On the relative and absolute strength
of a memory trace

DOUG ROHRER
University of California, San Diego, La Jolla, California

Subjects studied either an 8- or 16-word list and later recalled the items while a voice key recorded each response latency. The trials were partitioned by recall total in order to examine the means and distributions of both latencies and interresponse times as a function of recall total. Each analysis was consistent with the view that an item's absolute strength determines *whether* it is recalled whereas an item's relative strength determines *when* it is recalled. In addition, mean latency was effectively proportional to study list length yet independent of recall total. All of the analyses were consistent with the view that the set of study items is sampled according to a relative-strength rule until all items are found and that a sampled item is recovered into consciousness only when its absolute strength exceeds a fixed threshold.

The recall of multiple items from long-term memory is often held to be a competition of sorts in which an item's chances depend as much on the strengths of other items as it does on its own strength (Rundus, 1973; Shiffrin, 1970). According to this *relative-strength* rule, the probability of retrieving an item equals its strength divided by the sum of all the item strengths. Thus, a weak item with 10 equally weak peers has as much chance of being retrieved as a strong item with 10 equally strong peers. More paradoxically, an item that is the weakest of a few will often fare better than one that is the strongest of many because an item's absolute strength can be overshadowed by the sheer number of its competitors.

Models of free recall that invoke the relative-strength rule typically include two stages (e.g., Gronlund & Shiffrin, 1986; Raaijmakers & Shiffrin, 1980; Rohrer & Wixted, 1994; Wixted & Rohrer, 1994). First, items are *sampled* from a set of items according to the relative-strength rule, which, as described below in more detail, determines the latency of each response. Second, a sampled item is *recovered* into consciousness only if its strength exceeds some threshold. Thus, whereas all items will eventually be sampled (unless the search ends prematurely), only the strong items will be recovered, as consistent with the tip-of-the-tongue phenomenon (Brown & McNeill, 1966). Notably, this view of free recall predicts that recall latency and recall total should be independent. An item's latency reflects its relative strength and an item's likelihood of being recalled reflects its absolute strength. The predictions of this dissociation are now presented in more detail.

Relative Strength and Recall Latency

A very specific instance of the relative-strength model is the so-called random-search model (Bousfield, Sedgewick, & Cohen, 1954; McGill, 1963). This version of the relative-strength model assumes that all items have the same strength, whereas the relative-strength rule, of course, allows for variable strengths. The assumption of equal strengths has rightly been criticized as untenable because subjects typically recall items in a systematic order, not a random one (Murdock & Okada, 1970; Shiffrin, 1970). Such systematic recall orders are easily explained, of course, by item strength variability, as allowed by the relative-strength model.

The relative-strength model will be tested herein by analyzing the means and distributions of both latencies and interresponse times (IRTs). For each analysis, the equal-strength model will be fit to the data and the observed deviation will be compared to the deviation predicted by variable strengths. Fortunately, the mathematical nature of the predicted deviation for several of these analyses has previously been derived by Vorberg and Ulrich (1987) in order to "motivate researchers to test the model rigorously" (p. 3). Each of the analyses allows for more than one *qualitatively* distinct pattern of deviation from the equal-strength model and only one, of course, is predicted by the variable-strength model. Therefore, the presence of any one of these alternative deviations would provide evidence against the relative-strength model.

The relative-strength model can be envisioned as either a serial or parallel process (Vorberg & Ulrich, 1987). The serial version is presented here because it better serves the purpose of the present investigation, not because of any evidence in favor of it. According to this interpretation, items within a mental search set are randomly sampled one at a time at a constant rate. A sampled item is immediately recognized as either a not-yet-sampled item (and then recovered into consciousness) or a previously sampled item (and ignored). The probability of sampling a particular

item (say, item k) equals its strength divided by the sum of all item strengths ($s_k/\Sigma s_i$), which is the relative-strength rule. Regardless of whether item strengths vary or not, each item strength remains constant across time. Mathematically, the reciprocal of the sampling probability equals the item's expected latency (τ_k). For example, if the flip of a coin yields "heads" with probability $1/2$, then the number of flips required, on average, before "heads" appears equals the reciprocal of $1/2$, or 2. Likewise, if the probability of an item being sampled equals $1/6$, then the number of samples required, on average, before that item is sampled equals 6. Thus, if each sample requires, say, $1/2$ sec, then its observed latency equals 3 sec.

Table 1 illustrates the effect of relative strength on recall latency. In row 1a, each of the 6 items has strength 1. Thus, each item has a sampling probability of $1/6$ and an expected latency of 6. If each strength is doubled, as in row 1b, the expected latency of each item is unchanged because latency is a function of relative strength, not absolute strength. When strengths vary, expected latencies vary as well but are calculated in the same manner. For the variable strengths in row 2a, for example, the item with strength 0.5 has a sampling probability of $0.5/6$ or $1/12$ and, therefore, an expected latency of 12. On the other hand, the item with strength 1.5 has a sampling probability of $1.5/6$ or $1/4$ and, therefore, an expected latency of 4. Not surprisingly, the stronger items have faster expected latencies. If each item strength is doubled, as in row 2b, none of the item latencies change. In summary, an item's expected latency depends on its relative strength regardless of whether item strengths vary or not.

Even though each item's sampling probability remains constant throughout the recall period, the relative-strength model predicts that the observed rate of recall will decline asymptotically. This is simply explained by the declining number of not-yet-sampled items within the search set. In contrast, the recall of items from short-term store (i.e., after a zero retention interval) yields a rate of recall that is roughly constant. This difference serves to underscore the important point that the relative-strength model describes recall from long-term store (i.e., after a task-filled retention interval), not short-term store.

The equal-strength model has received considerable attention in recent years because of its heuristic interpretation of mean latency (Rohrer & Wixted, 1994; Rohrer, Wixted, Salmon, & Butters, 1995; Wixted & Rohrer, 1993, 1994). Specifically, the equal strength model predicts that

mean latency = search set size $*$ duration of sample.

Thus, if sampling rate is assumed to remain constant, mean latency measures search set size. According to the model, then, an experimental manipulation that doubled mean latency must have also doubled search set size. In Table 1, for example, the 3 items in row 3a yield a mean latency of 3 whereas the 6 items in row 3b yield a mean latency of 6.

Fortunately, this direct relationship between mean latency and search set size is effectively retained when item strengths vary. In Table 1, for example, the 6 items in row 1a have no variability and their mean latency equals 6. In row 2a, the 6 item strengths vary considerably and their mean latency equals 7, a relatively modest increase. Moreover, because an increase in variability can only increase mean latency, as proven in the Appendix, relative comparisons of mean latency across experimental conditions are only slightly distorted. That is, if the mean latencies for two experimental conditions can only increase, then the ratio of the two mean latencies remains relatively stable. As an illustration, the 3 equal-strength items in row 3a yield a mean latency of 3, and the 6 equal-strength items in row 3b yield a mean latency of 6. In contrast, the 3 variable-strength items in row 4a yield a mean latency of 3.67 and the 6 variable-strength items in row 4b yield a mean latency of 7. Thus, even though variability increases the mean latency for each of the two list lengths by roughly 20%, the ratio of the variable-strength mean latencies ($3.67/7 = .52$) is only 4% greater than the ratio of the equal-strength latencies ($3/6 = .50$).

