

Does Familiarity Change in the Revelation Effect?

Michael F. Verde and Caren M. Rotello
University of Massachusetts at Amherst

The revelation effect describes the increased tendency to call items “old” when a recognition judgment is preceded by an incidental task. Past findings show that d' for recognition decreases following revelation, evidence that the revelation effect is due to familiarity change. However, data from receiver operating characteristic curves from 3 experiments produced no evidence of changes in recognition sensitivity. The authors illustrate how the use of a single-point measure like d' can be misleading when familiarity distribution variances are unequal. Also investigated was whether the effect depends on the revelation materials used. Neither the memorability of the revelation items, their similarity to recognition probes, nor the difficulty of the task changed the size of the effect. Thus, the revelation effect is not the result of a memory retrieval mechanism and seems to be generic and all-or-nothing. These characteristics are consistent with response bias rather than familiarity change.

Watkins and Peynircioglu (1990) coined the term *revelation effect* to describe the increased tendency to call an item “old” when it is revealed in distorted or fragmentary form just prior to a recognition judgment. Subsequent work has shown that almost any incidental task performed prior to recognition will lead to a revelation effect. Typing the recognition probe backwards, generating synonyms, and counting syllables and characters all produce the effect (Luo, 1993; Westerman & Greene, 1998). The incidental task need not involve the recognition probe at all: The effect is produced by solving unrelated anagrams, word fragments, and arithmetic problems (Bornstein & Neely, 2001; Niewiadomski & Hockley, 2001; Westerman & Greene, 1996; 1998). The task must, however, include active processing that is distinct from the memory judgment itself. Simply inserting a delay or increasing the presentation time of the recognition probe is not sufficient (Luo, 1993; Westerman & Greene, 1998). In addition, the revelation effect has only been observed for judgments of recent episodic memory, including judgments of list frequency (Bornstein & Neely, 2001; Westerman & Greene, 1996) but not for lexical decision, estimation of linguistic frequency, or semantic evaluation (Frigo, Reas, & LeCompte, 1999; Watkins & Peynircioglu, 1990).

Critical to a theoretical interpretation of the revelation effect is whether the phenomenon is due to response bias or reflects a real change in the information retrieved from memory. We evaluate the latter possibility, the *familiarity change* hypothesis, in light of two conflicting pieces of evidence. First, previous findings suggest that the size of the revelation effect does not depend on the memory

qualities of the revelation item, the similarity between the revelation item and the recognition probe, or the difficulty of the revelation task. These properties rule out the most likely mechanisms of familiarity change and weigh against the plausibility of that hypothesis. On the other hand, prior signal detection analysis of the same data suggests that revelation may cause a change in memory sensitivity (Hicks & Marsh, 1998). This is clear evidence for familiarity change.

In three experiments, we show that several factors that should be critical to a familiarity change mechanism fail to affect the magnitude of the revelation effect. Moreover, analysis of receiver operating characteristic (ROC) curves suggests that sensitivity is unaffected by revelation. We argue that the existing evidence for sensitivity change may be an artifact of the inappropriate application of single-point measures of sensitivity such as d' . Taken together, our findings make a strong case against the familiarity change hypothesis.

Underlying Mechanisms

Familiarity is the nonspecific sense of “oldness” produced when an object matches the contents of memory. Whether recognition relies on the familiarity process alone or is supplemented by recollection (memory for specific information) remains a matter of debate. However, even from the latter perspective, evidence suggests that the revelation task primarily affects the familiarity process. For example, when process dissociation is used to separate the contributions of each process in recognition, revelation mainly affects the familiarity component (LeCompte, 1995). In addition, the revelation effect is often reduced or absent in tasks thought to rely heavily on recollection, such as associative recognition (Cameron & Hockley, 2000; Westerman, 2000). How might revelation affect familiarity? Three properties are critical to delineating the nature of this hypothetical mechanism: the similarity of the revelation item to memory, the similarity of the revelation item to the recognition probe, and the difficulty of the revelation task.

One possibility is that the revelation item recruits normal memory processes, leading to residual familiarity that carries over into the subsequent recognition judgment. If this were true, the size of

Michael F. Verde and Caren M. Rotello, Department of Psychology, University of Massachusetts at Amherst.

This research was supported in part by National Institutes of Health Research Grant MH60274-02 to Caren M. Rotello, and by National Institutes of Health Training Grant MH16745-19 to Michael F. Verde. We wish to thank John Reeder, Neil Macmillan, and several anonymous reviewers for their invaluable comments.

Correspondence concerning this article should be addressed to Michael F. Verde, Department of Psychology, Box 37710, University of Massachusetts, Amherst, Massachusetts 01003-7710. E-mail: mverde@psych.umass.edu

the revelation effect would be related to the similarity of the revelation item to memory. Several studies have directly compared different revelation materials. Peynircioglu and Tekcan (1993) varied the similarity of the revelation item to the study list by manipulating word frequency, category membership, and orthographic similarity. They found no differences in the size of the revelation effect. Niewiadomski and Hockley (2001) found no difference between word and number revelation items when studied items were words. On the other hand, Westerman and Greene (1998) failed to obtain a revelation effect when revelation and recognition materials were very different (numbers and words). Whittlesea and Williams (2001) found that words produced a larger revelation effect than nonwords (although reversing the presentation order of revelation and recognition items reversed the effect, suggesting the role of response bias). Taken together, there is (at best) weak evidence that the revelation effect is the product of normal memory processes.

