Journal of Experimental Psychology:
Learning, Memory, and Cognition
1992, Vol. 18, No. 2, 345-355

0278-7393/92/$3.00

Tests of the List-Strength Effect in Recognition Memory

Andrew P. Yonelinas
McMaster University
Hamilton, Ontario, Canada

William E. Hockley
Wilfrid Laurier University
Waterloo, Ontario, Canada

Bennet B. Murdock
University of Toronto
Toronto, Ontario, Canada

The list-strength effect arises when increasing the strength of some items in a list reduces memory for the remaining items. Here the list-strength effect was investigated under conditions of rapid visual presentation. Randomized and blocked formats were used for the mixed lists. Performance was measured with both yes–no and forced-choice recognition procedures. Overall no evidence for a list-strength effect in recognition was found except under conditions that may promote reverse rehearsal borrowing. Two experiments were conducted to determine why performance on the yes–no tests was greater than on the forced-choice tests. We found that repeated testing with the yes–no procedure promoted more effective encoding than the forced-choice procedure.

Does the strength of an item's competitors in a study list affect recognition for that item at test? One would expect that recognizing a given item would be easier when its competitors in the study list were on average weaker compared with when its competitors were stronger. This effect has been called the list-strength effect (Ratcliff, Clark, & Shiffrin, 1990).

A list-strength effect has been found with free recall (Tulving & Hastie, 1972). However, no such effect has been found for cued recall or recognition. Tulving and Hastie (1972, Experiment 1) showed that strengthening certain items in a list by repeating them reduced free recall for the remaining items in the list when the total number of list items was held constant. Ratcliff et al. (1990) also reported a list-strength effect for free recall when items were strengthened by increasing presentation duration at study. However, in several experiments, they did not find a similar effect for recognition or cued recall.

The list-strength effect is of particular interest in the case of recognition. Shiffrin, Ratcliff, and Clark (1990) argued that all the current global memory models (e.g., the MINERVA2 model of Hintzman, 1986; the SAM model of Gillund & Shiffrin, 1984; the TODAM model of Murdock, 1982; and the matrix model of Humphreys, Pike, Bain, & Tehan, 1989) predict a list-strength effect. That no such effect has been found seems to pose a problem in all these models. (However, see Murdock, 1991, for an alternative interpretation of the predictions of global memory models.)

The present experiments examined the effect of list strength on item recognition using rapid presentation rates. As in Ratcliff et al. (1990), the list-strength hypothesis was tested by presenting a study list that contained items that were stored with different levels of strength; this was called the mixed list. The strength with which an item was stored was controlled by varying the presentation duration at study. In the simplest case (Experiments 4, 5, and 6), there were only two different levels of strength; weak items were presented for a short duration, whereas strong items were presented for a longer duration. Performance on the mixed list was compared with two control conditions; these were called the pure lists. One pure list contained only strong items, and the other contained only weak items. If a list-strength effect is present, memory for strong items should be better in a mixed list than in a pure list because its competitors are on average weaker in the mixed list. Conversely, memory for weak items should be worse in the mixed list because its competitors are on average stronger in the mixed list.

A potential problem for the experiments of Ratcliff et al. (1990) is rehearsal borrowing. If rehearsal borrowing was not adequately controlled, it would have been possible for subjects to borrow rehearsal time from the stronger items and redistribute this rehearsal time to the weak items. This would decrease the difference between the strong and weak items and thus reduce any list-strength effect. Although Ratcliff et al. used a variety of means to prevent rehearsal borrowing, all of their experiments used slow presentation rates, which may have confounded their efforts. Single items were never presented for less than 1 s, and pairs of items were presented for durations as long as 6 s. Such a procedure provided more than enough time for rehearsal borrowing.

Murnane and Shiffrin (1991) attempted to test for the existence of rehearsal borrowing by using a final recognition test. They argued that if rehearsal borrowing masked a real list-strength effect, a negative list-strength effect should show up on a final recognition test. However, as Murdock (1991) pointed out, a final recognition test may not provide crucial evidence of rehearsal borrowing.

In the present experiments, the items were studied under conditions of rapid visual presentation to reduce effects of

This research was supported by Grant APA 146 to Bennet B. Murdock and Grant OGPIN 012 to William E. Hockley.

We thank the reviewers for many helpful comments and Roger Ratcliff for suggesting (insisting on) Experiment 6.

Correspondence concerning this article should be addressed to Andrew P. Yonelinas, Department of Psychology, McMaster University, Hamilton, Ontario, Canada L8S 4K1.

rehearsal strategies. Presumably rehearsal strategies take time. If words are presented at such a rapid rate that subjects only have enough time to read each word, there is no time left over for rehearsal. Presentation rates similar to those used in the current experiments were used by Crowder and Neath (1991) to eliminate the effects of rehearsal. A further advantage to using rapid presentation rates is that the list-strength effect, if it exists, should be easier to detect at faster presentation rates. Loftus (1974) showed that differences in presentation rate with rapid visual presentation have a greater effect on d' than those same differences at a slower presentation rate. Because the list-strength effect is measured by the interaction between item strength and type of list, increasing the differences in strength should increase the effect.

A disadvantage to using a rapid presentation rate is that it changes the basic conditions because the previous research in the area has been conducted using slower presentation rates. However, to test the list-strength effect, one requires only that items be stored at different levels of strength and not that they be stored at or above any absolute level of strength. Furthermore, experiments on rapid visual presentation of words show that recognition memory increases continuously with presentation duration for rates of 25 to 500 ms/item (Loftus, 1974). It is also important to note that none of the current models of recognition memory restrict their domain of application to a limited range of presentation rates. Weber (1988) applied the TODAM model to these presentation rates, and at least one of the current models does make explicit predictions about rapid presentation rates. The most current version of the SAM model appears to predict a list-strength effect under conditions of rapid presentation while predicting no effect at slower presentation rates (see Shiffrin et al., 1990, Figure 1).

The first experiment examined the effects of rapidly presented, randomly mixed lists on yes–no recognition performance. Lists of items were presented in a mixed format such that each list contained a number of item strengths mixed in a random order. In this way, the strength of an item could not be predicted by the strength of its predecessor, and encoding strategies could not be appropriately adjusted. Subsequent experiments investigated the list-strength effect by comparing performance on mixed lists to that of pure lists for both yes–no and forced-choice recognition and for both random and blocked mixed-list formats. The final two experiments were conducted to determine why in the earlier experiments performance was always greater when the yes–no procedure was used than when the forced-choice procedure was used.

