Memory & Cognition
2001, 29 (8), 1176-1184

Interrupting recognition memory: Tests of a criterion-change account of the revelation effect

WILLIAM E. HOCKLEY
Wilfrid Laurier University, Waterloo, Ontario, Canada

and

MARTY W. NIEWIADOMSKI
University of Toronto, Toronto, Ontario, Canada

The revelation effect is evidenced by an increase in positive recognition responses when the test probe is immediately preceded by an unrelated problem-solving task. As an alternative to familiarity-based explanations of this effect (Hicks & Marsh, 1998; Westerman & Greene, 1998), Niewiadomski and Hockley (2001) proposed a decision-based account in which it is assumed that the problem-solving task displaces the study list context in working memory, leading subjects to adopt a more liberal recognition criterion. In the present study, we show that the revelation effect is seen when the stimulus materials are pure lists of very rare words or nonwords. In contrast, for mixed lists of common words and very rare words or nonwords, the revelation effect is found for common words but disappears for very rare words and nonwords. We argue that, in mixed lists, the liberal decision bias following the revelation task and the criterion changes between common words and very rare words and nonwords serve to offset each other.

The revelation effect is a curious phenomenon of recognition memory in which performing a problem-oriented task immediately prior to a recognition judgment leads to an increased probability of classifying the test probe as "old." Most accounts of this effect have involved the assumption that the revelation task influences the familiarity of the following test probe (Hicks & Marsh, 1998; LeCompte, 1995; Luo, 1993; Peynircioğlu & Tekcan, 1993; Westerman & Greene, 1998). For example, Westerman and Greene (1998) proposed that the revelation task briefly activates additional information in memory that is not activated by the probe itself. This activation is summed with the activation produced by the probe, thereby increasing the overall activation levels of test probes in the revelation condition and thus increasing the hit and false alarm rates. Hicks and Marsh (1998) offered an alternative familiarity-based account. They suggested that the revelation task temporarily activates competing information in memory, which leads to a reduction in the signal-to-noise ratio for the test item. As a consequence, subjects adopt a more liberal decision criterion in the revelation condition, leading to an increase in the hit and false alarm rates.

This research was supported by an operating grant from the Natural Sciences and Engineering Research Council of Canada to the first author. The authors thank Melissa Wells for her assistance in data collection and Robert Greene for his suggestions concerning an earlier version of this paper. Correspondence should be addressed to W. E. Hockley, Department of Psychology, Wilfrid Laurier University, Waterloo, ON, N2L 3C5 Canada (e-mail: whockley@wlu.ca).

The fact that a wide variety of tasks can produce the revelation effect poses a strong challenge for familiarity-based accounts. Watkins and Peynircioğlu (1990) demonstrated that a number of tasks involving the probe item produce the revelation effect. The magnitude of the revelation effect is also very similar whether the revelation task involves the probe item or an unrelated item (Cameron & Hockley, 2000; Westerman & Greene, 1998). Westerman and Greene (1998) have also shown that several "revelation tasks" unrelated to the probe (e.g., memory-span, synonym-generation, letter-counting, anagrams of nonwords) can produce the revelation effect. Niewiadomski and Hockley (2001) have shown that a numerical addition task also produces the revelation effect and that the effect is approximately the same magnitude as a revelation effect involving verbal information (word anagrams). This finding would seem to be at odds with familiarity-based accounts of the revelation effect because such accounts should predict that the activation of additional information in memory that is related to a verbal study list should be greater for a verbal revelation task than for an arithmetic task.

Prull, Light, Collett, and Kennison's (1998) finding that the elderly do not show a revelation effect provides another difficulty for familiarity-based accounts of the revelation effect. Since older adults are usually more prone to manipulations that influence familiarity than are younger adults (e.g., Dywan & Jacoby, 1990), a reasonable prediction of a familiarity-based view of the revelation effect would be that the elderly should be at least as susceptible, if not more so, than younger adults to the revelation effect.

Niewiadomski and Hockley (2001) proposed that the revelation task might displace list-relevant information in working memory, and the temporary loss of contextual information leads subjects to adopt a more liberal criterion for the subsequent test probe. It is not fully understood how subjects set and adjust their decision criterion, but it is generally assumed that one basis involves an assessment of the overall memorability of the study list. For example, Stretch and Wixted (1998) have argued that subjects adopt a more liberal criterion for "easy" study lists (e.g., relatively short lists, or lists presented at a relatively slow rate) and a more conservative criterion for "difficult" lists (longer study lists, or lists presented at a faster rate). This provides a straightforward explanation of why both the hit rate is lower and the false alarm rate is greater for difficult lists than for easy lists (i.e., strength-based mirror effects). It is also generally assumed that subjects normally maintain a relatively consistent criterion throughout the test list.

