JOURNAL OF MATHEMATICAL PSYCHOLOGY: 3, 316-347 (1966)

Strength Models and Serial Position in Short-Term Recognition Memory[1]

WAYNE A. WICKELGREN

Massachusetts Institute of Technology, Cambridge, Massachusetts

AND

DONALD A. NORMAN

Harvard University, Cambridge, Massachusetts

A number of continuous strength models for memory are developed for and tested by an experimental study of recognition memory for three-digit numbers at all serial positions in lists of length two through seven. Empirical estimates of trace strength in different conditions, independent of response bias, are obtained by means of the operating characteristic. The principal theoretical findings are: (a) strength in short-term memory (STM) appears to decay exponentially with the number of subsequent items; (b) subjects report that they recognize an item if and only if strength in memory exceeds a criterion; (c) the first item of a list is remembered better than subsequent items because it receives a greater increment in strength in STM upon presentation, not because it decays more slowly in STM or because it acquires some strength in a long-term memory.

This paper has two general purposes. First, we develop several strength models of recognition memory. Second, we evaluate the fit of these models to the data from an experimental study of the serial position effect in recognition memory of items (digit triples) from lists of different lengths. We wish to account for the response probabilities for test items from each serial position in the lists. To describe these data we develop a model with two different psychological processes: memory and decision. Every item, whether presented or not, is assumed to have some strength in memory. The memory model describes the changes in memory strengths as each item in the list is presented. The decision model describes how the subject uses these strengths to choose his response.

[1] D.N. was supported by grant MH 08083-02, to the Harvard Center for Cognitive Studies, and W.W. was supported by grant MH 08890-01, from the National Institute of Mental Health, U.S. Public Health Service. We thank Marta Weigle (S4) for her assistance in analyzing the data and serving as a subject.

On each trial of the experiment, subjects listened to a list of L different items which was followed by a test item. The subject's task was to decide whether the test item had been presented in the previous L items. If so, he was to say "yes"; otherwise he was to say "no." The subject then stated his confidence in the correctness of the response. The items were selected so that every serial position in lists of length two through seven was tested; in about one-third of the test trials the test item had not appeared in the list.

Decision Models

We assume that every possible test item has some representation in memory to which we can assign a unidimensional measure of strength, even if that item has not yet been presented to the subject. When the test item is presented, the subject chooses his response by considering the strength of the test item in memory and his response biases. There are several ways in which he might do this, two of which are considered in this paper. The two decision models are, first, a criterion rule similar to the decision process of signal detection theory (Swets, Tanner, and Birdsall, 1961) and the paired-comparison model of Thurstone (1927), and, second, a ratio rule, coming from Luce's choice theory (1959, 1963).

The Criterion Rule

Let the output of the memory system for some test item i be a real number, t_i. Assume that the subject's biases determine a response criterion c and he responds *yes* (he says that he recognizes the test item) if and only if $t_i - c \geqslant 0$. Our models of the memory trace are algebraic so that the probabilistic properties of the responses are due entirely to variability in the decision process. We represent this variability by the addition of a random variable $\mathbf{X}$ (noise) to the decision process. It does not matter where the noise is considered to be added—e.g., in the transmission of memory trace strengths to the decision stage or in the location of the decision criterion—the end result is the same. With the addition of noise, the decision rule becomes:

$$\text{respond } \textit{yes} \text{ to item } i \text{ iff } [t_i - c + \mathbf{X}] \geqslant 0,$$

and so

$$\Pr(\text{yes} \mid i) = \Pr\{[t_i - c + \mathbf{X}] \geqslant 0\}. \tag{1}$$

This process is summarized by Fig. 1.

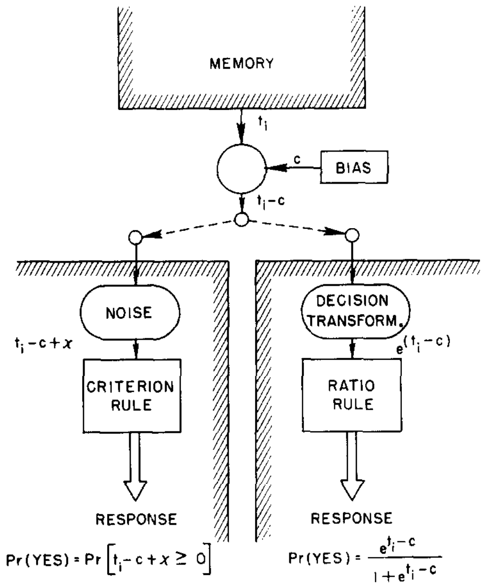

FIG. 1. Structural model of the memory system and two alternative decision systems.