Experiment 1

We began our investigation by examining recognition memory for rapidly presented items in which items of different strengths were randomly mixed within each list. Three different levels of item strength were tested: Strong items were presented for a long duration, weak items were presented for a short duration, and medium items were presented for an intermediate duration. Recognition was tested with a yes–no procedure. The purpose of the experiment was twofold. First, it was necessary to determine if differences in presentation duration at rapid presentation rates had a significant effect on recognition performance. This point is critical for the later tests of the list-strength effect because the effect is essentially the difference between the differences of strong to weak items in the mixed and pure lists. If the different levels of item strength are not sufficiently different from each other, there is no hope of detecting the list-strength effect. The second aim of this experiment was to determine if rehearsal borrowing was effectively reduced with the mixed-list procedure. If the weak items were borrowing rehearsal time from the strong items, the list-strength effect would be diminished. Such rehearsal borrowing should be most evident for adjacent items in the study list. If one is to increase the amount of rehearsal time for any given item, it is most likely that the rehearsal time would be borrowed from the adjacent items—either the preceding item or the following item—in the study list. If there is rehearsal borrowing present, then the strength of the prior item and the strength of the succeeding item should influence performance. This technique was also used by Murnane and Shiffrin (1991) to investigate rehearsal borrowing in mixed lists.

Method

Subjects. Seventeen subjects participated in the experiment. Subjects were enrolled in an introductory psychology course at the University of Toronto and received credit for participating.

Materials. Nine hundred sixty words were chosen randomly from the Toronto pool of 1,024 words for each session. Sixteen study lists containing 30 words each were selected, and the remaining words served as test lures.

Design. The experiment consisted of 16 successive recognition tests. Each test was made up of a study list of 30 words followed by a yes–no recognition test for words in that list. The strength of each word was controlled by varying its presentation duration. There were three levels of item strength: strong, medium, and weak. Each list contained 10 strong words, 10 medium words, and 10 weak words presented in a random order.

Procedure. The subjects were tested individually. Items were presented on a monochrome green monitor, and responses were collected on an IBM PC. Words appeared in uppercase letters in the center of the screen. The character size of the stimuli was approximately 5 × 5 mm. Viewing distance was approximately 0.5 m. Subjects were informed that they would be given 16 short recognition tests on a computer and were presented with a sample study list to familiarize them with the procedure.

Subjects received 16 recognition tests in one session. Thirty words were presented in each study list one at a time on a computer screen with a 50-ms mask (a row of asterisks) after each word. Each study list contained 10 strong items (words presented for 200 ms), 10 medium items (words presented for 100 ms), and 10 weak items (words presented for 50 ms). Items in each list were presented in a different random order. The rapid presentation rates were chosen to reduce the time available for rehearsal strategies. The fastest presentation rate, which was 10 words per second, was above the perceptual threshold, which is about 12 words per second (Potter, 1984).

Immediately after each list, a yes–no recognition test was carried out. The 30 list items and 30 new items were tested in a random order. Subjects responded at their own pace, pressing a key marked "yes" if they recognized the word and a key marked "no" otherwise. Each new item appeared 250 ms after the prior response. When the test was complete and subjects were ready for the next study list, they pressed an assigned key, and the presentation of the next list began.

Table 1
Yes–No Recognition Performance in Experiment 1

Strength	Hit rate	False alarm	d'	Response time (ms)
Strong (200 ms)	0.510	0.195	1.057	1,203
Medium (100 ms)	0.375	0.195	0.527	1,228
Weak (50 ms)	0.282	0.195	0.310	1,228

Results and Discussion

Table 1 presents accuracy and latency results for yes–no recognition. Hit rate, false-alarm, d', and average response time for old items are given for each condition. The d' values presented are based on the average d' for each subject condition for each list. Because it was a mixed-list design, the same false-alarm rate had to be used for all conditions. Analysis was based on d' and response time for old items calculated for each subject condition for each list.

The item strength (presentation duration) significantly affected d'. Performance increased with item strength, $F(2, 32) = 35.36, p < .001, MS_e = 2.03$. Reaction time was not affected by item strength ($F < 1$) and was not further analyzed.

Rehearsal borrowing was investigated by examining the effects of the strength of an item's predecessor in the study list. Table 2 presents performance as a function of item strength and strength of the preceding item in the study list. The strength of the preceding item did not significantly affect performance, $F(2, 32) = 2.12, p > .05, MS_e = 0.06$. However, in several cases, performance on an item was greater when the preceding item was weaker and poorer when the preceding item was stronger. A similar analysis was performed on the strength of an item's successor in the study list and revealed that the strength of the succeeding item did not significantly affect performance ($F < 1$). Although there was no significant borrowing between adjacent items in the study list, the possibility remains that items are borrowing from more distant items. To address this possibility, we analyzed the effects of the item two items before each item, but again we found no evidence of borrowing ($F = 1.50$).

To determine if there were systematic changes over study–test trials, a two-way (Study List × Strength) analysis of variance (ANOVA) was performed on d'. There was no evidence of any such effects. The effect of study list was not significant, $F(15, 768) = 1.42, p > .05, MS_e = 0.86$, nor was the Study List × Strength interaction significant, $F(30, 720) = 1.02, p > .05, MS_e = 0.85$.

Table 2
Yes–No Recognition Performance (d′) as a Function of an Item's Strength and the Strength of the Preceding Item in the Study List for Experiment 1

Item strength	Strength of predecessor		
	Weak	Medium	Strong
Weak	0.36	0.28	0.25
Medium	0.62	0.57	0.57
Strong	1.04	0.87	0.97
Mean	0.67	0.57	0.60

Table 3 presents performance as a function of input position and output position. Performance is greatest for the last items studied, next best for the initial items, and worst for the intermediate items. Performance declines gradually with output position. These findings are typical of most experiments of recognition memory.

This experiment shows that for mixed lists item strength as manipulated by presentation duration had a significant effect on yes–no recognition accuracy. The stronger the item, the better it was recognized. Furthermore, there was no statistically significant evidence of rehearsal borrowing between adjacent items when item strength was randomized within a study list. Finally, the general pattern of results was quite consistent with study–test results using slower presentation rates, suggesting that rapid serial visual presentation does not give qualitatively different results from slower presentation conditions.

Experiment 2

Experiment 1 demonstrated that differences in item strength within a mixed list affected yes–no recognition performance and that there was no significant rehearsal borrowing between neighboring items. In this second experiment, we tested the list-strength hypothesis by comparing the performance in mixed lists to that in pure lists. To conform with Experiment 1, we used three different presentation rates. Although previous experiments investigating the effect of list strength presented lists with only two levels of strength, the effect should be observable with lists with more than two levels of strength. In keeping with the previous experiment, we also used rapid presentation rates and randomly mixed lists to prevent rehearsal borrowing.

Method

Subjects and materials. There were 30 subjects from the same subject pool as in the previous experiment. Seven hundred twenty words were chosen randomly from the Toronto pool of 1,024 words for each session. Twelve study lists containing 30 words each were selected, and the remaining words served as test lures.