Performing an unrelated task prior to some of the test probes might cause subjects to forget their criterion setting, or the information on which the criterion setting is based. When the probe is then presented, subjects are faced with a recognition decision of uncertain difficulty, and they may adopt a more liberal criterion as a consequence. After responding to the probe, the context of the study list is regained, and subjects can establish a more appropriate criterion. In this view, any task that precedes the probe would produce the revelation effect as long as the task is of sufficient difficulty to displace the list-relevant information in working memory. Older subjects may not be as affected by such interruptions (Prull et al., 1998), because the elderly are less efficient in monitoring memory (e.g., McIntyre & Craik, 1987) and adjusting their decision criterion to meet task demands (Multhaup, 1995).

The following experiments were designed to provide a test of such a criterion-flux account of the revelation effect. The logic of our approach was based on Stretch and Wixted's (1998, p. 1394) discussion of Wixted's (1992) findings. Wixted compared recognition performance for very rare words (functionally nonwords) and high-frequency words. In contrast to the pattern of hits and false alarms (known as the *mirror effect*) that is typically observed between high- and low-frequency words (cf. Glanzer & Adams, 1985), Wixted found that both the hit rate and the false alarm rate were higher for very rare words than for high-frequency words. The overall level of discrimination for the two types of stimuli was, however, comparable. Stretch and Wixted suggested that the subjects in Wixted's study adopted a more liberal decision criterion for very rare words than for high-frequency words on the (erroneous) assumption that very rare words are less memorable.

If subjects adopt a more liberal recognition decision criterion for very rare words than they do for common words, and if the revelation effect also involves a liberal criterion change, then combining these two manipulations might lead to one of these effects canceling out the other. That is, if subjects adopt a more liberal criterion for very rare words than for common words, this might overshadow the liberal change in criterion that may occur following a revelation task. More generally, if two manipulations that are due to criterion changes are pitted against each other, one manipulation may outshine or dominate the other, and, thus, one manipulation may not be effective in the presence of the other.

In Experiment 1, we first replicated Wixted's (1992) finding that the hit and false alarm rates are greater for very rare words than for common words. Together, these results indicate that subjects adopt a more liberal decision criterion for very rare words than for common words. In Experiment 2A, we then found that very rare words are subject to the revelation effect when tested in pure lists. Since very rare words are functionally nonwords to subjects but may have unusual orthographic characteristics, the revelation effect for pure lists of pronounceable nonwords was examined in Experiment 2B. Finally, in Experiments 3, 4, and 5, we contrasted the joint effects of stimulus type (common words vs. very rare words or nonwords) and the revelation effect in a mixed-list design. If both the manipulations of stimulus type and the revelation effect involve changes in the recognition decision criterion, then, when these manipulations are combined, we predicted that one effect would attenuate or eliminate the other.

EXPERIMENT 1

In Experiment 1, recognition performance for mixed lists of common words and very rare words was examined. This experiment constituted a replication of Wixted's (1992) study with a different set of stimulus materials.

Method

Subjects. The subjects in all of the experiments participated for credit toward their introductory psychology course. There were 48 students in Experiment 1.

Apparatus and Stimuli. A pool of 120 very rare words (e.g., *aphonic*, *dubitate*, *nutate*, and *tunicle*) were chosen from the Web-based MRC Psycholinguistic Database (1997). The common words were randomly selected from a pool of 688 nouns derived from Paivio, Yuille, and Madigan (1968). The imageability rating for these words was 5.00 or greater, based on the 1–7 word-rating norms of Paivio et al..

Stimulus presentation and response recording in all experiments were controlled by IBM-compatible laboratory computers equipped with 17-in. monitors. The subjects responded by pressing the "/" key for an "old" recognition decision and the "z" key for a "new" decision. The keyboards were covered so that only the labeled response keys were exposed.

Procedure. The subjects completed one study–test trial. The study lists consisted of 30 common words and 30 very rare words. The first and last six presentations consisted of 3 common words and 3 very rare words and were treated as buffers for primacy and recency effects and were not tested. The study items were presented for 1.5 sec, with a 0.5-sec blank interval between presentations. The test lists consisted of 48 study items and 48 new items (lures), with half of each being common words and half very rare words. The order of both study and test presentations was randomly determined for each subject. The test list was subject-paced, with a 1-sec blank interval between each recognition response and the presentation of the next probe.

Table 1
Mean Proportions of "Old" Responses, Standard Deviations of the Means, and Signal Detection Estimates for Each Stimulus Condition of Experiment 1

	Test Probe							
	Old		New		A'		β''_D	
Condition	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Common words	.66	.17	.14	.11	.85	.08	.44	.46
Very rare words	.74	.14	.22	.13	.84	.07	.09	.52

Results and Discussion

The mean proportion of positive responses for old (hits) and new (false alarms) test probes for the two types of stimuli are presented in Table 1. A 2 (old vs. new test probes) × 2 (common vs. very rare words) analysis of variance (ANOVA) revealed a significant main effect of probe [$F(1,47) = 765.5, MS_e = 0.017, p < .001$]. Not surprisingly, the hit rate was greater than the false alarm rate. More importantly, there was also a significant main effect for word type [$F(1,47) = 14.62, MS_e = 0.021, p < .001$]. Both the hit rate and the false alarm rate were greater for very rare words than for common words. The probe × word type interaction did not approach significance [$F(1,47) < 1$].