Before we can solve Eq. 1, we need to know the probability density function of the noise distribution. Usually this is taken to be the normal density function with zero mean and standard deviation of σ, $N(0, \sigma)$. In this case

$$\Pr(\text{yes} \mid i) = \int_0^\infty N(t_i - c, \sigma) = \int_{(c-t)/\sigma}^\infty N(0, 1). \tag{2}$$

As is well known, the logistic function, $p = 1/[1 + e^{-(x-b)/a}]$, is very similar to the cumulative normal (Bush, 1963). If we let the probability density function for the noise be given by the derivative of the logistic we can get an explicit expression for response probability. Solving Eq. 1 using the logistic (replacing the variable x by $t_i - c$ and letting the mean, b, be zero and the "spread," a, be one) we find

$$\Pr(\text{yes} \mid i) = \frac{e^{(t_i - c)}}{1 + e^{(t_i - c)}}. \tag{3}$$

THE RATIO RULE

According to the ratio rule of Luce's choice theory (Luce, 1959, 1963), the probability that the subject responds "yes" to item i is

$$\Pr(\text{yes} \mid i) = \frac{w_i b}{1 + w_i b}, \tag{4}$$

where w_i depends upon strength and b is a bias parameter, w_i and $b \geqslant 0$. It turns out that it is possible to find a transformation of the output of the memory system (including the bias) that maps the difference $t_i - c$ onto the product $w_i b$ such that when the ratio rule is applied to the transformed strengths it is completely equivalent to the criterion rule with logistic noise.[2] The required transformation is found by equating Eqs. 3 and 4 and solving for $w_i b$:

$$\Pr(\text{yes} \mid i) = \frac{w_i b}{1 + w_i b} = \frac{e^{t_i - c}}{1 + e^{t_i - c}}$$

and $w_i b = e^{t_i - c}$.

The application of this transformation is illustrated in Fig. 1. When the decision process is the ratio rule, the output of the memory system must be exponentially transformed before the decision rule is applied. When the decision process is the criterion rule the transformation is not needed but, in its place, noise must be added. Of course, were $w_i b$ the output of the memory system, no transformation would be needed with the ratio rule but a logarithmic transformation would be needed with the criterion rule. The form of the exponential transform tempts one to consider that it applies separately to strength and criterion, but this is not necessarily true.

This is an interesting result. For "yes-no" recognition memory, the criterion decision rule and the ratio rule produce identical transformations of the response probabilities into underlying strengths in memory, under the following conditions:

1. the probability distribution of the noise is logistic, with equal variance for all items;

2. either

A. an exponential transformation is applied to the strength and bias variables before applying the ratio rule, or

B. a logarithmic transformation is applied to the strength and bias before applying the criterion rule.

For "yes-no" recognition memory, the differences in prediction between the ratio rule and the criterion rule with normally distributed noise (of constant variance for all items) result entirely from the differences between the cumulative normal and the logistic distributions. These differences are very small and show up mainly in the tails of the distributions (Bush, 1963). We do not attempt to distinguish between these two but simply use the normal distribution and criterion rule because of the availability of convenient tables (Elliott, 1964) and graph paper.

[2] The equivalence of criterion and ratio rules is discussed by Torgerson (1958, p. 201) for unbiased decisions (Thurstone's law of comparative judgment and the Bradley-Terry ratio rule). Proof that the logistic is the only distribution that leads to equivalence of the two rules is given by Burke and Zinnes (1965) and Adams and Messick (1957).

The equivalence of criterion and ratio rules does not mean that the two different scales they impose on the underlying memory (or sensory) system are equally desirable. We may wish to choose between the two decision processes on the basis of the mathematical simplicity of the memory model. A model that appears simple mathematically when used with one decision rule may appear to be much more complex—and therefore much less satisfactory—after it has been transformed for application with the other rule.

OPERATING CHARACTERISTICS

Our memory models describe the way memory strength varies with serial position and list length. We do not have a comparable model for the decision system to tell us how response bias might change with these variables. Yet in order to predict response probabilities we need to know both the memory strength and the bias value for each item that is tested. However, we can predict the relationship between two response probabilities without knowing the values of the biases, provided we are willing to assume that the two biases are the same.

In this paper we compare the response probability for an old item with the response probability for a new item. The probability of a "yes" response to an old test item, $p(y \mid o)$, is some function of the trace strength of the old item, t_o, and the bias, c_o: $p(y \mid o) = h(t_o, c_o)$. Similarly, the probability of a "yes" response to a new test item, $p(y \mid n)$, is the same function of the new item's trace strength and bias: $p(y \mid n) = h(t_n, c_n)$. If we assume that $c_o = c_n = c$, and if we can solve for the bias term, $c = g[t_n, p(y \mid n)]$, then we can determine the relationship between the response probabilities:

$$p(y \mid o) = h\{t_o, g[t_n, p(y \mid n)]\} = f[t_o, t_n, p(y \mid n)].$$

The function f is called the operating characteristic.