Design. The experiment consisted of 12 successive recognition tests. Each test was made up of a study list followed by a yes–no recognition test for words in that list. The two independent variables were item strength and list type. Item strength was controlled by varying the presentation duration at study. Items were either strong, medium, or weak. List type was either pure or mixed. In pure lists, each word in the study list had the same strength; all words were presented for the same duration. Mixed lists contained an equal number of items of each strength. Strength and type were crossed, and both were within-subject factors. The dependent variable was accuracy. Because there was no effect on response time in Experiment 1, it was not included as a dependent variable.

Procedure. Experiment 2 followed the same procedure as Experiment 1. Twelve lists of 30 words each were presented to each subject. Items were either strong, medium, or weak (presented for 200, 100, or 50 ms, respectively). Lists were either pure or mixed. There were two pure lists for each of the three different strengths and six mixed lists. Each mixed list contained 10 words presented for 50 ms, 10 words for 100 ms, and 10 words for 200 ms. Assignment of words to conditions was random. As in the previous experiment, there was a 50-ms mask after each item. The order of the presentation durations

Table 3
Yes–No Recognition Performance as a Function of Input Position and Output Position for Experiment 1

Input position	Hit rate	Output position	Hit rate
1–3	0.387	1–6	0.569
4–6	0.380	7–12	0.463
7–9	0.376	13–18	0.421
10–12	0.361	19–24	0.378
13–15	0.370	25–30	0.355
16–18	0.361	31–36	0.342
19–21	0.360	37–42	0.350
22–24	0.390	43–48	0.338
25–27	0.390	49–54	0.348
28–30	0.459	55–60	0.273

within each mixed list and the order of lists within a session was randomized independently for each subject. Immediately after the presentation of each study list, the yes–no recognition test was performed. The 30 list items and 30 new items were presented in a different random order for every test list.

Results and Discussion

Table 4 presents the average hit rate, false-alarm rate, and the average d' for each condition. Analysis is based on d' calculated for individual subject conditions.

The analysis shows a significant list-strength effect. The strong items produced greater performance in the mixed lists than in the pure lists, whereas the weak items produced poorer performance in the mixed lists than in the pure lists. The effect is represented by the Type of List × Strength interaction, $F(2, 58) = 5.50$, $p < .01$, $MS_e = 0.06$.

A measure used by Ratcliff et al. (1990) to represent the magnitude of the list-strength effect was the ratio of ratios. This is the ratio of the strong to weak items in the mixed lists divided by the ratio of the strong to weak items in the pure lists. Values greater than 1 represent the list-strength effect. We calculated the ratio of ratios to be 1.77.

There was a significant overall strength effect, $F(2, 58) = 51.50$, $p < .01$, $MS_e = 0.08$; longer presentation times produced higher recognition performance. This is in agreement with the findings of Experiment 1. The type of list did not affect overall performance, $F(1, 29) = 1.83$, $p > .05$, $MS_e = 0.06$. The average d' for mixed lists was .59 compared with .63 for pure lists.

Table 4
Yes–No Recognition Performance for Mixed and Pure Lists in Experiment 2

	Mixed list			Pure list		
Item strength	Hit rate	FA rate	d'	Hit rate	FA rate	d'
Strong (200 ms)	0.537	0.230	0.98	0.468	0.215	0.84
Medium (100 ms)	0.376	0.230	0.48	0.433	0.253	0.60
Weak (50 ms)	0.318	0.230	0.31	0.360	0.248	0.47

Note. FA = false alarm.

The possibility of rehearsal borrowing in the mixed lists was investigated by examining the effects of the strength of the preceding item in the study list. Table 5 presents performance as a function of item strength and the strength of the preceding item in the study list. The strength of the preceding item did not significantly affect performance ($F < 1$). However, as in the prior experiment, the strength of the weak items was inversely proportional to the strength of its predecessors. We return to this point in Experiment 6.

This experiment demonstrates that there is a significant list-strength effect for yes–no recognition when the items in the mixed list were randomly mixed. The effect was observed under conditions of rapid visual presentation with lists containing three different levels of item strength.

Experiment 3

Experiment 2 demonstrated that there was a list-strength effect for yes–no recognition. In this third experiment, we assessed the generality of this finding. Because the two most common methods of studying item recognition are the yes–no and forced-choice procedures, it was decided to determine whether the list-strength effect is present with a two-alternative forced-choice recognition test.

Method

Subjects and materials. There were 30 subjects from the same subject pool as in the previous experiments. The materials were the same as those used in the previous experiment.

Design and procedure. Experiment 3 was based on the same design and followed the same procedure as Experiment 2 except that the yes–no recognition test was replaced by a two-alternative forced-choice recognition test.

Immediately after the presentation of each study list, a two-alternative forced-choice recognition test was performed. The 30 studied items were tested in a random order. Two words appeared on the screen, one of which was in the study list and the other was not. Subjects indicated which item they thought was in the study list by pressing an appropriate key on the computer keyboard. The next pair appeared 250 ms after the response was made. When the test was complete and the subjects were ready for the next study list, they pressed an assigned key and the next test began.

Results and Discussion

Table 6 presents the performance on the forced-choice recognition test. The scores represent the proportion of items correctly recognized. The analysis is based on performance for individual subject conditions.

There was no evidence of a list-strength effect. This was reflected in the Type × Strength interaction, which was nonsignificant ($F < 1$). The performance on strong items was slightly greater in the mixed lists than in the pure lists. However, the same was true for the weak items. As in the previous experiment, there was a significant effect of strength, $F(2, 28) = 42.86$, $p = .01$, $MS_e = 0.01$, and no effect of list type ($F < 1$).

Table 5
Yes–No Recognition Performance (d') as a Function of an Item's Strength and the Strength of the Preceding Item in the Study List for Mixed Lists in Experiment 2

Item strength	Strength of predecessor		
	Weak	Medium	Strong
Weak	0.33	0.22	0.18
Medium	0.38	0.25	0.43
Strong	0.96	0.97	0.95
Mean	0.57	0.48	0.52

The experiment failed to find a list-strength effect when recognition was tested with a two-alternative forced-choice test. One factor that may have played a critical role is the relatively poor performance of the subjects. The average correct recognition rates for the different conditions varied between .535 and .660. Considering that chance performance on a two-alternative recognition task is .5, these scores are suspiciously close to the floor. The equivalent scores in terms of d' are .14 and .58, and these scores are considerably lower than the scores in the yes–no recognition test of Experiment 2 (see Table 4). To compare the performance on this test with that of the preceding yes–no test, the scores for each subject condition were converted to d' using Hacker and Ratcliff's (1979) conversion tables for M-alternative forced-choice tests. Performance was significantly poorer in the forced-choice recognition test than in the yes–no recognition test, $F(1, 348) = 59.21$, $p < .001$, $MS_e = 0.10$. For this reason, a second forced-choice test was performed in which an attempt was made to increase the overall performance.