We examined potential differences in discriminability and criterion placement by estimating A' and β''_D.[1] The means and standard deviations of these estimates are also presented in Table 1. Paired-samples *t* tests based on A' revealed there was no difference in discriminability between the common words and the very rare words ($t < 1$). The analysis of the β''_D estimates indicated that the subjects assumed a more liberal criterion for very rare words than for to common words [$t(47) = 3.77, p < .001$].

The results of Experiment 1 provide a replication of Wixted's (1992) findings. As Stretch and Wixted (1998) have argued, the most straightforward explanation for the higher hit and false alarm rates for very rare words than for common words is that subjects adopt a more liberal decision criterion for the very rare words. Such a response strategy is reasonable if it is assumed (albeit incorrectly) that very rare words are less memorable than common words.

EXPERIMENTS 2A AND 2B

The revelation effect for pure lists of very rare words was examined in Experiment 2A. Since very rare words are functionally nonwords to subjects but may have unusual orthographic characteristics, we also examined the revelation effect for pronounceable nonwords in Experiment 2B.

Method

Subjects. There were 51 subjects in Experiment 2A and 35 subjects in Experiment 2B.

Apparatus and Stimuli. The 140 eight-letter anagrams used in the following experiments were adapted from Gibson and Watkins (1988). All the anagrams were scrambled in the same order, with the solution sequence being 54687321 (i.e., *navigate* would appear as *giaetvan*). The very rare words were the same as those used in Experiment 1. The nonwords were selected from a pool of 500 pronounceable nonwords (e.g., *jerkin*, *anock*, *lomet*, and *truffin*) used by Ratcliff, Hockley, and McKoon (1985).

The "z" key was used for "new" responses, the "/" key was used for "old" responses, and the "y" key indicated problem solution. The keyboards were covered with opaque covers with only the response keys exposed. The words *old*, *new*, and *problem solved* appeared on the keyboard covers, located appropriately to indicate the purpose of each response key.

Procedure. Study lists consisted of 60 very rare words (Experiment 2A) or nonwords (Experiment 2B) selected randomly without replacement. The items were presented one at a time, in the center of the screen for 1.5 sec, with a 0.5-sec blank interval between them. The first and last six items were considered primacy and recency buffers and were not tested.

The recognition test comprised 96 probes: 48 were targets from the study list, and 48 were new items. Half of the old and new test probes were preceded by the anagram task. The order of the test presentations was random, with a different order for each subject.

The anagrams were presented one at a time in the center of the screen. The statement "Press top key when solved" appeared at the bottom of the screen. The answer key for the anagrams was always presented above that statement. The subjects voiced their anagram solutions aloud into a tape recorder and pressed the "problem solved" key to continue. This task was self-paced. There was a 1-sec blank interval before the next display.

The recognition probes were presented one at a time, in the center of the screen flanked by question marks. The words *old* and *new* appeared in the bottom right and left of the screen, respectively, to serve as a reminder of the recognition task and the response keys. This task was also self-paced. There was a 1-sec blank interval after a response before the next presentation.

The subjects were given two sets of instructions, one prior to the study list and the other prior to the test. At study, the subjects were asked to remember the items, and they were told that their memory for the items would be tested later. Immediately prior to the test list, the subjects were asked to respond "old" to the items presented in the study list and "new" to the items that had not been shown earlier. The instructions concerning the anagram task were then given.

Results and Discussion

The mean proportion of positive responses to old and new test probes in each test condition for each experiment are given in Table 2. The hit and false alarm rates were analyzed in 2 (old vs. new probes) × 2 (intact vs. anagram test conditions) ANOVAs. The analysis for very rare words showed that both the main effects of test probe [$F(1,50) = 257.8, MS_e = 0.024$] and test condition [$F(1,50) = 4.235, MS_e = 0.064, p = .045$] were reliable. Thus, the hit rate was greater than the false alarm rate, and these rates were higher in the anagram condition than in the intact condition. The interaction between these effects was not significant [$F(1,50) = 1.784, MS_e = 0.008$].

The same pattern of results was found for nonwords. The hit rate was greater than the false alarm rate [$F(1,34) = 133.24, MS_e = 0.032$], and the proportion of "old" responses was significantly greater in the anagram test condition

Table 2
Mean Proportions of "Old" Responses, Standard Deviations of the Means, and Signal Detection Estimates for Each Test Condition of Experiments 2A and 2B

	Test Probe							
	Old		New		A'		β''_D	
Condition	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
				Experiment 2A				
Intact	.66	.14	.29	.15	.77	.08	.11	.48
Revelation	.68	.15	.35	.19	.76	.10	−.01	.53
				Experiment 2B				
Intact	.66	.12	.30	.15	.77	.10	.13	.40
Revelation	.73	.15	.36	.18	.77	.10	−.19	.52

than in the intact condition [$F(1,34) = 11.88$, $MS_e = 0.015$]. The interaction between these variables was again not reliable ($F < 1$).

