Memory & Cognition
2001, 29 (8), 1130-1138

Interrupting recognition memory: Tests of familiarity-based accounts of the revelation effect

MARTY W. NIEWIADOMSKI and WILLIAM E. HOCKLEY
Wilfrid Laurier University, Waterloo, Ontario, Canada

The revelation effect is a puzzling phenomenon in which items on a recognition test are more likely to be judged as "old" when they are immediately preceded by a problem-solving task, such as anagram solution. The present experiments were designed to evaluate Westerman and Greene's (1998) and Hicks and Marsh's (1998) familiarity-based accounts of this effect. We found comparable revelation effects when probes were preceded by an anagram or a numerical addition task and when subjects performed either one or two of these tasks. Taken together, the results do not support familiarity-based accounts of the revelation effect but are consistent with a proposed decision-based interpretation (i.e., *criterion flux*), in which it is assumed that the revelation task displaces the study list context in working memory, leading subjects to adopt a more liberal recognition decision criterion, thereby increasing the hit and false alarm rates.

Watkins and Peynircioğlu (1990) found that, in recognition memory tasks, when items are preceded by a problem-solving task, such as doing an anagram, those items are more likely to be judged as "old," regardless of whether they have been presented in the study list (targets) or not (lures). This outcome was initially labeled the *revelation effect*, because the problem-solving task involved the probe word that was revealed in the solution to the task prior to the recognition test. Westerman and Greene (1998), though, have shown that the same effect occurs when the revelation task involves a word unrelated to the test probe. The revelation effect is thus quite puzzling since it is not predicted, nor easily explained, by current theories of recognition memory.

The revelation effect has been found with a wide variety of problem-oriented tasks, such as revealing the test word one letter at a time, as an anagram, or rotating the individual letters of a word, or the word as a whole, by varying degrees (Watkins & Peynircioğlu, 1990). Westerman and Greene (1998) also found that the revelation effect can be obtained when item recognition is preceded by letter memory-span tasks, letter-counting tasks, and synonym-generation tasks. The revelation effect occurs regardless of success on the problem-solving task (Westerman & Greene, 1998, Experiment 1), and the size of the effect is not influenced by the time and effort dedicated to the task (Luo, 1993; Peynircioğlu & Tekcan, 1993; Westerman & Greene, 1998). To date, no revelation effect has been found when the type of information involved in the revelation task is different from the type of information used for the recognition test. Westerman and Greene found that simple arithmetic problems and a digit-span task presented before verbal recognition probes did not produce a reliable revelation effect.

Recent studies (Cameron & Hockley, 2000; Westerman, 2000) have demonstrated another boundary condition for the revelation effect. It appears as though the revelation task influences episodic memory decisions based on familiarity, but not decisions that involve recall or recollection of the study episode. The assumption that the revelation task only influences familiarity-based decisions provides an explanation of two aspects of the revelation effect. First, the revelation effect is typically greater for false alarms than for hits (cf. Hicks & Marsh, 1998). This would be expected if false alarms are most often the result of familiarity-based decisions, whereas a proportion of hits are based on recollection rather than on familiarity. Second, Bornstein and Neely (2001) examined the revelation effect for judgments of item frequency, and in two of three experiments, found that the revelation effect became larger as actual frequency increased. Brown (1995) has shown that frequency judgments can be based either on a count of retrieved instances (enumeration) or on a more general familiarity-based strategy (estimation). If, as Brown suggests, subjects use an enumeration strategy more often when frequency is low and use an estimation strategy more often when frequency is high, the revelation effect should increase in magnitude with higher frequency values.

The conclusion that the revelation task influences familiarity-based decisions is also consistent with Wester-

This study was supported by an operating grant from the Natural Sciences and Engineering Research Council of Canada to W.E.H. Part of this study was carried out by M.W.N. for his Honours' thesis. The authors thank Melissa Wells for her help in data collection, Deanne Westerman and Zehra Peynircioğlu for their constructive reviews, and Robert Greene for suggesting Experiment 5, along with several comments that greatly improved this paper. M. W. Niewiadomski is now at the University of Toronto, Scarborough. Correspondence should be addressed to W. E. Hockley, Department of Psychology, Wilfrid Laurier University, Waterloo, ON, N2L 3C5 Canada (e-mail: whockley@ wlu.ca).

man and Greene's (1998) explanation of the revelation effect. Their account is based on the assumptions of global-matching models of memory (e.g., Gillund & Shiffrin, 1984; Hintzman, 1988; Murdock, 1993). In this view, recognition decisions are based on the overall similarity between the test item and the contents of episodic memory. Westerman and Greene further suggest that the revelation task briefly activates additional information in memory that is not activated by the probe itself. This additional activation is summed with the activation produced by the probe, thus increasing the overall activation of test probes preceded by a revelation task.

Hicks and Marsh (1998) have also proposed a familiarity-based account that assumes the revelation effect is due to a shift in the decision criterion. In their view, the revelation task temporarily activates competing alternatives, which in turn reduces the signal-to-noise ratio for the test item. Therefore, instead of increasing the familiarity of the probe, as proposed by Westerman and Greene (1998), Hicks and Marsh (1998) suggested that the problem-solving task reduces familiarity. As a consequence of the increased difficulty of the recognition decision in the revelation condition, subjects adopt a more liberal decision criterion, leading to an increase in the hit and false alarm rates.