Empirical support for this interpretation of mean latency as a measure of search set size derives from two studies in particular. First, an increase in study list length, which is

<div align="center">

Table 1
Illustrations of the Relative-Strength Model and Recovery Threshold

</div>

	s_1	s_2	s_3	s_4	s_5	s_6	s	τ_1	τ_2	τ_3	τ_4	τ_5	τ_6	τ
			Item Strengths				Mean Strength			Item Latencies				Mean Latency
1a.	1	1	1	1	1	1	1	6	6	6	6	6	6	6
1b.	2	2	2	2	2	2	2							
2a.	0.5	0.6	1.0	1.2	1.2	1.5	1	12	10	6	5	5	4	7
2b.	1.0	1.2	2.0	2.4	2.4	3.0	2							
3a.	1	1	1				1	3	3	3				3
3b.	1	1	1	1	1	1	1	6	6	6	6	6	6	6
4a.	0.5	1.0	1.5				1	6	3	2				$3\frac{2}{3}$
4b.	0.5	0.6	1.0	1.2	1.2	1.5	1	12	10	6	5	5	4	7
5a.	0.4	0.5	0.6	1.0	1.5	2.0	1	*	12	10	6	4	3	7
5b.	0.5	0.6	1.0	1.2	1.2	1.5	1	12	10	6	5	5	4	7
6a.	0.4	0.5	0.6	1.0	1.5	2.0	1	*	12	10	6	4	3	7
6b.	1.0	1.2	2.0	2.4	2.4	3.0	2	12	10	6	5	5	4	7

Note—All values are exact.   *Indicates item with strength below the recovery threshold of 0.5 (see text).

widely believed to increase the number of items within the mental search set, causes mean latency to increase by a roughly proportional amount (Rohrer & Wixted, 1994). The second line of evidence derives from a study of proactive interference (Wixted & Rohrer, 1993). In this experiment, subjects underwent a rapid succession of 3 trials with categorically related words (e.g., three fruits, three more fruits, and three more fruits). On each trial, subjects were required to recall only those words from the most recent study list. Recall totals declined with each subsequent trial (which is the usual finding), yet the mean latency of those responses increased from about 3 sec on Trial 1 to about 7 sec on Trial 3. Thus, given the widely held view that the mental search set includes all categorically related items from recently studied lists, the observed increase in mean latency is consistent with the view that mean latency reflects search set size.

Absolute Strength and Recall Total

A relative-strength sampling process by itself cannot account for most free recall phenomena. For example, if any sampled item is automatically recalled, then recall total would simply equal search set size. Therefore, because search set size determines recall latency, a relative-strength sampling process by itself would predict an intrinsic correlation between the measures of recall latency and recall total. In reality, however, recall latency and recall total are empirically independent, as demonstrated by the manipulations of study list length and proactive interference that were described above. The longer study lists increased both mean latency and mean total—a positive correlation—whereas the build-up of proactive interference increased mean latency but *decreased* mean total—a negative correlation. Incidentally, rather than measure recall total, one can instead calculate recall accuracy, the proportion of study items recalled, which is not redundant with recall total when study lists vary in length. However, recall latency and recall accuracy were also empirically dissociated in Rohrer and Wixted (1994).

If search set size does not equal recall total, then the search set obviously includes items that are not recalled. For this reason, relative-strength models include a process that determines whether a sampled item is recovered into consciousness. According to the simplest instantiation of this recovery stage, an item is recovered if its absolute strength exceeds some fixed threshold. In Table 1, for example, the recovery threshold is set at 0.5. Thus, the item in row 5a with strength 0.4 is below threshold and consequently not recalled. In contrast, all items in row 5b are above threshold.

Even though below-threshold items are not recalled, these items still affect retrieval because their presence affects the relative strengths of the above-threshold items. Therefore, the relative-strength model cannot be tested unless it specifies how many unrecalled items are in the search set. It is often implicitly assumed that the search set includes all of the study items, more or less, and several findings are consistent with this hypothesis. In the list length experiment described above, for example, mean latency, a

measure of search set size, increased by an amount that was proportional to the increase in list length. Likewise, mean latency increased with the build-up of proactive interference by an amount that was roughly proportional to the number of recently studied, categorically related words. Finally, free recall errors are quite rare, consistent with the view that the search set includes few, if any, extralist items.

Now, if recall latency reflects search set size (i.e., list length) and recall total reflects absolute strength, then it ought to be possible to manipulate recall total without affecting recall latency. Specifically, if each item's absolute strength is increased by the same proportion, then each item's relative strength should remain constant (as in rows 2a and 2b in Table 1). This prediction of the model is consistent with the results of an experiment in Rohrer and Wixted (1994) in which subjects studied 6-word lists at different presentation rates. Not surprisingly, longer study time produced considerably greater recall total, in keeping with the view that an increase in the mean absolute strength results in a greater number of above-threshold items. In spite of this robust effect on recall total, study time had no effect on mean latency, which is consistent with the relative-strength model's claim that recall latency reflects the number of study words. An illustration of this finding is given in Table 1. The "quickly studied" items in row 6a have mean strength 1 whereas the "well-studied" items in row 6b have mean strength 2. As shown, recall totals differ and mean latencies do not.

The relative-strength (with threshold) model can now be briefly summarized as follows. A mental search set that includes essentially all of the study items is sampled (in serial or in parallel) according to the relative-strength rule, and a sampled item is recovered into consciousness only when its absolute strength exceeds a threshold. Recall latency reflects the number of items within the search set (i.e., list length), whereas recall total reflects the number of *recoverable* items within the search set. Likewise, an item's relative strength determines when it is recalled, whereas an item's absolute strength determines whether it is recalled.

The model makes a counterintuitive prediction that will be tested by the present experiment. Specifically, mean latency should be independent of recall total *within* a given condition. For example, if subjects study 8 words, then the mean latency for trials in which 3 words are recalled should equal that for trials in which 6 words are recalled, even though latency is measured from the onset of the recall period. This result could not be easily explained by any model that defines the search set as a collection of eventually recalled items, as does the relative-strength model *without* recovery threshold. Such a finding is also counter to the widely held view that mean latency should be greater for trials with greater recall totals because a subject requires more time to give more responses.

In summary, the first aim of the present experiment is to examine whether the nature of free recall latencies and IRTs is consistent with the predictions of the relative-strength model. Second, the independence of relative strength (as measured by mean latency) and absolute strength (as mea-
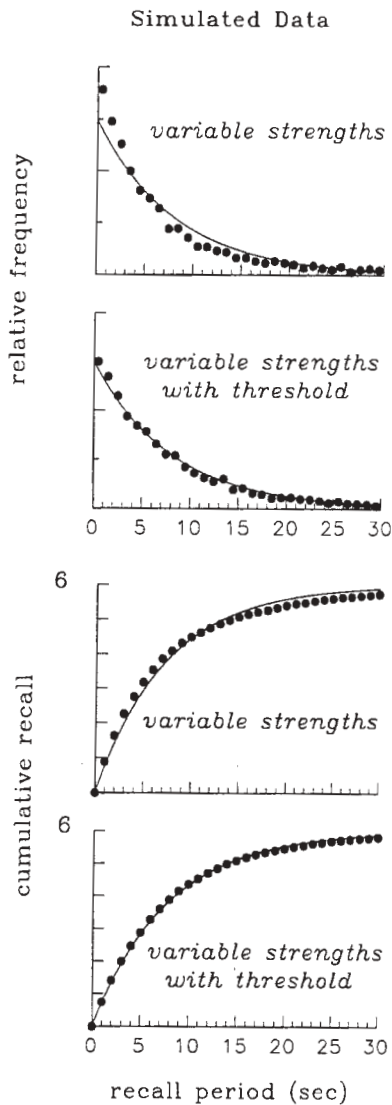
Simulated Data

Figure 1. Latency distributions and cumulative recall functions given by the variable-strengths simulation (row 5a, Table 1) and the variable-strengths-with-threshold simulation (row 5b, Table 1). Each plot includes the best fitting exponential distribution (Equation 1).

sured by recall total) will be tested by partitioning trials by recall total before performing each analysis.

METHOD

Subjects
Twenty-five University of California, San Diego, undergraduates participated for course credit.

Materials
For each subject, 28 eight-word lists were randomly chosen without replacement from a list of 224 four-letter monosyllabic words.

Procedure
Subjects were tested by computer in the presence of an experimenter. Each subject participated in 3 practice trials and 25 scored trials. Each trial began with a 2-sec get-ready prompt and the subsequent presentation of 8 study words, one at a time, for 3 sec each. Subjects read the words aloud. In the subsequent distractor task, 10 three-digit numbers appeared one at a time for 2 sec each and subjects read the 3 digits aloud in ascending order of value. Subjects were then prompted to recall, in any order, the words of the most recently studied list during a 30-sec recall period. A voice key and computer recorded each response latency to the nearest millisecond. An experimenter monitored these latencies via computer in order to record voice key false alarms (e.g., the recording of a cough), voice key misses (e.g., the failure to record a soft-spoken response), and subject errors. A 10-sec rest period preceded the next trial.