When the decision process uses the criterion rule with normally distributed noise, to solve for the bias, c, we define the tails normal deviate transformation on p, [TND (p)], such that

$$\text{TND}(p) = \text{TND} \int_a^\infty N(b, \sigma) = \frac{b - a}{\sigma}.$$

Taking the TND of both $p(y \mid o)$ and $p(y \mid n)$ as given by Eq. 2 (letting the standard deviation of the noise distribution be different for new and old items, just for the moment) we get

$$\text{TND}\, p(y \mid o) = \frac{t_o - c}{\sigma_o} \qquad \text{and} \qquad \text{TND}\, p(y \mid n) = \frac{t_n - c}{\sigma_n}.$$

Solving for c and equating we get the operating characteristic

$$\mathrm{TND}\, p(y \mid o) = \frac{\sigma_n}{\sigma_o} \mathrm{TND}\, p(y \mid n) + \frac{t_o - t_n}{\sigma_o}. \tag{5}$$

This operating characteristic is linear with a slope of σ_n/σ_o and intercept of $(t_o - t_n)/\sigma_o$ when we plot $\mathrm{TND}\, p(y \mid o)$ versus $\mathrm{TND}\, p(y \mid n)$ or, what is equivalent, plot the response probabilities on normal-normal probability coordinates. In our analyses we let $\sigma_o = \sigma_n = 1$ so that we expect the operating characteristic to have unit slope, and memory strengths have the standard deviation of the noise distribution as the unit of measurement.

When the decision process is either the criterion rule with logistic noise or the ratio rule we get the operating characteristic by solving Eq. 3 or 4. This gives

$$\frac{p(y \mid o)}{1 - p(y \mid o)} = \left(\frac{w_o}{w_n}\right) \left(\frac{p(y \mid n)}{1 - p(y \mid n)}\right),$$

where $w_o/w_n = e^{(t_o - t_n)}$. Plotting the ratio $p(y \mid o)/[1 - p(y \mid o)]$ versus $p(y \mid n)/[1 - p(y \mid n)]$ gives a linear operating characteristic which passes through the origin. An alternative expression comes from using the logistic equivalent of the TND, namely, logit p, which by definition is $\log [p/(1 - p)]$ (Bush, 1963). From Eq. 3, $\mathrm{logit}\, p(y \mid i) = t_i - c$. As usual the operating characteristic can be obtained by eliminating c so that

$$\mathrm{logit}\, p(y \mid o) = \mathrm{logit}\, p(y \mid n) + (t_o - t_n).$$

This operating characteristic is linear with a slope of unity and intercept $t_o - t_n$ when plotted on logit-logit probability coordinates.

Both the criterion and ratio decision processes predict very similar operating characteristics that are functions of the difference in trace strength between old and new items $(t_o - t_n)$. We can determine relative strengths of items directly from the operating characteristics, and it is this relative strength that is predicted by the various memory models that are considered later in the paper.

Equal Bias Assumption

In solving for the form of the operating characteristic it is necessary to assume that the subject's response bias is the same for the two conditions being compared. This assumption is reasonable because most of the possible biases of the subject apply equally to all experimental conditions. For example, bias for a particular stimulus item does not much matter, because each stimulus item has equal *a priori* likelihood of appearing in any experimental condition, and each item usually appears in several conditions. This type of bias would, therefore, appear as an unsystematic perturbation of bias values which would lower our strength estimates (by increasing σ) but have no effect on the form of our results. If the subject varies his bias with list length, no problem is created because we estimate strength values by comparing conditions within the same list length. It is difficult to see how the subject could vary his bias with serial position because knowing the serial position of an item is equivalent to knowing whether the item appeared in the list. Therefore, we assume that the bias is equal for items in all serial positions (including nonpresented items) of lists of the same length. This assumption could be invalid if subjects went through a two-stage process of first guessing the serial position of the test item (*assuming* the item had been in the list), letting this determine their bias, and then deciding whether the item had been in the list. However, this seems unlikely.

Method of Confidence Judgments

To plot the operating characteristic from data we need to observe the response probabilities as we vary response bias (c) but hold constant the memory strengths, t_o and t_n. If the subject responds either "yes" or "no," we get only one pair of response probabilities $[p(y \mid n), p(y \mid o)]$ for each experiment. To get more points on the operating characteristic we must do separate experiments, varying response bias but holding memory strengths constant. A more efficient method of estimating the operating characteristic is the method of confidence judgments, developed by Egan (1958) and Egan, Schulman, and Greenberg (1959). The subject is asked to section the decision continuum into N regions and respond by saying in which region the strength of the test item lies. He does this by responding both with a yes-no judgment and then a statement of his confidence in that response. Assuming that confidences correspond directly to strengths, a report of high confidence for a "no" response must represent a very weak strength, and a report of high confidence for a "yes" response a very strong strength. If the confidences range numerically from five (extremely confident) through one (no confidence) we assume that the strength continuum is partitioned as shown in Fig. 2. Note that for N categories there are $N - 1$ criteria. If we identify the boundary between confidence judgments with a bias or criterion value, we see that the probability that a subject responds Y3, Y4, or Y5, for example, is the same as the probability that he would say "yes" with a criterion at c_3. Thus, in the experimental situation illustrated in Fig. 2 with ten response categories, we can transform the results into the probability of a "yes" response for nine different criteria. The probability of "yes" for each of these nine criteria is the cumulative probability of giving any response-confidence pair to the right of each criterion. The operating characteristic comes from looking at these nine probabilities conditional upon an old test item plotted against the corresponding nine probabilities conditional upon a new test item.