The present experiments were designed to test predictions derived from these accounts and, more generally, predictions from any potential familiarity-based explanation of the revelation effect. One prediction concerned the question of whether multiple revelation tasks would have a cumulative effect—that is, would solving two different anagrams have a larger effect than solving only one anagram? Peynircioğlu and Tekcan (1993) showed that the difficulty of the revelation task (see also Westerman & Greene, 1998, Experiments 1 and 2), or the time spent on the task, does not affect the size of the revelation effect. These results could be taken to suggest that there would not be a difference in the size of the revelation effect when the revelation task involves two problems, as compared with one.

In contrast to the findings above, Watkins and Peynircioğlu (1990) found that the size of the revelation effect is correlated with the amount of revelation that was involved in the task. Bornstein and Neely (2001) have also reported a positive relationship between degree of distortion (the number of letter fragments) and the revelation effect for frequency estimation. More difficult revelation tasks might activate more additional information than might easier revelation tasks (cf. Westerman & Greene, 1998, p. 385). This would suggest that two different anagrams might activate more additional information than would one anagram, thereby producing a larger revelation effect.

Since it is not clear what information might be activated by solving an anagram problem, it is also not clear how much additional information, if any, would be activated by a second anagram task, over and above the activation produced by the first anagram task. In an attempt to address this problem, we compared a single-anagram condition with both a double-anagram condition and a condition that involved an anagram problem and a different, unrelated problem task. The second task in this condition was a numeric addition problem. This type of problem is sufficiently different from the anagram task in that it is unlikely that the two tasks would activate similar information.

The use of a numeric addition task would seem to be an odd choice in light of the fact that Westerman and Greene (1998, Experiment 6) failed to find a revelation effect with this type of task. In pilot work, however, we found evidence for a revelation effect when the preceding task involved addition problems. Before proceeding to the test of whether two tasks produce a greater revelation effect than does one, we report an experiment designed specifically to demonstrate that an addition task can produce a revelation effect when the to-be-remembered information is verbal.

EXPERIMENT 1

Method

Subjects. Forty-five introductory psychology students at Wilfrid Laurier University participated in Experiment 1 for course credit.

Apparatus and Stimuli. The study and test 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 to 7 scale for word rating norms of Paivio et al. (1968). The arithmetic problems consisted of the addition of two randomly generated three-digit numbers whose sum did not exceed 999.

Stimulus presentation and response recording were controlled by computers. Specific keys were assigned to specific responses: 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. The subjects voiced their problem solutions into a tape recorder before pressing the "y" key.

Procedure. The subjects were first shown a study list consisting of 60 random words. The words were presented one at a time, in the center of the screen for 3.5 sec, with a 0.5-sec blank interval between them. The first and last six words were considered primacy and recency buffers.

The recognition test comprised 96 probes: 48 were targets from the study list (excluding the buffers) and 48 were new (not studied) lures. The order of the study and test presentations was random, with a different order for each subject. All 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 reminders of the recognition task and the response keys.

Prior to presentation of a random half of the old (24) and half of the new (24) test items, the subjects were presented with the arithmetic addition task. The two numbers were presented on the same line in the center of the screen followed by an equals sign ("=") and a question mark ("?"). The statement "Press top key when solved" appeared at the bottom of the screen. The subjects voiced their solution to each addition problem aloud in the presence of a tape recorder and then pressed the key labeled *problem solved* to proceed. All aspects of the test were self-paced, with a 1-sec blank interval between presentations.

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 words and were told that their memory for

Table 1
Proportions of "Old" Responses for All Conditions in Experiments 1, 2, and 3

Condition	Hits *M*	Hits *SD*	False Alarms *M*	False Alarms *SD*
	Experiment 1			
Intact	.74	.13	.15	.13
Math	.74	.14	.22	.17
	Experiment 2			
Intact	.75	.16	.15	.13
Single anagram	.76	.18	.25	.19
Double anagram	.77	.20	.28	.19
Math–anagram	.75	.20	.27	.19
	Experiment 3			
Intact	.74	.16	.16	.16
Single anagram	.78	.14	.29	.20
Single math	.76	.15	.28	.21
Anagram–math	.80	.16	.31	.20

those words would be tested later. Following the study list, the subjects were asked to respond "old" to the words presented in the study list and "new" to the words that were not shown earlier. The instructions on how to perform the addition task were also given at this time.

Results and Discussion

The mean proportions of "old" responses for all conditions are given in Table 1. These results were submitted to a 2 (old vs. new probe) × 2 (intact vs. math test condition) within-subjects analysis of variance (ANOVA). An alpha level of .05 was used for all statistical tests in this study. Not surprisingly, there was a main effect of probe [$F(1,44) = 471.70$, $MS_e = 0.029$], indicating a greater rate of hits than false alarms. More importantly, there was also a main effect of test condition [$F(1,44) = 4.77$, $MS_e = 0.010$], demonstrating a significantly greater proportion of "old" responses in the math condition than in the intact condition. The two variables interacted significantly [$F(1,44) = 6.30$, $MS_e = 0.081$], showing that the revelation effect was greater for false alarms than for hits.