Latencies resulting from an extraneous noise (e.g., coughs) were removed. Of the 625 trials, 36 included a response that was not recorded by the voice key, and these trials were excluded from further analysis. In the remaining 589 trials, subjects recalled a total of 2,594 words, including 52 errors (2%). Because temporal issues
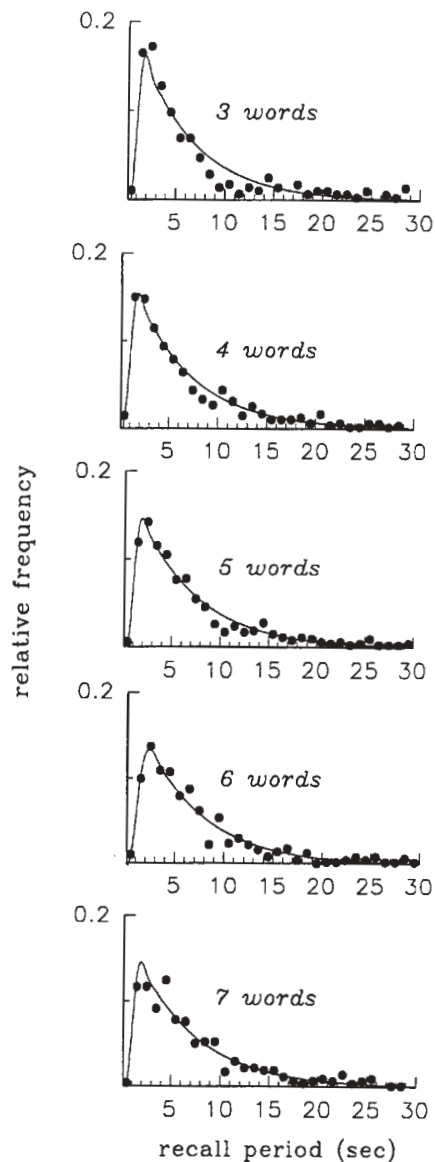

Figure 2. Latency distributions for recall totals 3–7 and the best fitting ex-Gaussian functions (Equation 2). The parameter estimates are given in Table 2.

were primary in this investigation, these errors were included in the analyses in order to avoid distortion of the IRTs that preceded or followed each error (see Rohrer & Wixted, 1994). The number of trials with recall totals 1 through 8, respectively, were 27, 56, 91, 120, 112, 85, 56, and 30. Analyses are presented for recall totals 3–7 only. The recall totals of 1 and 8 were too rare and a recall total of 2 yields only one IRT, thereby disallowing an examination of mean IRT growth.

Statistical Technique

The best fitting theoretical distributions were determined by maximum likelihood estimation (Maindonald, 1984; Ratcliff & Murdock, 1976). After the best fit was determined, a χ^2 goodness-of-fit statistic was calculated by combining bins with few responses (and reducing degrees of freedom) so that the expected value of each bin would exceed 5 (see Ratcliff & Murdock). Fits to nondistributional data were determined by minimizing either least squares or weighted least squares. Asymptotic standard errors (ASEs) for each parameter estimate were obtained from the Hessian matrix of second partial derivatives (Maindonald, 1984; Ratcliff & Murdock, 1976). All analyses were grouped across subjects, though subject heterogeneity was presumably minimized by partitioning trials by recall total before each analysis.

LATENCY DISTRIBUTIONS

If item strengths are equal, latencies will be distributed as an exponential,

$$e^{-t/\tau}/\tau, \tag{1}$$

where the only parameter, τ, equals mean latency (Bousfield & Sedgewick, 1944). Whereas the exponential distribution is monotonically declining, recall latency distributions, in practice, sharply ascend during the first 1 sec and then gradually decline throughout the remainder of the recall period (Rohrer & Wixted, 1994). Rohrer and Wixted demonstrated that this ascending arm results from the pause that precedes recall, not from an initial increase in rate of recall. In order to measure recall latency without including this initiation time, the latencies can be fit to a distribution that results when an exponential process is preceded by a roughly symmetric, unimodal distribution. If that unimodal distribution is arbitrarily assumed to be a normal distribution, the resulting distribution is ex-Gaussian (Hohle, 1965),

$$\frac{e^{-(t-\mu)/\tau+\sigma^2/2\tau^2}}{\tau\sqrt{2\pi}}\int_{-\infty}^{(t-\mu)/\sigma-\sigma/\tau}e^{-y^2/2}dy. \tag{2}$$

The normally distributed stage has mean μ and standard deviation σ, and the exponentially distributed stage has mean τ (see Rohrer & Wixted, 1994, for a complete de-

scription and derivation). Importantly, the sum of μ and τ necessarily equals the total mean latency. Throughout both the present article and in Rohrer and Wixted, estimates of μ equaled about 1 sec, whereas estimates of τ varied from about 3 to 12 sec, which is consistent with the view that μ simply represents an initiation of the recall process. This initiation time is not critical in the present study and the term *mean latency* will hereafter refer to τ. In sum, the ex-Gaussian is fit in order to obtain a more accurate measure of the exponential decline in latency distributions.

With the introduction of variable item strengths, the relative-strength model predicts that the tails of these latency distributions will deviate from exponentiality. In essence, the stronger-than-average items will be retrieved sooner than average whereas the weaker-than-average items will be retrieved later than average. In order to examine these deviations, two 1,000-trial simulations of the relative-strength model were performed, one with and one without a recovery threshold. In the *variable-strengths* simulation, the search set included 6 variable-strength items (0.4, 0.6, 0.6, 1.2, 1.2, 2.0) and all 6 items were recalled on each trial. In the *variable-strengths-with-threshold* simulation, the search set included 9 variable-strength items (0.4, 0.5, 0.7, 0.9, 0.9, 1.0, 1.0, 1.8, 1.8), but only the 6 strongest items were above the threshold of 0.8. Each set of item strengths has a mean strength of exactly 1 and a standard deviation of about 0.5. More importantly, each scenario yields an expected mean latency of exactly 8. Thus, even though the search set sizes differ, the observable measures of recall total and mean latency are equal.

Figure 1 presents the latency distributions given by the variable-strengths and variable-strengths-with-threshold simulations and the best fit exponential (Equation 1). As shown, the best fit exponential underestimates the data points for the first several bins and then overestimates them for most of the remainder of the recall period. Finally, though it cannot be seen in these plots, the curve slightly underestimates the data very late in the recall period. Note, however, that the inclusion of a recovery threshold almost entirely nullifies the magnitude of this deviation, as revealed by a comparison of the two plots. In effect, the very weak items that would have otherwise been recalled late in the recall period are no longer recalled, thereby decreasing the variability in the *observed* recall latencies.