The conversion from response-confidence pairs to criterion judgments is relatively straightforward. It is not quite so obvious how to transform the biases in the ratio rule into confidence judgments and we do not attempt to do so here.

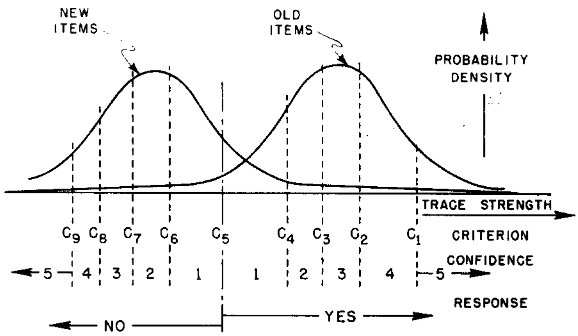

FIG. 2. Strength distributions for old and new items. The vertical lines illustrate locations of criteria for the various responses. The probability that a subject respond "no-1" to an old test item is the area under the curve for old items between criteria c_6 and c_5.

The confidence rating technique allows us to get nine points on the operating characteristic in one experiment but depends heavily upon the assumption that subjects can categorize their responses along the strength continuum in a stable manner. If the various criteria vary over the course of the experiment or are different for different test items the results may be inappropriate. Note, however, that variability in locating each criterion does not necessarily cause trouble. If we can characterize the variability of criterion c_j by the addition of noise, $\mathbf{Y}_j$, to the criterion location, then the decision rule is simply to choose a response to the right of c_j iff $[t_i - c_j - \mathbf{Y}_j + \mathbf{X} \geqslant 0]$. If the criterion variability is the same for all the criteria then we can simply combine the terms $\mathbf{X}$ and $\mathbf{Y}$ (as long as $\mathbf{X}$ and $\mathbf{Y}$ are not perfectly correlated) to form a new noise distribution and redefine our scale values so that the standard deviation of the new $\mathbf{X} + \mathbf{Y}$ distribution is taken as the unit. Thus, even with decision variability we can assume stationary criteria and added noise (or, for that matter, no added noise but variability in the criterion value).

MEMORY MODELS

In this section we discuss the ways that the psychological processes of acquisition, decay, and generalization might be represented by a set of assumptions about real-valued memory traces. Various subsets of these assumptions are combined into models that describe the way trace strengths in memory change with list length and serial position. As we have seen in the previous section, the observable datum, the operating characteristic, depends upon relative trace strengths $(t_0 - t_n)$ rather than absolute values. This limitation on what can be observed means that some of our models which are different in theory turn out to be experimentally indistinguishable.

Definitions

As each item in the experimental list is presented to the subject, the memory strengths of all the possible items are changed. Because items are assumed to be homogeneous, they only differ in the serial position in which they are presented in the experimental list. It is most convenient, therefore, to label the stimulus items by their location in the list; thus, when we speak of the strength of stimulus item x, we refer to the strength in memory of the item which is to be (or has been) presented in position x of a list. If there are L items in a list, x may have $L + 1$ values: any of the L list positions and the null position. If an item is not presented in the list we say it is a new item and let its (null) position, x, represented by *. As the stimulus list is presented, the strength of item x can be tested at any point. Two numbers are needed to specify any item: the serial position in which it is presented and the serial position at which its strength is sampled (which may be before it is to be presented).

Let $s(x, y)$, $\ell(x, y)$, and $t(x, y)$ be the strengths of the short-term (STM), long-term (LTM), and total memory traces for an item which is presented in position x and tested after y items have been presented, counting from the beginning of the list. If the list length is L, the first variable, presentation position, may have any value from * through L and the second variable, test position, runs from 0 through L. We use the letter "k" to refer to a position that specifically excludes the null position *, thus, $1 \leqslant k \leqslant L$. The difference in total strength between old and new items, $d(k, L)$, is $t(k,L) - t(*, L)$.

Assumptions about STM

The fundamental operations that we presume occur in STM are those of acquisition and decay. Presentation of an item causes the strength of its representation in memory to increase above its current strength. Presentation of other items causes its strength to decay. A third process, generalization, may also occur in STM so that, at times, items undergo partial acquisition.

Two different types of acquisition and generalization assumptions are examined: incremental and proportional acquisition. These assumptions lead to two different models, one with incremental acquisition and generalization assumptions, the other with proportional acquisition and generalization assumptions. Other reduced models are possible by systematic elimination of the various parameters.

