

## Recall Latency Following Pure- and Mixed-Strength Lists: A Direct Test of the Relative Strength Model of Free Recall

John T. Wixted, Holly Ghadisha, and Ross Vera  
University of California, San Diego

Strong items (e.g., those studied for a longer period of time) are not recalled faster than weak items in pure-strength lists. Although counterintuitive, this result is consistent with a relative strength model of free recall. In mixed-strength lists, by contrast, the relative strength model requires that strong items be recalled significantly faster than weak items. A considerable body of recent research on this issue suggests that, if anything, the opposite may be true. Four experiments reported here measured free-recall latency following pure- and mixed-strength lists. Recall latency for strong items was consistently shorter than that for weak items, but in mixed lists only. Moreover, as uniquely predicted by a relative strength model, in mixed lists, strong items were recalled more quickly than items from a pure-strength list of the same size, and weak items were recalled more slowly by a corresponding amount.

The idea that stronger memories come to mind more quickly and easily than weaker ones has been around for a long time. Evidence consistent with this idea was reported by Marbe (see Bousfield, Cohen, & Silva, 1956) around the turn of the century and by Bousfield (Bousfield & Barclay, 1950; Bousfield et al., 1956; Bousfield, Whitmarsh, & Esterson, 1958) in the 1950s. As intuitively obvious as this result may seem, it is clearly incorrect under certain conditions (cf. MacLeod & Nelson, 1984). For example, Rohrer and Wixted (1994) showed that strengthening words on a list by allowing extra study time increased the number of items recalled but did not affect the average time to recall. They also showed that under the right conditions (namely, a long list of strong items vs. a short list of weak items), strong items were recalled more slowly than weak items. When are intuitions about the relationship between memory strength and recall latency correct and when are they incorrect? Wixted and Rohrer (1994) argued that relative strength models of free recall effectively guide thinking about this issue.

### Theoretical Background

According to relative strength models of free recall, of which SAM (search of associative memory) is the best known example (Gillund & Shiffrin, 1984; Shiffrin, 1970), the probability that an item will be sampled from a search set is a function of its strength relative to the strength of the other items in the set. If each of six items in a search set has a

strength of 1, for example, the probability that a given item will be sampled on a given draw is equal to 1 divided by 6 (the strength of that item divided by the summed strength of all items in the search set). As described in more detail in the Appendix, the expected *recall latency* for a given item is, theoretically, inversely related to its relative strength. If items are sampled (with replacement) from a search set at a rate of one per second, for example, the expected recall latency for an item with a relative strength of 1/6 is simply 6 s. That is, on average, 6 s will elapse before the item is sampled for the first time.

For the idealized case in which all items on the list have exactly the same strength, expected recall latency averaged across all items is the same as the expected latency for a single item (6 s in this example). Note that if each item's strength is increased to 2, the probability that a given item will be sampled does not change because the item's strength (2) divided by summed strength (12) is still 1/6. Thus, for a given list length, mean recall latency should be the same whether all of the items are weak or all of the items are strong. This is basically the result reported by Rohrer and Wixted (1994).

A different result is predicted for strong and weak items in the mixed-strength situation. If three of the six items have a strength of 1 and the other three have a strength of 2 (perhaps because they were given extra study time), the summed strength of all six items is 9. Thus, the relative strength of a weak item is 1/9, and the relative strength of a strong item is 2/9 (or 1/4.5). Because recall latency is theoretically determined by relative strength, strong items should now be recalled more quickly than weak items. More specifically, if items are sampled with replacement at a rate of one per second, the mean recall latency for strong items should be 4.5 s, and the mean recall latency for weak items should be 9 s. An important implication of the relative strength model is that strong items in a mixed list should also be recalled more quickly and weak items should be recalled more slowly than items from a pure-strength list of the same size (for which mean recall latency would be 6 s in this example).

---

John T. Wixted, Holly Ghadisha, and Ross Vera, Department of Psychology, University of California, San Diego.

We thank Doug Rohrer for many helpful discussions concerning this research (and for working out the proof presented in the Appendix). We also thank Jose Dias and Hamed Farid for their help in data collection.

Correspondence concerning this article should be addressed to John T. Wixted, Department of Psychology, 0109, University of California, San Diego, La Jolla, California 92093-0109. Electronic mail may be sent via Internet to [jwixted@ucsd.edu](mailto:jwixted@ucsd.edu).

These examples illustrate the predictions of the relative strength model with respect to mean recall latency. How does the model explain the well-known effects of these manipulations on the more common dependent measure, recall probability? Without additional assumptions, the simple relative strength model predicts that all items on a list will be recalled given an infinitely long recall period. No matter how weak an item is relative to the other items in the search set, it will eventually be sampled and recalled. Why, then, are more items recalled from a pure-strong list than from a pure-weak list?

SAM accounts for probability of recall effects by relying, in part, on the notion of *recoverability* (Gillund & Shiffrin, 1984; Raaijmakers & Shiffrin, 1980). For example, the presentation of an item on a list presumably creates a representation that is included in the search set, but that representation may or may not be sufficiently intact to enable the participant to reproduce the item that created it. Some items, especially weak ones, may leave a trace that allows the retrieval of partial information at best (e.g., the first letter of the word, what the word rhymes with, etc.). Strong items are more likely to produce intact traces that facilitate recovery when they are sampled. Thus, although both strong and weak items should influence the dynamics of free recall (because both are members of the search set), strong items are more likely to actually be recalled than weak ones. Note that the predictions of the relative strength model with regard to recall latency are not affected by the proportion of items that happen to be recoverable. That is, pure-strong lists should have the same mean recall latency as pure-weak lists even though they have a larger proportion of recoverable items.

Before considering prior research bearing on the relative strength model, a brief discussion of the measurement of recall latency and its relationship to the mathematical form of cumulative recall is presented. Although these issues may appear to be separate, they are in fact intimately related to each other (and to the relative strength model itself).

### Measuring Recall Latency

In a free-recall experiment, every recalled word has associated with it a corresponding latency measured from the onset of the recall period. Mean recall latency is simply the average of those values. Although it may seem that average interresponse time (i.e., the average time between successive retrievals) would be a more natural measure, the mathematics of the situation suggests otherwise. The time course of free recall is adequately described by a two-parameter exponential function. When progress is plotted cumulatively throughout the recall period, the form of the recall function is  $R(t) = N(1 - e^{-\lambda t})$ , where  $R(t)$  is the cumulative number of items recalled up to time  $t$ ,  $N$  is the asymptotic number of items recalled, and  $\lambda$  is the rate of approach to asymptote (Bousfield & Sedgewick, 1944; Roediger, Stellon, & Tulving, 1977). The vast majority of studies concerned with free recall have used the number of items recalled (or, equivalently, recall probability) as the dependent measure. If the recall period is long enough to

allow performance to reach near asymptotic levels, this measure is basically equivalent to  $N$  in the exponential growth function. However,  $N$  items can be recalled quickly or slowly, and that property of performance is captured by  $\lambda$ . A large value of  $\lambda$  reflects a rapid rate of approach to asymptote, whereas a small value reflects a slow rate of approach to asymptote. This is the second basic property of recall performance, and it is the property of interest here. If cumulative recall is perfectly exponential in form, mean recall latency (hereafter denoted  $\tau$ ) is equal to  $1/\lambda$ . That is to say, rate of approach to asymptote and mean recall latency are two sides of the same coin. Thus, one can estimate  $N$  and  $\tau$  either by fitting an exponential function to the obtained data or by simply counting the number of items recalled ( $N$ ) and computing the average latency associated with those items ( $\tau$ ). An advantage of fitting an exponential function is that it allows one to estimate the projected value of  $N$  and the projected mean recall latency given an infinitely long recall period, but the conclusions drawn from a given experiment are generally the same regardless of which method is used.

Note that the mathematical form of retrieval is itself consistent with a relative strength model that assumes sampling with replacement from a search set (McGill, 1963). That is, in the ideal case in which all items in the search set have exactly the same strength (and the rate of sampling from the search set remains constant throughout the recall period), the predicted cumulative-recall function is the exponential equation given above. When the equal-strength assumption is relaxed, the form of recall is no longer predicted to be perfectly exponential, but the predictions of the relative strength model concerning recall latency remain largely unchanged (Rohrer, 1996). Thus, although ideal conditions are assumed when illustrating the predictions of this model, those predictions are by no means restricted to the ideal case.

### Prior Research on the Role of Relative Strength

Wixted and Rohrer (1993) and Rohrer and Wixted (1994) showed that  $\tau$  remains constant with study time and increases with both list length and the buildup of proactive interference. The results of these studies (all of which involved pure-strength lists) are consistent with the simplest relative strength model of free recall. As already described, a relative strength model predicts a constant value of  $\tau$  for pure-weak and pure-strong lists of the same size (because an item's relative strength does not change with study time even though its absolute strength does). With regard to list length, relative strength models predict a larger value of  $\tau$  for longer lists. Continuing with the earlier example, each item's relative strength on a six-item list is  $1/6$ , and mean recall latency should be 6 s in the ideal case. For a nine-item list, an individual item's relative strength is  $1/9$ , and mean recall latency should increase to 9 s. A similar explanation accounts for the growth in  $\tau$  with the buildup of proactive interference (Wixted & Rohrer, 1993).

Although these findings support the relative strength model using pure-strength lists, no previous studies have measured free-recall latency in mixed-strength lists. Studies

that have come closest to doing so measured the output order of items that varied in strength. Output order provides a rough measure of recall latency because items that are, on average, recalled earlier in the output sequence presumably (though not necessarily) have a lower mean recall latency than items recalled later in the output sequence. Studies of retrieval from semantic memory (in which participants are asked to name as many exemplars from a category as possible) consistently reveal that the most commonly produced items are also the ones most likely to be produced first (e.g., Bousfield & Barclay, 1950). Because the commonly produced items are presumably the strongest in some sense (e.g., they are the items with which participants have the most experience), this finding is in agreement with the relative strength model. Similarly, in single-trial free recall, primacy items (which are strengthened by extra rehearsal) tend to be output earlier than middle items (Bjork & Whitten, 1974; Bousfield et al., 1958).<sup>1</sup> In an effort to experimentally manipulate item strength, Tulving and Hastie (1972) and Hastie (1975) conducted several experiments in which some items on a list were presented once and others twice. On average, twice-presented (i.e., strong) items appeared earlier in the output sequence than once-presented (i.e., weak) items, although the effects were fairly small.