Results and Discussion

Figure 2 includes the latency distributions for recall totals 3–7 and the best fitting ex-Gaussian distributions. Be-

Table 2
Ex-Gaussian Fits of Latency Distributions (Figure 2)

Recall Total (n)	Initiation		Mean Latency		Goodness-of-Fit		
	μ	SE	τ	SE	N	χ^2	df
3	1.07	.08	5.58	.35	273	33.05*	13
4	1.05	.07	5.72	.27	480	18.87	17
5	1.16	.06	6.05	.26	560	25.68	17
6	1.25	.13	6.07	.30	510	38.95*	18
7	1.18	.07	6.22	.32	392	18.86	16

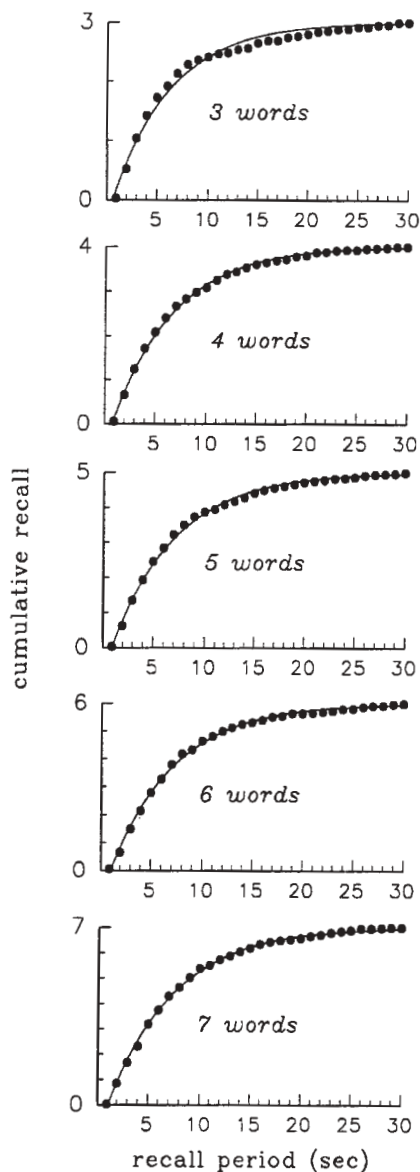Note—Initiation and latency are given in seconds.   *$p < .05$.

Figure 3. Cumulative recall growth for recall totals 3–7 and the best fitting cumulative exponentials (Equation 3). The parameter estimates are given in Table 3.

cause the number of response latencies varied across recall totals (i.e., across plots), these distributions were plotted as *relative* frequency distributions (i.e., as proportions) in order to facilitate the visual comparison of the shapes. Table 2 lists the absolute number of response latencies (N), the parameter estimates of μ and τ, and goodness-of-fit χ^2 values for each recall total (n). The sum of the χ^2 values across recall totals was less than the corresponding sum provided by the fits of the lognormal, the gamma, and the Weibull, each with an added third parameter to account for the time shift due to the initiation pause.

As reported in Table 2, the estimates of mean latency (τ) equaled about 6 sec for each recall total. Thus, even though latency is measured from the onset of the recall period, mean

latency was effectively independent of recall total. These estimates did, however, increase slightly as a function of recall total at an average rate of 0.16 sec per extra word. Though this increase is most likely real, the total increase (about 0.6 sec) is an order of magnitude less than the values themselves (about 6 sec). Thus, whereas recall total varied by more than a factor of 2, recall latency varied by less than 10%. In contrast, a factor-of-2 increase in *study* total increases latency by 100% (Rohrer & Wixted, 1994; see Introduction). In addition, the small increase in mean latency as a function of recall total may simply have resulted from the additional time needed to say each extra word aloud.

The deviation between the latency distributions in Figure 2 and the best fitting ex-Gaussians is very slight. In order to examine whether any systematic deviations exist, the deviations for each bin were summed across recall totals 3–7 (i.e., across all 5 plots), yielding a total deviation for each bin that was either positive (data > fit) or negative (data < fit). For bins 1 through 30, respectively, where bin 1 represents the initiation pause, these deviation sums were

(+)++++ −−−−− −−−−+ −−−+− +−+++ +++++.

Thus, there are long streaks or "runs" of deviations in the same direction that are clearly systematic. In fact, there are only 9 such runs, and a Wald-Wolfowitz runs test revealed this number to be significantly lower than that expected by chance ($z = -2.31, p < .05$, two-tailed). More importantly, this systematic deviation is practically identical to that observed in the relative-strength simulations in Figure 1. Specifically, the exponential tail underestimates the first few data points, overestimates the data until the tail reaches the floor, and then underestimates the data thereafter.

In sum, the mathematical nature of the observed latency distributions is consistent with the predictions of the relative-strength model when a recovery threshold is incorporated. In addition, the means of these distributions were effectively independent of recall total, which is consistent with the view that search set size remains roughly constant even though the number of recoverable items can vary.

CUMULATIVE RECALL CURVES

If item strengths are equal, the growth in cumulative recall as a function of time is a cumulative exponential,

$$n(1-e^{-(t-c)/\tau}), \tag{3}$$

where c represents the initiation pause that precedes retrieval and τ equals mean latency (Bousfield & Sedgewick, 1944). Because a cumulative recall curve is actually a cumulative latency distribution, the analyses of cumulative recall and (noncumulative) latency distributions are somewhat redundant. Nevertheless, cumulative recall curves are presented here so that the present experiment can be easily compared with experiments by other researchers who uniformly prefer the cumulative form (e.g., Bousfield et al., 1954; Gronlund & Shiffrin, 1986; Payne, 1986; Roediger, Stellon, & Tulving, 1977; Roediger & Thorpe, 1978). Cumulative recall has not, however, been analyzed after first partitioning trials by recall total.

Table 3
Cumulative Exponential Fits of
Cumulative Recall Total (Figure 3)

Recall	Initiation		Mean Latency		
Total (n)	c	SE	τ	SE	%VAF
3	.68	.14	5.71	.21	98.5
4	.83	.05	5.95	.07	99.8
5	.98	.07	6.21	.09	99.7
6	1.09	.06	6.21	.08	99.8
7	1.12	.05	6.37	.07	99.8

Note—Initiation and latency are given in seconds. %VAF, the percentage of variance accounted for by the fit.

Table 4
IRT Means (in Seconds) and Standard Deviations (in Seconds)
for Each Interresponse Interval

Recall	Interresponse Interval (i)					
Total (n)	$n-6$	$n-5$	$n-4$	$n-3$	$n-2$	$n-1$
			Means			
3					2.63	6.34
4				2.15	3.20	5.46
5			1.68	2.31	3.65	4.93
6		1.38	1.70	2.35	2.89	5.97
7	1.24	1.56	1.53	2.07	3.01	5.45
			Standard Deviations			
3					3.21	5.83
4				1.89	2.66	5.20
5			1.61	1.99	3.69	4.39
6		0.61	0.92	2.68	2.83	5.55
7	0.64	0.92	0.92	1.78	2.92	4.52

Note—The ith interresponse interval occurs between responses i and $i+1$.

As proven by Vorberg and Ulrich (1987), the relative-strength model predicts that the best fit cumulative exponential will initially underestimate the data points and overestimate the data thereafter. Once again, however, the presence of a recovery threshold reduces the extent of this deviation. As an illustration, the cumulative recall curves given by the variable-strengths and variable-strengths-with-threshold simulations described in the last section are shown in Figure 1 along with the best fitting cumulative exponentials (Equation 3).

Results and Discussion

As shown in Figure 3, the cumulative recall curves for recall totals 3–7 are well described by the cumulative exponential. Each fit accounted for more than 98% of the variance, as listed in Table 3. This table also includes the estimates of τ, each of which was roughly equal to the corresponding estimate of τ given by the ex-Gaussian fits of latency distributions. Incidentally, because the data are plotted cumulatively, the height of each data point must be greater than the one that precedes it. This lack of independence in cumulative data is responsible for the striking smoothness of these curves.

As was true with the latency distributions, the cumulative recall curves exhibited a small yet systematic deviation from the equal-strength model that is consistent with the prediction of the relative-strength model. As in the last sec-

tion, this was revealed by adding the deviations for each bin across recall totals (i.e., across plots), thereby yielding either a positive sum (data > fit) or negative sum (data < fit). Because the cumulative data are not independent, a statistical test on the number of runs like that performed in the last section was not possible. Nevertheless, the deviations for bins 5 through 10 were each positive, whereas the deviations for bins 11 through 25 were each negative. Thus, except for the first few and last few points, the best fit cumulative exponential roughly underestimated the first 10 bins and overestimated the next 15 bins. An identical pattern was observed in the simulated cumulative recall data in Figure 1. In summary, the results of the cumulative recall analysis confirm those of the latency distribution analysis.