These results were corroborated by analyses of the signal detection estimates. The means and standard deviations of the estimates of A' and β''_D are given in Table 2. In both Experiment 2A and Experiment 2B, no significant differences in discriminability (A') was observed (paired-samples *t*s < 1). Differences in criterion placement were indicated, however, in the β''_D analyses. In both experiments, the subjects assumed a more liberal criterion for test items that followed the revelation task. This difference approached significance in Experiment 2A [$t(50) = 1.87$, $p = .068$] and was statistically reliable in Experiment 2B [$t(34) = 3.81$, $p = .001$].

The results of Experiments 2A and 2B demonstrate that the revelation effect is observed for both very rare words and nonwords when these two types of stimuli are studied and tested in pure lists. In the following experiments, the revelation effect was examined for mixed lists of common words and very rare words (Experiment 3) and nonwords (Experiments 4 and 5).

EXPERIMENT 3

The purpose of Experiment 3 was to compare the revelation effect for both common words and very rare words when these two types of stimuli are mixed within the same study and test lists. If subjects use a different decision criterion for very rare words than they do for common words, as the results of Wixted's (1992) study and Experiment 1 suggest, and if the revelation effect also involves a criterion change, then one effect may cancel the other. If, on the other hand, the revelation effect does not involve a change in criterion, then the changes in hit and false alarm rates associated with the stimulus manipulation and the revelation task should both be observed when they are combined.

In Experiment 3, the revelation and intact test conditions were blocked. The order of common words and very rare words was random within each of the test blocks. This was done in an attempt to make the stimulus manipulation more salient than the revelation manipulation. That is, because the words and very rare words occurred in each list block of the test list, and because very rare words are distinctly different from the common words, we thought that the subjects might be more likely to change their decision criterion on the basis of the very rare words than on the basis of the preceding unrelated task.[2]

Method

Subjects. Seventy-three subjects participated.

Apparatus and Stimuli. The apparatus and stimuli were the same as those used in the previous experiments.

Procedure. The subjects were presented with a study list comprising 30 words and 30 very rare words in a random order. The items were shown 1.5 sec, with a blank interval of 1 sec between presentations. An equal number of words and rare words was shown in

Table 3
Mean Proportions of "Old" Responses, Standard Deviations of the Means, and Signal Detection Estimates for Each Test and Stimulus Condition of Experiment 3

	Test Probe							
	Old		New		A'		β''_D	
Condition	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Common words								
Intact	.70	.17	.20	.16	.82	.11	.27	.57
Revelation	.76	.13	.36	.20	.79	.10	−.23	.57
Very rare words								
Intact	.70	.18	.24	.15	.81	.11	.10	.57
Revelation	.66	.21	.24	.19	.80	.12	.18	.65

the first and last six positions of the study list and were not tested. The test list consisted of a total of 96 tests that were divided into four consecutive blocks. Within each block of 24 tests, there were 6 old and 6 common word tests and 6 old and 6 new very rare word tests. The order of tests within a block was random. The intact and revelation test conditions alternated between blocks, and the order of the blocks was alternated over subjects. All phases of the test list were subject-paced, with a 1-sec blank interval between each presentation.

Results and Discussion

The mean proportions of positive responses to old and new test probes for each stimulus and test condition are shown in Table 3. A 2 (test probe) × 2 (word type) × 2 (test condition) within-subjects ANOVA revealed significant main effects of probe type [$F(1,72) = 598.8, MS_e = 0.048$], word type [$F(1,72) = 6.05, MS_e = 0.053$], and test condition [$F(1,72) = 16.72, MS_e = 0.021$]. The main effects were qualified by significant interactions between probe and test condition [$F(1,72) = 11.421, MS_e = 0.017$] and test condition and word type [$F(1,72) = 53.689, MS_e = 0.013$]. The interactions involving probe × word type [$F(1,72) = 0.151, MS_e = 0.020$] and probe × test condition × word type [$F(1,72) = 1.480, MS_e = 0.016$] were not reliable.

To explore the significant interactions, separate analyses were performed for each stimulus type. For common words, the main of effects of probe [$F(1,72) = 484.6, MS_e = 0.030$] and test condition [$F(1,72) = 65.72, MS_e = 0.015$] and their interaction [$F(1,72) = 9.84, MS_e = 0.018$] were reliable. In the analysis for rare words, the main effect of probe [$F(1,72) = 371.1, MS_e = 0.038$] was significant, but the main effect of test condition [$F(1,72) = 1.49, MS_e = 0.018$] and the probe × test condition interaction [$F(1,72) = 2.84, MS_e = 0.015$] were not.

