False recognition of associates: How robust is the effect?

AUDREY A. TUSSING and ROBERT L. GREENE
Case Western Reserve University, Cleveland, Ohio

False recognition of nonpresented words that were strong associates of 12 words in a study list was examined. Six lists were read to subjects; each list contained the 12 strongest associates to a critical nonpresented word. False-alarm rates to the 6 critical nonpresented words were obtained under several different conditions. The manipulations included varying the level of processing done to the study lists, varying the recognition-test procedure, repeating each of the study lists three times, and mixing the words from the six study lists together. A reliable false-recognition effect for critical nonpresented words was obtained in all conditions. However, the effect was not impervious to all of the manipulations. Significantly lower false recognition was obtained when learning was incidental as well as when the words on the six lists were mixed together. Neither level of processing nor repetition significantly influenced false recognition. This last result is inconsistent with Hintzman's (1988) MINERVA 2 global memory model, but agrees with predictions from Shiffrin, Ratcliff, and Clark's (1990) SAM model.

The study of recognition memory over the last 10 years has been heavily influenced by a class of models known as global matching models (see, e.g., Gillund & Shiffrin, 1984; Hintzman, 1988; Murdock, 1982). These models posit that recognition of a test item is based on an index of familiarity computed for the item. This familiarity value is calculated by comparing the test item in a parallel fashion with all items in memory. A test item that is similar to several items in memory will have a high familiarity value and most likely will be called "old" on a recognition test. Clark and Gronlund (1996) noted that support for this type of model has come from studies examining false-alarm rates to certain types of nonstudied items. These studies have demonstrated that false alarms reliably occur to test items semantically similar to studied items. Subjects are much more likely to call a new item "old" if that item is related in meaning to several words on the study list. The false-alarm rate generally increases as a function of the number of list items that are semantically related to a distractor (Hintzman, 1988; Shiffrin, Huber, & Marinelli, 1995). This finding has been referred to as the global similarity effect.

The global similarity effect can easily be explained in terms of the global matching models. If memory for a given test item is determined by comparing that item to all studied items in memory, then false alarms should occur for test items that are similar to many items on the study list. Furthermore, as the number of study items sim-

ilar to a test item increases, false alarms for the test item will increase as well. Both of these predictions have been borne out in several studies examining global similarity effects (e.g., Hintzman, 1988; Shiffrin et al., 1995).

Roediger and McDermott (1995) have recently popularized a specific technique for examining global similarity effects on false alarms. The technique, modeled after that of Deese (1959), involves presenting subjects with short lists of related words. Each list is composed of 12 words that are associatively related to a single, nonpresented word. Because the global matching models see the probability of a positive response as being greatly affected by the number of similar items on the study list, they would predict that subjects would score many false alarms when these critical nonpresented words are included on a recognition test. These predictions have been supported by recent studies using the Deese paradigm (e.g., McDermott, 1996; Payne, Elie, Blackwell, & Neuschatz, 1996; Read, 1996; Roediger & McDermott, 1995).

Although a few studies using the Deese (1959) paradigm to study recognition memory have been reported (Payne et al., 1996; Roediger & McDermott, 1995; Schacter, Verfaellie, & Pradere, 1996), little is known about conditions that may influence the rate of false alarms in this paradigm. The present study used the paradigm to examine false alarms for the critical nonpresented words under several different conditions. The first condition involved a procedure similar to that used by Roediger and McDermott (1995, Experiment 1). We chose the six strongest lists from their materials. Each list was read aloud to subjects, who were then given a standard recognition test over all six lists. This condition was called the baseline condition, in that it provided us with a baseline measure of recognition performance for various kinds of items. Another condition involved repeating each of the

This research was conducted while A.A.T. was supported by the National Institute of Child Health and Human Development Mental Retardation Research Training Grant 5-T32-HD07176-16. Correspondence concerning this article should be addressed to A. A. Tussing, Department of Psychology, Case Western Reserve University, 10900 Euclid Avenue, Cleveland, OH 44106-7123 (e-mail: aat2@po.cwru.edu).

six lists three times. The third condition involved combining the six related lists into one longer list in random format, such that related words from each list were separated by other words. McDermott (1996) reported a tendency for false free recall of nonpresented critical items to decrease when the words related to those items were spread apart. The fourth condition was another repetition condition using the single mixed list as opposed to each of the lists in blocked format. McDermott found a decrease in recall of the critical lures when the mixed list was repeated as opposed to when it was presented only once. However, in McDermott's study, repetition of the study list was also accompanied by repetitions of the recall test. We wished to see how repetition of the mixed list affected critical false alarms when subjects were not repeatedly tested on the list.