Experiment 4

In Experiment 3, we failed to find a list-strength effect with a forced-choice recognition test. Because the overall performance in that experiment was extremely poor, it could be that the absence of the list-strength effect was due to floor effects. To test this possibility, we attempted to increase the performance level in a forced-choice test by making two changes to the study conditions. First, the study lists were shortened because it has been shown that recognition performance is better for shorter lists (Gillund & Shiffrin, 1984). Second, the number of different levels of item strength was reduced. It was hoped that reducing the number of levels of item strength might make the effect easier to detect. To preserve the same level of reliability, the number of lists per subject was increased.

Table 6
Forced-Choice Recognition Performance for Mixed and Pure Lists in Experiment 3

Item strength	Proportion recognized	
	Mixed list	Pure list
Strong (200 ms)	0.660	0.655
Medium (100 ms)	0.586	0.596
Weak (50 ms)	0.539	0.535

Method

Subjects and materials. There were 30 subjects from the same subject pool as in the previous experiments. The materials consisted of 800 words chosen randomly from the Toronto pool of 1,024 words. Twenty study lists containing 20 words each were selected, and the remaining words served as test lures.

Design and procedure. Experiment 4 was based on the same design and followed the same procedures as the previous experiment with the following exceptions: (a) The medium strength condition was dropped, (b) the list length was shortened from 30 to 20 items, and (c) the number of lists in each session was increased from 12 to 20.

The experiment consisted of 20 forced-choice recognition tests. Word strength was either strong (200 ms) or weak (50 ms). There were 5 pure lists of weak items, 5 pure lists of strong items, and 10 mixed lists. Immediately after the presentation of each list, a two-alternative forced-choice recognition test was given.

Results and Discussion

Table 7 presents the results of the forced-choice recognition test. Scores represent the proportion of items correctly recognized. Analysis is based on performance for individual subject conditions.

The average performance in this experiment was .630, which was a slight improvement over the previous experiment in which the overall performance was .595. Thus, reducing list length from 30 to 20 and reducing the number of levels of strength from three to two improved overall performance somewhat. However, there was still no evidence of a list-strength effect with the forced-choice procedure. Recognition for strong items was slightly higher in the pure list than in the mixed list, whereas recognition for weak items was slightly worse in the pure list than in the mixed list, which indicates a negative list-strength effect. However, the trend was not significant. The Type × Strength interaction did not approach significance, $F(1, 29) = 1.22$, $p > .05$, $MS_e = 0.002$. As in the previous experiments, there was a significant effect of strength, $F(1, 29) = 138.64$, $p = .001$, $MS_e = 0.004$, and no effect of list type ($F < 1$).

This experiment failed to find a list-strength effect with a two-alternative forced-choice recognition test. However, because the overall performance in this experiment was poor, as was that of the previous experiment, the absence of the effect may be due to the insensitivity of the test. Why the performance on the forced-choice tests is so poor relative to that of the yes–no test in Experiment 2 is not clear because previous studies showed the equivalency of forced-choice and yes–no procedures (Green & Moses, 1966). This problem was examined in the final two experiments of this study.

Experiment 5

In Experiment 2, in which we used a yes–no recognition test, we found a list-strength effect. In Experiments 3 and 4, in which we used a forced-choice recognition test, we failed to find a list-strength effect. Is the list-strength effect in yes–no recognition reliable? To find out, in this experiment we returned to the yes–no procedure. This experiment was the same as Experiment 2 with two exceptions. First, the number

Table 7
Forced-Choice Recognition Performance for Mixed and Pure Lists in Experiment 4

Item strength	Proportion recognized	
	Mixed list	Pure list
Strong (200 ms)	0.688	0.702
Weak (50 ms)	0.568	0.562

of items per list was reduced from 30 to 20. Second, the number of different levels of item strength was reduced from three to two. This makes Experiment 5 comparable to Experiment 4 in the same way Experiment 2 was comparable to Experiment 3.

Method

Subjects and materials. There were 30 subjects from the same subject pool as in the previous experiments. The materials consisted of 800 words chosen randomly from the Toronto pool of 1,024 words. Twenty study lists contained 20 words each, and the remaining words served as test lures.

Design and procedure. The experiment was based on the same design and followed the same procedure as the previous experiment except that the forced-choice recognition test was replaced by a yes–no recognition test.

The experiment consisted of 20 yes–no recognition tests, each of which consisted of 20 study items. Word strength was either strong (200 ms) or weak (50 ms). There were 5 pure lists of weak items, 5 pure lists of strong items, and 10 mixed lists of strong and weak items. Immediately after the presentation of each list, subjects were given a yes–no recognition test.

Results and Discussion

Table 8 presents d' for each experimental condition. Analysis is based on d' calculated for individual subject conditions.

There was evidence of a list-strength effect because the performance for strong items was greater in the mixed lists than in the pure lists, and the performance for weak items was greater in the pure lists than in the mixed lists. The ratio of ratios was 1.30. However, the effect as measured by the Strength × Type interaction was not statistically significant, $F(1, 29) = 3.45$, $.10 > p > .05$, $MS_e = 0.07$. As in the previous experiments, there was a significant effect of strength, $F(1, 29) = 208.84$, $p < .001$, $MS_e = 0.08$, and no effect of list type ($F < 1$).

In the course of further analysis, we found that a list-strength effect was present in all but the initial study positions. The effect of study position was examined by blocking items into four groups of five serial positions and calculating the means for each experimental condition within each block. This is shown in Figure 1; the list-strength effect is represented by an interaction between strength and type. The interaction is absent in the first block (Study Positions 1 through 5) but present thereafter. A further analysis was performed in which *responses for the initial* 5 study items were excluded. Performance on Study Items 6 through 20 is presented in Table 8. An ANOVA performed on d' revealed a stronger and now significant list strength effect, $F(1, 29) = 12.14$, $p < .01$, $MS_e = 0.07$. The ratio of ratios was 1.48.

A similar analysis in which the initial five study items were excluded was performed for the previous experiments. In Experiments 3 and 4, in which there was no overall list-strength effect, excluding the initial items did not have an effect. However, in Experiment 2, in which there was an overall list-strength effect, dropping the initial study items did increase the magnitude of the effect. This was reflected in an increase in the F ratio for the Strength × Type interaction from 5.50 to 6.81. Although the increase was not as large as it was in the present experiment, this would be expected because the lists in the earlier experiment were longer, and the effect of the initial study items played a smaller role in the overall effect.