The pattern of results in Table 3 suggests that a revelation effect was found for common words, but not for very rare words. Paired-samples t tests confirmed this observation. For common words, the hit rate [$t(72) = 3.209, p = .002$] and the false alarm rate [$t(72) = 7.815, p < .001$] were significantly greater in the anagram test condition than in the intact condition. In contrast, for very rare words, the hit rate [$t(72) = 1.894, p = .062$] and the false alarm rate [$t(72) = 0.232, p = .817$] did not reliably differ between the anagram and intact test conditions. (Although the difference in hit rates approached significance, this trend was in the opposite direction to the revelation effect.) Thus, a revelation effect was found for common words, but not for very rare words.

To examine differences in discriminability and criterion placement, A' and β''_D estimates were calculated. Means and standard deviations are presented in Table 3. A 2 (intact vs. revelation condition) × 2 (common vs. rare word type) within-subjects ANOVA on A' revealed that the subjects were better able to discriminate old items from new items in the intact condition than in the revelation condition [$F(1,72) = 6.67, MS_e = 0.008$]. No differences in discriminability between the two word types were found, and the two variables did not interact (Fs < 1).

The same analysis for β''_D estimates showed a main effect of test condition [$F(1,72) = 16.78, MS_e = 0.19$]. The main effect of word type did not approach significance, but the interaction between word type and test condition was reliable [$F(1,72) = 43.35, MS_e = 0.14$]. For common words, the criterion was significantly more liberal in the revelation condition than in the intact condition [$t(72) = 7.365, p < .001$]. For very rare words, the estimates of criterion placement did not differ reliably between the intact and revelation conditions [$t(72) = 1.257, p = .213$].

It was predicted that if two effects that are the result of criterion changes were pitted against each other, then one effect might overshadow the other. Experiment 3 showed just this result. The revelation effect for very rare words that was observed in the pure-list design of Experiment 2A disappeared when very rare words were presented in mixed lists with common words.

EXPERIMENT 4

Experiment 4 was designed to determine whether the pattern of results found in Experiment 3 could be replicated when the stimulus manipulation involved common words versus pronounceable nonwords.

Method

Subjects. Thirty-eight subjects participated.

Apparatus and Stimuli. The apparatus and stimuli were the same as those used in the previous experiments.

Table 4
Mean Proportions of "Old" Responses, Standard Deviations of the Means, and Signal Detection Estimates for Each Test and Stimulus Condition of Experiment 4

	Test Probe							
	Old		New		A'		β''_D	
Condition	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Common words								
Intact	.77	.15	.33	.22	.80	.12	−.19	.60
Revelation	.82	.14	.46	.21	.78	.11	−.49	.20
Nonwords								
Intact	.71	.19	.26	.15	.81	.10	.04	.42
Revelation	.63	.20	.23	.15	.79	.10	.26	.62

Procedure. The procedure was identical to that of Experiment 3, except that the nonwords from Experiment 2B were used instead of very rare words, and the presentation rate at study was 2.5 sec per item.

Results and Discussion

The mean proportions of positive responses for each stimulus and test condition are summarized in Table 4. These results were analyzed in a 2 (old vs. new probes) × 2 (intact vs. anagram test condition) × 2 (word vs. nonwords) within-subjects ANOVA. As expected, the hit rate was greater than the false alarm rate [$F(1,37) = 297.47$, $MS_e = 0.044$]. There was also a main effect of stimulus type [$F(1,37) = 25.24$, $MS_e = 0.056$], indicating that the hit and false alarm rates were higher for words than for nonwords. Stimulus type also interacted significantly with test condition [$F(1,37) = 15.35$, $MS_e = 0.024$]. Finally, the interaction between test probe and test condition approached significance [$F(1,37) = 3.93$, $MS_e = 0.020$, $p = .055$].

To explore the interactions, separate analyses were performed for each stimulus type. The analysis for words showed main effects of test probe [$F(1,37) = 165.85$, $MS_e = 0.037$] and test condition [$F(1,37) = 10.60$, $MS_e = 0.028$]. The interaction between these variables was not reliable. The same analysis for nonwords also showed a difference between hit and false alarm rates [$F(1,37) = 243.75$, $MS_e = 0.029$]. In contrast to words, however, the greater proportion of "old" responses was made in the intact condition rather than in the revelation condition [$F(1,37) = 5.47$, $MS_e = 0.019$]. The interaction between test probe and test condition did not approach significance [$F(1,37) = 1.18$, $MS_e = 0.016$].

Two separate 2 (intact vs. revelation task) × 2 (word vs. nonword) ANOVAs were performed on the A' and β''_D estimates (see Table 4 for means and standard deviations). The A' analysis revealed no significant main effects, and the variables did not interact. The β''_D analysis, however, showed a significant main effect of stimulus type [$F(1,37) = 18.69$, $MS_e = 0.49$] and a significant interaction between stimulus type and test condition [$F(1,37) = 12.07$, $MS_e = 0.21$]. For words, the criterion estimate was reliably lower in the revelation condition than in the intact condition [$t(37) = 2.859$, $p = .007$]. In contrast, for nonwords, the criterion estimate was significantly higher in the revelation condition than in the intact condition [$t(37) = 2.161$, $p = .037$].