Potential differences in discriminability and criterion shift were examined by estimating A' and β''_D.[1] The mean estimates are presented in Table 2. A one-way ANOVA was performed on each of these measures. There was a small but reliable decrease in discriminability in the math condition relative to the intact condition [$F(1,44) = 7.54$, $MS_e = 0.003$]. This decrease was accompanied by a significantly more liberal criterion in the math condition [$F(1,44) = 6.52$, $MS_e = 0.016$].

The results of Experiment 1 demonstrate that a revelation effect can be obtained when the preceding task involves addition problems. This finding contrasts with Westerman and Greene's (1998, Experiment 6) failure to observe a significant revelation effect using an addition task, although the pattern of their results was in the same direction as the results of our Experiment 1. The different outcomes of these experiments could be due to differences between the two types of addition tasks; however, more likely, they simply reflect sampling variability.

Having demonstrated that an addition problem can also produce the revelation effect, we now turn to the question of whether two preceding tasks can produce a greater revelation effect than can one task. In Experiment 2, test probes were presented intact or were preceded by one anagram problem, two separate anagram problems, or an addition task followed by an anagram problem.

EXPERIMENT 2

Method

Subjects. Forty-five students participated in Experiment 2 for course credit.

Apparatus and Stimuli. The 140, eight-letter anagrams used in the experiment 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 answer key for the anagrams was presented below every anagram.

Procedure. The experimental procedure employed was nearly identical to that of Experiment 1. Each subject was exposed to four conditions that varied as a function of test probe manipulation. The old and new test probes were presented intact (not preceded by either an anagram or a math problem), preceded by one unrelated anagram task (anagram), preceded by two unrelated anagrams (double anagram), or preceded by an addition problem followed by an anagram (math–anagram). This yielded 12 tests in each condition.

Results and Discussion

The mean proportions of "old" responses for old and new tests in each test condition are presented in Table 1. The proportions of "old" responses were submitted to a 2 (old vs. new probe) × 4 (intact vs. single-anagram vs. double-anagram vs. math–anagram test condition) within-subjects ANOVA.

As expected, hit rates were significantly greater than were false alarm rates [$F(1,44) = 391.27$, $MS_e = 0.061$]. The main effect of test condition was also significant [$F(3,132) = 5.07$, $MS_e = 0.018$], as was the interaction between probe and test condition [$F(3,132) = 5.83$, $MS_e = 0.062$].

To explore the probe × condition interaction, all test conditions were compared with each other, for each probe

Table 2
Mean Estimates of A' and β''_D for All Conditions in Experiments 1, 2, and 3

Condition	A' *M*	A' *SD*	β''_D *M*	β''_D *SD*
	Experiment 1			
Intact	.87	.07	.33	.54
Math	.84	.09	.11	.55
	Experiment 2			
Intact	.87	.08	.25	.56
Single anagram	.84	.09	.02	.70
Double anagram	.83	.09	−.12	.66
Math–anagram	.82	.11	−.04	.60
	Experiment 3			
Intact	.87	.08	.30	.62
Single anagram	.82	.11	−.10	.61
Single math	.82	.12	−.01	.59
Anagram–math	.82	.11	−.21	.58

type, in a series of two-tailed, paired-samples t tests. It was found that no significant differences existed among the test conditions in hit rates. The intact condition, however, did differ significantly from the single-anagram condition in false alarms [$t(44) = 3.98$]. There was also a significant difference in false alarms between the intact and double-anagram conditions [$t(44) = 5.38$], as well as between the intact and math–anagram conditions [$t(44) = 4.71$]. In all three revelation conditions, the false alarm rate was lower for the intact tests than for the other conditions, demonstrating the revelation effect.

Critical to this experiment was the difference between the double-problem conditions and the single-anagram condition. A significantly greater proportion of "old" responses in the double conditions would indicate a potential incremental property of the revelation effect. There was, however, no reliable difference in false alarm rates between double-anagram and single-anagram conditions [$t(44) = 0.383$] or between math–anagram and single-anagram conditions [$t(44) = 0.748$], indicating that no such incremental effect occurred.

These results were further corroborated by signal detection measures. The mean estimates of A' and β''_D for each test condition are presented in Table 2. The A' estimate for the intact condition was significantly higher than for the single-anagram condition [$t(44) = 2.73$], the double-anagram condition [$t(44) = 2.90$], and the math–anagram condition [$t(44) = 4.05$]. The differences between the single-anagram, double-anagram, and math–anagram conditions were not statistically reliable. The same pattern of results was observed in the β''_D estimates. The criterion placement in the intact condition was significantly more conservative than in the single-anagram condition [$t(44) = 2.09$], the double-anagram condition [$t(44) = 3.40$], and in the math–anagram condition [$t(44) = 2.83$]. The double-problem conditions did not reliably differ.