An alternative possibility is that the processing of the revelation item, rather than the characteristics of the item, leads to familiarity increase. If so, one might expect that the effect would be greater when the revelation and memory tasks engage similar processes by targeting similar materials. Evidence for this has been inconsistent (Niewiadomski & Hockley, 2001; Westerman & Greene, 1998). Moreover, one might expect that a revelation task requiring more cognitive resources would exacerbate the interaction between revelation and memory tasks. However, Niewiadomski and Hockley (2001) showed that the size of the revelation effect does not change when recognition is preceded by two revealed items rather than one. Using anagrams as revealed items, Peynircioglu and Tekcan (1993) found an analogous result: There is no correlation between revelation effect size and anagram completion time.

In summary, evidence suggests that the familiarity-change mechanism is not tied to normal memory retrieval processes, may not be material specific, and produces an all-or-nothing rather than a graded effect. These data impose strong constraints that challenge the plausibility of a familiarity-change mechanism. In the experiments to follow, we gathered additional evidence to support this characterization of the literature.

Signal Detection Model

Although the qualitative evidence seems to implicate a decision process, quantitative evidence from signal detection analysis supports the familiarity-change hypothesis. Hicks and Marsh (1998) calculated sensitivity measures for 32 successful replications of the revelation effect and found a statistically reliable difference between revelation ($d' = 0.81$) and no revelation ($d' = 0.90$) conditions.¹ A change in memory sensitivity implies that familiarity (either mean strength or variability), and not just response bias, has changed.

In signal-detection theory, the ability to discriminate old from new items depends on the distance between strength distributions as well as their variance. In the standard recognition task, the binary decision ("old" vs. "new") yields a pair of hit and false-alarm rates per condition. Determining sensitivity from such sparse information requires assumptions about underlying representation; in the standard signal-detection model, the familiarity distributions are assumed to be equal variance Gaussian. Given these assumptions, the appropriate measure of memory sensitivity is d' , the

difference between the means of the distributions in units of the common standard deviation (z scores),

$$d' = z(H) - z(F). \quad (1)$$

If the revelation manipulation leads to a decrease in d' , as suggested by the Hicks and Marsh (1998) meta-analysis, then at least one distribution has moved along the familiarity dimension or changed in variability. However, a number of studies have shown that the new item distribution typically has a smaller standard deviation than the old item distribution in item recognition tasks (Glanzer, Kim, Hilford, & Adams, 1999; Hirshman & Hostetter, 2000; Ratcliff, Sheu, & Gronlund, 1992). When the equal variance assumption is violated, conclusions based on d' can be misleading.

The key to resolving this problem is a more complete description of recognition performance: the receiver operating characteristic (ROC) curve. The decision criterion may be placed at any point along the strength axis. Each of these hypothetical points yields a pair of hit and false-alarm rates. An ROC curve is a plot of the hit and false-alarm rate pairs that would result from moving the criterion from right to left along the familiarity axis. Thus, ROC curves describe how performance changes as a result of response bias when sensitivity is held constant.

Transforming the ROC curve into z -score units results in a z -ROC (or *normal-normal*) plot that has two relevant properties. Assuming Gaussian distributions, the z -ROC is linear, with slope equal to the ratio of the standard deviations of the new and old item distributions (Lockhart & Murdock, 1970). In Figure 1, the z -ROC lines L_{chance} , L_A , and L_B describe different levels of sensitivity when new and old item familiarity have equal variance (slope = 1). L_{chance} describes chance performance or zero sensitivity. Because d' is defined as the vertical or horizontal distance between a (z_H , z_F) point and the chance line, all points on L_A have $d' = d'_A$, and all points on L_B have $d' = d'_B$, where $d'_A > d'_B$. Consider now L_R , which describes equal sensitivity when old item variance is greater than new item variance (slope < 1). If L_R describes the true performance curve in a recognition test, then Points a and b represent the same degree of sensitivity (they fall on the same z -ROC function). However, Point a implies $d' = d'_A$, whereas Point b implies $d' = d'_B$. Suppose that the increase in hits and false alarms following revelation is the product of a more liberal response bias. When slope < 1, d' will systematically decrease as bias becomes more liberal, as is clear in Figure 1. Thus, there is a simple alternative explanation for the decrease in d' following revelation observed in the Hicks and Marsh (1998) meta-analysis.

In the experiments to follow, we replaced the binary ("old" vs. "new") judgment typically used in revelation experiments with a confidence rating on a scale ranging from 1 (*very sure new*) to 6 (*very sure old*). With confidence ratings, one can plot hit and false-alarm rates at several points on the ROC curve and thus observe the z -ROC slope empirically. This information allows the

¹ Our own calculation of d' based on the hit and false-alarm rates provided by Hicks and Marsh (1988) in their meta-analysis yielded values different from those reported by the authors; notably, their reported values for the revelation conditions had mean $d' = 0.58$, whereas ours had mean $d' = 0.81$. Nevertheless, the difference between revelation and no revelation condition d' remained statistically reliable, $t(31) = 4.28$, $p < .001$.