The fifth, sixth, and seventh conditions involved varying the level of processing done to the lists at study. Some evidence suggests that false memory for the critical nonpresented items may depend on how their corresponding study list is encoded. Read (1996) examined false recall of a critical item when its corresponding study list was processed in one of several ways. One group of subjects was asked to remember the exact ordering of the words, one group was asked to rehearse the words in an elaborate way, and a third group was asked to engage in maintenance rehearsal. Manner of encoding affected false recall of the critical item: false recall was lower in the serial order encoding condition than in either rehearsal condition. However, Read did not find a significant difference in false recall between the elaborate and maintenance rehearsal conditions; maintenance rehearsal produced as much false recall as did elaborate rehearsal. Elaborate rehearsal presumably involves deeper, more meaningful processing than does maintenance rehearsal. If one assumes that deeper processing of the study list causes the critical nonpresented word to "come to mind" more easily than it does when the list is processed shallowly, false recall should be higher in the deep-processing condition. Inconsistently with these predictions, Read did not obtain higher false recall of the critical word in the elaborate condition. Our fourth, fifth, and sixth conditions were designed to further explore the effects of level of processing on false recognition.

In the eighth condition, subjects took six separate recognition tests; each of these tests corresponded to one of the study lists. Each test was given immediately after its corresponding study list was presented. We wished to see if subjects were still susceptible to false recognition when they were tested immediately after each list.

METHOD

Subjects

All subjects were students taking introductory psychology courses at Case Western Reserve University. They participated in order to fulfill a course requirement.

Materials

The study material consisted of six of the lists reported in Roediger and McDermott's (1995) Appendix. These stimuli were originally taken from Deese (1959) but contained a few word substitutions deemed more appropriate by Roediger and McDermott. Each list consisted of 12 words designed to elicit a critical nonpresented word. For each critical nonpresented word, the 12 strongest associates were used. The 12 words were arranged in terms of associative strength, with the strongest associate to the critical nonpresented word in Position 1 and the weakest associate in Position 12. Hence, the following study list would correspond to the critical word SLEEP: *bed, rest, awake, tired, dream, wake, snooze, blanket, doze, slumber, snore, nap*. The five other lists consisted of the 12 strongest associates of the critical nonpresented words CHAIR, MOUNTAIN, NEEDLE, ROUGH, and SWEET. The order of presentation of the lists remained constant across conditions.

The same recognition test was used in seven of the eight conditions. The test consisted of 66 items, 42 of them being distractors taken from the Thorndike and Lorge (1944) word-frequency norms. High-frequency (50 or more occurrences per million) abstract words were chosen as unrelated distractors. Eighteen of the test items were words from each of the study lists. Three words from each of the six lists were chosen according to serial position; words occurring in Positions 1, 8, and 10 of each list were used. The remaining six test items were the critical nonpresented words corresponding to each list. The test was given in a written format whereby each test item was listed randomly on a sheet of paper, with the restriction that words from the same list had to have at least three other words between them. To the right of each word were the choices "old" or "new"; the subjects were required to circle one of the choices. Two of the critical lures appeared on the test before the three instances of their related study words; four of the lures appeared between instances of their related words.

In the eighth condition of the experiment, subjects were given a separate recognition test immediately after each list. Each of the six tests was constructed in the same fashion. Each test consisted of 11 items, 7 of them being distractors taken from the same pool as the distractors on the longer recognition test. In other words, the same set of 42 distractors used in the 66-item recognition test was used in these six tests, with 7 distractors assigned randomly to each of the six tests. The same studied items used in the longer test were also used in each shorter test; words from Positions 1, 8, and 10 were used in each test. The remaining item in each of the six tests was the critical nonpresented word corresponding to each study list. All 11 items in each test were arranged randomly on a sheet of paper in the same format as the items in the longer recognition test. The critical lure appeared between its corresponding study items on four tests, before its study items on one test and after its study items on one test.