Initial strength. At the start of each trial all items have an initial strength in STM of α_0:

$$s(x, 0) = \alpha_0 , \qquad 0 \leqslant \alpha_0 .$$

Decay. Presentation of an item that is different from the item presented in position x causes the strength of x in STM to decrease to some proportion φ of its strength:

$$s(x, k) = \varphi s(x, k - 1), \qquad 0 \leqslant \varphi \leqslant 1, \qquad x \neq k.$$

ACQUISITION. We consider two possible forms for the acquisition of items in STM: incremental and proportional. In both cases acquisition occurs in addition to decay. The presentation of an item has two effects: first, its memory strength is decreased by the decay assumption; second, its strength is increased by the acquisition process.

According to the *incremental acquisition* assumption, presentation of an item adds a constant amount, α, to the decaying strength of the item. Thus, strength in memory increases with each presentation to some limit set by the rate of decay:

$$s(k, k) = \alpha + \varphi s(k, k-1), \qquad 0 \leqslant \alpha.$$

According to the *proportional acquisition* assumption there is some maximum bound on strengths. Presentation of an item increases its strength by some proportion of the distance between the decaying strength value and the maximum strength. We study only the case where the proportion is unity. In this case trace strength in STM reaches its maximum value, α, in one trial:

$$s(k, k) = \alpha, \qquad 0 \leqslant \alpha.$$

Generalization. We assume that generalization causes items similar to the presented item to undergo partial acquisition. We examine two types of generalization corresponding to our two types of acquisition assumptions. Let σ, the generalization parameter, be the degree of similarity between any two stimulus items, $0 \leqslant \sigma \leqslant 1$. In the experiments reported in this paper, similarity is assumed to be constant over all pairs of nonidentical stimulus items.

First, by analogy with the incremental assumption for acquisition, let the presentation of an item increase the strength of similar items by some constant amount. Thus, we obtain *incremental generalization*:

$$s(x, k) = \sigma\alpha + \varphi s(x, k-1), \qquad x \neq k, \qquad 0 \leqslant \alpha, \qquad 0 \leqslant \sigma \leqslant 1.$$

This assumption becomes identical to the decay assumption when $\sigma = 0$ and identical to the acquisition assumption when $\sigma = 1$.

Second, by analogy to the proportional acquisition assumption, let the presentation of an item increase the strength of similar items by some proportion σ of the distance remaining between the present trace strength and the maximum possible strength, α Thus, we obtain *proportional generalization*:

$$s(x, k) = \sigma[\alpha - \varphi s(x, k-1)] + \varphi s(x, k-1), \qquad x \neq k, \qquad 0 \leqslant \alpha, \qquad 0 \leqslant \sigma \leqslant 1.$$

This assumption becomes identical to the decay assumption when $\sigma = 0$ and identical to the acquisition assumption when $\sigma = 1$.

Assumptions about LTM

Although it is possible to consider LTM in as much detail as we have considered STM, there seems little reason to do so since in our experiment the LTM component is certain to be less important than the STM component. Therefore we consider only one model for LTM in which there is an initial strength in LTM of λ_0, incremental acquisition, no generalization, and perfect retention over the duration of the trial. The assumptions of the model are:

$$\ell(x, 0) = \lambda_0, \qquad 0 \leqslant \lambda_0,$$

$$\ell(x, k) = \ell(x, k-1), \qquad x \neq k.$$

and

$$\ell(k, k) = \lambda + \ell(x, k-1), \qquad 0 \leqslant \lambda.$$

Combination of STM and LTM

We assume that total memory strength is an *additive combination* of strength in STM and strength in LTM.

$$t(x, k) = s(x, k) + \ell(x, k).$$

We have no justification, except mathematical simplicity, for choosing additive combination of traces over other combinations. LTM plays such a small part in our experiment that it does not seem wise to investigate alternative formulations. What seems important is to show that the present strength models can encompass both STM and LTM in a single consistent framework.

Primacy

Early items in a list are often remembered better than items in the middle of the list. No model built out of the previous assumptions will predict this. There are many ways to account for primacy with a strength model. One could assume that the level of acquisition in STM is a function of serial position, or decay in STM could be different for the first item(s) than for later ones.

As one hypothesis, we consider primacy in our experiment to be the result of greater learning at the time of presentation, in STM, or LTM, or both. Various intuitive reasons can be given for this, such as more intense or prolonged attention to the first items, but whatever the reason, such a primacy effect can be described by assuming that the acquisition parameters, α and λ, are functions of serial position, k. Since most of the primacy effect is on the first item, and since we do not have a theory giving us the explicit function, we simply approximate it by a step function: let $\alpha(k)$ and $\lambda(k)$ be estimated separately for the first position of the list so that $\alpha(k)$ and $\lambda(k)$ are α_1 and λ_1 for $k = 1$ and simply α and λ for $k \geqslant 2$.

