonomic Science, 26,* 335–337.

Dreistadt, R. (1969). The use of analogies and incubation in obtaining insights in creative problem solving. *Journal of Psychology, 71,* 159–175.

Duchek, J. M., & Neely, J. H. (1989). A dissociative word-frequency × levels-of-processing interaction in episodic recognition and lexical decision tasks. *Memory & Cognition, 17,* 148–162.

Foss, D. J. (1982). A discourse on semantic priming. *Cognitive Psychology, 14,* 590–607.

Gernsbacher, M. A. (1984). Resolving 20 years of inconsistent interactions between lexical familiarity and orthography, concreteness, and polysemy. *Journal of Experimental Psychology: General, 113,* 256–281.

Glanzer, M., & Adams, J. K. (1985). The mirror effect in recognition memory. *Memory & Cognition, 13,* 8–20.

Joordens, S., & Besner, D. (1992). Priming effects that span an intervening unrelated word: Implications for models of memory representation and retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 483–491.

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

Masson, M. E. J. (1991). A distributed memory model of context effects in word identification. In D. Besner & G. W. Humphreys (Eds.), *Basic processes in reading: Visual word recognition* (pp. 233–263). Hillsdale, NJ: Erlbaum.

McNamara, T. P. (1991). *Theories of priming: I. Associative distance and list context.* Manuscript submitted for publication.

Murray, H. G., & Denny, J. P. (1969). Interaction of ability level and interpolated activity (opportunity for incubation) in human problem solving. *Psychological Reports, 24,* 271–276.

Neely, J. H., & Keefe, D. E. (1989). Semantic context effects on visual word processing: A hybrid prospective–retrospective processing theory. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 24, pp. 207–248). San Diego, CA: Academic Press.

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.

Olton, R. M. (1979). Experimental studies of incubation: Searching for the elusive. *Journal of Creative Behavior, 13,* 9–22.

Olton, R. M., & Johnson, D. M. (1976). Mechanisms of incubation in creative problem solving. *American Journal of Psychology, 89,* 617–630.

Reder, L. M., & Ritter, F. E. (1992). What determines initial feeling of knowing? Familiarity with question terms, not with the answer. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 435–451.

Schulman, A. I. (1976). Memory for rare words previously rated for familiarity. *Journal of Experimental Psychology: Human Learning and Memory, 2,* 301–307.

Silveira, J. M. (1971). Incubation: The effect of interruption timing and length on problem solution and quality of problem processing (Doctoral dissertation, University of Oregon, 1971). *Dissertation Abstracts International, 32,* 5500B. (University Microfilms No. 72-9560)

Smith, S. M., & Blankenship, S. E. (1989). Incubation effects. *Bulletin of the Psychonomic Society, 27,* 311–314.

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

Received October 9, 1990
Revision received July 19, 1991
Accepted September 27, 1991 ■