In both of the mixed stimulus lists of Experiments 3 and 4, a revelation effect was found for common words. In contrast, no revelation effect was observed for very rare words in Experiment 3, and a "negative revelation effect" was found for nonwords in Experiment 4. The latter result was quite unexpected; therefore, Experiment 5 was designed to see whether this pattern of results could be replicated.

EXPERIMENT 5

Method

Subjects. Sixty-four subjects participated.

Apparatus and Stimuli. The apparatus and stimuli were the same as those used in Experiment 4.

Procedure. The procedure was identical to that of Experiment 4, except that the study presentation rate was 1.5 sec per item.

Results and Discussion

The mean proportions of positive responses for each stimulus and test condition of Experiment 5 are summarized in Table 5. These results were analyzed in a 2 (old vs. new probes) × 2 (intact vs. anagram test condition) × 2 (word vs. nonword) within-subjects ANOVA. As usual, the hit rate was greater than the false alarm rate [$F(1,63) = 492.40$, $MS_e = 0.046$]. There were also significant main effects of test condition [$F(1,63) = 12.10$, $MS_e = 0.025$] and type of stimulus [$F(1,63) = 11.78$, $MS_e = 0.051$]. The main effects were qualified by significant interactions between probe and test condition [$F(1,63) = 8.71$, $MS_e = 0.016$] and between test condition and stimulus type [$F(1,63) = 19.30$, $MS_e = 0.018$]. The interaction between probe and type of stimulus approached significance [$F(1,63) = 3.92$, $MS_e = 0.018$]. The three-way interaction between these variables was not reliable ($F < 1$).

Separate analyses were performed for each stimulus type. The analysis for words showed main effects of test probe [$F(1,63) = 306.2$, $MS_e = 0.033$], test condition [$F(1,63) = 31.35$, $MS_e = 0.021$], and their interaction [$F(1,63) = 7.24$, $MS_e = 0.015$]. There was a reliable revelation effect

Table 5
Mean Proportions of "Old" Responses, Standard Deviations of the Means, and Signal Detection Estimates for Each Test and Stimulus Condition of Experiment 5

	Test Probe							
	Old		New		A'		β''_D	
Condition	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Common words								
Intact	.70	.19	.26	.19	.80	.11	.13	.61
Revelation	.76	.18	.40	.20	.78	.09	−.28	.60
Nonwords								
Intact	.70	.20	.23	.16	.81	.12	.17	.57
Revelation	.67	.17	.25	.16	.80	.11	.18	.54

for words, and this effect was greater for false alarms than for hits.

For nonwords, only the difference between hit and false alarm rates [$F(1,63) = 401.9$, $MS_e = 0.032$] was reliable; the main effect of test condition [$F(1,37) < 1$, $MS_e = 0.022$] and the interaction between test probe and test condition [$F(1,63) = 1.79$, $MS_e = 0.022$] did not approach significance. Thus, no revelation effect was observed for nonwords.

The results of the signal detection analyses corroborated the above findings. Descriptive statistics can be found in Table 5. A 2 (intact vs. revelation condition) × 2 (word vs. nonword) ANOVA on A' showed no reliable differences in discriminability, although the main effect of test condition approached significance [$F(1,63) = 3.70$, $MS_e = 0.008$, $p = .59$]. Performance tended to be more accurate for the intact condition than for the revelation condition. The ANOVA performed on the β''_D estimates showed significant main effects of test condition [$F(1,63) = 11.63$, $MS_e = 0.19$] and word type [$F(1,63) = 9.01$, $MS_e = 0.40$]. The two variables interacted significantly [$F(1,63) = 19.09$, $MS_e = 0.18$]. The criterion estimate for words was significantly more liberal in the revelation condition than in the intact condition [$t(63) = 5.675$, $p < .001$]. For nonwords, the estimates of criterion placement did not differ reliably between the intact and revelation conditions [$t(63) = 0.580$, $p = .564$].

The results of Experiment 5 replicated the principal findings of Experiments 3 and 4. A revelation effect was observed when the test probes were common words, but no revelation effect was found for nonwords. The significant "reverse revelation effect" that was found for nonwords in Experiment 4 was not replicated, and this aspect of the results of Experiment 4 quite likely represents chance variation.

GENERAL DISCUSSION

We examined and contrasted two different manipulations that influence the probability of a positive response on a recognition test: the revelation effect, and stimulus comparisons between common words and very rare words and nonwords. Experiment 1 showed that, in mixed lists, the hit and false alarm rates were greater for very rare words than for common words, thus replicating the results of Wixted's (1992) study. Experiments 2A and 2B demonstrated that, when studied and tested in pure lists, very rare words and nonwords were both subject to the revelation effect. The finding of a revelation effect for very rare words and nonwords was expected because, although this effect has not previously been tested for such stimuli, there was no reason to believe that such stimuli would be exempt from this effect.