The results of these studies are consistent with the idea that the memory traces most likely to be accessed are also the ones likely to be accessed first. However, tracking output order provides a weak test of the relative strength model, which makes rather more specific predictions about mean recall latency. For example, mean recall latency for strong and weak items in a mixed list can be meaningfully compared with mean recall latency for items in a pure-strength list, but no such comparison is possible with mean output position as a dependent measure. That is, to say that strong items are output earlier than weak items in a mixed list is not to say that strong items are recalled faster than (and weak items slower than) items from a pure-strength list of the same size (which is what the relative strength model predicts). One reason for performing the mixed-strength experiments described later (Experiments 2, 3, and 4) was to subject the relative strength model to a test of these key predictions.

A second reason to examine these issues is that a plethora of recent studies, which also used output order as a dependent measure, suggests that weak items are actually recalled before strong items much of the time, exactly the opposite of what a relative strength model predicts (e.g., Brainerd, 1995; Brainerd, Reyna, Harnishfeger, & Howe, 1993; Brainerd, Reyna, Howe, & Kevershan, 1990). This effect, termed *cognitive triage*, is assumed to occur because the memory system is aware of the vulnerability of weak items and makes those items available to conscious awareness first (before they are lost). Strong items can wait to be retrieved because they are less likely to be affected by output interference while the weak items are being recalled. The fact that cognitive-triage theory and the data collected in support of it are directly contrary to the predictions of a relative strength model suggests that additional work is

needed to sort out the dynamics of free recall in the mixed-strength situation.

The first two experiments described below were designed to test critical predictions of the basic relative strength model using pure-strength (Experiment 1) and mixed-strength (Experiment 2) lists. These two experiments used a very common method of strengthening, namely, item repetition. Experiments 3 and 4 were designed to test the relative strength model using a procedure that has been most commonly used to test the predictions of cognitive-triage theory, namely, multitrial free recall. In this procedure, the same list is presented for study several times in succession, with recall tested after each list. Thus, after the first list presentation, the study lists can be regarded as mixed-strength lists, with strong items being those that were recalled on the prior trial and weak items being those that were not.

### Experiment 1

Rohrer and Wixted (1994) showed that mean recall latency remained unchanged when study time for a list of six words was varied over a fairly wide range (even though recall probability changed considerably). This counterintuitive finding has not been replicated, and the first experiment reported here was designed to do so using item repetition as a means of strengthening. Theories differ as to whether a second presentation of an item strengthens the original representation or creates another copy of it. However, in either case, the predictions of the relative strength model are the same. For example, even if a second presentation of six items creates a second copy of each, the probability of an individual item being sampled is  $2/12$  or  $1/6$  (the same as when a single presentation is used). Thus, recall latency should still remain unchanged when the number of item presentations is varied (although recall probability should change considerably). Because item repetition was the method of strengthening used to produce mixed-strength lists in Experiment 2, it seemed important to first evaluate the accuracy of this prediction for pure-strength lists. Participants in this experiment studied six-item lists in which the items were presented once, twice, or three times in a randomly arranged order. A filled distractor task followed each list to minimize the contribution of items retrieved from short-term store. Each list was followed by a 30-s recall period during which responses were timed.

### Method

**Participants.** The participants were 15 undergraduates of the University of California, San Diego, who were enrolled in an introductory psychology course. Participation in the experiment satisfied a course requirement.

<sup>1</sup> Recency items are also recalled early in the output sequence if recall is immediate. However, that is presumably because those items are still in short-term store when the signal to begin recalling is given, not because they are stronger than other items in long-term store (e.g., Craik, 1970).

**Materials and design.** Lists were constructed by drawing words randomly without replacement from a pool of 200 monosyllabic nouns. A different random order was used for every participant. A session involved 15 study-test trials consisting of 6 items presented once, twice, or three times. Each repetition condition occurred five times during a session, and the order in which conditions occurred was randomized uniquely for each participant. For lists involving repeated items, repetitions were arranged randomly. If, for example, a list involved three presentations of each word (6 types, 18 tokens), the 18 tokens were arranged in a random sequence to form the list. Thus, repeated items were sometimes widely separated and sometimes adjacent, depending on the outcome of the random ordering.

**Procedure.** Participants were tested individually and were informed that the purpose of the experiment was to test their memory for words. A session consisted of 3 practice trials (1 trial for each repetition condition) and 15 experimental trials (5 trials for each repetition condition). A trial was initiated by a warning tone. Following the tone, words were presented one at a time at the center of a computer screen. Each word was displayed for 1 s and was followed by a 0.5-s interstimulus interval. Following the presentation of the last item, a distractor task was initiated. This task involved a series of 12 three-digit numbers displayed one at a time at the center of the screen for 1.5 s each. The participant was asked to read the three digits comprising each number out loud in ascending order. Thus, the number 6 8 2 was to be read "2 6 8." Following the 18-s distractor task, the signal to begin recalling items from the list was given. Participants were instructed to verbally recall as many items as possible in any order they wished. The recall period lasted 30 s, after which the next list was presented.

Recall responses were timed with a voice-activated relay attached to a millisecond timer. A research assistant was present at all times to take note of occasional misses (i.e., instances in which a response was missed by the voice key) and false alarms (i.e., instances in which the voice key was activated by an extraneous noise). Sessions were tape-recorded as well so that responses missed by the voice key could be timed after the session. Intrusions (i.e., recalled words that were not on the study list) occurred rarely and were excluded from analysis. That is, only recall latencies for correct items were analyzed. Note that including latencies for incorrect items would have had no bearing on the final conclusions.

## Results and Discussion

As expected, the proportion of items recalled increased monotonically as the number of presentations increased,  $F(2, 28) = 51.30$ ,  $MSE = 0.005$ . (An alpha level of .05 was used throughout.) The question of interest was whether recall latency was also affected. Figure 1 shows the obtained free-recall latency distributions for each condition. Each distribution shows the number of items recalled (summed over participants) in each 1-s bin of the 30-s recall period. Thus, the first point represents the number of words that were recalled in the first second of the recall period, the second point represents the number of words recalled in the next second of the recall period, and so on. Although the data shown in Figure 1 could have been plotted cumulatively, we preferred to analyze the noncumulative functions because they display data points that are independent.

The smooth curves in Figure 1 represent the best fitting

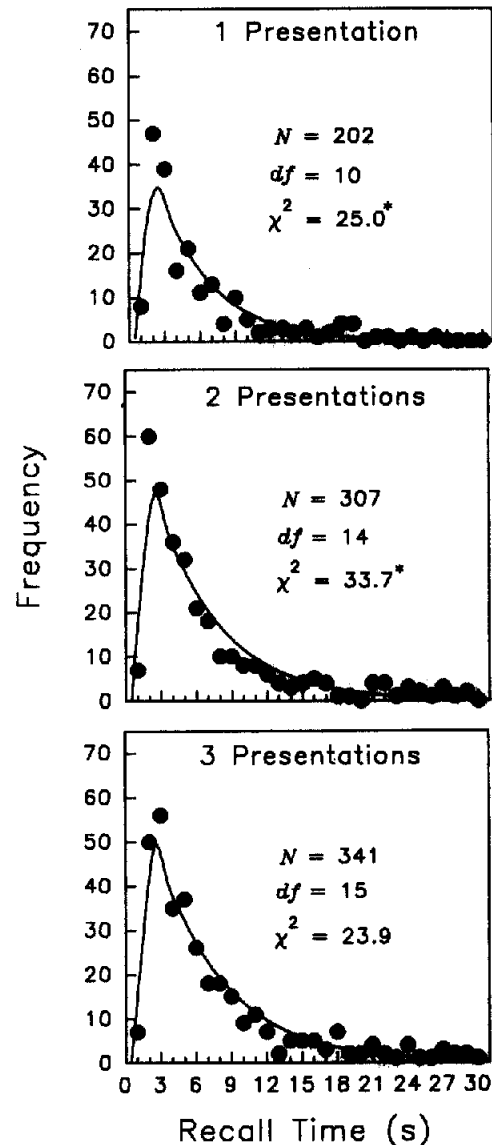


Figure 1. Recall-latency distributions for trials in which list items were presented once, twice, or three times. The solid curves represent the best fitting ex-Gaussian. The asterisk indicates a statistically significant deviation from the ex-Gaussian.

ex-Gaussian distribution, which has previously been shown to provide an adequate description of free-recall latency data (Rohrer & Wixted, 1994; Wixted & Rohrer, 1993). The ex-Gaussian distribution, which was fit using maximum likelihood estimation, involves three free parameters,  $\mu$ ,  $\sigma$ , and  $\tau$ . The first two represent the mean and standard deviation of a Gaussian stage of retrieval. This stage presumably reflects a brief period of time following the distractor task during which participants shift attention from recoding digits to recalling words. The values of these parameters were always small ( $\mu$  was typically about 1.25 s) and did not change systematically across conditions (which is true of prior free-recall studies as well). Thus, only values

of  $\tau$ , the mean of the exponential stage, are reported here. It is this parameter that essentially equals the inverse of  $\lambda$  in the cumulative exponential discussed earlier.<sup>2</sup>

Some deviation from the ex-Gaussian is evident early on, but the tails of the distributions (which determine the values of  $\tau$ ) appear to be adequately described by the exponential component. Table 1 shows the maximum likelihood parameter estimates of  $\tau$  for each condition, along with the corresponding proportion of items recalled. As already noted, mean recall probability increased monotonically with the number of item presentations. Unexpectedly, mean recall latency increased monotonically as well, with strong items paradoxically requiring slightly more time to recall than weaker items. However, a  $t$  test performed on the  $\tau$  values from the one- and three-presentation conditions did not quite reach statistical significance,  $t(24) = 1.90$ ,  $p = .069$ . This  $t$  test, like those described below, was based on the asymptotic standard errors associated with the maximum likelihood parameter estimates (Ratcliff & Murdock, 1976; Ratkowsky, 1983). This marginally significant increase in recall latency contrasts with the constant recall-latency data reported by Rohrer and Wixted (1994), who used increased study time (rather than item repetition) to strengthen items. However, it should be noted that even if the effect is real, the increase in recall latency is rather small compared with the large increase in the proportion of items recalled (which ranged from .37 to .63). When recall probability is varied over a similar range by varying list length instead of list strength, the effect on recall latency is much more dramatic, typically changing by a factor of more than two (Rohrer & Wixted, 1994).