IRT GROWTH

Because the rate of recall declines throughout the recall period, mean IRT (the average time between successively recalled items) necessarily increases as a function of output position. Thus, whereas latencies can be grouped across output position, a meaningful analysis of IRTs requires the partitioning of data by interresponse interval (i). As a matter of notation, the IRT between the first and second responses is designated as the first IRT ($i = 1$). Thus, a trial with n responses yields $n-1$ IRTs. As first noticed by Albert (1968), the partitioning of IRTs by both
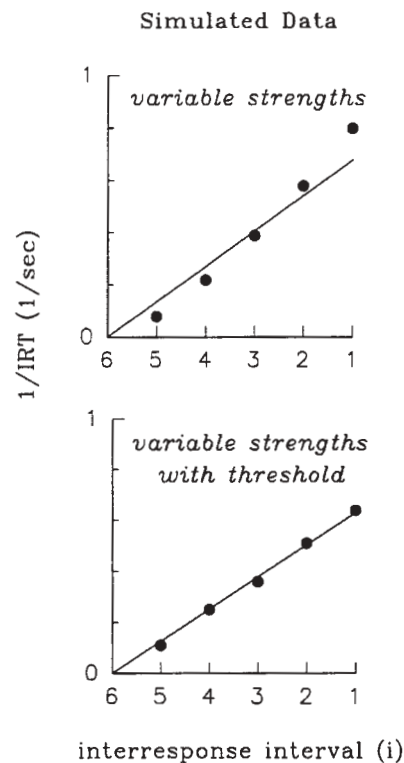

Figure 4. Values of 1/interresponse time (IRT) given by the variable-strengths simulation (row 5a, Table 1) and the variable-strengths-with-threshold simulation (row 5b, Table 1). Each plot includes the best fitting one-parameter lines (Equation 5).
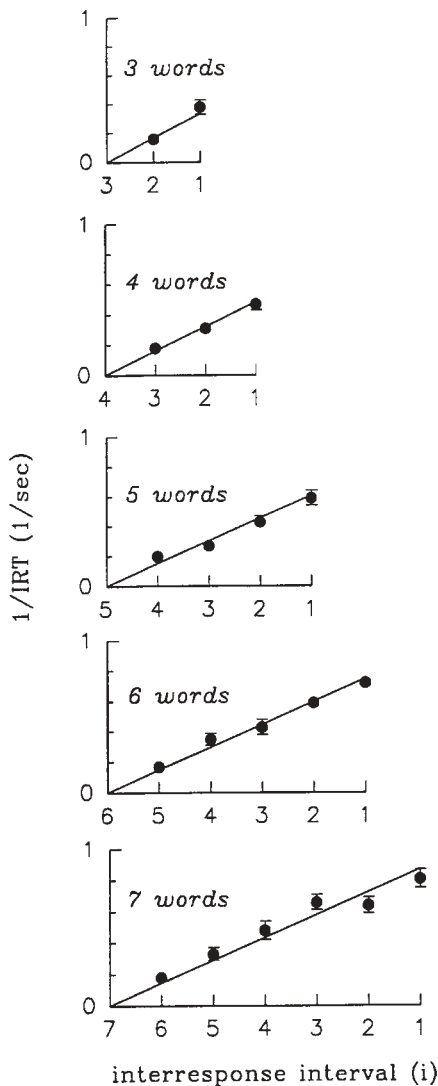
Figure 5. 1/interresponse time (IRT) growth as a function of reverse output position for recall totals 3–7 and the best fitting one-parameter lines (Equation 5). The estimates of τ given by these fits are given in Table 5 and the mean IRTs are given in Table 4. Error bars represent ±1 standard error.

recall total (n) and interresponse interval (i) produces a systematic pattern, and that pattern was observed in the present data as well. As shown in Table 4, each of the last IRTs (which are essentially equal) was greater than each of the second-to-last IRTs, and each of the second-to-last IRTs (which are themselves essentially equal) was greater than each of the third-to-last IRTs, and so forth. Thus, mean IRT depends on the number of not-yet-recalled items.

This pattern of IRT growth falls out of a more specific prediction of the equal-strength model. If IRT_i represents the mean IRT for the interresponse interval i, then

$$IRT_i = \tau/(n-i), \qquad (4)$$

where n equals recall total and τ equals mean latency (McGill, 1963). Thus, the last IRT ($i = n-1$) equals $\tau/1$,

the second-to-last IRT ($i = n-2$) equals $\tau/2$, and so on, regardless of recall total (n).

The purposes of the present study are better served by fitting an equation that is mathematically equivalent to Equation 4,

$$1/IRT_i = (n-i)/\tau. \qquad (5)$$

If the 1/IRT values given by this equation are placed in *reverse* output position ($i = n-1, n-2, \ldots, 1$), then the 1/IRT values equal $1/\tau, 2/\tau, \ldots, (n-1)/\tau$. Thus, the equal-strength model predicts that values of 1/IRT, when placed in reverse output position, increase as a one-parameter line that passes through the "origin" with slope $1/\tau$. As an illustration, Figures 4 and 5 (which are discussed in more detail below) each includes one of these one-parameter lines.

When item strengths vary, this linearity in 1/IRT no longer holds. As proven by Vorberg and Ulrich (1987), the relative-strength model produces 1/IRT values that increase with positive acceleration as a function of reverse output position. (Technically, these authors proved that the line segment between any 1/IRT value and the origin necessarily lies above all of the intermediate points, a slightly stronger claim than positively accelerated growth.) As an illustration, Figure 4 includes the 1/IRT values given by the variable-strengths and variable-strengths-with-threshold simulations that were described in the section on latency distributions. Once again, the presence of a recovery threshold almost entirely nullifies the deviation caused by variable strengths.

Results and Discussion

As shown in Figure 5, the increase in 1/IRT as a function of reverse output position is well described by the predicted one-parameter line with slope $1/\tau$ (Equation 5). Because the slopes of these lines were roughly equal, the estimates of τ (mean latency) given by these fits were roughly independent of recall total, as reported in Table 5. Thus, mean latency and recall total were again independent.

However, these estimates of τ were slightly greater than those given by the ex-Gaussian fits of latency distributions, especially for the recall total of 7. Because the predicted one-parameter lines have slope $1/\tau$, these larger-than-expected τ values reflect smaller-than-expected slopes. An examination of Figure 5 reveals that these smaller-than-expected slopes are clearly the result of the last 2 points (i.e.,

Table 5
One-Parameter Linear Fits of 1/IRT (Figure 5)

Recall Total (n)	Mean Latency		%WVAF
	τ	SE	
3	5.94	.35	94.1
4	6.19	.18	97.0
5	6.58	.29	85.1
6	6.69	.20	98.5
7	6.87	.34	94.9

Note—Latency is given in seconds. %WVAF, the percentage of the weighted variance accounted for by the fit where each weight equaled the reciprocal of the variance for that data point.

interresponse intervals 1 and 2). That is, these 1/IRT values appear to be less than that expected given the trend established by the other values. Notably, these apparently distorted values correspond to the smallest mean IRTs (Table 4).

One explanation for these smaller-than-expected 1/IRT values is given by the model's simplifying assumption of instantaneous recovery. Recovery must require some small amount of time, of course, but the assumption is effectively inconsequential when measuring latencies that typically average 5 sec or so. However, because the IRTs that precede the first few responses of a trial with at least 5 responses are typically less than 2 sec (Table 4), the assumption of instantaneous recovery is more problematic. To illustrate this potential problem, let each IRT $= S + R$, where S represents the duration of the sampling stage and R represents the constant duration of the recovery stage. When IRTs are long and the sampling stage is much longer than the brief recovery stage (say, $S = 20R$), the value of 1/IRT is about 95% as great as $1/S$. However, when IRTs are very brief (say, $S = 4R$), the value of 1/IRT is only 80% as great as $1/S$. Thus, as shown in Figure 5, 1/IRT values for brief IRTs would be less than expected. This explanation is consistent with the analyses of IRT distributions in the next section.

In summary, the growth in mean IRT is, for the most part, solely determined by a single quantity—mean latency—and this growth is independent of recall total. Moreover, because mean latency can be calculated directly by simply averaging response latencies, the single-parameter lines in Figure 5 were, in a sense, predicted rather than fit.