Procedure

Baseline condition. Twenty-nine subjects were seated in a small testing room and were given a pencil and blank sheet of paper. The instructions for the experiment were read aloud to them. They were informed that they would be hearing six lists of 12 words each, and that their memory for the lists would be tested after the last list was read. The nature of the memory test (recognition) was not specified. The subjects were then told that, prior to being given the test, they would be required to complete a 45-sec arithmetic task. They were informed that a number would be read aloud to them and that they should write it down. They would then be required to count backward from the number by a specified amount, writing down as many numbers as they could until told to stop. After the instructions were given, the experimenter read the lists aloud at the rate of one word every 1.5 sec. The experimenter indicated the beginning of each list by saying "List 1," "List 2," and so forth. All lists were read in succession with no pausing between lists. After the last list was read, the subjects were given the arithmetic task. Following this, they were given the 66-item recognition test and were given as much time as they needed to complete it.

Repetition condition. Twenty-five subjects were given the same instructions as those in the baseline condition, with one exception. Instead of being told that they would hear 6 short lists of words, subjects

Table 1
Mean Performance Measures for the Eight Conditions

Condition	CFA	FA	Hits	A'	B''_D	$FA-A'$
Baseline	.90	.07	.80	.92	.53	.96
Repetition	.91	.04	.86	.95	.57	.97
Mixed list	.74	.14	.83	.91	.12	.87
Mixed repetition	.67	.09	.90	.95	.14	.90
Pleasantness rating	.82	.13	.89	.93	.00	.91
Letter counting	.73	.18	.91	.92	-.34	.86
Vowel judgment	.81	.15	.85	.91	.04	.91
Immediate test	.93	.03	.91	.97	.37	.97

Note—CFA = critical false alarms; FA = unrelated false alarms. A' = a measure of accuracy in recognition performance; B''_D = a measure of subjects' criterion placement; $FA-A'$ = a measure of discrimination between critical and unrelated false alarms.

in this condition were told that they would hear 18 short lists of words. They were not told that the 18 lists were actually 6 lists repeated three times each. Prior to reading the lists, the experimenter informed the subjects that their memory for the words would be tested. Instructions for the arithmetic task were then given. The experimenter read each list aloud at a rate of 1.5 sec per word after indicating the beginning of each list by saying "List 1," "List 2," and so on. Once the sixth list was read, the experimenter went back to the first list and read it again, this time calling it "List 7." The lists were all repeated in the same order, and the experimenter again returned to the first list, this time calling it "List 13." The lists were read a third time in the same order. After the 18th list was read, the subjects completed the arithmetic task. Following that task, the subjects were given the same recognition test used in the baseline condition.

Mixed-list condition. Thirty-five subjects were given instructions similar to those given in the previous two conditions, except that they were told that they would hear a single list of 72 words. This 72-word list consisted of the six lists (of 12 words each) used in the other conditions. The order of the words in this longer list was randomized, such that the related words from each of the six smaller lists were spread apart. Words from the same list had to be separated by at least three other words. The categorical nature of the 72-word list was not revealed to the subjects. The list was read aloud at a rate of 1 word every 1.5 sec. After hearing the list, the subjects completed the arithmetic task and then took the same recognition test used in the other conditions.

Mixed-repetition condition. This condition was similar to the first repetition condition except that the presentation format of the six study lists was modified. Instead of presenting each individual list three times (blocked format), we presented a single list in which the words of the six lists were mixed together (random format). This list was identical to that used in the mixed-list condition. Twenty-three subjects were told that they would be hearing three lists of 72 words each, and that they should try to remember them for a memory test to be given afterwards. As in the first repetition condition, they were not told that they would actually be hearing repetitions of the same list. Following these instructions, the experimenter read aloud to them the 72-word list, calling it "List 1." After the last word in the list was read, the experimenter returned to the beginning of the list, this time calling it "List 2" and reading it again. This procedure was followed through the third repetition of the list. After the list was read a third time, the subjects completed the arithmetic distractor task and then received the 66-item recognition test.

Pleasantness-rating condition. Twenty subjects served in this condition. The procedure was similar to that of the baseline condition except that different instructions were given before the lists were read. The subjects were told that they would hear six short lists of words; they were not told that their memory for the words would be tested. Instead, they were required to do a semantic orienting task during study. The subjects were given a sheet of paper consisting of 72 blank lines, which were broken down into six blocks of blank lines numbered 1–12. The title of each of the six lists ("List 1," "List 2," etc.) directly preceded each block. A 7-point pleasantness scale was printed at the top of the page. As each word was read, the subjects were instructed to rate how

pleasant each word seemed to them and to write a number from the scale on the blank line corresponding to each word. The words were read at the 1.5-sec rate. After the last word was read, the subjects completed the arithmetic task and were then given the recognition test on the words.