A question that sometimes arises in connection with an analysis of recall latency concerns the possible role of articulation time. That is, perhaps recall latency would have been faster for stronger items, but the effect was masked by the extra time required to articulate the additional items recalled. However, articulation time should have a significant effect on recall latency only when participants are recalling at their maximum rate. Monosyllabic words require only a few hundred milliseconds to articulate, whereas mean recall latency was approximately 5 s. Thus, the absolute rate of recall was well below maximum. An analysis recently offered by Rohrer (1996) cast further doubt on a significant role for articulation time. He computed mean recall latency following lists of 8 or 16 items partitioned on the number of items recalled per trial. In both list-length conditions, participants sometimes recalled 4 items, sometimes 5, sometimes 6, and so on. Mean recall latency was

Table 2  
*Mean Recall Latency ( $\tau$ ) Values Partitioned on Number of Items Recalled Per Trial*

Number recalled	Presentation	
	1	2/3
2	3.48	3.68
3	4.36	5.48
4	4.59	5.90
5		5.30
6		5.93

Note. Empty cells indicate that those outcomes happened too rarely to obtain an estimate of  $\tau$ .

nearly constant within a list-length condition regardless of the number of items output. (Actually, it increased about 200 msec per item.) However, mean recall latency was always much longer following 16-item lists (about 11 s regardless of output total) than following 8-item lists (about 6 s regardless of output total). Thus, the number of items actually articulated per trial had a minimal effect on mean latency, whereas the number of items actually studied (which, theoretically, determined search set size) had a strong effect.

A similar analysis was performed on the data reported here, and the same result was observed. That is, mean recall latency did not change much as a function of the number of items actually recalled. Moreover, recall latency was still, if anything, slightly longer in the multiple-presentation conditions relative to the single-presentation condition. The results of an analysis performed on data partitioned by the number of items recalled per trial are shown in Table 2. (Data from the two- and three-presentation conditions were pooled because they yielded similar results.) Even when matched by the number of items actually recalled on a given trial, strong items were not recalled faster than weak items. Thus, the intuitively appealing notion that strong items will be recalled faster than weak items, independent of the strength of the other items on the list, is shown by these results to be incorrect.

In spite of the small increase in recall latency with increasing list strength, these data reinforce the conclusion that greater strength does not quicken recall latency for pure-strength lists. According to the relative strength model, the reason why recall latency did not decrease when the memory traces were strengthened by repetition is that all of the competitors were strengthened as well. Thus, the odds of sampling any one trace did not change. By contrast, the reason why the probability of recall increased substantially is that the extra study time changed the status of some items in the search set from nonrecoverable to recoverable (such

Table 1  
*Estimated Mean Recall Latency ( $\tau$ ) and Proportion of Items Recalled ( $p$ ) From Experiment 1*

Presentation	$\tau$	ASE	$p$
1	4.48	0.33	.37
2	5.36	0.32	.57
3	5.79	0.33	.63

Note. ASE = asymptotic standard error of parameter estimates.

<sup>2</sup> These data were also analyzed by fitting the exponential distribution (with least squares) to the Vincentized group reaction-time distributions (cf. Ratcliff, 1979). With the number of quantiles fixed at 10, the percentage of data variance accounted for was never lower than 90% across the four experiments. Conclusions based on the resulting estimates of  $\tau$  were identical to those based on the ex-Gaussian fits to the pooled reaction-time distributions.

that now a sampled trace actually supports the recovery of the item that created it).

## Experiment 2

The second experiment was designed to assess free-recall latency for strong and weak items in mixed lists relative to pure lists. Participants were exposed to study lists consisting of (a) six once-presented words, (b) nine once-presented words, and (c) three once-presented and three twice-presented words (i.e., six types, nine tokens). The nine-item list was included to evaluate the theoretical effect of presenting items twice in the mixed list. If a second presentation doubles the strength of an item (or, equivalently, creates a second copy of the item in the search set), then the dynamics of recall for weak items in the mixed list should be identical to the dynamics of recall for words from the nine-item list. In both cases, relative strength would be 1/9.

The main prediction of the relative strength model is that weak items will be recalled more slowly than strong items in the mixed list and that their values will straddle the value obtained from the six-item list. This kind of test has not been previously performed because the few relevant studies used relative output position as a dependent measure (e.g., Hastie, 1975; Tulving & Hastie, 1972). As such, no comparison with a pure-strength list was possible.

In contrast to the relative strength model, cognitive-triage theory often makes the counterintuitive prediction that some weak items will be recalled prior to strong items (Brainerd, 1995). In general, that prediction applies to multitrial free recall (in which some items are strengthened by virtue of being recalled), rather than to the single-trial free-recall task we used (in which some items were strengthened during list presentation). We nevertheless considered the possibility that weak items might be retrieved first by examining output order in addition to recall latency. The cognitive-triage pattern actually obtained by Brainerd and his colleagues (Brainerd et al., 1990) was weak-to-strong-to-weak (i.e., weak items are recalled first, followed by strong items, followed by additional weak items). As such, clear predictions about recall latency cannot be derived. That is, because some weak items are recalled quickly (prior to strong items) and others are recalled more slowly (well after strong items), the overall mean recall latency for weak items may or may not differ from that of strong items. Thus, output order for weak and strong items was also examined to determine whether or not a triage effect occurred (which could happen even if latency data supported the relative strength model).

The design of this experiment also provides another test of the *list-strength effect* in free recall (Ratcliff, Clark, & Shiffrin, 1990). The list-strength effect refers to the observation that weak (i.e., once-presented) items in a mixed list are less likely to be recalled than weak items in a pure list. In other words, the presence of strong items on the list in some way interferes with the recall of weak items. If that effect occurred here, then the recall of weak items from the mixed-strength list would be impaired relative to the recall of (the equally weak) items from the six-item list.

## Method

**Participants.** The participants were 15 undergraduates who were drawn from the same pool as before.

**Materials and design.** Lists were constructed by drawing words randomly without replacement from a pool of 200 monosyllabic nouns. A different random order was used for every participant. A session involved six study-test trials consisting of two six-item pure-strength lists, two nine-item pure-strength lists, and two six-item mixed-strength lists. The order of list presentation was randomly determined. In the two pure-strength lists, each item was presented once (i.e., these were pure-weak lists). In the mixed-strength list, three words were presented once (the weak items), and three were presented twice (the strong items). Items to be presented twice were selected randomly prior to the start of a trial. The nine tokens that made up the mixed-strength list were then arranged in a random order for presentation. Thus, repeated items were sometimes widely separated and sometimes adjacent, depending on the outcome of the random ordering.

**Procedure.** Participants were tested individually. A session consisted of three practice trials (one trial for each type of list) and six experimental trials (two trials for each type of list). A trial was initiated by a warning tone. Following the tone, words were presented one at a time at the center of a computer screen. Each word was displayed for 2 s and was followed by a 0.5-s interstimulus interval. Following the presentation of the last item, the distractor task was initiated. (This task was the same as that used in Experiment 1.) The recall period lasted 30 s, after which the next list was presented. Recall responses were timed as in Experiment 1.

## Results and Discussion

**Recall probability.** The strength manipulation in the mixed list (strong vs. weak) and the list-length manipulation (six vs. nine items) both had the anticipated effects on the probability of recall. Strong items in the mixed list were more likely to be recalled than weak items (.73 vs. .50, respectively),  $t(14) = 5.60$ , and the probability of recall for the six-item list exceeded that for the nine-item list (.58 vs. .50, respectively),  $t(14) = 2.78$ . In addition, a list-strength effect was observed for recall probability. That is, the probability of recall for weak items in the pure-weak six-item list (.58) exceeded that for weak items in the mixed-strength six-item list (.50),  $t(14) = 2.30$ . This effect was first observed some time ago by Tulving and Hastie (1972), and the effect was recently replicated by Ratcliff et al. (1990). Thus, its appearance here is not surprising.

**Recall latency.** Figure 2 shows the obtained free-recall latency distributions for each condition. The distributions for the strong and weak items are shown separately, but those items were recalled during the same recall periods following the mixed-strength lists. For these maximum likelihood fits of the ex-Gaussian,  $N$  (the number of items ultimately recalled) was allowed to vary as a free parameter, using the procedure described by Maindonald (1984). This was done because some of the distributions (e.g., for the weak items) were not quite at asymptote by the end of the recall period. By allowing  $N$  to vary as a parameter (symbolized by  $\hat{N}$ ), the ex-Gaussian fits yield a projected number of items recalled (given infinite recall time) and a projected latency for all items that would eventually be recalled. Note that fixing  $N$  at the number of items actually

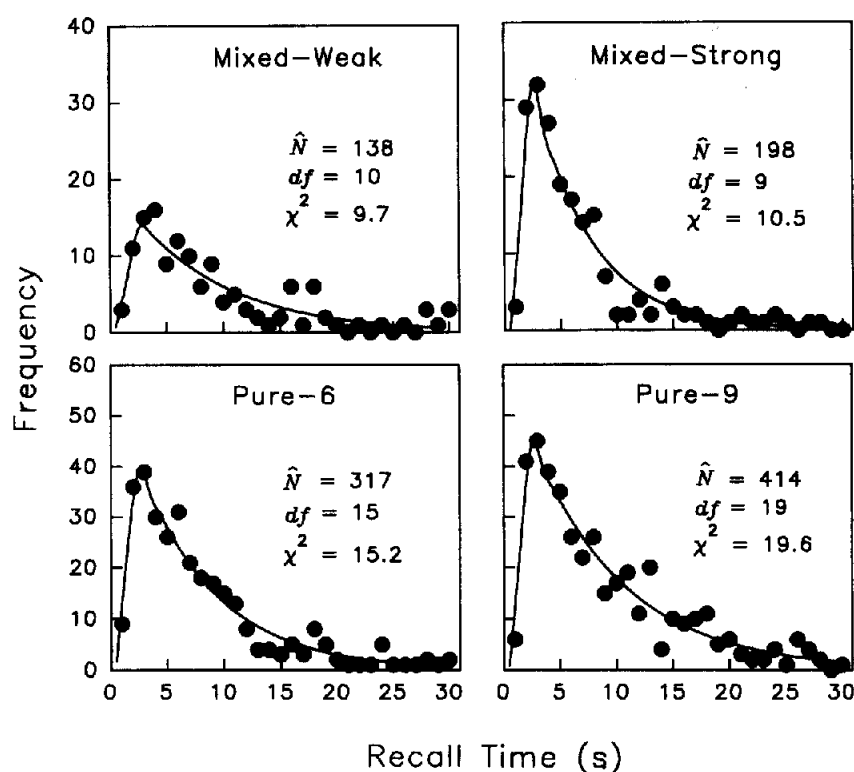


Figure 2. Recall-latency distributions for strong and weak items from the mixed-strength list and from the six- and nine-item lists. The solid curves represent the best fitting ex-Gaussians.  $\hat{N}$  = estimated number of items recalled given an infinite recall period.

recalled results in slightly poorer fits, but no conclusions are affected.