According to signal detection theory, there are two ways in which a manipulation can lead to an increase in both the hit rate and the false alarm rate in one condition relative to another. For one, a manipulation can cause an increase in the familiarity of items in Condition A relative to Condition B. In such a case, the underlying distributions for new and old items in Condition A would be shifted to the right on the familiarity dimension relative to the distributions for items in Condition B. If subjects adopt a common decision criterion, then the hit and false alarm rates would be higher for items in Condition A than for items in Condition B. A prime example of such a manipulation is the familiarity effect (Greene, 1999) in which preexposing some items prior to study and test leads to an increase in the hit and false alarm rates of these items relative to items that were not preexposed.

A second reason for a difference in hit and false alarm rates between two conditions would be if subjects use different decision criteria for the items in the two different conditions. That is, if subjects adopt a consistently more liberal decision criterion for items in Condition A and a more conservative criterion for items in Condition B, then the hit and false alarm rates would be greater in Condition A than in Condition B. This logically requires that subjects can distinguish between the items in each condition and that they have some reason for adopting a different decision criterion for each condition. Stretch and Wixted (1998) have offered such an explanation for the differences in hit and false alarm rates between very rare words and common words. A criterion-change explanation of this difference makes good sense for two reasons. First, the difference between very rare words and common words is a salient one that subjects cannot miss. It is also quite understandable that subjects may believe that very rare words (or nonwords) are more difficult to remember than common words because of their lack of experience with such stimuli, and, as a consequence, they adopt a more liberal criterion for these items. The second reason favoring such a criterion-change explanation is that there are no good grounds for assuming that very rare words or nonwords would have a higher familiarity value than would common words if both classes of stimuli are studied under similar circumstances. Indeed, it would be more reasonable to assume that any difference in familiarity would favor common words.

Distinguishing between familiarity-based (Hicks & Marsh, 1998; Westerman & Greene, 1998) and decision-based (Niewiadomski & Hockley, 2001) explanations of the increase in hit and false alarm rates produced by the revelation effect is more problematic, because these views make very similar predictions. We reasoned that if the congruent changes in hit and false alarm rates associated with very rare words and nonwords versus common words and the revelation effect are both due to criterion changes, then pitting these two manipulations against each other might lead to one manipulation overpowering the other. This is what we observed in Experiments 3, 4, and 5. When the revelation effects for both common words and very rare words or nonwords were compared together, a revelation effect for common words, but not for very rare words or nonwords, was found. This stands in contrast to the results of Experiments 2A and 2B, which showed that these stimuli are subject to the revelation effect when they are studied and tested by themselves.

The finding that the revelation effect for very rare words and nonwords is eliminated when they are tested in mixed lists with common words poses a challenge for familiarity-based explanations of the revelation effect. We see no obvious explanation for why the presence of common words would eliminate any effects of familiarity caused by the revelation task on very rare words or nonwords but still leave intact such effects on common words. If the unrelated revelation task serves to temporarily increase (Westerman & Greene, 1998) or decrease (Hicks & Marsh, 1998) the familiarity of the subsequent test probe, why should it matter for nonwords but not words what other types of items are in the list?

The results of Experiments 3, 4, and 5 are, however, consistent with the view that the changes in hit and false alarm rates between common words and very rare words or nonwords and between interrupted versus noninterrupted recognition decisions are mediated by criterion changes. When these two manipulations were combined, the criterion changes associated with very rare words and nonwords and the criterion changes that normally give rise to the revelation effect offset each other.

The revelation effect is a particularly robust phenomenon that can be produced by a wide variety of unrelated tasks from letter counting (Westerman & Greene, 1998) to numeric addition (Niewiadomski & Hockley, 2001). Three limitations to the generality of this effect have been shown. First, the revelation effect is restricted to decisions based on episodic memory (Frigo, Reas, & LeCompte, 1999; Watkins & Peynircioğlu, 1990). Second, it is limited to episodic memory decisions based largely on familiarity, but not decisions that involve recall or recollection (Cameron & Hockley, 2000; Westerman, 2000). Finally, in the present study, we have demonstrated an unusual, but illuminating, exception to the revelation effect: Very rare words and nonwords that are susceptible to the revelation effect when tested alone do not show this effect when they are mixed with common words, although the revelation effect still occurs for the common words. This pattern of results was derived from predictions based on Niewiadomski and Hockley's (2001) criterion-flux account of the revelation effect. In this view, it is assumed that during the course of the recognition test, an interrupting task causes a temporary loss of the study-list context that makes the subsequent recognition decision of uncertain difficulty. Subjects adopt a more liberal decision criterion as a consequence. This effect does not occur for very rare words and nonwords when they are tested together with common words, because the criterion change associated with these types of stimuli eclipses the criterion change that typically gives rise to the revelation effect. The present results thus provide support for the criterion-flux account of the revelation effect. The strategy of contrasting or pitting together two different manipulations that may be due to criterion changes may also be an effective technique for testing criterion-based accounts of other phenomenon of recognition memory.