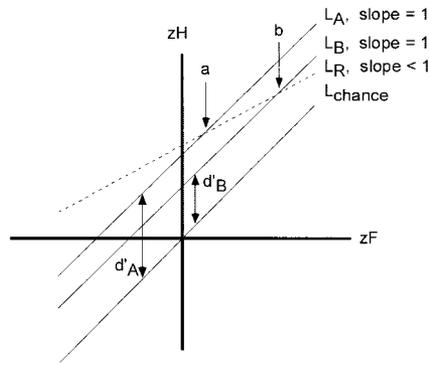


Figure 1. Examples of z -transformed ROC curves when the familiarity distributions of new and old items have equal variance (slope = 1; L_A , L_B , L_{chance}) and when old item variance exceeds that of new items (slope < 1; L_R). Point a describes a greater degree of sensitivity than Point b when slope = 1, but the same degree of sensitivity when slope < 1.

use of the sensitivity measure d_a , an alternative to d' that allows for possible differences in variance between familiarity distributions:

$$d_a = [2/(1 + \text{slope}^2)]^{1/2}[z(H) - \text{slope} \cdot z(FA)]. \quad (2)$$

In z space, d_a is the average of the horizontal and vertical distances between a point and the chance line (Macmillan & Creelman, 1991). The measures d_a and d' share the same unit scale, and when slope = 1 they yield the same value.

A change in d_a following revelation would provide strong support for the familiarity change hypothesis. Finding no change in d_a would not rule out familiarity change, but it would mean that familiarity is changing so precisely that both the distance between the means of the new and old item distributions as well as their relative variance remain constant. More importantly, observing no change in d_a is a requirement of any claim that the revelation effect is purely a product of response bias.

Experiment 1

A goal common to all of the present experiments was to determine whether the revelation manipulation leads to a decline in memory sensitivity. In addition, each of the experiments looked at one of the qualitative characteristics of a hypothetical mechanism for familiarity change. In Experiment 1, we examined the relationship between the revelation item and memory.

Method

Subjects. Forty undergraduates from the University of Massachusetts at Amherst participated for course credit.

Materials and design. Stimuli were drawn from a pool of 300 eight-letter nouns of low frequency (<101/million; Kučera & Francis, 1967). The study list consisted of 135 words (125 critical words and 10 fillers used as primacy and recency buffers). The test list consisted of 150 recognition probes, half from the study list and half new words. An additional 12 practice trials, created from filler items and new words, were placed at the beginning of the test. During the test, some recognition probes were preceded by an anagram created by scrambling the letters of an eight-letter word. Every anagram could be unscrambled with the key: 54687321 (the first letter of the anagram was the fifth letter in the unscrambled form, the second letter of the anagram was the fourth letter in the unscrambled form,

and so forth; for example, the anagram *etmtpnoc* would be solved to reveal *contempt*). Every anagram was unique, and no word appeared as both an anagram and a recognition probe. One third of both old and new recognition probes were preceded by an anagram created from a word that had appeared in the study list (old revelation condition). One third were preceded by an anagram created of a word that had never appeared before (new revelation condition). Finally, one third served as the no-anagram controls (no revelation condition).

The assignment of words to list position and condition was randomized for each subject. List creation, stimulus presentation, and response collection were all computer controlled. Subjects were assigned to individual Pentium-2 PCs and testing rooms.

Procedure. The experiment consisted of a 50-min session, divided into a study phase followed by a test phase. During the study phase, subjects were shown a 135-word study list and instructed to learn the list for an upcoming memory test. On each study trial, a single word appeared in the center of the computer screen for 2,000 ms, followed by a 500-ms blank interval.

The test phase began with 12 practice trials (from which no data were collected), followed by 150 critical trials. Each trial was preceded by a fixation line, + + + + +, displayed in the center of the screen for 500 ms. During the test, subjects were told to expect two types of probes: anagram probes and recognition probes. With recognition probes, a word appeared in the center of the screen with the prompt *confidence? (1–6)* below it. Subjects were to decide whether the word had appeared in the study list. Judgments were made on a 6-point scale ranging from 1 (*very sure new*) to 6 (*very sure old*). The recognition probe remained on the screen until a response was made. A 1,500-ms blank interval followed the response. Some recognition probes were preceded by an anagram probe. The goal was to unscramble the anagram and type the resulting word using the keyboard. A 1,000-ms blank interval followed response completion. Subjects were given the anagram key and instructed to use it to ensure 100% accuracy when solving the anagrams.

Results

For the anagrams, typing errors were defined as those in which no more than 25% of the letters were incorrect additions, omissions, or transpositions. More serious errors were rare (<1%) for either type of anagram, not surprising given that subjects were provided the solution key. Anagram completion was not analyzed further. Recognition performance was analyzed in two ways. First, for each of the three experimental conditions, confidence ratings 1–3 were combined to form the “new” response category, and ratings 4–6 were combined to form the “old” response category. This yielded overall hit and false-alarm rates that were submitted to an analysis of variance (ANOVA) in order to describe the general trends in the data. Second, the confidence ratings were used to construct an ROC curve for each experimental condition for each subject using the maximum likelihood estimation procedure (see Appendix for full confidence-rating data). These curves provided individual measures of z -ROC slope and sensitivity (d_a). An alpha level of .05 was used for all statistical tests.