Letter-counting condition. Twenty-five subjects were told that six short lists of words would be read aloud to them. They were told that immediately after each word was read, they were to count how many letters each word consisted of. Once they had determined the number of letters, they were to write the number on the appropriate blank line on the answer sheet. The answer sheet was very similar to the one used in the semantic-processing condition. Seventy-two blank lines corresponded to each of the words presented. The words were read at the 1.5-sec rate. The subjects were not told that their memory for the words would be tested. After the last word was read, the subjects completed the arithmetic task and then received the 66-word recognition test.

Vowel-judgment condition. The procedure in this condition was similar to that used in the letter-counting condition. Twenty-five subjects were told that six short lists of words would be read aloud to them. However, this time they were told that immediately after each word was read, they were to determine whether or not the first letter was a vowel. They were then told to write their response ("Y" for *yes* and "N" for *no*) on the appropriate line of their answer sheet. Seventy-two blank lines corresponded to each of the words presented. The words were read at a 1.5-sec rate. The subjects were not informed that their memory for the words would be tested. After the last word was read, they completed the arithmetic task and then received the 66-word recognition test.

Immediate-test condition. Twenty-two subjects were told that six short lists of words would be read aloud to them. They were then informed that their memory for each list would be tested immediately after the list was read. Instructions for the arithmetic task were given, and the subjects were told that they would be required to do the task before completing each of the six recognition tests. The words in each list were read at the 1.5-sec rate. After the first list was read, the subjects performed the arithmetic distractor task for 45 sec (this task was identical to that used in the other conditions). Following the distractor task, the subjects were given the 11-item recognition test corresponding to List 1. After everyone had completed the test, List 2 was read aloud. The subjects then performed another arithmetic distractor task after which they took the recognition test corresponding to List 2. This procedure was followed through all six lists.

RESULTS

Table 1 contains measures of recognition performance for each of the eight conditions. The first three columns of the table list the proportions of critical false alarms (i.e., positive responses to the critical nonpresented items), false alarms to unrelated distractors, and hits.

A significant false-recognition effect was obtained in all eight conditions of the experiment, as evidenced by a

greater number of false alarms for critical nonpresented words than for unrelated distractors. The smallest difference in false alarms between critical and unrelated distractor items was found in the mixed-repetition condition where the critical false-alarm rate (67%) still greatly exceeded the unrelated false-alarm rate [9%; $t(22) = 10.08$, $p < .001$].

The crucial analyses involved comparing performance across the eight conditions. All eight conditions were conducted on samples from the same population, although two conditions (the vowel-judgment and mixed-repetition conditions) were tested approximately 6 months after the others. Critical false alarms, unrelated false alarms, and hits in the conditions were compared using analyses of variance. A significant main effect of condition was found for each measure: critical false alarms [$F(7,196) = 5.27$, $MS_e = .04$], unrelated false alarms [$F(7,196) = 5.39$, $MS_e = .01$], and hits [$F(7,196) = 5.27$, $MS_e = .01$]. Our main interest was in how critical false alarms in each of the seven other conditions compared with those in the baseline condition. The three lowest rates of critical false alarms were found in the mixed-list, mixed-repetition, and letter-counting conditions. These rates were found to be significantly less than the critical false-alarm rate in the baseline condition using Newman-Keuls post hoc tests. The letter-counting, mixed-list, and mixed-repetition conditions can be differentiated from the other conditions in another way. Only in these three conditions were there significantly more hits than critical false alarms [$t(24) = -4.36$, $p < .001$ (letter-counting condition); $t(22) = 3.54$, $p < .01$ (mixed-repetition condition); and $t(34) = 2.35$, $p < .01$ (mixed-list condition)], indicating again that false recognition of critical distractors was reduced in these conditions.

A surprising result was the high hit rate found in the letter-counting condition (91%). This high hit rate was accompanied by a high rate of false alarms to unrelated distractors (18%), suggesting that some correction of performance might be in order. The fourth and fifth columns in Table 1 report measures of memory accuracy and response criterion for each of the conditions. The accuracy measure was A' , which ranges between 0 and 1, with a value of .5 representing chance performance (Donaldson, 1992). This measure was calculated using the hit rate to presented items and the false-alarm rate to unrelated distractors. Note that performance was quite high in all conditions using this measure. Nevertheless, performance did differ significantly between the conditions [$F(7,196) = 4.90$, $MS_e = .01$]. Not surprisingly, the immediate-test condition (where accuracy would be helped by both a shorter retention interval and a shorter list length for each test) and the first blocked repetition condition led to the highest accuracy; interestingly, these conditions also led to high levels of false alarms to the critical related distractors. Therefore, high accuracy in discriminating between old items and unrelated distractors does not necessarily imply an ability to discriminate between old items and related distractors. Accuracy did not differ significantly between the deep- and shallow-

processing conditions; presumably, the fast presentation rates used did not allow subjects sufficient time to carry out the processing that would have led to the typical depth-of-processing effect.