A similar effect of study position also occurred in a free-recall study by Tulving and Hastie (1972). In their second experiment, judging by their serial position curves, the first four or five items in the study list did not show the list-strength effect that was present for the remaining items in the list. However, in their experiment, the pure lists contained a larger number of words than did the mixed lists, so the comparison to our results is not perfect. Regrettably, in Tulving and Hastie's first experiment in which they presented pure lists with the same number of items as the mixed lists, the serial position curves were too noisy to confirm the same pattern of data.

The experiment provided some evidence of a list-strength effect. However, overall the effect was not significant. Analysis showed that although the effect was not present for the initial items in the list, there was evidence of a list-strength effect for the later items in the list.

Experiment 6

In the previous experiments, the list-strength effect was tested using randomly mixed lists. Although there was no evidence of weak items borrowing from strong items, there was some evidence of reverse borrowing (strong borrowing from weak) in Experiments 1 and 2 (see Tables 2 and 5). This was particularly true of weak items, and if such borrowing were occurring, it could produce or inflate any apparent list-strength effect. In this experiment, we attempt to eliminate any possibility of reverse rehearsal borrowing by using a blocked format for the mixed lists. In the blocked format, mixed lists are presented such that items of the same strength were grouped together. If borrowing does occur, it would most often be between items of the same strength and thus would

Table 8
Yes–No Recognition Performance (d′) for Mixed and Pure Lists in Experiment 5

Item strength	Mixed list		Pure list	
	d'	Performance[a]	d'	Performance[a]
Strong (200 ms)	1.25	1.31	1.17	1.11
Weak (50 ms)	0.42	0.41	0.51	0.54

[a] For items in Study Positions 6 through 20.

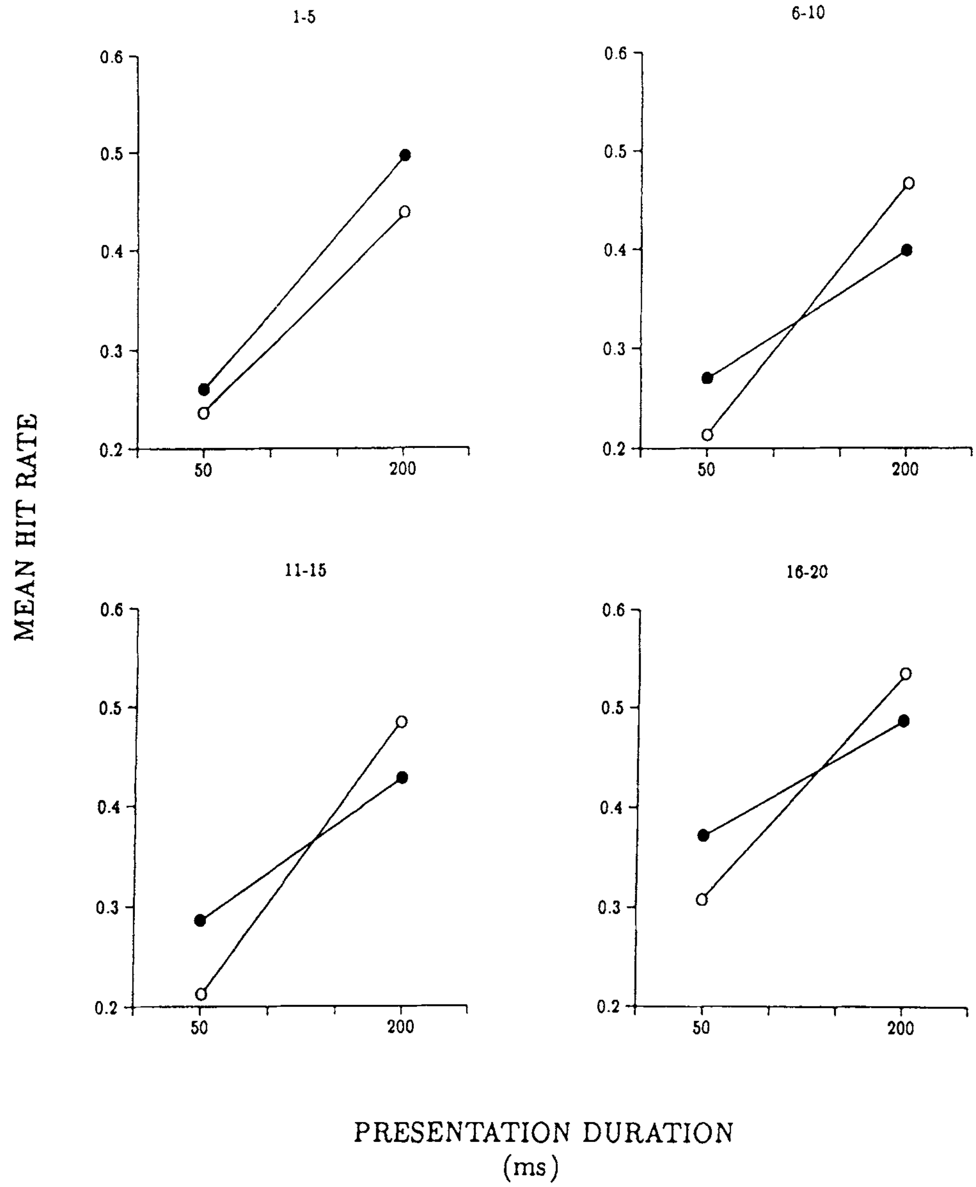

Figure 1. Recognition performance for Experiment 5 as a function of study position, type of list, and presentation duration. (The open circles represent mixed lists, and the filled circles represent pure lists. Mean hit rate is presented for each block of five study positions.)

not influence the list-strength effect. The procedure is identical to that used by Ratcliff et al. (1990) except that rapid presentation rates are used. If reverse rehearsal borrowing were responsible for the list-strength effect in the previous experiments, and if the blocked format minimizes or eliminates rehearsal borrowing, then our list-strength effect should decrease or disappear.

Method

Subjects and materials. Twenty-four subjects participated in this experiment. Subjects were recruited from Wilfrid Laurier's voluntary participant pool. The materials consisted of 768 words chosen randomly from the Toronto pool of 1,024 words. Twelve study lists containing 32 words each were compiled, and the remaining words served as test lures.

Design and procedure. The experiment was based on the same design and followed the same procedure as the previous experiment except that the mixed lists were presented in a blocked format.

The experiment consisted of 12 yes–no recognition tests each consisting of 32 study items. Word strength was either strong (200 ms) or weak (50 ms), and there was always a 50-ms interstimulus interval between items. There were three pure lists of weak items, three pure lists of strong items, and six mixed lists of strong and weak items. The mixed lists were such that items of the same strength were blocked together. Half the time, there were 4 strong items, then 12 weak items, then 12 strong items, and then 4 weak items. The rest of the mixed lists used the same schedule with the item strengths reversed. Immediately after the presentation of each list, subjects were given a yes–no recognition test for the items in that list.