Mean hit and false-alarm rates are reported in Table 1. Solving an anagram just prior to recognition led to a revelation effect: There was an increase in both hits and false alarms. However, the size of the effect did not depend on whether anagrams were constructed from studied words or novel words. These conclusions were supported by a 2 (probe type: old, new) \times 3 (condition: old, new, no revelation) repeated measures ANOVA that yielded significant main effects for both probe type, $F(1, 39) = 199.18$, $MSE = 0.04$, $p < .01$, and condition, $F(2, 78) = 15.58$,

Table 1
Hit and False-Alarm (FA) Rates for Experiments 1–3

| Revelation condition | Hit | FA |
|----------------------|-----|-----|
| Experiment 1 | | |
| Old | .77 | .42 |
| New | .77 | .40 |
| No revelation | .71 | .31 |
| Experiment 2 | | |
| Word | .79 | .46 |
| Number | .76 | .45 |
| No revelation | .71 | .32 |
| Experiment 3 | | |
| Eight digit | .76 | .38 |
| Three digit | .76 | .41 |
| No revelation | .69 | .28 |

$MSE = 0.01$, $p < .01$. The interaction of probe type and condition was not significant, $F(2, 78) = 1.93$, $MSE = 0.01$, *ns*.

Signal detection analyses revealed that the z -ROC slope was less than 1 in each condition (old = 0.85, new = 0.82, no revelation = 0.79). These slopes did not differ reliably, $F(2, 78) = 0.73$, $MSE = 0.05$. In addition, there was a significant change in decision criterion c_a from a relatively unbiased placement in the no revelation condition to a more liberal bias in the revelation conditions (no revelation = -0.04 , revelation = -0.26); $t(39) = 4.75$, $p < .001$.² Together, these findings support our contention that use of a single-point sensitivity measure such as d' is inappropriate and misleading in this paradigm. Instead, we compared sensitivity across conditions using d_a . Revelation condition had no significant effect on d_a (old = 1.09, new = 1.11, no revelation = 1.08); $F(2, 78) = 0.08$, $MSE = 0.10$, *ns*.

Discussion

One version of the familiarity-change hypothesis is that the revelation item makes contact with representations in memory, generating familiarity that is misattributed to the recognition probe (Westerman & Greene, 1998). The results of this experiment do not support that hypothesis. On the contrary, the fact that both old and new anagrams produced revelation effects of the same magnitude suggests that the memorability of particular items has nothing to do with the revelation effect. In other words, the effect is not a by-product of the normal recognition process. If residual familiarity is introduced into the system by revelation, it is by means of some other channel. The analysis of d_a uncovered no evidence that revelation affected memory sensitivity, but analysis of decision criterion c_a revealed a more liberal response bias in the revelation conditions.

Experiment 2

If the memory qualities of the revelation item are not important, perhaps the similarity of the processes engaged by the revelation task and by the recognition task are critical to the revelation effect. In Experiment 2, the revelation task consisted of unscrambling

either a novel, eight-letter word or an eight-digit number. The recognition probe was always a word.

Method

Subjects. Twenty-two undergraduates from the University of Massachusetts at Amherst who had not been involved in Experiment 1 participated for course credit.

Materials and design. Word stimuli were drawn from the pool of nouns used in Experiment 1. The study list consisted of 76 words (66 critical words and 10 fillers used as primacy and recency buffers). The test list consisted of 132 recognition probes, half from the study list and half new words. An additional 12 practice trials, with old words drawn only from the filler items, were placed at the beginning of the test. During the test, some recognition probes were preceded by either a word or a number anagram. Word anagrams were formed from novel eight-letter words. Number anagrams were randomly generated eight-digit numbers that were described to the subject as numbers that had been scrambled and needed to be unscrambled. The key used by subjects to rearrange the letters or numerals of the anagram was identical for both types: 54687321. Every anagram was unique, and no word appeared as both an anagram and a recognition probe. One third of both old and new recognition probes were preceded by a word anagram (word condition). One third were preceded by a number anagram (number condition). Finally, one third served as the no-anagram controls.

The assignment of words to list position and condition was randomized for each subject. List creation, stimulus presentation, and response collection were all computer controlled. Subjects were assigned to individual Pentium-2 PCs and testing rooms.

Procedure. The procedure was identical in most respects to that used in Experiment 1. It differed only in the length of study and test lists and the types of anagrams used for the revelation conditions.

Results

There were very few serious errors (<1%) in the completion of anagrams of either type. As before, recognition performance was analyzed first in terms of overall hit and false-alarm rates and then by examining the ROC curves constructed from recognition confidence ratings.

Mean hit and false-alarm rates are reported in Table 1. Revelation led to an increase in both hits and false alarms, but the size of the increase did not depend on whether anagrams were words or numbers. These conclusions were supported by a 2 (probe type: old, new) \times 3 (condition: word, number, no revelation) repeated measures ANOVA that yielded significant main effects for both probe type, $F(1, 21) = 112.50$, $MSE = 0.04$, $p < .01$; and condition, $F(2, 42) = 15.21$, $MSE = 0.01$, $p < .01$. The interaction of probe type and condition was not significant, $F(2, 42) = 1.68$, $MSE = 0.01$, *ns*.