IRT DISTRIBUTIONS

The equal-strength model predicts that IRTs will be distributed as an exponential,

$$e^{-t/\mathrm{IRT}_i}/\mathrm{IRT}_i, \tag{6}$$

where IRT_i again represents the mean IRT for interresponse interval i (McGill, 1963). As before, IRT_1 denotes the mean IRT between the first and second response. Because IRT_i increases with output position, as described in the last section, IRT distributions cannot be meaningfully collapsed across interresponse intervals.

Incidentally, the model's prediction of both exponential latencies and exponential IRTs appears contradictory because any response latency is a sum of IRTs, yet the sum of exponentially distributed stages is not exponentially distributed. However, the equal-strength model predicts exponential latency distributions only when responses are grouped across output position. If items are instead partitioned by output position, the resulting latency distributions are described by a particular type of generalized-gamma distribution that is discussed in more detail in McGill (1963). The take-home story can be described briefly: If responses are collapsed across output position, latencies are exponential and IRTs are not, but if responses are par-

titioned by output position, latencies are not exponential and IRTs are.

Though IRT distributions given by the recall of items from long-term memory (i.e., after a retention interval) have not been reported, it is doubtful that IRTs will indeed be exponential. First, the exponential is monotonically declining, whereas RT distributions, in practice, begin with an ascending arm. Second, the equal-strength model predicts exponential IRTs only when the recovery process is instantaneous. However, if the duration of this recovery (R) is relatively brief, constant, and independent of the sampling stage
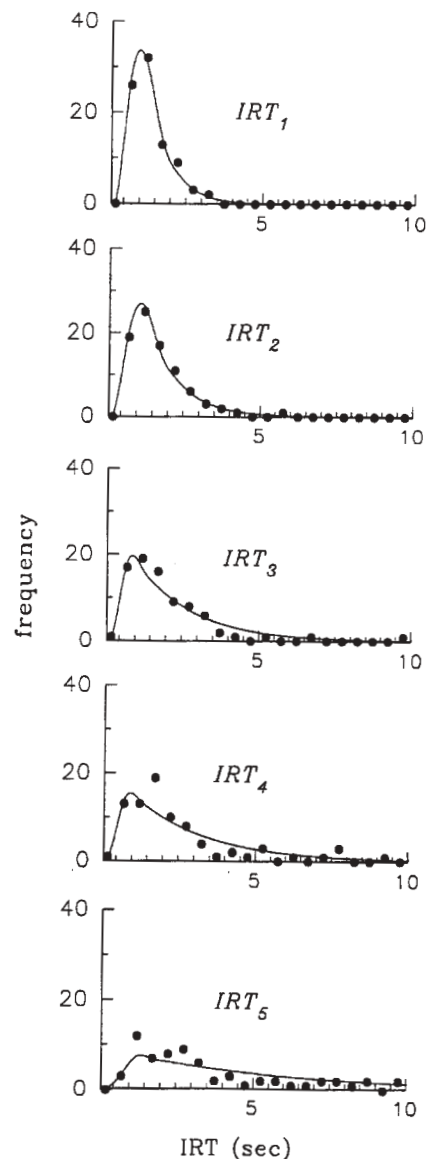

Figure 6. Interresponse time (IRT) distributions for output positions 2–6 in trials with recall totals of 6 and the best fitting ex-Gaussian distributions. The parameter estimates are given in Table 6.

Table 6
Ex-Gaussian Fits of IRT Distributions for Each Recall Total
and Output Position (Figure 6)

Recall Total (n)	Interresponse Interval (i)	Recovery		Sampling		Goodness-of-Fit		
		R	SE	S	SE	N	χ^2	df
3	1	.59	.05	2.03	.22	91	12.48	6
	2	.57	.08	5.75	.61	91	4.43	11
4	1	.60	.04	1.54	.15	120	9.06	5
	2	.58	.06	2.63	.25	120	15.01	8
	3	.63	.13	4.86	.46	120	20.39	12
5	1	.60	.03	1.09	.11	112	4.65	2
	2	.78	.08	1.53	.16	112	6.87	5
	3	.67	.09	2.99	.30	112	32.44*	9
	4	.84	.13	4.09	.41	112	16.98	10
6	1	.73	.05	.65	.08	85	1.56	2
	2	.73	.06	.96	.12	85	.85	3
	3	.58	.06	1.79	.20	85	5.40	4
	4	.57	.08	2.31	.26	85	10.41	5
	5	.79	.12	5.18	.57	85	7.25	10
7	1	.66	.10	.59	.13	56	3.29	1
	2	.67	.10	.88	.15	56	3.91	1
	3	.62	.06	.94	.14	56	2.73	2
	4	.60	.06	1.47	.20	56	8.04	3
	5	.58	.07	2.42	.33	56	5.62	5
	6	.59	.22	4.88	.69	56	5.94	4

Note—Recovery and sampling are given in seconds.   *$p < .05$.

(S), then the resulting IRT distributions would exhibit a brief ascent and a longer, exponentially declining tail.

When strengths vary, the relative-strength model predicts nonexponential IRT distributions (McGill, 1963; Vorberg & Ulrich, 1987). For example, if two items have already been recalled, the expected duration of IRT_2, which precedes the third item, depends on the strengths of the not-yet-recalled items. Because words are recalled in different orders on different trials, the strengths of the not-yet-recalled items will vary from trial to trial. Therefore, the expected IRT will depend on which words have already been recalled and will therefore also vary from trial to trial. In short, the deviation from exponentiality in both IRT distributions and latency distributions is caused by a variability in expected values. Therefore, variable item strengths yield IRT distributions that deviate from exponentiality in the same manner as latency distributions, as was shown in Figure 1. Specifically, the best fit exponential initially underestimates the data and then overestimates the data for most of the remainder of the distribution tail.

Results and Discussion

After partitioning the trials by recall total ($n = 3, 4, 5, 6, 7$) and then partitioning IRTs by interresponse interval ($i = 1, \ldots, n-1$), 20 separate distributions were obtained ($2 + 3 + 4 + 5 + 6$). For each distribution, IRTs were grouped into 500-msec bins. As a representative sample, the five IRT distributions for the recall total of 6 are shown in Figure 6. All of the IRT distributions include an ascending arm; in fact, less than 1% of the IRTs were faster than 500 msec.

Each IRT distribution reached a peak at about 0.6 sec and then declined thereafter (Figure 6). This pattern is consistent with the view that each interresponse interval includes a sampling stage (S) and a constant recovery stage (R), as discussed in the previous section. If the recovery stage is assumed to be normally distributed, then these IRT distributions should be ex-Gaussian (Equation 2), which, as noted above, describes the sum of an exponentially distributed stage and a normally distributed stage. For this reason, the ex-Gaussian was fit to the observed IRT distributions. Table 6 includes the parameter estimates and χ^2 values given by these fits. Rather than use the parameters μ and τ (Equation 2) and risk confusing these estimates with those given by the ex-Gaussian fits of latency distributions, μ and τ were replaced by R and S, respectively. As shown in Table 6, values of R remain constant while values of S increase with output position. Though the ex-Gaussian fits were consistent with the view that each interresponse interval includes both a sampling stage of variable duration and a recovery stage of constant duration, this interpretation should be considered tentative. All of the fits were quite noisy and alternative explanations certainly exist.

With regard to the deviation predicted by variable item strengths, the IRT distributions were consistent with those given by the simulations, though the evidence is again not overwhelming. As illustrated by the plots in Figure 6, the exponential tails of the ex-Gaussian underestimated the data for the first one third of the exponential tail and overestimated the data for the remainder of the tail.