Results and Discussion

Table 9 presents the results of the yes–no recognition test. Scores represent average hit rates, false-alarm rates, and d's for each condition. Analysis is based on performance (d') for individual subject conditions.

There was no significant list-strength effect as measured by the Strength × Type interaction, $F(1, 23) = 1.48$, $p > .05$, $MS_e = 0.04$. The ratio of ratios was 1.18. Performance on the pure lists was greater than that on the mixed lists for both weak and strong items. This was reflected in the significant effect of the type of list, $F(1, 23) = 9.95$, $p < .01$, $MS_e = 0.06$. As in the previous experiments, the effect of item strength was significant, $F(1, 23) = 100.36$, $p < .01$, $MS_e = 0.05$.

No evidence of rehearsal borrowing between blocks in the mixed lists was found. To investigate the possibility of rehearsal borrowing, we used a technique used by Ratcliff et al. (1990) whereby we analyzed the hit rates of the middle four items within each block of 12 items of the same presentation duration. Any rehearsal borrowing that would involve these items would come from items of the same strength and thus not adversely influence the list-strength effect. The mixed/pure ratio of ratios was unaffected, suggesting that rehearsal borrowing was not a problem.

It should be noted that these results replicate recent findings reported by Ratcliff and McKoon (1991), who also failed to find a list-strength effect in two experiments that used a blocked mixed-list format and rapid sequential presentation.

The results of Experiment 6 demonstrated that in the blocked design in which rehearsal borrowing was minimized or eliminated, there was no evidence of a list-strength effect. The results of Experiments 2 and 5, in which we found evidence of a list-strength effect, can be explained in terms of reverse rehearsal borrowing. In the randomly mixed lists used in those experiments, subjects might have borrowed rehearsal time from the weak items and redistributed it to the strong items. This would have produced an apparent list-strength effect. Support for this came from our conditional analysis of Experiment 1 and 2, in which we found that performance on weak items was inversely proportional to the strength of its predecessor. Although the effect of the strength of the previous item was not significant, perhaps the experiments were not powerful enough to make the effect statistically reliable.

The reverse rehearsal borrowing interpretation of the list-strength effect in Experiments 2 and 5 suggests that the use of extremely rapid presentation does not wholly eliminate the problem of rehearsal borrowing. The differences observed between the yes–no and forced-choice procedures may also be due to encoding strategies. We examine this issue in the last two experiments.

Table 9
Yes–No Recognition Performance for Mixed and Pure Lists in Experiment 6

	Mixed list			Pure list		
Item strength	Hit rate	FA rate	d'	Hit rate	FA rate	d'
Strong (200 ms)	0.628	0.417	0.58	0.576	0.308	0.79
Weak (50 ms)	0.489	0.417	0.18	0.520	0.413	0.29

Note. FA = false alarm.

Yes–No and Forced-Choice Procedures

Two final experiments were conducted to determine why the yes–no procedure produced better performance than did the forced-choice procedure. In Experiments 2 and 3, in which three levels of item strength were tested, the performance with the yes–no procedure was better than that with the forced-choice procedure. Furthermore, a list-strength effect was found with the yes–no procedure, and no effect was found with the forced-choice procedure. The same results were found in Experiments 4 and 5, in which two levels of item strength were tested. Why should performance differ for the two types of procedure?

Previous work by Green and Moses (1966) showed that performance on the two types of test was equivalent. They compared performance on a forced-choice procedure to that of a rating task. The rating task was similar to a yes–no procedure except that subjects were required to rate on a numeric scale how sure they were that an item was in the previous study list. They reported that recognition memory for nonsense syllables was equivalent when measured with either the forced-choice procedure or the rating task. However, Norman and Wickelgren (1969) found that recognition for digit pairs was slightly greater with the yes–no procedure than with the two-alternative forced-choice procedure.

The following experiments investigated several factors that may have been responsible for the difference in performance between the two test procedures. The first was the item strength or the rate at which items were presented at study. It is possible that the difference in performance on the two test procedures is present only for items presented at rapid rates because in the Green and Moses study items were presented at a relatively slow rate compared with those used in the current study.

The second factor that was investigated was the physical similarity between study and test. Because the study stimuli in the yes–no procedure are typically identical to the test stimuli (one word at a time is studied and one word at a time is tested), this procedure might be more sensitive to the physical match between study and test than the forced-choice procedure in which the test stimulus involves more than just the study item. If the performance on the yes–no procedure benefited from the effects of study-test similarity, we should be able to reduce that advantage by varying the degree of physical similarity of the study and test stimuli.

The third factor investigated was whether subjects knew at the time of study what type of test they would receive. In the previous experiments in this study, subjects were informed of the type of test they would receive and were repeatedly tested with the same test procedure. However, in the Green and Moses study, each subject received both yes–no and forced-choice tests in a random order such that they did not know what type of test procedure would be used for any given list. If subjects encoded the study lists differently for the two types of test, this could account for the differences in levels of performance that we found.

In Experiment 7, performance on the two types of test procedure was tested over a range of presentation rates. As with the previous experiments, each subject was tested repeatedly with only one procedure. In Experiment 8, each subject received both test procedures presented in a random order such that they did not know at the time of study what type of test they would receive.

Experiment 7

In this experiment, we investigated the effects of item strength and study–test similarity on yes–no and forced-choice recognition memory; each subject was repeatedly tested with the same test procedure. The presentation rate of items at study was varied to determine if the advantage of the yes–no over the forced-choice procedure seen in the previous experiments was general across a range of presentation rates. The physical similarity of stimuli at study and test was manipulated to determine if the increased performance of the yes–no procedure was due to heavier reliance on study–test similarity.

Method

Subjects and materials. Twenty-four subjects participated in this experiment. Each student was enrolled in an introductory psychology course at the University of Toronto and received credit for participating in the study. Seven hundred twenty words were chosen randomly from the Toronto pool of 1,024 words for each session. Eighteen study lists containing 20 words each were selected, and the remaining words served as test lures.

Design. The experiment consisted of 18 successive recognition tests. Each test was made up of a study list of 20 words followed by either a yes–no or a forced-choice recognition test for words in that list. Strength was manipulated by varying presentation rate. Six different levels were tested. All of the items in a list were presented for the same duration. Study–test similarity was manipulated by varying the letter case in which the words were presented. All words were studied in uppercase and tested in either uppercase or lowercase. Item strength was crossed with type of test and study–test similarity. The type of test was a between-subjects factor, whereas strength and similarity were within-subjects factors. There were 24 cells, each containing 360 observations.

Procedure. The procedure was the same as that used in the previous experiments with the following exceptions. Subjects were informed that they would be given 18 short recognition tests on a computer. Twenty words were presented in each study list one at a time on a computer screen with a 50-ms mask (a row of asterisks) between each word. There were six different presentation durations tested: 50, 100, 200, 400, 800, and 1,600 ms per item. All the items in a list were presented for the same duration, and each subject was presented with three of each type of list. The lists were presented in a random order.