The criterion measure reported is B'_D (Donaldson, 1992), which ranges from 1 to -1 , with zero indicating no bias and negative numbers indicating a liberal response bias (e.g., a tendency to give positive responses). This criterion measure was calculated using the proportions of positive responses to presented items and to unrelated distractors. Criteria differed significantly among the conditions [$F(7,196) = 6.48$, $MS_e = .36$]. Subjects in most conditions were generally conservative in their responding but adopted a liberal response criterion in the letter-counting condition; this liberal criterion undoubtedly played a role in this condition's surprisingly high proportion of hits and relatively high proportion of false alarms to unrelated distractors.

One other measure of performance supports the finding of reduced false recognition in certain conditions. For each condition, we calculated a measure of false-alarm discrimination, called $FA-A'$. This measure was calculated in the same way as A' (overall accuracy of performance), except that the critical false-alarm rate was used in place of the hit rate. Specifically, $FA-A'$ reflects the degree to which subjects give more "old" responses to the critical lures than to the unrelated lures. $FA-A'$ can take on any value between 0 and 1, with values close to 1 indicating the tendency to give more "old" responses to critical lures than to unrelated lures. $FA-A'$ differed significantly across conditions [$F(7,196) = 8.49$, $MS_e = .01$]. Newman-Keuls tests revealed that $FA-A'$ was significantly lower in five of the eight conditions than in the baseline condition. Subjects in the pleasantness-rating, letter-counting, vowel-judgment, mixed-list, and mixed-repetition conditions were less likely to say "old" to the critical lures than subjects in the other conditions. The differences in $FA-A'$ provide further evidence of decreased false recognition in those conditions.

Some caution should be exercised in interpreting the comparisons of each condition with the baseline condition because some comparisons involve several differences in procedure. For example, the immediate-test condition differed from the baseline condition in both the retention interval and the testing procedure. Six separate tests were used, and the retention interval was shorter. The levels-of-processing conditions varied from the baseline condition in terms of both how the material was encoded as well as whether or not subjects expected a memory test. Although we do not believe that these differences in procedure compromise the validity of our results, caution should nonetheless be exercised in their interpretation.

DISCUSSION

Several empirical conclusions can be based on these data. One conclusion is that false recognition is a very robust phenomenon; there were far higher false-alarm rates to the critical (related) distractors than to the unrelated distractors in all eight conditions. However, the level of false recognition (as defined by high false-alarm rates to the

related distractors) can be manipulated. False recognition was lower when the words in each set of associates were spread apart than when they were presented together; this was true whether the words were presented once or three times and replicates a (nonsignificant) trend in the free-recall study of McDermott (1996).

False recognition also seemed to be lower in those conditions where the study lists were learned incidentally. The proportion of false alarms to related distractors was significantly lower in the letter-counting condition than in the baseline condition. When analyses were done on the *FA-A'* measure, which reflects differential responding to related and unrelated distractors, false recognition was lower in all three of the incidental-learning conditions (letter-counting, vowel-judgment, and pleasantness-rating) than in the baseline condition. There are several possible explanations for this pattern. One possibility is that false recognition is enhanced by rehearsal processes; if subjects do not see a need to rehearse, this may reduce false recognition. Alternatively, given the rapid presentation rates that were used here, the incidental orienting tasks may have functioned as distractor activities that competed in some way with processes that contributed to false recognition.

Two of the null results merit further discussion. Level of processing here did not influence false recognition, as shown by the failure of the pleasantness-rating (semantic) condition to differ from either the letter-counting or vowel-judgment conditions. In this study, interpretation of this null result is complicated by the failure to observe the standard level-of-processing effect on hit rates. However, it should be noted that Read (1996) found equivalent effects of maintenance and elaborative rehearsal on false recall. At least, it is clear that semantic processing is not required for substantial false recognition to occur.