As in Experiment 1, signal detection analyses revealed that the z -ROC slopes were less than one in each condition (word = 0.77, number = 0.90, no revelation = 0.77) and did not differ reliably, $F(2, 42) = 1.56$, $MSE = 0.08$, *ns*. In addition, there was a

² Bias measure c_a is an alternative to c given unequal variance (Macmillan & Creelman, 1991). It equals

$$- \{2^{1/2} \text{slope} / [(1 + \text{slope}^2)^{1/2} (1 + \text{slope})]\} [z(H) + z(F)].$$

With equal variance, $c_a = c$. Conclusions were not changed when c was used instead of c_a .

significant change in decision criterion c_a from a relatively unbiased placement in the no revelation condition to a more liberal bias in the revelation conditions (no revelation = -0.07 , revelation = -0.35); $t(21) = 4.26$, $p < .01$. We compared sensitivity across conditions using d_a , and found no significant differences (word = 1.03 , number = 1.03 , no revelation = 1.07); $F(2, 42) = 0.11$, $MSE = 0.14$, *ns*.

Discussion

The observation that the revelation effect did not differ for numerical and word anagrams suggests that the revelation effect is generic and not material specific. Considering the findings of Niewiadomski and Hockley (2001), we believe it is clear that a revelation effect can result from materials very different from those used in the recognition test. Westerman and Greene's (1998) discrepant finding is puzzling, but it now seems likely to have been due to factors other than the similarity of materials across tasks. As in Experiment 1, analysis of decision criterion c_a revealed a more liberal bias in the revelation conditions, and analysis of d_a uncovered no evidence that revelation affected memory sensitivity.

Experiment 3

In Experiments 1 and 2, the specific qualities of the revelation items had no bearing on the size of the revelation effect. In the final experiment, we turned to qualities of the task itself. In Experiment 3, we manipulated task difficulty by varying the length of the anagram used in the revelation task. Numerical anagrams were either eight-digit strings or three-digit strings. The three-digit strings could be unscrambled by simply reversing the order of the digits, which subjects reported was a trivial task.

Method

Subjects. Twenty-one undergraduates from the University of Massachusetts at Amherst who had not been involved in Experiment 1 or 2 participated for course credit.

Materials and design. Word stimuli were drawn from the pool of nouns used in Experiment 1. The study list consisted of 76 words (66 critical words and 10 fillers used as primacy and recency buffers). The test list consisted of 132 recognition probes, half from the study list and half new words. An additional 12 practice trials, with old words drawn only from the filler items, were placed at the beginning of the test. During the test, some recognition probes were preceded by either eight-digit or three-digit number anagrams. Number anagrams were randomly generated eight-digit or three-digit numbers. The keys used to unscramble the numeral strings were 54687321 (eight-digit) and 321 (three-digit). Every anagram was unique. One third of both old and new recognition probes were preceded by an eight-digit number anagram (eight-digit condition). One third were preceded by a three-digit number anagram (three-digit condition). Finally, one third served as the no anagram controls.

The assignment of words to list position and condition was randomized for each subject. List creation, stimulus presentation, and response collection were all computer controlled. Subjects were assigned to individual Pentium-2 PCs and testing rooms.

Procedure. The procedure was identical in most respects to that used in Experiment 1. It differed only in the length of study and test lists and the types of anagrams used for the revelation conditions.

Results

One subject failed to complete the anagrams and was removed from the analysis. Otherwise, there were very few serious errors ($<1\%$) in anagram completion.

Mean hit and false-alarm rates are reported in Table 1. Revelation led to an increase in both hits and false alarms, but the size of this increase did not depend on whether anagrams were eight-digit or three-digit numbers. These conclusions were supported by a 2 (probe type: old, new) \times 3 (condition: eight-digit, three-digit, no revelation) repeated measures ANOVA that yielded significant main effects for both probe type, $F(1, 19) = 105.60$, $MSE = 0.04$, $p < .01$, and condition, $F(2, 38) = 10.69$, $MSE = 0.01$, $p < .01$. The interaction of probe type and condition was not significant.

As in Experiments 1 and 2, signal-detection analyses revealed that the z -ROC slope was less than one in each condition, eight-digit = 0.84 , three-digit = 0.85 , no revelation = 0.81 , and did not differ reliably across conditions, $F(2, 38) = 0.18$, $MSE = 0.06$, *ns*. In addition, there was a significant change in decision criterion c_a from a relatively unbiased placement in the no revelation condition to a more liberal bias in the revelation conditions: no revelation = 0.05 , revelation = -0.23 ; $t(19) = 3.90$, $p = .01$. We compared sensitivity across conditions using d_a , and found no significant differences: eight-digit = 1.02 , three-digit = 1.10 , no revelation = 1.06 ; $F(2, 38) = 0.30$, $MSE = 0.09$, *ns*.