The question of interest is what effect the various manipulations had on mean recall latency ( $\tau$ ). Table 3 shows the maximum likelihood parameter estimates of  $\tau$  for each condition, along with the corresponding proportion of items recalled. The first point to note is that mean recall latency for the nine-item list was longer than that for the six-item list (8.12 s vs. 6.73 s, respectively). This list-length effect for recall latency has been observed before (Rohrer & Wixted, 1994) and is consistent with the basic relative strength model.

The most important result for present purposes is that the mean recall latency for strong items in the mixed list was significantly shorter than that for weak items (5.15 s vs. 8.33 s, respectively),  $t(19) = 2.85$ . The effect of item strength on

recall latency is actually clear from an examination of the latency distributions in Figure 2. Virtually all of the strong items were recalled within 20 s, but weak items were still being recalled toward the end of the 30-s recall period. This result is clearly consistent with the basic relative strength model.

Comparing the latency measures from the mixed condition with the measures from the two pure-strength lists reveals several interesting and theoretically significant results. First, just as predicted by the relative strength model, strong items from the mixed list ( $M_S$ ) were recalled more quickly and weak items ( $M_W$ ) more slowly than items from a pure-strength list of length 6 ( $P_6$ ), although only the former effect was statistically significant,  $t(24) = 2.39$ . Note that this statistical test is conservative because it does not take into account the fact that the observed pattern (which from fastest to slowest can be represented as  $M_S$ - $P_6$ - $M_W$ ) is one of six possible patterns ( $M_S$ - $M_W$ - $P_6$ ,  $M_W$ - $M_S$ - $P_6$ ,  $M_S$ - $P_6$ - $M_W$ ,  $M_W$ - $P_6$ - $M_S$ ,  $P_6$ - $M_S$ - $M_W$ , and  $P_6$ - $M_W$ - $M_S$ ) and thus would occur by chance with a probability of only .167 (1/6). With each participant's mean recall latency values, the predicted pattern (i.e., mean latency scores in the  $M_S$ - $P_6$ - $M_W$  order) was observed in 7 out of 15 participants,  $p = .007$  by a binomial test. No other pattern appeared more than three times in the remaining 8 participants.<sup>3</sup>

Table 3  
Estimated Mean Recall Latency ( $\tau$ ) and Proportion of Items Recalled ( $p$ ) From Experiment 2

Condition	$\tau$	ASE	$p$
Mixed strength			
3 weak	8.33	1.02	.50
3 strong	5.15	0.43	.73
Pure strength (weak)			
6 items	6.73	0.49	.58
9 items	8.12	0.55	.50

Note. ASE = asymptotic standard error of parameter estimates.

<sup>3</sup> A linear trend analysis performed on each participant's mean recall-latency scores ( $M_S$ ,  $P_6$ , and  $M_W$  conditions, respectively) was also highly significant,  $F(1, 14) = 8.03$ ,  $MSE = 4.03$ .

Table 3 also shows that the recall latency for the weak items in the mixed list essentially matched recall latency for the (equally weak) items in the nine-item list (8.33 s vs. 8.12 s, respectively). This result is consistent with the idea that the extra presentation of strong items in the mixed list doubled the strength of those items. That is, if the weak items had a strength of 1 and the strong items had a strength of 2, then the relative strength of a weak item would be 1/9 (where 9 is the sum of the strengths of all six items). Similarly, the strength of an item in a list of nine pure-strength items is also 1/9 (whether those items are all strong or, as was true here, all weak). This result also happens to be consistent with a multiple-copy model that assumes the extra presentation of a strong item creates another copy of that memory trace (rather than doubling the strength of the original trace). The two models, one of which assumes a doubling of strength and the other of which assumes the creation of a new copy, are mathematically identical.

**Output order.** Although the recall-latency data agreed with the predictions of the relative strength model, it is still possible that the weak-to-strong-to-weak output order was observed. This could occur, for example, if the long recall latencies associated with the weak items that appeared late in the output sequence were long enough to more than offset the short recall latencies associated with the weak items that appeared early in the output sequence (i.e., before the strong items). To investigate this possibility, the mean number of weak and strong items recalled at each output position was computed. The data were Vincentized by computing the proportion of strong and weak items recalled in each fourth of the output protocol for each participant and then averaging over participants. Fourths were used because all participants recalled at least some items in Output Positions 1 through 4 following six-item mixed-strength lists. The results of this analysis are presented in Table 4. Of the strong items that were recalled, 34% were recalled in the first quartile. By contrast, of the weak items that were recalled, only 21% were recalled in the first quartile. The reverse was true of last quartile where 28% of the weak responses occurred and only 13% of the strong responses occurred. Overall, the data indicate that the strong items tended to appear early in the output sequence, and the weak items tended to appear late in the output sequence (with no hint of a cognitive-triage effect). In other words, like the recall-latency data, the output-order data are consistent with the relative strength model.

Triage effects have never been examined using number of presentations as a measure of strength. Instead, strength is

usually determined on the basis of whether or not an item was recalled previously in a multitrial free-recall procedure. It is possible that the relative strength model provides accurate predictions when strength is defined by the amount of study an item receives during list presentation, whereas a triage effect is observed when strength is defined in other ways. The next two experiments investigated this possibility.

### Experiment 3

In a multitrial free-recall procedure, the same list of words is presented repeatedly (in a different random order each time), with recall tested following each presentation. In this case, strength was defined not by the number of times an item was presented (as in the previous two experiments) but by whether or not the item was recalled on the previous trial. Items recalled on the previous trial (and again on the current trial) were considered strong, whereas items being recalled for the first time were considered weak. That prior recall strengthens items in terms of increasing their future probability of recall is well established (Wheeler & Roediger, 1992).

The present experiment used a multitrial free-recall task to assess whether recall latency of previously recalled items differs from recall latency of items being recalled for the first time. If strategic factors are not involved, a relative strength model predicts that previously recalled items will tend to be recalled first (because those items are inherently stronger for the participant for idiosyncratic reasons and because they were strengthened by virtue of the fact that they were previously recalled). Vincentized output order measures were also computed to detect the possibility of a weak-to-strong-to-weak output pattern even if the latency data agreed with the relative strength model.

### Method

**Participants.** The participants were 18 undergraduates drawn from the same pool as before.

**Materials and design.** Lists were constructed by drawing words randomly without replacement from a pool of 200 monosyllabic nouns. A different random order was used for every participant. A session involved the presentation of five lists of 15 words (each presented three times).

**Procedure.** Participants were tested individually. Each of five 15-item lists was presented for three successive study-test trials, and the words were presented in a different random order each time. A trial was initiated by a warning tone. Following the tone, words were presented one at a time at the center of a computer screen. Each word was displayed for 2 s (with no interstimulus interval). Following the presentation of the last item, the distractor task was initiated. (This task was the same as that used in Experiment 1.) The recall period lasted 20 s, after which the next list was presented. Recall responses were timed as before.

### Results and Discussion

**Recall probability.** As expected, the mean proportion of items recalled increased from Trial 1 to Trial 3. On Trial 1, the proportion of items recalled was .25. On Trials 2 and 3, the proportion increased to .45 and .53, respectively. Of the items recalled on Trial 2, 43% were strong in the sense that

Table 4  
*Proportion of Weak and Strong Items Recalled in Each Output Quartile for the Mixed-Strength List*

Quartile	Weak	Strong
1	.210	.341
2	.269	.324
3	.244	.203
4	.276	.132



they were also recalled on Trial 1, and 57% were weak. On Trial 3, 65% were strong (having also been recalled on Trial 2), and 35% were weak.

**Recall latency.** Figure 3 shows the recall-latency distributions from Trials 1 through 3. The distributions for Trials 2 and 3 are plotted separately for strong and weak items, but these items were recalled during the same recall period. Because none of the items were recalled prior to Trial 1 (obviously), none of the items recalled on Trial 1 can be classified as strong (i.e., the first trial involves a pure-weak list). For these fits,  $N$  was again allowed to vary as a free parameter because it was clear that participants were not finished recalling weak items by the end of the recall period. Although recall was not complete in some cases, the distributions appear to be fairly well constrained by the data.