Experiment 2 showed a reliable revelation effect. This effect was greater and significant only for false alarms, a finding consistent with the results of Hicks and Marsh's (1998) meta-analysis of the revelation effect. More importantly, no incremental effects were found, as demonstrated by the lack of a difference between the single-anagram, double-anagram, and math–anagram conditions. As mentioned earlier, an incremental effect would support familiarity-based interpretations of the revelation effect because a familiarity-based effect could exhibit additive properties. Since conclusions drawn from Experiment 2 would be based on a null result, we thought it advisable to replicate the lack of a difference between the one- and the two-problem tasks in Experiments 3, 4, and 5. In Experiment 3, single-anagram and single-math tasks were compared with an anagram-plus-math task.

The design of Experiment 3 also allowed for a comparison of the revelation effect for single-addition and single-anagram problems, as well as for the combination of both tasks. Most familiarity-based accounts of the revelation effect would not predict that the revelation effect for an addition task would be equivalent to the revelation effect for a verbal task, such as anagram solution. This is because a verbal revelation task is more likely to activate information related to the verbal items on the study list than is a nonverbal revelation task.

EXPERIMENT 3

Method

Subjects. Forty-two introductory psychology students participated for course credit.

Apparatus and Stimuli. The apparatus and stimuli were identical to those used in Experiment 2.

Procedure. The procedure was identical to that of Experiment 2, with the exception of the composition of the recognition test conditions. In Experiment 3, the four conditions in the test list were intact, single anagram, single math, and anagram–math.

Results and Discussion

The mean proportions of "old" responses are summarized in Table 1. The analysis performed was a 2 (old vs. new probe) $\times$ 4 (intact, single-anagram, single-math, and anagram–math test condition) within-subjects ANOVA. As expected, there was a main effect of probe [$F(1,41) = 332.31$, $MS_e = 0.067$], as well as a main effect of test condition [$F(3,123) = 13.33$, $MS_e = 0.013$]. These two variables interacted significantly [$F(3,123) = 3.45$, $MS_e = 0.014$]. To evaluate this interaction, the differences in hits and false alarms were examined between all recognition test conditions. There was only one significant difference in hits among all the conditions: the subjects in the anagram–math condition responded "old" to a greater number of old probes than did the subjects in the intact condition [$t(41) = 2.17$]. The same effect between these two conditions was also found for new probes [$t(41) = 5.47$]. Furthermore, the false alarm rate was greater in the single-anagram condition than in the intact condition [$t(41) = 5.44$], indicating a revelation effect. A similar effect was obtained with the single-math manipulation, as demonstrated by comparing the false alarm rate between the intact and single-math conditions [$t(41) = 4.49$].

The mean differences in hit and false alarm rates among the three revelation conditions did not differ reliably. In other words, the mean hit and false alarm rates for single-anagram items were statistically equivalent to the single-math and anagram–math rates. This indicates that all of the problem tasks elicited comparable effects on the probes that followed them.

Mean estimates of A' and β''_D are given in Table 2. Analysis of the A' estimates revealed that discrimination in the intact condition was again higher than that in the single-anagram condition [$t(41) = 2.97$], the anagram–math condition [$t(41) = 2.86$], and the single-math condition [$t(41) = 3.33$]. No other comparisons approached significance. The β''_D analysis indicated significantly more conservative responding in the intact condition than in the single-anagram condition [$t(41) = 4.76$], the anagram–math condition [$t(41) = 5.29$], and the single-math condition [$t(41) = 3.15$]. The only other reliable difference was found in the comparison of the single-math condition with the anagram–math condition, where the criterion was set at a more liberal level in the double-problem condition [$t(41) = 2.45$].

Experiment 3 confirmed that it is possible to obtain a revelation effect with the use of an arithmetic revelation task and, furthermore, that single-addition and single-anagram tasks produce comparable revelation effects. The anagram–math condition did not differ reliably from the single-math or single-anagram conditions in either hit or false alarm rates. Nevertheless, the hit rate was slightly greater in the anagram–math condition, and this was the only revelation condition in which the hit rate was reliably greater than in the intact condition. In Experiments 2 and 3, the subjects performed both anagram and addition tasks. It is possible that performing both types of tasks might have attenuated potential differences between the tasks. In Experiment 4, the anagram and addition problem conditions where compared between subjects. The subjects were presented either with single and double anagrams or with single and double addition problems.

EXPERIMENT 4

Method

Subjects. Ninety introductory psychology students participated for course credit.

Apparatus and Stimuli. The apparatus and stimuli were the same as in the preceding experiments.

Procedure. The subjects were randomly divided into two groups ($n = 45$). The *anagram* group was presented with intact words, single anagrams, and double anagrams throughout the test, and the *math* group was given intact words, single-math problems, and double-math problems. The test lists consisted of 24 old and 24 new intact tests, 12 old and 12 new single-anagram or single-math tests, and 12 old and 12 new double-anagram or double-math tests. All other aspects of this experiment were identical to those of Experiment 3.