Discussion

Both eight-digit and three-digit anagrams produced robust revelation effects that did not differ in size. This is consistent with past findings suggesting that the revelation effect is all-or-nothing rather than a graded function of the amount of effort required by the incidental task. As in Experiments 1 and 2, analysis of c_a indicated that a more liberal response bias was used in the revelation conditions, and analysis of d_a uncovered no evidence that revelation affected memory sensitivity.

General Discussion

We compared the revelation effect produced by different types of materials: old and new word anagrams in Experiment 1, word and number anagrams in Experiment 2, and eight-digit and three-digit number anagrams in Experiment 3. Solving an anagram prior to a recognition judgment consistently led to an increased tendency to call a recognition probe "old." However, the size of this effect was the same regardless of anagram type.

Measures of Sensitivity

The question of sensitivity change is critical to a theoretical interpretation of the revelation effect. A change in sensitivity means that the shape or the distance between familiarity distributions has changed. We constructed ROC curves on the basis of recognition confidence ratings, allowing us to measure sensitivity while accounting for unequal variance. In none of the three experiments was there evidence that d_a was affected by revelation (see Table 2). In a final effort to detect a sensitivity effect, we pooled the data from all subjects ($N = 82$) and compared the revelation condition (collapsing the separate revelation types in each experiment) to the no revelation condition. As before, there was no

Table 2
Summary of Signal Detection Measures in Experiments 1–3

| Condition | d_a | d' | Slope |
|---------------|-------|------|-------|
| Experiment 1 | | | |
| No revelation | 1.08 | 1.19 | 0.79 |
| Revelation | 1.12 | 1.06 | 0.81 |
| Difference | -0.04 | 0.13 | 0.02 |
| Experiment 2 | | | |
| No revelation | 1.07 | 1.10 | 0.77 |
| Revelation | 1.02 | 0.96 | 0.78 |
| Difference | 0.05 | 0.14 | 0.01 |
| Experiment 3 | | | |
| No revelation | 1.06 | 1.17 | 0.80 |
| Revelation | 1.08 | 1.07 | 0.79 |
| Difference | -0.02 | 0.10 | -0.01 |

Note. The revelation condition is based on collapsing the two separate revelation conditions within an experiment. Difference = no revelation – revelation.

reliable difference between the revelation ($d_a = 1.08$) and no revelation ($d_a = 1.09$) conditions, $t(81) = 0.06$.

Hicks and Marsh (1998) noted a reliable trend for d' to decrease following revelation. However, d' carries with it the assumption that distributions of new and old items have equal variance. The z -ROC slopes in the revelation tasks reported here, as well as those observed generally in item recognition (Ratcliff et al., 1992), indicate that the equal variance assumption is unjustified. Values of d' for each experiment can be found in Table 2.³ We noted that when sensitivity is constant but the z -ROC slope is less than 1, the value of d' necessarily decreases as response bias becomes more liberal, as it did in the revelation conditions of each experiment. The difference in d' may not always be large enough to detect empirically; it depends on factors such as the z -ROC slope and the difference in bias between conditions for individual participants. However, across the three experiments, there was a systematic decrease in d' following revelation, and the size of these differences was similar to that observed in the Hicks and Marsh meta-analysis. Pooling across the three experiments, we found that the differences between revelation ($d' = 1.04$) and no revelation ($d' = 1.16$) conditions was significant, $t(81) = 2.09$, $p < .05$. These results are consistent with our observation that the decision criterion c_a was more liberally placed in the revelation condition of each experiment.

The finding that revelation does not lead to a change in sensitivity does not rule out familiarity change, but it does impose strong constraints: If revelation increases the familiarity of both new and old items, it does so in a way that maintains both the distance between the means of the signal distributions and the relative variance of these distributions. The finding also means that an alternative explanation based solely on response bias is now viable.

Familiarity Change or Response Bias?

The revelation effect does not depend on the memorability of the revelation items, the similarity between the revelation item and

the recognition probe, or the difficulty of the revelation task. These characteristics argue against mechanisms having to do with residual familiarity or the interaction of overlapping cognitive processes. They seem more plausibly attributed to a strategic response bias. The fact that sensitivity remains constant is also consistent with the change in response bias that we observed in each experiment.

Two additional findings provide converging evidence for response bias. Whittlesea and Williams (2001) found that words produced a larger revelation effect than nonwords, a pattern consistent with familiarity change because words are generally more familiar than nonwords. However, they found the opposite result when the recognition probe preceded the revelation task and the “old–new” judgment of the probe was delayed until after the revelation task. Hockley and Niewiadomski (2001) found a revelation effect with lists composed entirely of either rare words or nonwords. However, when rare words and nonwords were intermixed with common words in study and test lists, the revelation effect was observed only for the common words. The difficulty in isolating familiarity and bias effects when there is a change in hit and false-alarm rates is that both effects may be present. If familiarity and bias change in opposite directions, one effect may hide the other. Thus, while these findings clearly implicate a decision bias, they do not rule out the possibility that an increase in hits and false alarms caused by an increase in familiarity was hidden by a decrease in hits and false alarms caused by a conservative bias shift. The constraints we have outlined are useful in that they actively argue against the presence of familiarity change.