A quick examination of the recall-latency distributions immediately reveals what the maximum likelihood parameter estimates underscore: Strong items were recalled faster than weak items (the latter of which were still being recalled by the end of the 20-s recall period). The estimated  $\tau$  values are shown in Table 5. On both Trial 2 and Trial 3, weak items (those being recalled for the first time) had a much longer

Table 5

*Estimated Mean Recall Latency ( $\tau$ ) Values From Experiment 3*

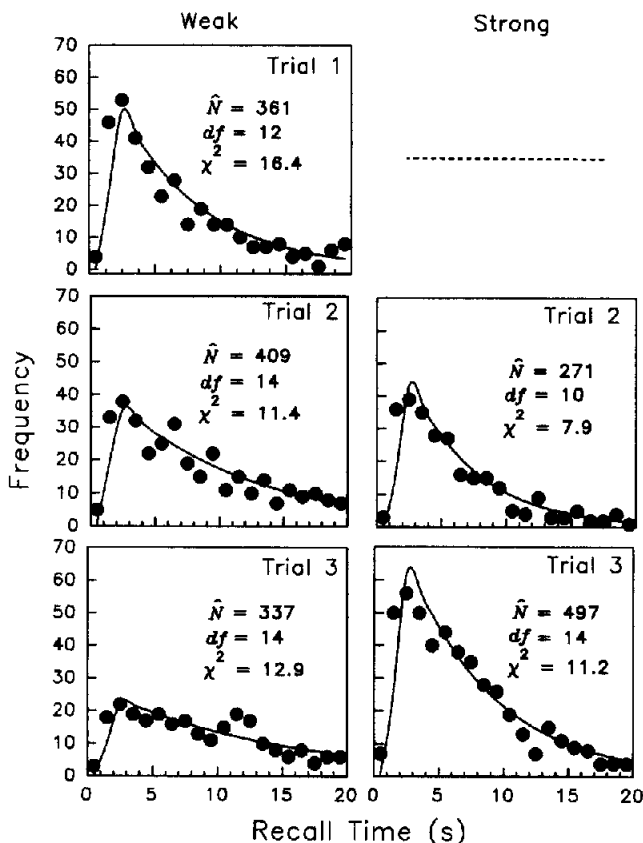
Trial	Weak		Strong	
	$\tau$	ASE	$\tau$	ASE
1	6.27	0.53	—	—
2	10.33	1.10	5.15	0.44
3	13.45	1.79	6.57	0.49

*Note.* Dashes indicate that by definition, strong items do not exist on Trial 1. ASE = asymptotic standard error of parameter estimates.

recall latency than strong items (those that were also recalled on the prior trial). This finding is reminiscent of data from the hypermnnesia literature in which participants were asked to recall the same list several times in succession (though the list was not repeated). Roediger and Thorpe (1978) tracked cumulative recall by using this procedure and found that on the second recall test, items recalled on the first test were recalled faster than items that were newly recalled on the second test.

The data from Trial 2 also replicated the compelling evidence from Experiment 2 in favor of a relative strength model in that strong items were recalled faster than (and weak items slower than) the pure-strength items from Trial 1. That is, whereas the value of  $\tau$  on Trial 1 was 6.27 s, the corresponding values for strong and weak items on Trial 2 were 5.15 and 10.33 s, respectively. On Trial 3, however, the data did not conform exactly to the predictions of the relative strength model. As the number of strong items in the search set increased, recall latency for both weak and strong items should have increased relative to Trial 2. The weak items should have been further delayed because the search set on Trial 3 contained a higher proportion of strong items relative to Trial 2, thereby placing the weak items at an even greater disadvantage. The strong items should also have been delayed relative to Trial 2 as the list approached being a pure-strong list (at which point recall latency should have matched that of Trial 1). However, whereas recall latency did increase for both weak and strong items on Trial 3, the value for the strong items already slightly surpassed the value obtained on Trial 1. This detail was not predicted by the relative strength model, and it resembles the unexpected (and relatively small) increase in recall latency for pure-strong lists relative to pure-weak lists in Experiment 1. Nevertheless, overall, the results support the idea that recall latency is a function of an item's strength relative to other items in the list.

**Output order.** The data presented above show no evidence of a cognitive-triage effect. Once again, however, the possibility remains that a weak-to-strong-to-weak output order was actually obtained but was masked because the later weak retrievals pulled mean recall latency for those items above that for strong items. To investigate this possibility, output order (which is the usual method used to identify the cognitive-triage pattern) was examined in detail. The data were Vincentized by computing the proportion of



**Figure 3.** Recall-latency distributions for Trials 1 through 3. For Trials 2 and 3, the data are separated on the basis of whether or not the items were recalled (strong) or not (weak) on the preceding trial. Recall-latency distributions for weak and strong items are shown in the left and right columns, respectively. The solid curves represent the best fitting ex-Gaussians.  $N$  = estimated number of items recalled given an infinite recall period.

strong and weak items recalled in each fourth of the output protocol as before. The results of this analysis are shown in Table 6. Once again, the data show no evidence of cognitive triage and instead show just what the relative strength model would predict. On Trial 2, 32% of the strong items that were recalled appeared in the first quartile. By contrast, only 24% of the weak items that were recalled appeared in that quartile. On Trial 3, the same pattern was apparent, with a higher proportion of strong items appearing early in the output sequence.

**Conflicting findings.** Although our results exhibit strong output priority, a number of prior studies have produced the opposite result. Brainerd et al. (1990) and Brainerd, Reyna, et al. (1993) presented considerable evidence in support of cognitive-triage theory and concluded that their counterintuitive results contradict a relative strength model of free recall. Their analyses of output order reliably showed a weak-to-strong-to-weak pattern (i.e., weak-output priority) in contrast to the strong-to-weak order consistently obtained here. Most of the studies showing the triage effect used a multitrial free-recall procedure, but even that procedure did not reproduce the effect here. Using a very similar procedure, Shuell and Keppel (1968) also did not find evidence for a weak-to-strong output order. How are these two apparently opposite sets of findings to be reconciled?

A study by Rundus (1974) may help to shed some light on this issue. That experiment also used a multitrial free-recall procedure and also found evidence for a weak-to-strong output order on Trials 2 and 3, just as Brainerd and his colleagues (Brainerd, Reyna, et al., 1993) usually have. By asking participants to rehearse overtly during list presentation, however, Rundus was able to show that the most rehearsal on Trials 2 and 3 was devoted to a subset of the items not recalled on Trial 1 (i.e., to a subset of the weak items). In particular, participants tended to devote a large amount of rehearsal to the weak items that happened to appear at the beginning of the list. The weak items that appeared in the middle or at the end of the list received, if anything, less rehearsal than the strong items appearing in those list positions. The rehearsal data from this experiment are reproduced in Table 7.

It is important to note that the weak items that received extra rehearsal were precisely the ones output early in the recall sequence on Trials 2 and 3. This was true even though the analysis excluded any item rehearsed in the 10 s prior to recall (i.e., even when the analysis excluded items that may have been retrieved from short-term store). On the basis of

Table 6  
*Proportion of Items Recalled in Each Output Quartile in Experiment 3*

Quartile	Trial 2		Trial 3	
	Weak	Strong	Weak	Strong
1	.236	.317	.181	.300
2	.283	.369	.287	.355
3	.291	.213	.299	.211
4	.190	.100	.232	.133

Table 7  
*Mean Normalized Number of Rehearsals as a Function of Serial Position, From Rundus (1974)*

Serial position block	Trial 2		Trial 3	
	Strong	Weak	Strong	Weak
1	5.32	6.60	4.60	7.80
2	3.23	3.22	3.07	3.52
3	2.98	2.34	2.63	2.60
4	2.76	1.91	2.74	2.09

*Note.* Serial Position Blocks 1 and 4 represent the first and last 5 serial positions, respectively, and Blocks 2 and 3 represent the middle 20 serial positions. Adapted from "Output Order and Rehearsal in Multi-Trial Free Recall," by D. Rundus, 1974, *Journal of Verbal Learning and Verbal Behavior*, 13, p. 660. Copyright 1974 by Academic Press. Adapted with permission.

this result, Rundus (1974) concluded that it is a mistake to think of the items recalled early in the output sequence as being weak. Instead, they were transformed into strong items during list presentation by means of differential rehearsal. The weak items that were not differentially rehearsed appeared late in the output sequence (presumably because they remained weak). This finding appears to be another example of a well-known phenomenon: The initial items of a list tend to receive the most elaborative processing (accounting for much of the primacy effect), and they tend to be retrieved early in the recall sequence (Bousfield et al., 1958). The only difference here, apparently, is that weak items at the beginning of the list were rehearsed to an even greater extent than strong items at the beginning of the list.

Why was the same weak-to-strong pattern not observed by Shuell and Keppel (1968) or in Experiment 3 here? Probably because the list items were presented at a relatively fast rate, thereby decreasing the opportunity to differentially rehearse. Whereas Rundus (1974) presented items at a rate of one item every 5 s, we (like Shuell & Keppel, 1968) presented them at a rate of one item every 2 s. In fact, we used relatively fast presentation times in this experiment (as well as in Experiments 1 and 2) for the very reason that this serves to minimize uncontrolled strength variations resulting from surreptitious rehearsal. That fast presentation times accomplish this is clear from both common sense and prior research. Brodie and Prytulak (1975), for example, exposed participants to lists of 18 words and asked them to rehearse aloud during list presentation. The words were presented at a rate of one item every 1.25, 2.5, or 5 s. The results of this study are reproduced in Figure 4.

Obviously, rehearsal time per item decreased substantially as the rate of presentation increased. Less obviously, the amount of rehearsal associated with items in the early serial positions (the positions in which weak items tend to be transformed into strong ones) dropped especially fast with increases in the rate of presentation. From these data, one can reasonably assume that differential rehearsal of weak items was less likely to occur in our experiment (in which words appeared at a rate of one every 2 s) than in the one reported by Rundus (1974). As with Rundus's research, Brainerd and his colleagues (Brainerd, Reyna, et al., 1993)

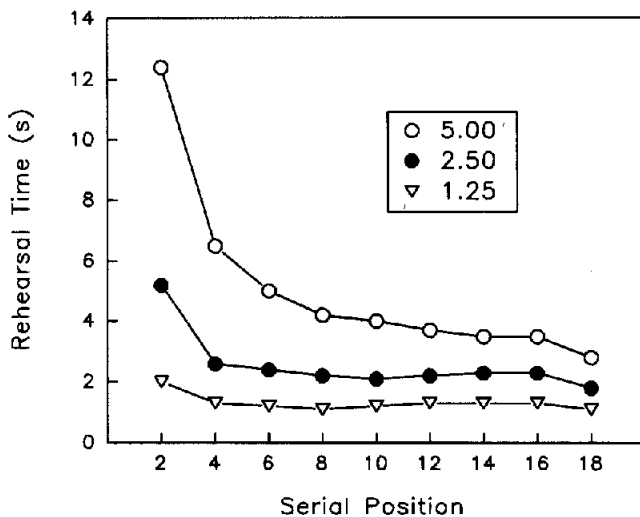


Figure 4. Mean rehearsal time as a function of serial position and presentation time. The data are from "Free Recall Curves: Nothing but Rehearsing Some Items More or Recalling Them Sooner?" by D. A. Brodie and L. S. Prytulak, 1975, *Journal of Verbal Learning and Verbal Behavior*, 14, p. 553. Copyright 1975 by Academic Press. Adapted with permission.

often used presentation rates of one item every 5 s (conditions that are more conducive to special processing) and always used a rate of at least one item every 3 s, which may explain why they more reliably obtained the weak-to-strong output pattern. The final experiment reported here tested this notion by exposing participants to lists that were presented rapidly or slowly in a multitrial free-recall procedure.