REFERENCES

Cameron, T. E., & Hockley, W. E. (2000). The revelation effect for item and associative recognition: Familiarity versus recollection. *Memory & Cognition*, **28**, 176-183.

Donaldson, W. (1993). Accuracy of d' and A' as estimates of sensitivity. *Bulletin of the Psychonomic Society*, **31**, 271-274.

Dywan, J., & Jacoby, L. (1990). Effects of aging on source monitoring: Differences in susceptibility to false fame. *Psychology & Aging*, **5**, 379-387.

Frigo, L. C., Reas, D. L., & LeCompte, D. C. (1999). Revelation with presentation: Counterfeit study list yields robust revelation effect. *Memory & Cognition*, **27**, 339-343.

Gibson, J. M., & Watkins, M. J. (1988). A pool of 1,086 words with unique two-letter fragments. *Behavior Research Methods, Instruments, & Computers*, **20**, 390-397.

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

Greene, R. L. (1999). The role of familiarity in recognition. *Psychonomic Bulletin & Review*, **6**, 309-312.

Hicks, J. L., & Marsh, R. L. (1998). A decrement-to-familiarity interpretation of the revelation effect from forced-choice tests of recognition memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **24**, 1105-1120.

Hockley, W. E., & Niewiadomski, M. W. (2000). [The revelation effect when conditions are blocked at test]. Unpublished raw data.

LeCompte, D. C. (1995). Recollective experience in the revelation effect: Separating the contributions of recollection and familiarity. *Memory & Cognition*, **23**, 324-334.

Luo, C. R. (1993). Enhanced feeling of recognition: Effects of identifying and manipulating test items on recognition. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **19**, 405-413.

McIntyre, J. S., & Craik, F. I. (1987). Age differences in memory for item and source information. *Canadian Journal of Psychology*, **41**, 175-192.

MRC Psycholinguistic Database (On-line). (1997). Available at: http://www.psy.uwa.edu.au/MRCDataBase/uwa_mrc.htm.

Multhaup, K. S. (1995). Aging, source, and decision criteria: When false fame errors do and do not occur. *Psychology & Aging*, **10**, 492-497.

Niewiadomski, M. W., & Hockley, W. E. (2001). Interrupting recognition memory: Tests of familiarity-based accounts of the revelation effect. *Memory & Cognition*, **29**, 1130-1138.

Paivio, A., Yuille, J. C., & Madigan, S. A. (1968). Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology Monographs*, **76**, 1-25.

Peynircioğlu, Z. F., & Tekcan, A. (1993). Revelation effect: Effort or priming does not create the sense of familiarity. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **19**, 382-388.

Prull, M. W., Light, L. L., Collett, M. E., & Kennison, R. F. (1998). Age-related differences in memory illusions: Revelation effect. *Aging, Neuropsychology, & Cognition*, **5**, 147-165.

Ratcliff, R., Hockley, W., & McKoon, G. (1985). Components of activation: Repetition and priming effects in lexical decision and recognition. *Journal of Experimental Psychology: General*, **114**, 435-450.

Stretch, V., & Wixted, J. T. (1998). On the difference between strength-based and frequency-based mirror effects in recognition memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **24**, 1379-1396.

Watkins, M. J., & Peynircioğlu, Z. F. (1990). The revelation effect: When disguising test items induces recognition. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **16**, 1012-1020.

Westerman, D. L. (2000). Recollection-based recognition eliminates the revelation effect in memory. *Memory & Cognition*, **28**, 167-175.

Westerman, D. L., & Greene, R. L. (1996). On the generality of the revelation effect. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **22**, 1147-1153.

Westerman, D. L., & Greene, R. L. (1998). The revelation that the revelation effect is not due to revelation. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **24**, 377-386.

Wixted, J. T. (1992). Subjective memorability and the mirror effect. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **18**, 681-690.

NOTES

1. A', like d', is an estimate of discriminability that is theoretically independent of the decision criterion. A' varies from 0 to 1, with .5 representing chance performance. A' is equivalent to percent correct on a two-alternative forced-choice recognition test. A', in contrast to d', is a slightly better measure of discriminability when criterion changes occur (Donaldson, 1993). β''_D is the measure of the decision criterion associated with A' and ranges from -1 to 1. Positive values reflect conservative performance, and negative values indicate liberal responding.

2. We should emphasize, however, that blocking the revelation and intact test conditions does not attenuate the standard revelation effect. We have found, in an unpublished study (Hockley & Niewiadomski, 2000), a reliable revelation effect for words when the intact and revelation conditions were blocked at test in the same manner as the experiments reported here. This result is consistent with Westerman and Greene's (1996, Experiment 2) demonstration of the revelation effect when the intact and revelation conditions were contrasted between subjects and lists.

(Manuscript received June 27, 2000;
revision accepted for publication July 16, 2001.)