Hicks and Marsh (1998) argued for the presence of both familiarity and bias effects on the basis of findings from two-alternative forced-choice (2AFC) recognition. In two experiments, they found that revealing one of the recognition probes had no effect on the subsequent memory judgment. When the memory task was made more difficult, first by introducing trials containing both new or both old probes, and then by inserting a delay between study and test, they found that revealing one of the probes made it less likely to be chosen. According to Hicks and Marsh, this *antirevelation effect* showed that revelation reduces familiarity. Because this would lead to a decrease in hits and false alarms in the old–new recognition task (contrary to empirical findings), the familiarity reduction was assumed to be coupled with a liberal criterion shift that led to an even larger increase in hits and false alarms.

Hicks and Marsh (1998) described the familiarity reduction as a reduction in the signal-to-noise ratio. This implies a reduction in sensitivity, for which we found no evidence in the present data. An alternative explanation for the antirevelation effect is suggested by

³ Another single-point estimate of sensitivity, A' , is an estimate of the area under the ROC curve. Like d' , A' has the weakness that it is inaccurate when the equal variance assumption is violated (Donaldson, 1993), as it is in our data. In addition to predicting symmetric ROC curves, A' has several other problematic characteristics, one of which deserves mention: When performance is high, A' takes on characteristics of a threshold process (e.g., curvilinear z -ROCs; Macmillan & Creelman, 1996), contrary to what is typically observed empirically. Because the shape of the ROC implied by the use of A' (symmetric ROC, curvilinear z -ROC) is inconsistent with our data, and because there is no unequal variance correction for A' , we do not consider it further. (See Pastore, Crawley, Berens, & Skelly, in press, for additional arguments against the use of A' .)

evidence that subjects sometimes attempt to counter the effects of priming. In a study by Jacoby and Whitehouse (1989), subjects became reluctant to call an item "old" when they were aware that it had been preceded by an identical prime. Similarly, Huber, Shiffrin, Quach, and Lyle (2002; Huber, Shiffrin, Lyle & Ruys, 2001) found that subjects showed a preference against an identically primed item in a 2AFC identification task when prime duration was sufficiently long. The effect was fragile and could be reversed under different conditions, which might explain the discrepancy between preference for identically primed items in revelation studies involving old–new recognition and preference against identically primed items sometimes observed in 2AFC recognition.

If the revelation effect is purely the result of a change in response bias, there remains the question of why such a bias shift occurs. Niewiadomski and Hockley (2001) suggested that disruption from the revelation task causes the subject to temporarily forget the criterion setting called for by the experimental context. Why this would consistently lead to more liberal responding is unclear. We suggest another possibility, mentioned also by Hicks and Marsh (1998), related to Hirshman's (1995) finding that strengthening memory leads to a more conservative response bias. The intuitive explanation is that as the memory judgment becomes easier, subjects adopt a higher standard of performance. The corollary is that, as the memory judgment becomes more difficult, subjects become more lenient in what they will call "old." It is clear from our sensitivity measures that revelation does not actually increase the difficulty of recognition. However, subjects may believe that it does and shift their decision criteria accordingly.

References

- Bornstein, B. H., & Neely, C. B. (2001). The revelation effect in frequency judgment. *Memory & Cognition*, 29, 209–213.
- 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.
- Frigo, L. C., Reas, D. L., & LeCompte, D. C. (1999). Revelation without presentation: Counterfeit study list yields robust revelation effect. *Memory & Cognition*, 27, 339–343.
- Glanzer, M., Kim, K., Hilford, A., & Adams, J. K. (1999). Slope of the receiver-operating characteristic in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25, 500–513.
- 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, and Cognition*, 24, 1105–1120.
- Hirshman, E. (1995). Decision processes in recognition memory: Criterion shifts and the list-strength paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 302–313.
- Hirshman, E., & Hostetter, M. (2000). Using ROC curves to test models of recognition memory: The relationship between presentation duration and slope. *Memory & Cognition*, 28, 161–166.
- Hockley, W. E., & Niewiadomski, M. W. (2001). Interrupting recognition memory: Tests of a criterion-change account of the revelation effect. *Memory & Cognition*, 29, 1176–1184.
- Huber, D. E., Shiffrin, R. M., Lyle, K. B., & Ruys, K. I. (2001). Perception and preference in short-term word priming. *Psychological Review*, 108, 149–182.
- Huber, D. E., Shiffrin, R. M., Quach, R., & Lyle, K. B. (2002). Mechanisms of source confusion and discounting in short-term priming 1: Effects of prime duration and prime recognition. *Memory & Cognition*, 30, 745–757.
- Jacoby, L. L., & Whitehouse, K. (1989). An illusion of memory: False recognition influenced by unconscious perception. *Journal of Experimental Psychology: General*, 118, 126–135.
- Kučera, F., & Francis, W. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.
- LeCompte, D. (1995). Recollective experience in the revelation effect: Separating the contributions of recollection and familiarity. *Memory & Cognition*, 23, 324–334.
- Lockhart, R. S., & Murdock, B. B., Jr. (1970). Memory and the theory of signal detection. *Psychological Bulletin*, 74, 100–109.
- Luo, C. R. (1993). Enhanced feeling of recognition: Effects of identifying and manipulating test items on recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 13, 405–413.
- Macmillan, N., & Creelman, C. D. (1991). *Detection theory: A user's guide*. New York: Cambridge University Press.
- Macmillan, N., & Creelman, C. D. (1996). Triangles in ROC space: History and theory of "nonparametric" measures of sensitivity and response bias. *Psychonomic Bulletin and Review*, 3, 164–170.
- Niewiadomski, M. W., & Hockley, W. E. (2001). Interrupting recognition memory: Tests of familiarity-based accounts of the revelation effect. *Memory & Cognition*, 29, 1130–1138.
- Pastore, R. E., Crawley, E. J., Berens, M. S., & Skelly, M. (in press). Nonparametric A' and other modern misconceptions about signal detection theory. *Psychonomic Bulletin & Review*.
- Peynircioglu, Z. F., & Tekcan, A. I. (1993). Revelation effect: Effort or priming does not create the sense of familiarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19, 382–388.
- Ratcliff, R., Sheu, C., & Gronlund, S. D. (1992). Testing global memory models using ROC curves. *Psychological Review*, 99, 518–535.
- Watkins, M. J., & Peynircioglu, Z. F. (1990). The revelation effect: When disguising test items induces recognition. *Journal of Experimental Psychology: Learning, Memory, and 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, and Cognition*, 22, 1147–1153.
- Westerman, D., & Greene, R. L. (1998). The revelation that the revelation effect is not due to revelation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24, 377–386.
- Whittlesea, B. W. A., & Williams, L. D. (2001). The discrepancy-attribution hypothesis: I. The heuristic basis of feelings and familiarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27, 3–13.