Table 3
Proportions of "Old" Responses for All Conditions in Experiments 4 and 5

Condition	Hits *M*	Hits *SD*	False Alarms *M*	False Alarms *SD*
Experiment 4				
Intact (A)	.70	.18	.20	.11
Single anagram	.74	.18	.29	.18
Double anagram	.81	.13	.34	.18
Intact (M)	.76	.13	.19	.13
Single math	.79	.17	.29	.17
Double math	.80	.16	.27	.19
Experiment 5				
Single anagram				
Intact	.70	.17	.25	.15
Revelation	.76	.11	.37	.15
Double anagram				
Intact	.73	.12	.26	.13
Revelation	.78	.14	.37	.15
Single math				
Intact	.70	.15	.23	.17
Revelation	.74	.14	.36	.17
Double math				
Intact	.69	.15	.21	.13
Revelation	.79	.13	.34	.17

Note—A and M correspond to the intact conditions associated with the anagram conditions and the math conditions, respectively, in Experiment 4.

Results and Discussion

The mean proportions of "old" responses for each condition and group are presented in Table 3. The proportions of "old" responses were submitted to a 2 (anagram vs. math group) × 2 (old vs. new probe) × 3 (intact vs. single-problem vs. double-problem recognition test condition) mixed ANOVA. The overall performance in the anagram and math groups did not differ reliably [$F(1,88) < 1$, $MS_e = 0.058$]. There was a main effect of probe [$F(1,88) = 812.52$, $MS_e = 0.042$], indicating greater rates of hits than for false alarms. There was also a significant difference in the type of test condition presented [$F(2,176) = 22.76$, $MS_e = 0.017$]. This variable interacted with probe [$F(2,176) = 3.52$, $MS_e = 0.012$] and with group [$F(2,176) = 3.52$, $MS_e = 0.017$]. There was no interaction between group and probe [$F(1,88) = 2.94$, $MS_e = 0.042$] or between group, probe, and test condition [$F(2,176) < 1$, $MS_e = 0.012$].

To explore the test condition × group interaction, the results for each group were submitted to two separate 2 (old vs. new probe) × 3 (intact vs. single- vs. double-stimulus test condition) within-subjects ANOVAs. In the anagram group, the expected probe main effect was significant [$F(1,44) = 328.93$, $MS_e = 0.046$], as was the main effect for test condition [$F(2,88) = 16.60$, $MS_e = 0.020$]. The two variables did not interact. The false alarm rate was greater in the single-anagram condition than in the intact condition [$t(44) = 3.99$]. The difference in hit rates did not reach significance [$t(44) = 1.29$]. Both the hit rate [$t(44) = 3.38$] and the false alarm rate [$t(44) = 5.53$] were greater in the double-anagram condition than in the intact condition. There was also a significantly greater proportion of hits in the double-anagram condition than in the single-anagram condition [$t(44) = 2.45$], but the difference in false alarms rates was not significant [$t(44) = 1.87$].

In the math group, there was a significant main effect of probe [$F(1,44) = 502.30$, $MS_e = 0.039$], as well as a main effect of test condition [$F(2,88) = 8.08$, $MS_e = 0.014$], with no interaction. The false alarm rates were greater in the single-math condition [$t(44) = 4.45$] and in the double-math condition [$t(44) = 3.84$] than in the intact condition. No reliable effect was found for hit rates, and there was no reliable difference between the single- and double-math conditions (all ts < 1).

To investigate the initial group × test condition × probe interaction, two additional analyses were conducted. First, a 2 (anagram vs. math group) × 3 (intact vs. single-problem vs. double-problem recognition test condition) mixed ANOVA was carried out on old probes. No reliable difference was found between the anagram and math groups. There was, however, a main effect of test condition [$F(2,176) = 6.95$, $MS_e = 0.016$]. The two variables did not interact. The main effect of the test condition was analyzed using two-tailed, paired-samples t tests. A significant difference was found between the intact condition and the double-stimulus condition [$t(89) = 3.59$], as well as between the single- and double-stimulus conditions [$t(89) = 2.02$].

A similar 2 × 3 mixed ANOVA was then carried out on new probes. The results mirrored those of old items, since there was no difference between the anagram and math groups ($F < 1$). The main effect of test condition was significant [$F(2,176) = 24.84$, $MS_e = 0.013$], and the two variables did not interact. The significant main effect was subjected to further analysis using two-tailed, paired-samples t tests. The results indicated that the false alarm rate was significantly lower in the intact condition than in the single-task condition [$t(89) = 5.99$] and also lower than in the double-task condition [$t(89) = 6.63$]. There was no significant difference between the two task conditions [$t(89) = 1.06$].

Mean estimates of A' and β''_D are presented in Table 4. The estimates were analyzed within groups, using simple three-level, one-way ANOVAs. In the anagram condition, there was no difference in discriminability between the intact items, items preceded by a single anagram, and items preceded by two consecutive anagrams [$F(2,88) = 1.66$, $MS_e = 0.01$]. There was, however, a significant difference associated with the criterion-shift estimate, β''_D, [$F(2,88) = 13.75$, $MS_e = 0.21$]. Paired-samples t tests further revealed that the significant result was due to reliable differences between the intact condition and the single-anagram condition [$t(44) = 3.71$], as well as between the intact condition and the double-anagram condition [$t(44) = 4.87$]. In both cases, the β''_D estimate was greater in the intact condition than in the other two conditions, indicating a more conservative response bias. There was no statistical difference between single- and double-anagram conditions [$t(44) = 1.92$].