One brief, final test of the relative-strength model concerns the variability of the IRT distributions. When item strengths are equal, the coefficient of variation (CV) for each IRT—the standard deviation divided by the mean—equals 1, as with any exponential distribution. Vorberg and Ulrich (1987) proved that variable strengths necessarily yield IRT CVs that are greater than 1. The values are still very close to 1, however, and even a bimodal distribution
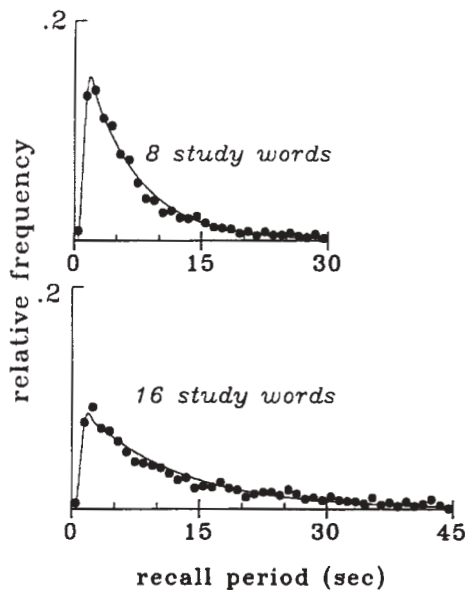
Figure 7. Latency distributions collapsed across recall totals for the 8- and 16-word experiments and the best fitting ex-Gaussian (Equation 2). The parameter estimates are given in Table 7.

of item strengths yields CVs between 1 and 1.2. The IRT means and IRT standard deviations for the present data are presented in Table 4. In contrast to the prediction, the mean CV was 0.85 and only 3 of the 20 exceeded 1.0. However, given the view that each IRT includes a recovery stage of constant duration, these CVs are underestimates. When the CVs were recalculated by dividing the standard deviations by the estimates of S (rather than $S + R$), the new mean CV equaled 1.16 and 16 of the 20 values were greater than 1.0.

In summary, the nature of IRT distributions is consistent with the relative strength model, but the evidence is less convincing than that given by the analyses of latency distributions and IRT growth. In general, analyses of IRT distributions are difficult because IRTs must be partitioned by both recall total and output position, thereby requiring a large number of response times. This probably explains their absence from the literature.

GENERAL DISCUSSION

The existence of a recovery threshold provides one explanation of why previous tests of the equal-strength model

have been successful despite the fact the item strengths cannot be equal. As illustrated by the simulations in Figures 1 and 4, the deviation from the equal-strength model that results from variable item strengths is essentially nullified by a recovery threshold. In effect, the recovery threshold prohibits the overt recall of weak, below-threshold items that are sampled late in the recall period, thereby decreasing the variability in the strengths of the *recalled* items. More generally, scenarios such as these illustrate a difficulty that plagues model testing. Because one feature of a model can either enhance or negate the effects of another, any test of a model with two or more features can only be tested as a whole. If the model fails that test, none of the model's features can be logically ruled out as being uniquely responsible for the failure.

The use of curve fitting as a means to "rule in" a model is an often-abused methodology, but the tests of the relative-strength model presented herein avoid many of the usual pitfalls. First, the model is quite simple. If the brief, 1-sec pause that precedes recall is ignored, both latency distributions (noncumulative or cumulative) and mean IRT growth can be described by a single parameter—mean latency (τ). Second, the parameter estimate of mean latency has a psychological interpretation. If an experimental manipulation affects mean latency, then search set size must have increased as well. Third, and most importantly, the parameter estimate of mean latency also has an *empirical* meaning. That is, mean latency can be directly calculated by simply averaging response latencies; for the data described above, mean latency equaled about 6 sec for each recall total. Therefore, the functions shown in Figures 2, 3, and 5 could have been plotted without first fitting them to the data. In short, the model predicts both the qualitative and quantitative nature of latency distributions and mean IRT growth.

The 16-Word Experiment

Because the experiment described herein included only one study list length, the generalization of these findings to other list lengths warrants consideration. The most commonly used list lengths range from only 3 words—as in the classic proactive interference experiments—to as many as 15 or 20 words (though much longer lists have been used). Thus, the 8-word list used in the present experiment lies roughly in the middle of this range. Of course, it makes little sense to examine the independence of mean latency and recall total for 3-word study lists because so few recall totals are possible. There is, however, good reason to extend the present analyses to longer list lengths.

Table 7
Ex-Gaussian Fits of Latency Distributions Collapsed Across Recall Total
for the 8-Word and 16-Word Experiments (Figure 7)

Study	Initiation		Mean Latency		Goodness-of-Fit		
Total	μ	SE	t	SE	N	χ^2	df
8	1.12	.03	5.98	.12	2,594	86.53*	25
16	1.10	.04	11.11	.26	1,898	95.83*	38

Note—Initiation and latency are given in seconds.   *$p < .05$.

Table 8
Ex-Gaussian Fits of Latency Distributions for 16-Word Experiment
(Figure 8)

Recall Total (n)	Initiation		Mean Latency		Goodness-of-Fit		
	μ	SE	τ	SE	N	χ^2	df
4	1.02	.16	10.12	.89	132	20.07	13
5	1.16	.08	9.20	.67	190	20.97	15
6	1.18	.11	10.64	.64	282	29.57	21
7	0.99	.12	10.69	.58	350	33.72	23
8	0.99	.15	11.78	.80	224	25.77	20
9	1.08	.15	12.30	.90	189	25.55	18
10	1.65	.35	11.77	1.09	130	10.89	14
11	1.39	.36	12.59	1.31	99	12.07	11
12	1.20	.19	10.95	1.20	84	12.07	11

Note—Initiation and latency are given in seconds.

Therefore, an additional 25 subjects were tested with 16-word study lists. Other than a longer recall period (45 sec) and fewer trials per subject (12), the procedure in this experiment was identical to that in the 8-word experiment. Subjects recalled an average of 6.33 words per trial, or 1,898 words in all, and 3.7% of these responses were errors. Two brief analyses are presented here. Latencies are analyzed when trials are grouped across recall total, as in Rohrer and Wixted (1994), and when trials are partitioned by recall total, as in the present study.

Figure 7 includes latencies collapsed across all possible recall totals for both the 8- and 16-word experiments as well as the best fitting ex-Gaussian distributions. These fits closely resemble those observed in Figure 2. In addition, the exponential tail of each best fit function slightly underestimates the first few data points and then overestimates the data for most of the remainder of the tail. The same pattern was observed in the latency distribution given by the variable-strength-with-threshold simulation shown in Figure 1.

Though the latency distributions given by the 8- and 16-word experiments are both described by the same function, their means differed considerably. As listed in Table 7, both estimates of initiation time (μ) equaled about 1 sec, whereas the estimate of mean latency (τ) for the 16-word experiment was almost double that for the 8-word experi-

ment—11.11 versus 5.98. Of course, subjects in the 16-word experiment had a longer recall period—45 versus 30 sec—yet an examination of the latency distributions in Figure 7 reveals that the rate of recall for both groups of subjects had essentially reached that floor by the end of the recall period. Besides, the increase in mean latency with longer study lists is both a robust and unsurprising effect that has previously been established (see Introduction).

In keeping with the findings of the 8-word experiment, mean latency and recall total were also effectively independent in the 16-word experiment. As listed in Table 8, estimates of mean latency given by fits of the ex-Gaussian for recall totals 4–12 are roughly constant. (These latency distributions for these recall totals were the only ones with more than 75 responses.) As in the 8-word experiment, mean latency increased by an average of about 200 msec per extra study word. Of course, as noted earlier, greater recall totals require subjects to voice more words aloud. Perhaps not coincidentally, 200 msec is approximately the time needed to voice a monosyllabic word. In short, the independence of mean latency and recall total appears to hold across a relatively wide range of recall totals. As noted earlier, this independence is perhaps counter to the natural intuition because latency is measured from the beginning of the recall period.