As the other hypothesis we let the rate of decay in STM be a (step) function of serial position so that $\varphi(k) = \varphi_1$ for $k = 1$ and simply φ for $k \geqslant 2$. Primacy effects on decay in STM could result from rehearsal or consolidation that occurs after presentation and so might vary with serial position.

Selection of Models for Theoretical Investigation

A very large number of memory models can be constructed by putting together the various possible combinations of our assumptions. There is no reason to discuss them all in this paper. Instead we have selected a number of models for discussion, guided by the fact that decay in STM is by far the largest effect under consideration. Everything else is a second-order effect compared to this decay. We use as our basic model the STM Decay model, constructed of a minimum of assumptions: no primacy, no generalization, and no LTM. We compare this model with a few more complicated models that include one or more of these other processes. The empirical predictions of these models are derived and certain equivalences are demonstrated.

STM Decay Model

This is the basic model from which all the others build. In it we ignore primacy, generalization, and LTM. There are three versions of the decay model, depending upon our choice of initial strength and acquisition assumptions. Regardless of the acquisition axiom, decay is exponential, $s(x, k) = \varphi s(x, k - 1)$.

Strictly speaking, the decay should be referred to as "geometric," because it is discrete. We choose to use the term "exponential decay" for two reasons. First, it has been commonly used in the past to refer to forgetting in discrete-trial experiments. Second, we do not think that the forgetting occurs instantaneously at discrete times, but rather that it occurs continuously over the trial. Geometric decay is the discrete analog of exponential decay, and the discrete representation is simply more convenient for the present experiment. It should be noted that the assumption of exponential decay is neutral with respect to whether the decay occurs with the passage of time or as the result of interfering activity.

If we make the simplifying assumption that $s(x, 0) = 0$, then $s(*, L) = 0$ for all L, and $s(k, k) = \alpha$ no matter which acquisition assumption we choose. Thus,

$$d(k, L) = s(k, L) = \alpha\varphi^{L-k}.$$

Assuming nonzero initial strength, $s(x, 0) = \alpha_0 > 0$, and the constant-increment assumption about acquisition, we get the same final equation for $d(k, L)$,

$$\begin{aligned} d(k, L) &= s(k, L) - s(*, L) \\ &= \alpha\varphi^{L-k} + \alpha_0\varphi^L - \alpha_0\varphi^L \\ &= \alpha\varphi^{L-k}. \end{aligned}$$

However, with the one-trial proportional assumption and nonzero initial strength, the form of the final equation is changed to:

$$d(k, L) = \alpha\varphi^{L-k} - \alpha_0\varphi^L.$$

We shall not consider further this more complicated decay model.

If normal (or logistic) noise is added to the trace strengths specified by the STM Decay Model, one obtains a normal (or logistic) density function with exponentially decaying mean and constant standard deviation. When the criterion rule is used to convert trace strengths into response probabilities and when trace strength starts out substantially above the criterion, the probability of responding "yes" to the test item with that strength will start out with a probability very close to unity, drop very slowly while the left tail of the distribution is moving past the criterion, then drop rapidly as the center of the distribution moves past the criterion, and finally drop very slowly as the right tail moves past the criterion. That is to say, for any given list length, the serial position curve of response probability will be S-shaped.

Incremental Model

In this model we assume that presentation increases the strength of an item in both STM and LTM by some amount above its previous strength. Generalization in STM is also assumed, but it will not change the form of the final equation for $d(k, L)$. The assumptions of the model are:

$$s(x, 0) = \alpha_0 \qquad \ell(x, 0) = \lambda_0 .$$

$$s(x, k) = \sigma\alpha + \varphi s(x, k-1), \quad x \neq k \qquad \ell(x, k) = \ell(x, k-1), \quad x \neq k.$$

$$s(k, k) = \alpha + \varphi s(k, k-1) \qquad \ell(k, k) = \lambda + \ell(k, k-1),$$

so that

$$s(*, L) = \sigma\alpha\left(\frac{1-\varphi^L}{1-\varphi}\right) + \varphi^L\alpha_0 \qquad \ell(*, L) = \lambda_0$$

$$s(k, L) = \sigma\alpha\left(\frac{1-\varphi^{L-k}}{1-\varphi}\right) + \varphi^{L-k}s(k, k) \qquad \ell(k, L) = \lambda_0 + \lambda$$

$$s(k, L) = (1-\sigma)\,\alpha\varphi^{L-k} + \sigma\alpha\left(\frac{1-\varphi^L}{1-\varphi}\right) + \varphi^L\alpha_0 .$$

Thus,

$$d(k, L) = s(k, L) + \lambda_0 + \lambda - s(*, L) - \lambda_0$$

$$d(k, L) = (1-\sigma)\,\alpha\varphi^{L-k} + \lambda.$$

The result is independent of initial strength in both STM and LTM. Furthermore, so long as similarity is held constant and LTM is negligible ($\lambda \simeq 0$), the expression for $d(k, L)$ is experimentally indistinguishable from that obtained in the simple decay model. That is to say, generalization has no effect on the equation for d as a function of k and L. However, had we varied similarity, the incremental model would have predicted the coefficient of φ^{L-k} to vary. For the present experiment, though, the above derivation demonstrates that we may ignore any effects of *incremental* generalization.