#### Experiment 4

Participants in this experiment were exposed to four lists of 15 items. Each list was presented for three successive study-test trials, just as in Experiment 3. For two lists, the words were presented at a rapid rate (one word per 1 s) on all three trials, and for two other lists, the words were presented at a slow rate (one word per 5 s) on all three trials. The prediction was that strong items (those recalled on previous trials) would be recalled earlier than weak items in the fast condition only. In the slow condition, some weak items were expected to be recalled before strong items because participants would be given the opportunity to differentially rehearse.

#### Method

**Participants.** The participants were 36 undergraduates drawn from the same pool as before.

**Materials and design.** Lists were constructed by drawing words randomly without replacement from a pool of 200 monosyllabic nouns. A different random order was used for every participant.

**Procedure.** A session involved the presentation of four lists of 15 words (each presented three times). Two lists were presented at a fast rate (one word per 1 s) and two at a slow rate (one word every 5 s) in a randomly determined order. The recall period was 45 s

instead of 20 s. Otherwise, the procedure was identical to that of Experiment 3.

#### Results and Discussion

**Recall probability.** The mean proportion of items recalled increased from Trial 1 to Trial 3 in both the fast and slow conditions. In the fast condition, the mean proportions of items recalled were .234, .442, and .557 for Trials 1, 2, and 3, respectively. The corresponding values for the slow condition were .462, .658, and .738.

On Trials 2 and 3 of the fast condition, .430 and .663 of the recalled items were strong in the sense that they were also recalled on the immediately preceding trial. On Trials 2 and 3 of the slow condition, .582 and .754 of the recalled items were strong.

**Recall latency.** Although the predictions of interest in this experiment pertain mainly to output order, the recall-latency distributions were analyzed as well. The main results are summarized in Figures 5A and 5B, which show the recall-latency distributions for the fast and slow conditions, respectively. Closed symbols are used to represent data from the fast condition (Figure 5A), and open symbols are used to represent data from the slow condition (Figure 5B) for the sake of clarity. As before, the fits appear to be adequate. Table 8 presents the estimates of recall latency ( $\tau$ ) that correspond to the distributions shown in Figures 5A and 5B. The results from the fast condition replicate the results from Experiment 3 in every respect. On Trial 2, strong items were recalled faster than (and weak items slower than) the pure-strength items from Trial 1. That is, whereas the value of  $\tau$  on Trial 1 was 8.80 s, the corresponding values for strong and weak items on Trial 2 were 6.08 and 12.05 s, respectively. On Trial 3, recall latency for strong items was still much shorter than that for weak items (8.80 and 13.85 s, respectively), but these values were equal to or exceeded the value obtained on Trial 1. On the whole, these results are very much in accord with the predictions of the relative strength model.

The latency results from the slow condition show the same pattern, although the effects are attenuated. On Trial 2, the latency values for strong and weak items (10.40 and 12.13 s, respectively) straddle the latency value obtained on Trial 1 (11.09 s). On Trial 3, overall latencies increased, and the advantage for strong items was more apparent (11.21 vs. 15.56 s).

Note that in this experiment and in Experiment 3, weak and strong items were defined by whether or not they were recalled on the prior trial. However, strength can also be defined by rate of presentation. On Trial 1, for example, the fast and slow manipulation can be construed as a pure-strength manipulation (as in Experiment 1). That manipulation increased the proportion of items recalled from .234 to .462 (or from 3.51 words per trial to 6.93 words per trial). An ideal relative strength model predicts a constant recall latency, except perhaps for a 250-ms or so increase for each extra word recalled. The actual values were 8.80 s and 11.09 s for the fast and slow conditions, respectively. Thus, as with Experiment 1, the pure-strength manipulation once again

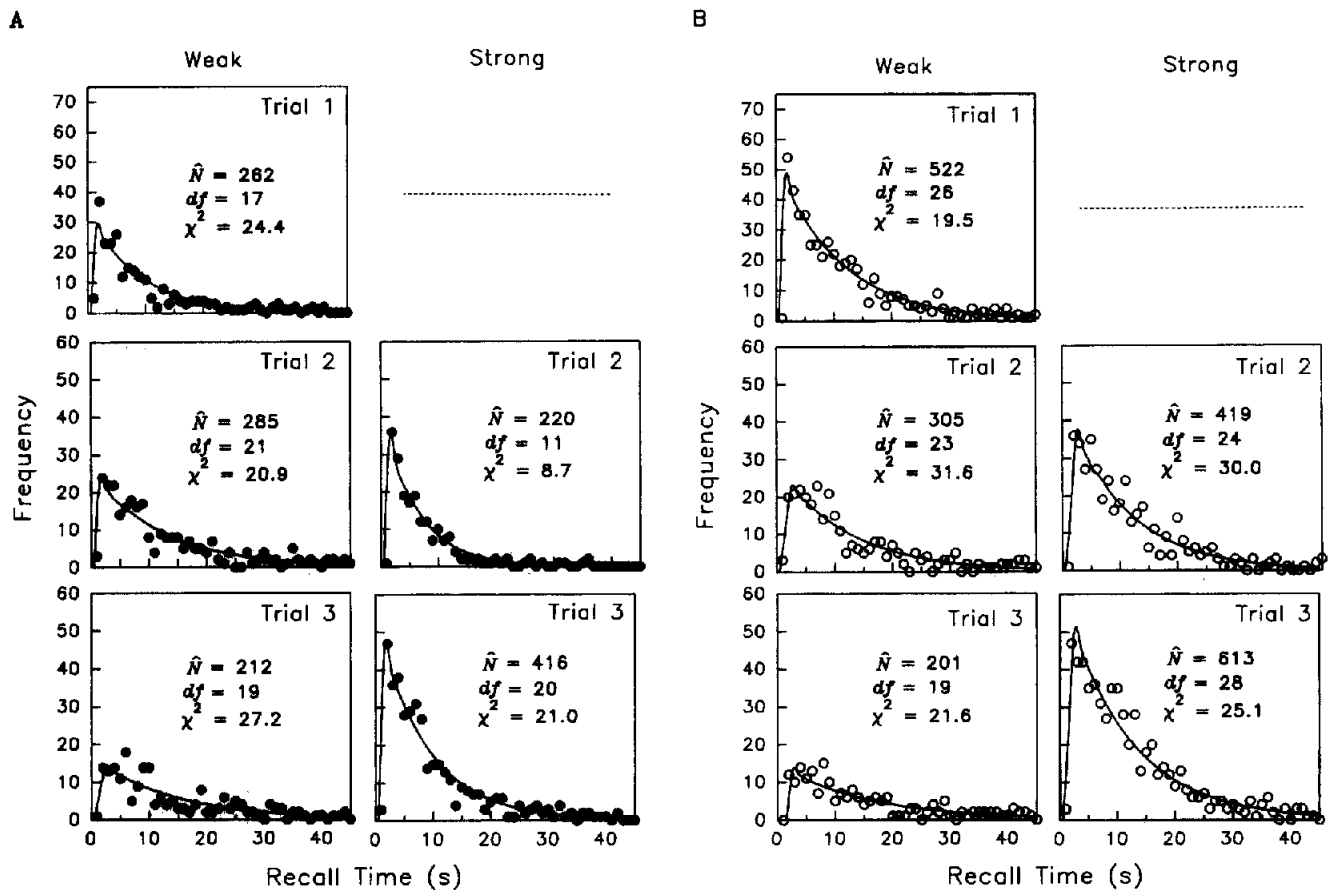


Figure 5. A: Recall-latency distributions for Trials 1 through 3 of the fast condition in Experiment 4. For Trials 2 and 3, the data are separated on the basis of whether the items were recalled (strong) or not (weak) on the preceding trial. Recall-latency distributions for weak and strong items are shown in the left and right columns, respectively. The solid curves represent the best fitting ex-Gaussians. B: Recall-latency distributions for Trials 1 through 3 of the slow condition in Experiment 4. Results for weak and strong items are shown in the left and right columns, respectively.  $\hat{N}$  = estimated number of items recalled given an infinite recall period.

increased recall latency to some extent—the opposite of what intuition might suggest would happen. Conceivably, this result occurred because participants were inclined to search for a longer period of time in the slow condition because they were overall more successful. Participants in

the fast condition may have sometimes given up the search prematurely (because relatively few items were coming to mind), thereby lowering average recall latency. In any case, as before, the pure-strength manipulation did not decrease mean recall latency.

Table 8  
Estimated Mean Recall Latency ( $\tau$ ) Values  
From Experiment 4

Condition and trial	Weak		Strong	
	$\tau$	ASE	$\tau$	ASE
Fast				
1	8.80	0.61	—	—
2	12.05	0.92	6.08	0.43
3	13.85	1.39	8.80	0.49
Slow				
1	11.09	0.60	—	—
2	12.13	0.93	10.40	0.63
3	15.56	1.65	11.21	0.58

Note. Dashes indicate that by definition, strong items do not exist on Trial 1. ASE = asymptotic standard error of parameter estimates.