Immediately after the presentation of each list, a recognition test was performed. Half of the subjects were tested with a yes–no procedure and the others with a forced-choice procedure. The subjects were randomly assigned to the test procedures.

Results and Discussion

Table 10 presents the recognition performance in terms of d' for Experiment 7. For the yes–no procedure, hit rate and false-alarm rate were used to calculate d'. For the forced-choice procedure, the proportion of items recognized was converted to d' using conversion tables (Hacker & Ratcliff, 1979). Scores represent performance for words tested in uppercase and lowercase. Scores are presented for the yes–no and the forced-choice procedures for items presented for 50, 100, 200, 400, 800, and 1,600 ms. Analysis was performed on the calculated d' scores for each subject condition.

Item strength (presentation duration) significantly affected performance, $F(5, 110) = 72.06$, $p < .001$, $MS_e = 0.25$. Performance increased with item strength. The type of test also significantly affected performance, $F(1, 22) = 11.04$, $p < .01$, $MS_e = 0.90$. Performance was greater when recognition was tested with a yes–no procedure than with a forced-choice procedure ($d' = 1.32$ vs. .95). This is in agreement with the results of the previous experiments. Moreover, this experiment demonstrates that the effect is not limited to items presented at a rapid rate but is general across a wide range of presentation rates. The superiority of the yes–no performance was present for all presentation durations and for similar and nonsimilar study–test items. The only exception was the 100-ms similar condition.

Words that were similar at study and test produced slightly better performance than those that were nonsimilar. However, this trend was not significant ($F < 1$). The superiority of words that were similar at study and test was present for all conditions except for the 50-, 100-, and 200-ms conditions, which were tested with a yes–no procedure. This reversal of the general similarity effect is reflected in the marginally significant Type of Test × Similarity interaction, $F(1, 22) = 6.95$, $p < .05$, $MS_e = 0.25$. The finding that recognition performance is, in general, slightly better for words which were similar at study and test could be explained in terms of encoding specificity, which suggests that recognition performance should be greatest when study and test stimuli are highly similar (Tulving & Thomson, 1974). However, why this similarity effect

Table 10
Recognition Performance (d') Measured With Yes–No and Forced-Choice Procedures as a Function of Presentation Duration and Study–Test Similarity for Experiment 7

Presentation duration (ms)	Test procedure			
	Yes–No		Forced choice	
	Similar	Nonsimilar	Similar	Nonsimilar
50	0.55	0.71	0.32	0.05
100	0.51	0.88	0.60	0.39
200	0.80	1.20	0.72	0.67
400	1.60	1.57	1.20	1.00
800	1.95	1.90	1.63	1.16
1,600	2.17	1.98	1.82	1.83

was not present for the weak items in the yes–no test is not clear.

Experiment 8

In the previous experiment, we found that performance was greater for the yes–no than for the forced-choice procedure across a wide range of item strengths. However, it was possible that subjects differentially encoded for the two types of test procedure. In this experiment, we eliminated this possibility by presenting each subject with a random mixture of both types of test procedure such that they did not know at the time of the study what type of test they would receive. If the difference in performance between the two procedures was due to differential encoding, we would expect to see the difference in performance disappear.

As in the previous experiment, study–test similarity was manipulated. However, in this experiment only two levels of item strength were tested.

Method

Subjects and materials. Nineteen subjects from the same subject pool as the previous experiment participated in the experiment. Eight hundred words were chosen randomly from the Toronto pool of 1,024 words for each session. Twenty study lists containing 20 words each were selected, and the remaining words served as test lures.

Design. The experiment consisted of 20 successive recognition tests. Each test was made up of a study list of 20 words followed by either a yes–no or a forced-choice recognition test for words in that list. The order of the tests was randomized. There were two levels of item strength tested, and items were either similar or nonsimilar at study and test. For each subject, item strength was crossed with type of test and study–test similarity. Thus, there were 8 cells, each containing 950 observations.

Procedure. The procedure was the same as that used in the previous experiments with the following exceptions. Subjects were informed that they would be given 20 short recognition tests on a computer and were familiarized with both the yes–no and forced-choice test procedures. Twenty words were presented in each study list one at a time on a computer screen with a 50-ms mask (a row of asterisks) between each word. Ten of the study lists were presented at a fast rate (100 ms per item), and 10 were presented at a slow rate (1,000 ms per item). The lists were presented in a random order.

Immediately after the presentation of each list, a recognition test was performed. Half of the lists were tested with a yes–no procedure and the others with a forced-choice procedure. The order of the tests was randomized. Although all the words were studied in uppercase, half were tested in uppercase and the others were tested in lowercase.

Results and Discussion

The treatment and analysis of the data were the same as that of the previous experiment. Table 11 presents the recognition performance in terms of d' for Experiment 8. Scores represent performance for words tested in uppercase and lowercase. Scores are presented for the yes–no and the forced-choice procedures for items presented for 100 ms and 1,000 ms. Analysis was performed on the calculated d' scores for each subject condition.

The only variable to affect performance significantly was presentation duration, $F(1, 18) = 1307.88$, $p < .001$, $MS_e =$ 0.51. Performance was greater for items presented for a long duration than for items presented for a short duration. All other factors failed to affect performance ($F < 1$ for all factors except the Type of Test × Similarity interaction in which $F = 2.64$ and the Duration × Similarity interaction in which $F = 2.07$). Performance for the yes–no procedure was equivalent to that of the forced-choice procedure regardless of presentation duration and study–test similarity.

Table 11
Recognition Performance (d') Measured With Yes–No and Forced-Choice Procedures as a Function of Presentation Duration and Study–Test Similarity for Experiment 8

Presentation duration (ms)	Test procedure: Yes–No		Forced choice	
	Similar	Nonsimilar	Similar	Nonsimilar
100	0.36	0.52	0.45	0.54
1,000	1.76	1.84	1.79	1.68

As we predicted, when the opportunity for differential encoding was removed, the two test procedures produce equivalent results. These results are in agreement with those of Green and Moses (1966), who reported the equivalence of the two procedures. Neither the presentation duration nor the study–test similarity significantly affected the performance.

Why should subjects differentially encode for the two types of test procedure? One possibility is that the subjects found the yes–no test to be a more demanding task and thus were motivated to put more effort into encoding. Anecdotal evidence came from the debriefing session after the final experiment. Several subjects commented that they found the yes–no procedure to be the more difficult of the two, arguing that with the forced-choice task they had a 50–50 chance of guessing correctly. Although the same argument also holds for the yes–no procedure, the subjects' failure to realize this would lead them to believe that the yes–no procedure was a more difficult task. If the yes–no procedure was perceived as more difficult, then subjects tested repeatedly with that test might be motivated to put more effort into encoding.