Proportional Model

In this model we assume that strengths are bounded by 0 and α. Presentation of an item increments it to full strength and presentation of a similar item increments it some proportion of the distance between its present strength and full strength. The assumptions of the model are:

$$s(x, 0) = \alpha_0 \qquad \ell(x, 0) = \lambda_0 ,$$

$$s(x, k) = (1 - \sigma)\,\varphi s(x, k-1) + \sigma\alpha, \quad x \neq k \qquad \ell(x, k) = \ell(x, k-1), \quad x \neq k,$$

$$s(k, k) = \alpha \qquad \ell(k, k) = \lambda + \ell(k, k-1),$$

so that

$$s(*, L) = \sigma\alpha \left\{\frac{1 - [\varphi(1-\sigma)]^L}{1 - \varphi(1-\sigma)}\right\} + [\varphi(1-\sigma)]^L \alpha_0 \qquad \ell(*, L) = \lambda_0 ,$$

$$s(k, L) = \sigma\alpha \left\{\frac{1 - [\varphi(1-\sigma)]^{L-k}}{1 - \varphi(1-\sigma)}\right\} + [\varphi(1-\sigma)]^{L-k} \alpha \qquad \ell(k, L) = \lambda_0 + \lambda,$$

and, finally,

$$d(k, L) = s(k, L) + \ell(k, L) - s(*, L) - \ell(*, L) = A\theta^{L-k} + B\theta^L + \lambda,$$

where

$$\theta = \varphi(1 - \sigma),$$

$$A = \alpha\,\frac{(1-\sigma)(1-\varphi)}{1-\theta} \quad \text{and} \quad B = \frac{\sigma\alpha}{1-\theta} - \alpha_0 .$$

Both the incremental and proportional models reduce to the simple decay model when the parameters, α_0, σ, and λ are zero. Proportional generalization yields a different form of equation for $d(k, L)$ than the incremental assumptions by introducing a term of the form $B\theta^L$.

STM Acquisition-Primacy Mcdel

In addition to the assumptions of the incremental model we assume that presentation increments the strength in STM of the first item by α_1 and later items by α. The assumption that α is a function of k has no effect on the form of the final equation for $d(k, L)$. Hence, we obtain:

$$d(k, L) = (1 - \sigma)\, \alpha(k)\, \varphi^{L-k}.$$

As noted before, the incremental model is indistinguishable from the simple decay model in the present experiment so we may write the above equation in the simpler form:

$$d(k, L) = \alpha(k)\, \varphi^{L-k}.$$

STM Decay-Primacy Model

This model is designed to account for primacy by assuming that the first item decays at a slower rate in STM than do later items. With the assumptions of the incremental model,

$$d(k, L) = \alpha[\varphi(k)]^{L-k}.$$

Incremental LTM-Primacy Model

In this model we attempt to account for primacy by assuming that the first item is incremented by presentation to a greater strength in LTM than are subsequent items. With the assumptions of the incremental model,

$$d(k, L) = \alpha\varphi^{L-k} + \lambda(k).$$

STM-LTM Primacy Model

Although it involves the largest number of parameters (five), this model is the more general version of the incremental acquisition-primacy models. It assumes that the first item may acquire greater strength in both STM and LTM. Thus,

$$d(k, L) = \alpha(k)\, \varphi^{L-k} + \lambda(k).$$

Selection of Models to be Tested

Selection of models to be tested was guided by two considerations. First, models that make equivalent predictions of our data obviously cannot be distinguished from one another, so only the simpler model is considered to be tested. For example, we shall call the model whose final equation is $d(k, L) = \alpha\varphi^{L-k}$ the STM Decay model, model, but it is also the STM Incremental model. Second, we do not want to consider models which require too many parameters to be estimated. So, for example, we do not combine the proportional generalization assumption with any of the primacy assumptions, on the grounds that improvement in fit is extremely unlikely to compensate for the extra parameter, when so many parameters have already been estimated from the data. The models to be tested for goodness-of-fit and their equations for $d(k, L)$ are as follows:

	$d(k, L)$
STM Decay	$\alpha\varphi^{L-k}$
STM Proportional	$\alpha\varphi^{L-k} + \gamma\varphi^{L}$
STM Acquisition-Primacy	$\alpha(k)\,\varphi^{L-k}$
STM Decay-Primacy	$\alpha[\varphi(k)]^{L-k}$
LTM Acquisition-Primacy	$\alpha\varphi^{L-k} + \lambda(k)$
STM-LTM Primacy	$\alpha(k)\,\varphi^{L-k} + \lambda(k)$.