**Output order.** The main motivation for this experiment was to test the hypothesis that strong items would appear early in the output sequence in the fast condition and that the reverse would be true in the slow condition. Table 9 presents the relevant Vincentized output order data. Once again, the data show no evidence of cognitive triage in the fast condition and instead show just what the relative strength model predicts. On Trial 2, 31% of the strong items that were recalled appeared in the first quartile. By contrast, only 14% of the weak items that were recalled appeared in that quartile. On Trial 3, the same pattern was apparent. On both trials, a preponderance of weak items occupied the last output quartile. In the slow condition, a very different pattern emerges. On Trial 2, 26% of the weak items and 26% of the strong items appeared in the first quartile. That is, no advantage for strong items was observed. On Trial 3, a slight advantage for the strong items was again apparent (.220 vs.

Table 9  
Proportion of Items Recalled in Each Output Quartile  
in Experiment 4

Condition and quartile	Trial 2		Trial 3	
	Weak	Strong	Weak	Strong
Fast				
1	.144	.312	.153	.283
2	.304	.326	.286	.289
3	.212	.237	.261	.241
4	.339	.125	.300	.186
Slow				
1	.256	.256	.220	.245
2	.269	.295	.250	.283
3	.242	.241	.237	.257
4	.233	.208	.293	.215

.245). Thus, for the first time in the present series of experiments, no clear advantage for strong items occurred in a mixed-strength situation.

Figure 5 shows the results from the first quartile averaged across Trials 2 and 3. An analysis of variance (ANOVA) performed on these data revealed main effects for strength,  $F(1, 35) = 7.99$ ,  $MSE = 0.029$ , as well as a significant interaction between strength and rate of presentation,  $F(1, 35) = 7.62$ ,  $MSE = 0.020$ . This interaction merely underscores what is apparent in the figure: Strong items dominate the early output positions in the fast condition but not in the slow condition. Indeed, a  $t$  test performed on the data from the slow condition in Figure 6 did not even approach significance. Findings like these (i.e., no advantage for either strong or weak items) have been reported before and have been taken to refute the basic assumption of the relative strength model that strong items are recalled sooner than weak items (Brainerd, Reyna, Howe, & Kevershan, 1991, Experiment 3).

The significant interaction illustrated in Figure 6 confirms that rate of presentation is an important factor to consider when testing the predictions of the relative strength model

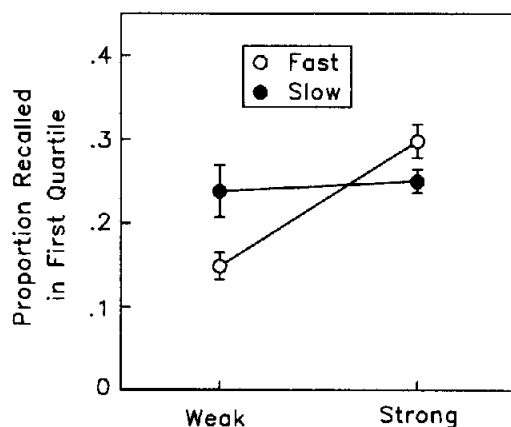


Figure 6. Proportion of strong and weak items recalled in the first output quartile in the fast and slow conditions of Experiment 4. The error bars represent standard errors.

with multitrial free recall. When the rate of presentation is slow, the usual advantage of strong items over weak ones disappears. The most likely explanation for this result is that in the slow condition only, participants were able to strengthen previously unrecalled items by devoting extra rehearsal to some of them. In the fast condition, differential rehearsal was much more difficult, especially in light of the fact that participants were required to read each word aloud as it appeared on the screen.

On the other hand, the use of a slow presentation rate still did not result in weak items being output sooner than strong items, which is the pattern most often observed by Brainerd and his colleagues (Brainerd et al., 1990). Perhaps the effect would have been observed had additional trials (beyond three) been conducted. Brainerd et al. (1991) showed that the triage effect is relatively small after the first few study-test trials and becomes increasingly apparent up to six study-test trials. In our experiment, the rehearsal that theoretically served to strengthen weak items in the slow condition for some reason did not actually strengthen them beyond that of the strong items.

Although we did not actually see a weak-item advantage in the slow condition, we did continue to see a preponderance of weak items in the last quartile in the slow condition (as shown in Table 8). This result has also been commonly observed by Brainerd and his colleagues (Brainerd et al., 1990). Our interpretation of the overall pattern of results from the slow condition is that the weak items that appeared in the first quartile along with the strong items were the ones that received differential rehearsal. The weak items that occurred late in the output sequence were the ones that did not.

### General Discussion

The present set of experiments was concerned mainly with evaluating the most basic relative strength model of free recall with regard to its predictions about whether and when strong items are recalled faster than weak items. The relative strength model predicts that in pure-strength lists, strong items will not be recalled faster than weak items but that in mixed-strength lists, they will be. Moreover, in the latter case, strong items should be recalled faster than (and weak items slower than) items from a pure-strength list of the same size. The only way to test these predictions is by measuring free-recall latency. Experiment 1 showed that, indeed, strong items were not recalled faster than weak items in pure-strength lists. (If anything, the strong items were recalled somewhat more slowly.) In Experiment 2, the recall latency for strong items (5.15 s) was shorter than that of weak items (8.33 s) in a mixed-strength list, whereas the corresponding pure-strength list yielded an intermediate latency value (6.73 s). This is exactly the pattern of results predicted by the relative strength model. In Experiment 3, items recalled (and presumably strengthened) on previous trials of a multitrial free-recall test were recalled faster than items recalled for the first time (i.e., weak items). Moreover, on the second study-test trial, the recall latency for strong items was shorter than (and for weak items was longer than) the recall latency for items recalled on the first trial, which

involved a pure-strength list. Again, this distinctive pattern of results is uniquely predicted by a relative strength model, which is the kind of model that forms the core of SAM's account of free recall.

Measures of output order following the mixed-strength lists in Experiments 2 through 4 tended to corroborate the conclusions based on recall latency. That is, a relatively high percentage of strong items tended to occupy the first output quartile, but a relatively low percentage of weak items appeared there. The opposite was true of the last output quartile. An exception occurred in Experiment 4 when the list items in a multitrial free-recall procedure were presented at a slow rate. In that case, the percentage of strong and weak items occupying the first output quartile was the same. On the surface, the absence of an advantage for strong items seems inconsistent with the predictions of the relative strength model.

Most prior research has evaluated the effect of strength on output order using relatively slow presentation rates in the context of multitrial free recall (see Brainerd, 1995, for a review). Much of that research shows that early in the output sequence, weak items actually have an advantage over strong items, exactly the opposite of what a relative strength model predicts. The present research suggests that rate of item presentation may be largely responsible for this. When items were presented rapidly, the usual strong-item advantage was robust. When items were presented slowly, we found no strong-item advantage (although we did not actually see a weak-item advantage). One possible reason for this pattern is that for slowly presented lists only, participants have an opportunity to differentially rehearse previously unrecalled items, thereby transforming them into strong ones. Thus, findings that are often taken to disconfirm the relative strength model may not contradict that theory after all (cf. Rundus, 1974).

Except in the case of multitrial free recall with slowly presented items, the results reported here appear to be inconsistent with the predictions of cognitive-triage theory (Brainerd, 1995). That theory states that some weak items are made available to conscious awareness prior to strong items following error-success feedback (as in multitrial free recall). Theoretically, this pattern reflects a basic retrieval mechanism that is designed to maximize the efficiency of free recall by giving priority to the more vulnerable traces. The following quotation from Brainerd et al. (1990) captures the essence of this model:

Output interference is minimized at the onset of a recall test, which encourages children to read out weaker words. As such words are recalled, however, output interference quickly builds, thereby selectively inhibiting the recall of further weak words. This encourages a switch to stronger words, which produces rapid growth of episodic activation and causes the output interference level to stabilize. Eventually, the system of episodic relationships becomes so thoroughly primed by the recall of stronger words that the stage is set for a return to recalling the remaining weak words. (p. 443)

That study involved children, but other research by Brainerd and his colleagues (e.g., Brainerd et al., 1991) shows that the

same effect is observed in adults. Although this theory might be able to explain why weak items overcame their usual disadvantage in the slow condition of Experiment 4, it cannot easily explain the strong-item advantage that is observed when list items are presented at a rapid rate. Perhaps a modification of the theory will eventually be able to do so. For the moment, a simpler explanation is that recall proceeds according to relative strength, as assumed by SAM, but weak items are differentially strengthened by rehearsal when lists are presented in such a way as to allow that to happen.

Brainerd and his colleagues (Brainerd et al., 1990, 1991; Brainerd, Reyna, et al., 1993) have repeatedly addressed this weak-item rehearsal account of cognitive triage and have always rejected it. Recently, for example, they investigated this issue experimentally by specifically instructing children to rehearse all of the items not recalled on the prior trial of a multitrial free-recall task and to avoid rehearsing items that were recalled on the prior trial (Brainerd, Olney, & Reyna, 1993). That is, participants were specifically instructed to devote special processing to the weak items, which were underlined on Trial 2 for easy identification. If differential rehearsal of weak items accounts for their early output, then this manipulation should exaggerate the effect. Far from increasing the weak-to-strong output pattern, they found that this manipulation almost eliminated it. Partly on the basis of this result, Brainerd, Olney, and Reyna (1993) argued that differential processing cannot account for the burst of weak items that ordinarily occurs early in the output sequence.

Although counterintuitive, this finding is actually just what a relative strength account would predict. A concrete example helps to illustrate why this is so. Imagine that 8 items of a 16-item list were recalled on Trial 1. For purposes of this hypothetical example, the 8 nonrecalled (and therefore weak) items are assigned a strength of 1, and the 8 recalled (and therefore strong) items are assigned a strength of 2. On Trial 2, suppose that participants devote special processing to 2 of the weak items that happen to appear in the early serial positions, thereby increasing their strengths to 4. Because these formerly weak items are now the strongest items, they should appear early in the output sequence. The remaining weak items (which still have a strength of 1) appear late in the output sequence. This example is consistent with what Rundus (1974) found and with the theory he advanced to explain that finding.