Repetition also failed to affect false recognition. There are two relevant comparisons here, namely, the repetition condition compared with the baseline condition and the mixed-repetition condition compared with the mixed-list condition; in neither case did repetition lead to an increase in false recognition. Like all null results, our findings must be viewed cautiously. However, it should be noted that our data resemble earlier findings reported by Shiffrin et al. (1995), who found that false alarms were influenced by the number of related words presented on a study list but not by the number of times each of those related words was presented.

This null result is somewhat difficult for most theoretical accounts of false recognition to explain. For example, Roediger and McDermott (1995) noted that the concept of implicit associative responses (Underwood, 1965) could explain false memory in this paradigm. This concept assumes that the perception of a study-list word could bring related words to mind and that subjects later have difficulty in discriminating between list words and these implicit associative responses. Still, one would have imagined that presenting the study list three times would triple the opportunity for implicit associative responses to occur. This should have led to an increase in false recognition, which it failed to do.

Global matching models assume that recognition decisions are based on the total similarity of a test item to the contents of memory. Because the critical distractors are related to many studied items, these models could explain the high levels of false recognition observed here. However, most such models would also predict that repetition of the list words would lead to an increase in critical false alarms. One of the clearest such cases is Hintzman's (1988) MINERVA 2 model, where repetitions of items lead to the storage of multiple memory traces; because repetition of studied items would increase the number of traces that overlap in features with the critical distractors, this model would predict that repetition would increase false recognition. However, one model that would not make this faulty prediction is the Search of Associative Memory (SAM)

model of Gillund and Shiffrin (1984), as revised by Shiffrin, Ratcliff, and Clark (1990). In this model, Shiffrin et al. assume that repetition of a list item merely strengthens the trace of that item. The probability of a false alarm to a particular distractor is influenced by the number of related traces that are contained in memory but not by the strength of these traces; indeed, the claim that the strength of one item does not affect recognition of other items appears to be one of the developing generalizations of memory research (Murnane & Shiffrin, 1991; Ratcliff, Clark, & Shiffrin, 1990; Shiffrin et al., 1995).

REFERENCES

- CLARK, S. E., & GRONLUND, S. D. (1996). Global matching models of recognition memory: How the models match the data. *Psychonomic Bulletin & Review*, **3**, 37-60.
- DEESE, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology*, **58**, 17-22.
- DONALDSON, W. (1992). Measuring recognition memory. *Journal of Experimental Psychology: General*, **121**, 275-277.
- GILLUND, G., & SHIFFRIN, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review*, **91**, 1-65.
- HINTZMAN, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review*, **95**, 528-551.
- MCDERMOTT, K. B. (1996). The persistence of false memories in list recall. *Journal of Memory & Language*, **35**, 212-230.
- MURDOCK, B. B. (1982). A theory for the storage and retrieval of item and associative information. *Psychological Review*, **89**, 609-626.
- MURNANE, K., & SHIFFRIN, R. M. (1991). Word repetitions in sentence recognition. *Memory & Cognition*, **19**, 119-130.
- PAYNE, D. G., ELIE, C. J., BLACKWELL, J. M., & NEUSCHATZ, J. S. (1996). Memory illusions: Recalling, recognizing, and recollecting events that never occurred. *Journal of Memory & Language*, **35**, 261-285.
- RATCLIFF, R., CLARK, S. E., & SHIFFRIN, R. M. (1990). List-strength effect: I. Data and discussion. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **16**, 163-178.
- READ, J. D. (1996). From a passing thought to a false memory in 2 minutes: Confusing real and illusory events. *Psychonomic Bulletin & Review*, **3**, 105-111.
- ROEDIGER, H. L., III, & MCDERMOTT, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **21**, 803-814.
- SCHACTER, D. L., VERFAELLIE, M., & PRADERE, D. (1996). The neuropsychology of memory illusions: False recall and recognition in amnesic patients. *Journal of Memory & Language*, **35**, 319-334.
- SHIFFRIN, R. M., HUBER, D. E., & MARINELLI, K. (1995). Effects of category length and strength on familiarity in recognition. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **21**, 267-287.
- SHIFFRIN, R. M., RATCLIFF, R., & CLARK, S. E. (1990). List-strength effect: II. Theoretical mechanisms. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **16**, 179-195.
- THORNDIKE, E. L., & LORGE, I. (1944). *The teacher's word book of 30,000 words*. New York: Teachers College Press, Columbia University.
- UNDERWOOD, B. J. (1965). False recognition produced by implicit verbal responses. *Journal of Experimental Psychology*, **70**, 122-129.

(Manuscript received October 27, 1996;
revision accepted for publication April 30, 1997.)