(Appendix follows)

Appendix

Confidence Rating Data

Table A1
Average Proportion of Responses at Each Confidence Rating in Experiment 1

| Condition and confidence rating | New item | Old item |
|---------------------------------|----------|----------|
| Old revelation | | |
| 1 | 0.165 | 0.044 |
| 2 | 0.228 | 0.080 |
| 3 | 0.189 | 0.106 |
| 4 | 0.180 | 0.143 |
| 5 | 0.145 | 0.166 |
| 6 | 0.093 | 0.461 |
| New revelation | | |
| 1 | 0.162 | 0.042 |
| 2 | 0.233 | 0.094 |
| 3 | 0.208 | 0.099 |
| 4 | 0.166 | 0.135 |
| 5 | 0.137 | 0.176 |
| 6 | 0.092 | 0.454 |
| No revelation | | |
| 1 | 0.214 | 0.055 |
| 2 | 0.275 | 0.104 |
| 3 | 0.203 | 0.127 |
| 4 | 0.128 | 0.143 |
| 5 | 0.115 | 0.142 |
| 6 | 0.065 | 0.427 |

Note. Confidence ratings range from 1 (*very sure new*) to 6 (*very sure old*).

Table A3
Average Proportion of Responses at Each Confidence Rating in Experiment 3

| Condition and confidence rating | New item | Old item |
|---------------------------------|----------|----------|
| Eight digit | | |
| 1 | 0.183 | 0.039 |
| 2 | 0.226 | 0.070 |
| 3 | 0.212 | 0.130 |
| 4 | 0.169 | 0.134 |
| 5 | 0.132 | 0.175 |
| 6 | 0.078 | 0.452 |
| Three digit | | |
| 1 | 0.175 | 0.041 |
| 2 | 0.232 | 0.102 |
| 3 | 0.184 | 0.098 |
| 4 | 0.175 | 0.141 |
| 5 | 0.152 | 0.186 |
| 6 | 0.082 | 0.432 |
| No revelation | | |
| 1 | 0.236 | 0.068 |
| 2 | 0.214 | 0.105 |
| 3 | 0.266 | 0.139 |
| 4 | 0.170 | 0.173 |
| 5 | 0.064 | 0.134 |
| 6 | 0.050 | 0.382 |

Note. Confidence ratings range from 1 (*very sure new*) to 6 (*very sure old*).

Table A2
Average Proportion of Responses at Each Confidence Rating in Experiment 2

| Condition and confidence rating | New item | Old item |
|---------------------------------|----------|----------|
| Word | | |
| 1 | 0.119 | 0.054 |
| 2 | 0.191 | 0.081 |
| 3 | 0.227 | 0.077 |
| 4 | 0.164 | 0.133 |
| 5 | 0.207 | 0.210 |
| 6 | 0.092 | 0.443 |
| Number | | |
| 1 | 0.190 | 0.066 |
| 2 | 0.200 | 0.086 |
| 3 | 0.150 | 0.084 |
| 4 | 0.175 | 0.111 |
| 5 | 0.180 | 0.212 |
| 6 | 0.104 | 0.440 |
| No revelation | | |
| 1 | 0.183 | 0.068 |
| 2 | 0.254 | 0.084 |
| 3 | 0.238 | 0.138 |
| 4 | 0.165 | 0.117 |
| 5 | 0.104 | 0.151 |
| 6 | 0.056 | 0.442 |

Note. Confidence ratings range from 1 (*very sure new*) to 6 (*very sure old*).

Received July 26, 2002
Revision received February 23, 2003
Accepted March 14, 2003 ■