The same analyses were carried out on the math group, and the results mirrored those of the anagram group. First, the A' estimate was analyzed using a three-level, one-way ANOVA performed on intact probes, single-math, and double-math items. No significant differences were found [$F(2,88) = 2.43$, $MS_e = 0.006$]. The same statistical test performed on the β''_D estimate yielded a significant result [$F(2,88) = 7.61$, $MS_e = 0.194$], which was due to more conservative responding in the intact condition than in both the single-math condition [$t(44) = 3.56$] and the double-math condition [$t(44) = 2.71$]. The two math conditions did not differ [$t(44) = 1.02$].

The results of Experiment 4 showed that the revelation effect was comparable in the single-anagram and single-math conditions. In contrast to Experiments 2 and 3, however, in Experiment 4, the revelation effect was significantly greater for the subjects who solved two anagrams than for those who solved only one. No such cumulative effect was observed in the math conditions.

The results of Experiment 4 differ from those of Experiments 2 and 3 in another regard as well. In Experiment 4, the revelation effect in the double-anagram condition was observed in both hit and false alarm rates, whereas in Experiments 2 and 3, the revelation effect was seen only in the false alarms. The presentation time used in Experiments 2 and 3 (3.5 sec) was longer than in most of the previous experiments that have examined the revelation effect. Landau (2001) has recently shown that the magnitude of the revelation effect is reduced when study time is increased. Of the 32 experiments included in Hicks and Marsh's (1998) meta-analysis of the revelation effect, only three experiments did not find a substantial effect for hits, and the presentation rate in two of these three experiments (LeCompte, 1995) was also longer (3 sec) than in the other experiments.[2] Longer study times may lead to a greater proportion of hits being based on recollection (i.e., retrieval of specific details of the study episode) rather than on familiarity, and decisions based on recollection are not susceptible to the revelation effect (cf. Cameron & Hockley, 2000; Westerman, 2000).

Experiment 5 was designed as a replication of Experiment 4 in which the four revelation task conditions examined in Experiment 4 were all contrasted between subjects. Thus, there were four groups of subjects, and the same task (single or double anagram, or single or double addition problem) was presented in each test list. In addition, the presentation rate for words at study was reduced to 1.5 sec per item in an attempt to increase the effects of the revelation tasks on the hit rates.

Table 4
Mean Estimates of A' and β''_D for All Conditions in Experiments 4 and 5

Condition	A' *M*	A' *SD*	β''_D *M*	β''_D *SD*
Experiment 4				
Intact (A)	.82	.12	.23	.51
Single anagram	.79	.16	−.08	.60
Double anagram	.82	.08	−.28	.65
Intact (M)	.86	.08	.12	.56
Single math	.83	.11	−.23	.62
Double math	.84	.09	−.14	.67
Experiment 5				
Single anagram				
Intact	.80	.11	.11	.53
Revelation	.78	.09	−.27	.41
Double anagram				
Intact	.82	.08	.01	.46
Revelation	.76	.12	−.35	.57
Single math				
Intact	.82	.10	.20	.51
Revelation	.77	.10	−.20	.49
Double math				
Intact	.82	.09	.19	.52
Revelation	.81	.08	−.26	.53

Note—A and M correspond to the intact conditions associated with the anagram conditions and the math conditions, respectively, in Experiment 3.

EXPERIMENT 5

Method

Subjects. A total of 132 introductory psychology students participated for course credit. They were randomly assigned to one of the four experimental conditions.

Apparatus and Stimuli. The apparatus and stimuli were identical to those used in the previous experiments.

Procedure. The procedure was identical to that in Experiment 4, except that the different revelation task conditions were presented between subjects, and the presentation rate at study was reduced to 1.5 sec per item.

Results and Discussion

The mean proportions of "old" responses for all conditions are given in Table 3. These results were submitted to a 2 (old vs. new probe) × 2 (intact vs. revelation task) × 4 (single-anagram vs. double-anagram vs. single-math vs. double-math group) mixed ANOVA. As usual, there was a main effect of probe [$F(1,128) = 861.00$, $MS_e = 0.029$], indicating a greater rate of hits than false alarms. There was also a main effect of test condition [$F(1,128) = 66.35$, $MS_e = 0.005$], demonstrating a significantly greater proportion of "old" responses to probes preceded by a revelation task than to intact probes. The probe variable interacted significantly with test condition [$F(1,128) = 22.34$, $MS_e = 0.008$]. Paired-samples t tests indicated that the interaction was due to a greater proportion of false alarms for revelation probes than for intact probes [$t(131) = 9.71$], as opposed to the hit rates for revelation probes compared with intact probes [$t(131) = 4.17$]. As the t tests indicate, however, a reliable revelation effect was observed for both old and new probes.