A comparison of performance across Experiments 7 and 8 provided further support for this claim. Performance on the yes–no procedure was greater when subjects were repeatedly tested with that procedure than when they were randomly tested with both procedures. However, the performance on the forced-choice procedure was unchanged. Comparing performance on the 100-ms condition in the last two experiments, we found that performance on the yes–no procedure was greater with repeated testing (.67) than with random testing (.43), but performance on the forced-choice procedure was the same for repeated (.49) and random testing (.48).

Furthermore, an examination of performance over the test session in Experiment 7 suggested that performance on the yes–no procedure gradually improved with repeated testing whereas forced-choice performance did not. Although performance over the test session was rather noisy, the only subjects to exhibit an increase in performance over the test session were those who were repeatedly tested with the yes–no procedure.

General Discussion

The primary purpose of these experiments was to investigate the effect of list strength on item recognition at rapid presentation rates. Overall, the results of this study do not support the claim that there is a list-strength effect in recognition memory. In two experiments using a forced-choice procedure, there was no evidence of a list-strength effect. Although there was some evidence of an effect when recognition was tested with a yes–no procedure, the possibility that this effect was due to rehearsal borrowing could not be ruled out. A final test in which a blocked design was used demonstrated that when the possibility of rehearsal borrowing was removed no list-strength effect was observed.

The results of these experiments are in agreement with those of Ratcliff et al. (1990), Murnane and Shiffrin (1991), and Shiffrin and Murnane (1991), who reported the absence of a list-strength effect when items were presented for a relatively long duration at study. Furthermore, they replicated Ratcliff and McKoon's (in press) failure to find a list-strength effect at rapid presentation rates.

A secondary question addressed in this study was the relationship between the yes–no and forced-choice test procedures. Rather unexpectedly, the initial experiments revealed that the yes–no procedure produced greater performance than the forced-choice procedure. Furthermore, the list-strength effect was present only when recognition was tested with the yes–no procedure. Two final experiments showed that the overall difference in performance was due to more effective encoding in the case of the yes–no procedure over the forced-choice procedure. It was suggested that subjects perceived the yes–no procedure to be more difficult than the forced-choice procedure and thus put more effort into encoding. Because the performance levels in the forced-choice experiments were lower than that of the experiments using the yes–no procedure, the absence of a list-strength effect could be attributed to the insensitivity of the test. Conversely, the enhanced encoding in the case of the yes–no procedure may be related to the appearance of the list-strength effect with that test procedure. This would be so if the enhanced encoding included rehearsal borrowing from weak to strong items. However, this final point cannot be verified in the current study.

From an experimental point of view, the results seem quite clear. Under a variety of experimental conditions, there does not seem to be a list-strength effect in recognition memory. Thus, these data confirm and extend the results of Murnane and Shiffrin (1991), Shiffrin and Murnane (1991), and Ratcliff et al. (1990). From a theoretical point of view, according to the interpretation provided by Shiffrin et al. (1990), the absence of a list-strength effect is inconsistent with most of the global memory models because they predict a list-strength effect. However, for an alternative interpretation of the list-strength effect in terms of the global memory models, see Murdock (1991). The present findings are also problematic for the SAM model that predicts a list-strength effect under conditions of rapid presentation while predicting no effect at slower presentation rates (Shiffrin & Murnane, 1991).

References

Crowder, R. G., & Neath, I. (1991). The microscope metaphor in human memory. In W. E. Hockley & S. Lewandowsky (Eds.), *Relating theory and data: Essays on human memory in honor of Bennet B. Murdock* (pp 111–125). Hillsdale, NJ: Erlbaum.

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

Green, D. M., & Moses, F. L. (1966). On the equivalence of two recognition measures of short-term memory. *Psychological Bulletin, 66,* 228–234.

Hacker, M. J., & Ratcliff, R. (1979). A revised table for d' for *M*-alternative forced choice. *Perception and Psychophysics, 26,* 168–170.

Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review, 93,* 411–428.

Humphreys, M. S., Pike, R., Bain, J. D., & Tehan, G. (1989). Global matching: A comparison of the SAM, MINERVA II, MATRIX and TODAM models. *Journal of Mathematical Psychology, 33,* 36–67.

Loftus, G. R. (1974). Acquisition of information from rapidly presented verbal and nonverbal stimuli. *Memory & Cognition, 2,* 545–548.

Murdock, B. B. (1982). A theory for the storage and retrieval of item and associative information. *Psychological Review, 89,* 609–626.

Murdock, B. B. (1991, November). *The list-strength effect (LSE) and global matching models.* Paper presented at the Psychonomics Society Meeting, San Francisco.

Murnane, K., & Shiffrin, R. M. (1991). Word repetitions in sentence recognition. *Memory and Cognition, 19,* 119–130.

Murnane, K., & Shiffrin, R. M. (1991). Interference and the representation of events in memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 855–874.

Norman, D. A., & Wickelgren, W. A. (1969). Strength theory of decision rules and latency in retrieval from short-term memory. *Journal of Mathematical Psychology, 6,* 192–208.

Potter, M. C. (1984). Rapid serial visual presentation (RSVP): A method of studying language processing. In D. E. Kieras & M. A. Just (Eds.), *New methods in reading comprehension research* (pp. 91–118). Hillsdale, NJ: Erlbaum.

Ratcliff, R., Clark, S., & Shiffrin, R. M. (1990). The list-strength effect: I. Data and discussion. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 163–178.

Ratcliff, R., & McKoon, G. (1991). Using ROC data and priming results to test global memory models. In W. E. Hockley & S. Lewandowsky (Eds.), *Relating theory and data: Essays on human memory in honor of Bennet B. Murdock* (pp. 279–296). Hillsdale, NJ: Erlbaum.

Shiffrin, R. M., & Murnane, K. (1991). Composition distribution and interference in memory. In W. E. Hockley & S. Lewandowsky (Eds.), *Relating theory and data: Essays on human memory in honor of Bennet B. Murdock* (pp. 331–346). Hillsdale, NJ: Erlbaum.

Shiffrin, R. M., Ratcliff, R., & Clark, S. (1990). The list-strength effect: II. Theoretical mechanisms. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 179–195.

Tulving, E., & Hastie, R. (1972). Inhibition effects of intralist repetition in free recall. *Journal of Experimental Psychology, 92,* 297–304.

Tulving, E., & Thomson, D. M. (1974). Encoding specificity and retrieval processes in episodic memory. *Psychological Review, 80,* 352–373.

Weber, E. U. (1988). Expectation and variance of item resemblance distributions in a convolution-correlation model of distributed memory. *Journal of Mathematical Psychology, 32,* 1–43.

Received March 20, 1991
Revision received July 29, 1991
Accepted July 30, 1991 ■