Now imagine that instead of a free-rehearsal procedure, a controlled-rehearsal procedure was implemented on Trial 2. That is, just as in the study conducted by Brainerd, Olney, and Reyna (1993), imagine that participants were required to distribute their rehearsal across all eight weak items instead of concentrating their efforts on the two weak items occupying early serial positions. Because rehearsal would be distributed, the mean increase in strength per item rehearsed should be correspondingly reduced compared with the free-rehearsal case. If all of the weak items were increased in strength to only 2 each (such that they now matched the strong items), then the experimental manipulation would

have served to create what is essentially a pure-strength list. In that case, the nominally weak items would have no advantage over the nominally strong items. As such, neither a weak-to-strong nor a strong-to-weak pattern would be observed (which was essentially the result reported by Brainerd, Olney, and Reyna, 1993).

Even if the weak items were strengthened beyond the strong items in the controlled-rehearsal procedure, the relative advantage of weak items over strong ones in the first output quartile would still be reduced relative to the free-rehearsal case. That is, many formerly weak items with a strength of, say, 3 would have less of an advantage over the strong items in terms of early output priority than a few formerly weak items with a strength of 4. A modified controlled-rehearsal procedure that required participants to concentrate their rehearsal efforts only on the first two weak items that appeared in the list should once again yield the weak-to-strong-to-weak output pattern (one about as strong as that usually observed if that is about what participants usually do anyway). That experiment has not yet been performed.

The issues discussed above are certainly open to debate. To take one example, weak output priority is often observed in second-grade children. Although some argue that young children engage in active processing of list items (e.g., Sodian, Schneider, & Perlmutter, 1986; Steinmetz & Battig, 1969), others disagree (e.g., Brainerd, 1995). The point of this discussion is not to finally settle the question of relative strength versus cognitive triage but to raise the possibility that the relative strength model is more viable than many recent findings would appear to suggest. Further research into this issue seems warranted given that the one major model of memory designed to handle the dynamics of free recall (SAM) depends on the relative strength rule.

The main implication of the experiments reported here is that in spite of recent evidence to the contrary, the relative strength model offers a compelling account of the dynamics of free recall. Indeed, when strength is varied by study time (Rohrer & Wixted, 1994) or by item repetition (as was done here), the predictions of the basic relative strength model with respect to mean recall latency are remarkably accurate. Strong items are not recalled faster than weak items in pure-strength lists, but strong items are recalled faster in mixed-strength lists. Moreover, strong items in a mixed list are recalled faster than (and weak items slower than) items from a pure-strength list of the same size. All of these findings suggest that to a large extent, recall operates according to a relative strength rule.

## References

- Bjork, R. A., & Whitten, W. B. (1974). Recency-sensitive retrieval processes in long-term free recall. *Cognitive Psychology*, 6, 173-189.
- Bousfield, W. A., & Barclay, W. D. (1950). The relationship between order and frequency of occurrence of restricted associative responses. *Journal of Experimental Psychology*, 40, 643-647.
- Bousfield, W. A., Cohen, B. H., & Silva, J. G. (1956). The extension of Marbe's law to the recall of stimulus words. *American Journal of Psychology*, 69, 429-433.
- Bousfield, W. A., & Sedgewick, C. H. W. (1944). An analysis of restricted associative responses. *Journal of General Psychology*, 30, 149-165.
- Bousfield, W. A., Whitmarsh, G. A., & Esterson, J. (1958). Serial position effects and the "Marbe effect" in the free recall of meaningful words. *Journal of General Psychology*, 59, 255-262.
- Brainerd, C. J. (1995). Interference processes in memory development: The case of cognitive triage. In F. N. Dempster & C. J. Brainerd (Eds.), *Interference and inhibition in cognition* (pp. 105-139). San Diego, CA: Academic Press.
- Brainerd, C. J., Olney, C. A., & Reyna, V. F. (1993). Optimization versus effortful processing in children's cognitive triage: Criticisms, reanalyses, and new data. *Journal of Experimental Child Psychology*, 55, 353-373.
- Brainerd, C. J., Reyna, V. F., Harnishfeger, K. K., & Howe, M. L. (1993). Is retrievability grouping good for recall? *Journal of Experimental Psychology: General*, 122, 249-268.
- Brainerd, C. J., Reyna, V. F., Howe, M. L., & Kevershan, J. (1990). The last shall be first: How memory strength affects children's retrieval. *Psychological Science*, 1, 247-252.
- Brainerd, C. J., Reyna, V. F., Howe, M. L., & Kevershan, J. (1991). Fuzzy-trace theory and cognitive triage in memory development. *Developmental Psychology*, 27, 351-369.
- Brodie, D. A., & Prytulak, L. S. (1975). Free recall curves: Nothing but rehearsing some items more or recalling them sooner? *Journal of Verbal Learning and Verbal Behavior*, 14, 549-563.
- Craik, F. I. M. (1970). The fate of primary memory items in free recall. *Journal of Verbal Learning and Verbal Behavior*, 9, 143-148.
- Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review*, 91, 1-67.
- Hastie, R. (1975). Intralist repetition in free recall: Effects of frequency attribute recall instructions. *Journal of Experimental Psychology: Human Learning & Memory*, 104, 3-12.
- MacLeod, C. M., & Nelson, T. O. (1984). Response latency and response accuracy as measures of memory. *Acta Psychologica*, 57, 215-235.
- Maindonald, J. H. (1984). *Statistical computation*. New York: Wiley.
- McGill, W. J. (1963). Stochastic latency mechanisms. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. 1, pp. 309-360). New York: Wiley.
- Raaijmakers, J. G. W., & Shiffrin, R. M. (1980). SAM: A theory of probabilistic search of associative memory. In G. H. Bower (Ed.), *Psychology of learning and motivation: Advances in research and theory* (Vol. 14, pp. 207-262). New York: Academic Press.
- Ratcliff, R. (1979). Group reaction time distributions and an analysis of distribution statistics. *Psychological Bulletin*, 86, 446-461.
- Ratcliff, R., Clark, S. E., & Shiffrin, R. M. (1990). List-strength effect: I. Data and discussion. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16, 163-178.
- Ratcliff, R., & Murdock, B. B. (1976). Retrieval processes in recognition memory. *Psychological Review*, 83, 190-214.
- Ratkowsky, D. A. (1983). *Nonlinear regression modeling: A unified practical approach*. New York: Marcel Dekker.
- Roediger, H. L., III, Stellan, C. C., & Tulving, E. (1977). Inhibition from part-list cues and rate of recall. *Journal of Experimental Psychology: Human Learning and Memory*, 3, 174-188.
- Roediger, H. L., III, & Thorpe, L. A. (1978). The role of recall time in producing hypermnnesia. *Memory & Cognition*, 6, 296-305.

- Rohrer, D. (1996). On the relative and absolute strength of a memory trace. *Memory & Cognition*, 24, 188–201.
- Rohrer, D., & Wixted, J. T. (1994). An analysis of latency and interresponse time in free recall. *Memory & Cognition*, 22, 511–524.
- Rundus, D. (1974). Output order and rehearsal in multi-trial free recall. *Journal of Verbal Learning and Verbal Behavior*, 13, 656–663.
- Shiffrin, R. M. (1970). Memory search. In D. A. Norman (Ed.), *Models of human memory* (pp. 375–447). New York: Academic Press.
- Shuell, T. J., & Keppel, G. (1968). Item priority in free recall. *Journal of Verbal Learning and Verbal Behavior*, 7, 969–971.
- Sodian, B., Schneider, W., & Perlmutter, M. (1986). Recall, clustering, and metamemory in young children. *Journal of Experimental Child Psychology*, 41, 395–410.
- Steinmetz, J. I., & Battig, W. F. (1969). Clustering and priority of free recall of newly learned items in children. *Developmental Psychology*, 1, 503–507.
- Tulving, E., & Hastie, R. (1972). Inhibition effects of intralist repetition in free recall. *Journal of Experimental Psychology*, 92, 297–304.
- Wheeler, M. A., & Roediger, H. L. (1992). Disparate effects of repeated testing: Reconciling Ballard's (1913) and Bartlett's (1932) results. *Psychological Science*, 3, 240–245.
- Wixted, J. T., & Rohrer, D. (1993). Proactive interference and the dynamics of free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19, 1024–1039.
- Wixted, J. T., & Rohrer, D. (1994). Analyzing the dynamics of free recall: An integrative review of the empirical literature. *Psychonomic Bulletin & Review*, 1, 89–106.

## Appendix

### Expected Recall Latency

If the probability of sampling an item on any given draw is  $p$ , and the rate of sampling is one per second, then the probability that the item will have a latency of 1 s is  $p$ . The probability that the item will have a latency of 2 s is  $(1 - p) \times p$  (i.e., the probability that it was not sampled on the first draw multiplied by the probability that it was sampled on the second draw). Similarly, the probability that the item will have a latency of 3 s is  $(1 - p)^2 \times p$  (i.e., the probability that it was not sampled on the first two draws multiplied by the probability that it was sampled on the third draw). Mean recall latency ( $\tau$ ) is given by the sum of all possible latencies, each multiplied by its probability of occurrence:

$$\tau = 1p + 2pq + 3pq^2 + 4pq^3 \dots, \quad (1)$$

where  $q = 1 - p$ . We now show why  $\tau = 1/p$ . Multiplying both sides of Equation 1 by  $q$  yields

$$q\tau = 1pq + 2pq^2 + 3pq^3 \dots \quad (2)$$

Subtracting Equation 2 from Equation 1 yields  $\tau - q\tau = p + pq +$

$pq^2 + pq^3 \dots$ , or  $(1 - q)\tau = p + pq + pq^2 + pq^3 \dots$ . Because  $p = 1 - q$ , this equation may be written as  $p\tau = p + pq + pq^2 + pq^3 \dots$ . Dividing both sides by  $p$  produces

$$\tau = 1 + q + q^2 + q^3 \dots, \quad (3)$$

and multiplying both sides by  $q$  yields

$$q\tau = q + q^2 + q^3 \dots \quad (4)$$

Subtracting Equation 4 from Equation 3 yields  $\tau - q\tau = 1$  or  $(1 - q)\tau = 1$ , which, substituting  $p$  for  $1 - q$ , becomes  $p\tau = 1$  and, finally,  $\tau = 1/p$ .

Received August 8, 1995

Revision received June 21, 1996

Accepted July 31, 1996 ■