The most relevant test in this experiment was the comparison of the revelation effects across the four groups. Finding a greater revelation effect for probes preceded by two tasks than for probes preceded by a single revelation task would provide support for familiarity-based accounts of the revelation effect. No such differences were found. We observed no statistically significant difference between the four conditions [$F(3,128) = 1.31$, $MS_e = 0.037$]. Furthermore, there were no interactions between the conditions and any other variables, indicating that the performance of the four groups was statistically equivalent.

These results are further supported by the analysis of the signal detection estimates A' and β''_D. These means are presented in Table 4. A 2 (A' intact vs. A' revelation task) × 4 (single-anagram vs. double-anagram vs. single-math vs. double-math group) ANOVA revealed only a main effect of revelation on discriminability [$F(1,128) = 16.10$, $MS_e = 0.004$]. The two variables did not interact. Paired-sample t tests revealed that the main effect of the revelation task was due to significantly better discrimination of old items from new ones in the intact condition than in the single-math group [$t(32) = 2.97$] and the double-anagram group [$t(32) = 3.25$]. The difference between the intact and revelation conditions in the double-math group was not significant [$t(32) = 0.567$].

The estimates of β''_D suggested that the subjects adopted a more liberal criterion following a revelation task. A 2 (intact vs. revelation task) × 4 (single-anagram vs. double-anagram vs. single-math vs. double-math group) ANOVA confirmed a main effect of the revelation task on criterion placement [$F(1,128) = 65.81$, $MS_e = 0.027$]. The main effect for group was not significant ($F = 1.03$), nor was the interaction between the two variables. To further explore the main effect of criterion, paired-samples t tests were performed. In all four groups, the criterion was significantly more liberal for items preceded by a revelation task than for intact items.[3]

In contrast to Experiments 1, 2, and 3, in Experiment 5, a revelation effect was observed for both hits and false alarms in all conditions, with the effect being greater for false alarms. Thus, the results of Experiment 5 illustrate the typical revelation effect pattern as defined by Hicks and Marsh's (1998) meta-analysis. Experiment 5 differed from the previous experiments in that the presentation rate at study was reduced to 1.5 sec per item. Considered together, the results of the present experiments are consistent with Landau's (2001) demonstration that the magnitude of the revelation effect is greater when study time is reduced, particularly for hits. This result is in accord with the view that the revelation effect influences only familiarity-based decisions, because reducing study time would decrease the opportunity to encode information that would support recollection and thus increase the proportion of responses based on familiarity.

More importantly, the results of Experiment 5 confirm the two major findings of the previous experiments. First, a revelation effect was found for the numerical addition task, and this effect was similar in magnitude to the revelation effect observed for the anagram task. Second, the revelation effect was comparable when the subjects performed one or two tasks prior to the recognition probe. That is, two tasks do not reliably increase the magnitude of the revelation effect beyond that produced by one task.

GENERAL DISCUSSION

The present study was designed to test predictions derived from familiarity-based accounts of the revelation effect. One prediction that was tested in Experiments 2, 3, 4, and 5 was that, if a problem-solving task influences the familiarity of the test probe, two such tasks could influence familiarity even further. To test this possibility, the effect of solving two separate tasks was compared with the effect of solving only one task on the subsequent recognition probe. Because of the possibility that a second task might not activate very much in the way of additional information that was not activated by the first task (such as might be the case for two anagram solution tasks), we also included two very different tasks (anagram solution and numeric addition) in the double-task condition. It seems reasonable to assume that the activation arising from these two very different tasks would not share very much in common. Nevertheless, in all but one comparison, two revelation tasks yielded the same statistical effect as one task. The exception was the double-anagram condition in Experiment 4. Five other comparisons, however, did not show this outcome. Thus, the effects of preceding problem-solving tasks on subsequent recognition decisions are not generally cumulative in nature.

Although finding a cumulative effect of revelation tasks would provide support for familiarity-based accounts, the failure to observe such an effect is not strong evidence against the familiarity view. It is possible that the additional activation produced by the first task declines during the second problem-solving task, so that the net effect of the activation produced by two problem-solving tasks is approximately equivalent to the activation generated by one task. Although this explanation of our failure to find a consistent cumulative effect of multiple tasks is a possibility, it raises a number of questions concerning the time course of the activation produced by the revelation task. If the activation produced by the problem-solving task decays over time, why then does the amount of time spent on the problem-solving task not affect the size of the revelation effect (Peynircioğlu & Tekcan, 1993; Westerman & Greene, 1998)? If the decline in the activation produced by the revelation task is not due to passive decay but rather is a result of some mechanism or process that resets activation levels, why is the activation level reset between different problem-solving tasks but not reset between the revelation task and the recognition probe?

Another problem for familiarity-based views of the revelation effect is the finding that a numeric addition task also produces a reliable revelation effect. Thus, the revelation effect is more general than previously thought. This result challenges the conclusion reached by Westerman and Greene (1998, Experiment 6), who did not find a reliable effect with their numeric revelation task. It is important to note, however, that their results showed a trend similar to our results in proportion, but their total revelation effect was smaller. It is also puzzling that Westerman and Greene did not find a revelation effect for a number memory-span task (Experiment 7). This is curious, since one would expect that, although this task involves numbers, subjects would presumably rehearse them verbally.