Finally, a comparison of the results for the two experiments illustrates an intriguing prediction of the model. As
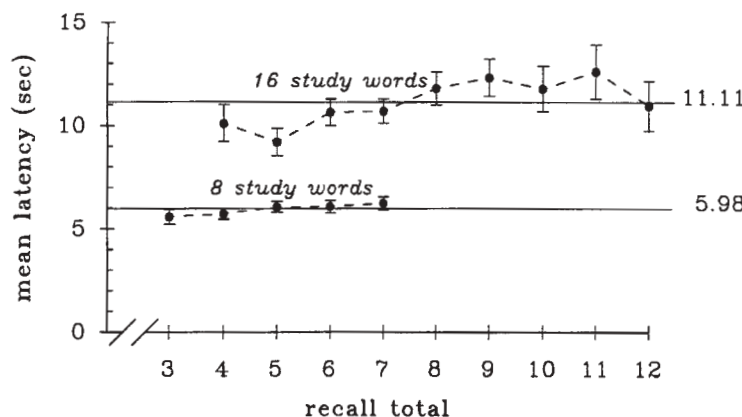
Figure 8. Mean latency (τ) as a function of recall total for both the 8- and 16-word experiments (Table 8). Error bars represent ±1 standard error.

shown in Figure 8, mean recall latency is determined by study total, not recall total. These findings contrast with the deeply held intuition that the time required to recall a set of items is determined by the number of words that must be spoken. In addition, these findings are not easily explained by any theory that construes recall total as a measure of search set size.

Summary

Three primary findings have been reported. First, the primary heuristic value of the equal-strength model, its interpretation of mean latency as a measure of search set size, was seen to be effectively retained when the assumption of equal item strengths was relaxed. As illustrated in Table 1, the varying of item strengths has only a small effect on mean latency and, because mean strength can only increase with variability (see Appendix), relative comparisons of mean latency remain relatively stable. Second, the means and distributions of both latencies and IRTs were shown to be quite adequately described by the relative-strength model when a recovery threshold is present. Third, the partitioning of trials by recall total in both the 8- and 16-word experiments revealed that mean latency was roughly proportional to list length yet independent of recall total.

Theoretically, all three of the findings are predicted by a simple two-stage model of free recall. According to the model, a search set composed of all study items is sampled according to the relative-strength rule until all items are found, and a sampled item is recovered into consciousness only when its absolute strength exceeds a fixed threshold. In short, the retrieval of an item from long-term memory is governed by both its relative and absolute strength.

REFERENCES

ALBERT, D. (1968). Freies Reproduzieren von Wortreihen als stochastische Entleerung eines Speichers [Free recall of word sequences as stochastic depletion of storage]. *Zeitschrift für experimentelle und angewandte Psychologie*, **15**, 564-581.

BOUSFIELD, W. A., & SEDGEWICK, C. H. (1944). An analysis of restricted associative responses. *Journal of General Psychology*, **30**, 149-165.

BOUSFIELD, W. A., SEDGEWICK, C. H., & COHEN, B. W. (1954). Certain temporal characteristics of the recall of verbal associates. *American Journal of Psychology*, **67**, 111-118.

BROWN, R., & MCNEILL, D. (1966). The "tip of the tongue" phenomenon. *Journal of Verbal Learning & Verbal Behavior*, **5**, 325-337.

GRONLUND, S. D., & SHIFFRIN, R. M. (1986). Retrieval strategies in recall of natural categories and categorized lists. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **12**, 550-561.

HOHLE, R. H. (1965). Inferred components of reaction times as function of foreperiod duration. *Journal of Experimental Psychology*, **69**, 382-386.

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.

MURDOCK, B. B., JR., & OKADA, R. (1970). Interresponse times in single-trial free recall. *Journal of Experimental Psychology*, **86**, 263-267.

PAYNE, D. G. (1986). Hypermnesia for pictures and words: Testing the recall level hypothesis. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **12**, 16-29.

RAAIJMAKERS, J. G., & SHIFFRIN, R. M. (1980). SAM: A theory of probabilistic search of associative memory. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 14, pp. 207-262). New York: Academic Press.

RATCLIFF, R., & MURDOCK, B. B., JR. (1976). Retrieval processes in recognition memory. *Psychological Review*, **83**, 190-214.

ROEDIGER, H. L., III, STELLON, C. C., & TULVING, E. (1977). Inhibition from part-list cues and rate of recall. *Journal of Experimental Psychology: Human Learning & Memory*, **3**, 164-188.

ROEDIGER, H. L., III, & THORPE, L. A. (1978). The role of recall time in producing hypermnesia. *Memory & Cognition*, **6**, 296-305.

ROHRER, D., & WIXTED, J. T. (1994). An analysis of latency and inter-response time in free recall. *Memory & Cognition*, **22**, 511-524.

ROHRER, D., WIXTED, J. T., SALMON, D. P., & BUTTERS, N. (1995). Retrieval from semantic memory and its implications for Alzheimer's disease. *Journal of Experimental Psychology: Human Learning & Memory*, **21**, 1127-1139.

RUNDUS, D. (1973). Negative effects of using list items as recall cues. *Journal of Verbal Learning & Verbal Behavior*, **12**, 43-50.

SHIFFRIN, R. M. (1970). Memory search. In D. A. Norman (Ed.), *Models of human memory* (pp. 375-447). New York: Academic Press.

VORBERG, D., & ULRICH, R. (1987). Random search with unequal search rates: Serial and parallel generalizations of McGill's model. *Journal of Mathematical Psychology*, **31**, 1-23.

WIXTED, J. T., & ROHRER, D. (1993). Proactive interference and the dynamics of retrieval. *Journal of Experimental Psychology: Learning, Memory, & 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

This proof shows that, for a search set with N items, mean latency is minimized when each item has the same strength. According to the variable-strength model, the probability of sampling item i equals $s_i / \sum s_j$ (see Introduction). Therefore, τ_i, the expected latency of item i, equals the reciprocal of this ratio ($\sum s_j / s_i$), when latency is measured in samples. Thus, if s equals the mean strength,

$$\text{mean latency } (\tau) = \{\sum[(\sum s_j)/s_i]\}/N$$
$$= \{\sum[(Ns)/s_i]\}/N$$
$$= \sum(s/s_i). \tag{A1}$$

If the strengths do not vary (all $s_i = s$), then Equation A1 can be simplified as follows.

$$\text{mean latency } (\tau) = \sum(s/s)$$
$$= N. \tag{A2}$$

Therefore, in order to show that mean latency is minimized when item strengths are constant, it is sufficient to show that Equation A1 > Equation A2, or

$$\sum(s/s_i) > N,$$
$$\sum[(s_i-d_i)/s_i] > N \text{ where deviation } d_i = s_i-s,$$
$$\sum(1-d_i/s_i) > N,$$
$$\sum(1)-\sum(d_i/s_i) > N,$$
$$N-\sum(d_i/s_i) > N,$$
$$\sum(d_i/s_i) < 0,$$
$$\sum[d_i/(s+d_i)] < 0. \tag{A3}$$

The summation in Equation A3 can be written as the sum of three summations—one for positive d_i, one for negative d_i, and one for zero d_i, giving

$$\sum_{d_i>0}[d_i/(s+d_i)] + \sum_{d_i<0}[d_i/(s+d_i)] + \sum_{d_i=0}[d_i/(s+d_i)] < 0,$$

$$\sum_{d_i>0}[d_i/(s+d_i)] + \sum_{d_i<0}[d_i/(s+d_i)] + 0 < 0,$$

$$\sum_{d_i>0}[d_i/(s+d_i)] - \sum_{d_i<0}[|d_i|/(s-|d_i|)] < 0$$

because $d_i = -|d_i|$ for $d_i < 0$,

$$\sum_{d_i>0}[d_i/(s+d_i)] < \sum_{d_i<0}[|d_i|/(s-|d_i|)]. \tag{A4}$$

Thus, if Equation A4 is true, the desired result is obtained. The left side of Equation A4,

$$\sum_{d_i>0}[d_i/(s+d_i)]$$

$$< \sum_{d_i>0}(d_i/s)$$

$$= [\sum_{d_i>0}(d_i)]/s$$

$$= [\sum_{d_i<0}|d_i|]/s \text{ because } \sum d_i = 0$$

$$= \sum_{d_i<0}(|d_i|/s)$$

$$< \sum_{d_i<0}[|d_i|/(s-|d_i|)].$$

Thus, Equation A4 is true. For N items within a search set, mean latency (τ) is minimized when each item has the same strength.