The revelation effect we observed for the addition task was also comparable in magnitude with that of the anagram task. This finding is more problematic for familiarity-based approaches to the revelation effect. In this view, the processes associated with performing the revelation task serve to activate information related to the study list that is not activated by the test probe. This additional activation either increases the familiarity of the subsequent test probe (Westerman & Greene, 1998) or increases the activation of competing information that reduces the signal-to-noise ratio of the test probe (Hicks & Marsh, 1998). Either of these accounts must predict that the activation of information related to a verbal study list should be greater for a verbal revelation task than for an arithmetic task, because a verbal task should be much more likely to activate information related to the memory traces of the verbal items on the study list. Thus, an anagram revelation task should produce a larger effect than should a nonverbal arithmetic task. We have consistently failed to observe such a difference.

The above findings pose a number of problems for familiarity-based explanations of the revelation effect. If we dismiss this approach, how can the revelation effect be explained? The simplest and most parsimonious explanation is that interrupting recognition with an unrelated and irrelevant problem task temporally induces subjects to adopt a more liberal decision criterion, resulting in an increase in hits and false alarms. But why should subjects change their criterion? Hicks and Marsh (1998) have suggested that the revelation task activates competing alternatives or leaves residual noise in working memory (Hicks & Marsh, 1999), which serves to reduce the signal-to-noise ratio of the probe. To compensate for a more difficult decision, subjects adopt a more liberal criterion. The fact that an addition problem has the same effect as an anagram task does, though, pose a problem for the assumption that the revelation task activates competing alternatives, since it is reasonable to expect that the competitors activated by a verbal task would interfere more than would those activated by an addition problem. Both types of tasks could, however, add noise to working memory or displace list-relevant information. The temporary loss of context of the study list might induce subjects to adopt a more liberal criterion for a subsequent test probe.

In the revelation task, subjects are interrupted during the course of the test list. These interruptions might cause subjects to forget their criterion setting or the information or study context that the criterion setting is based on. When the probe is then presented, subjects must try to reestablish the list context and set a decision criterion. We assume that this process is not fully completed for the first recognition decision following the unrelated task, and, faced with a decision of uncertain difficulty, subjects adopt a more liberal criterion as a consequence. When the study context is more fully reinstated for subsequent intact recognition tests, subjects are able to set a more appropriate criterion.

This *criterion-flux* explanation is supported by the β''_D analyses conducted for all the present experiments. Every single condition that induced the revelation effect also exhibited a significant and quite substantial drop in the criterion-placement estimate. Criterion flux can also provide a reasonable account of all the present findings. Both addition and anagram problem tasks would be expected to give rise to the revelation effect, since both tasks would be sufficiently demanding to displace the list-relevant information from working memory. Two problem tasks would not generally result in a larger effect than only one preceding task, because, usually, one task would be sufficient to displace list context from working memory.

The greatest challenge for familiarity-based explanations of the revelation effect is the fact that a wide range of interpolated tasks have been shown to produce the effect. Watkins and Peynircioğlu (1990) first demonstrated that a variety of problem-oriented tasks involving the probe item produce the revelation effect. Westerman and Greene (1998) went on to show that a variety of tasks unrelated to the probe (e.g., memory-span, synonym-generation, letter-counting, anagrams of nonwords) can also produce the revelation effect. The present study extends the generality of the revelation effect to include numeric addition problems. The processes and patterns of activation that are invoked by these quite different tasks do not have any straightforward relation to the probe or, indeed, to the information encoded from the study list more generally. According to the criterion-flux interpretation, any interpolated task that is sufficiently demanding of cognitive resources to displace list-relevant information from working memory should serve to produce the revelation effect. The criterion-flux account, thus, offers a simple and parsimonious alternative to familiarity-based explanations as to why interrupting recognition memory with an unrelated and irrelevant task serves to increase the likelihood of an "old" response to the subsequent test probe.

REFERENCES

Bornstein, B. H., & Neely, C. B. (2001). The revelation effect in frequency judgment. *Memory & Cognition*, **29**, 209-213.

Brown, N. R. (1995). Estimation strategies and the judgment of event frequency. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **21**, 1539-1553.

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.

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.

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review*, **91**, 1-67.

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.

Hicks, J. L., & Marsh, R. L. (1999). *On empirical constraints to theories of the revelation effect in recognition memory*. Unpublished manuscript.

Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace model. *Psychological Review*, **95**, 528-551.

Landau, J. D. (2001). Altering the balance of recollection and familiarity influences the revelation effect. *American Journal of Psychology*, **114**, 425-437.

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.

Murdock, B. B. (1993). TODAM2: A model for storage and retrieval of item, associative, and serial-order information. *Psychological Review*, **100**, 183-203.

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.

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

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 thank Deanne Westerman for pointing this out to us.

3. The paired-sample t test values for β''_D in Experiment 5 were as follows: $t(32) = 4.07$ (single anagram); $t(32) = 3.92$ (double anagram); $t(32) = 4.01$ (single math); and $t(32) = 4.21$ (double math).

(Manuscript received March 21, 2000;
revision accepted for publication July 16, 2001.)