EXPERIMENT

Procedure

Our aim was to examine in detail the serial position effect in recognition memory by minimizing the effects of response interference and getting operating characteristics from each individual subject for each possible position of the test item in each list length tested. We tested our subjects individually and made only one test after each list presentation. We used the confidence-rating method to get operating characteristics because of its greater efficiency compared to other methods.

We started at the smallest list length of any interest (two items) and tested every list position in every length up to seven items, getting 75 observations at each point by each of four subjects. The probability of a new test-tem was about .33 at each list length. The items were three-digit numbers (with some restrictions to be described later) and the recognition test occurred immediately after the end of the list. A trial consisted of the auditory presentation of L digit-triples at the rate of one triple per second, followed by a tone about .25 second long, followed by the test-triple. The subject had 5 seconds in which to decide whether the test-triple had been in the list just presented, answering "yes" or "no" and stating his confidence on a scale from "1" (least confident) to "5" (most confident).

Design

The digit lists and new test-triples were made from pseudo-random numbers prepared on a digital computer. The lists were recorded on magnetic tape, and no session lasted more than one hour including one 5-10 minute break and two 20-second pauses. The digits were all pronounced at the rate of one three-digit number per second; 0 was pronounced "oh." The selection of digit-triples was governed by the following restrictions:

1. No two digits in a triple could be the same.

2. Ascending and descending sequences were not allowed;

a. $i - 1, i, i + 1$, or

b. $i + 1, i, i - 1$, where $i = 1, 2, \cdots 8, 9$ and $9 + 1 = 1 - 1 = 0$.

3. The alternating sequences were not allowed;

a. $i, i + 2, i + 4$, or

b. $i + 4, i + 2, i$, where $i = 1, 2, 3, 4, 5$.

4. The sequences 369, 963, 248, 842 were not allowed.

5. No triple could be repeated on a tape within 500 items of its previous presentation.

Lists of only one length were included in each session, so subjects always knew the list length. The serial position of the test item (including *) was randomized over trials. For the longer lists it was necessary to use two or three tapes for each condition, giving a total of 12 tapes. There were five practice lists at the beginning of each session and three practice lists after the break. The first 1-hour session was also practice. After the practice session each subject went through the 12 tapes three times, with a different random order for each subject and for each time through the 12 tapes.

Subjects

The subjects were four Harvard undergraduates, two male and two female. They were paid $1.50 per hour for their services.

RESULTS

Correct and False Recognition Rates

Correct recognition rates for items in each serial position, for the six different list lengths, and for each of the four subjects are presented in Table 1 and Fig. 3. False recognition rates for new items presented after lists of each length are also presented in Table 1 and Fig. 3. Each correct recognition point comes from 75 observations and each false recognition point from 75 to 264 observations, depending on the list length. By using the parameter estimates in Table 3, the lines are the best fitting predictions of the STM Acquisition-Primacy model.

TABLE 1

CORRECT AND FALSE RECOGNITION PROBABILITIES

			$p(\text{Yes} \mid k, L)$				
k	L	N[a]	S1	S2	S3	S4	Average
*	2	75	.000	.000	.000	.013	.003
1		75	1.0	1.0	1.0	1.0	1.0
2		75	1.0	1.0	1.0	1.0	1.0
*	3	114	.026	.000	.026	.000	.013
1		75	.907	.920	.933	1.0	.940
2		75	.920	.987	1.0	.987	.973
3		75	1.0	1.0	1.0	1.0	1.0
*	4	150	.080	.060	.087	.080	.077
1		75	.560	.987	.787	.880	.803
2		75	.587	.800	.667	.973	.757
3		75	.973	.920	.920	1.0	.953
4		75	1.0	1.0	.987	1.0	.997
*	5	189	.106	.042	.079	.101	.082
1		75	.640	.880	.573	.787	.720
2		75	.653	.627	.707	.867	.713
3		75	.773	.800	.787	.973	.833
4		75	.987	.987	.907	1.0	.970
5		75	1.0	1.0	1.0	1.0	1.0
*	6	225	.071	.044	.076	.093	.071
1		75	.307	.613	.387	.507	.453
2		75	.320	.507	.320	.600	.437
3		75	.533	.760	.680	.907	.720
4		75	.720	.813	.747	.973	.813
5		75	.960	.973	.987	.987	.977
6		75	1.0	1.0	.987	1.0	.997
*	7	264	.155	.091	.129	.170	.136
1		75	.347	.533	.373	.387	.410
2		75	.227	.387	.253	.467	.333
3		75	.427	.547	.400	.747	.530
4		75	.640	·640	.507	.867	.663
5		75	.907	.853	.800	1.0	.890
6		75	.961	.974	.949	1.0	.971
7		75	1.0	1.0	1·0	1.0	1.0

[a] The N is for each subject; the average is based on $4N$ trials.

The correct recognition results show a substantial recency effect: recognition increases rapidly with serial position. There is some primacy effect which is masked in this plot, as will be demonstrated in subsequent analyses.

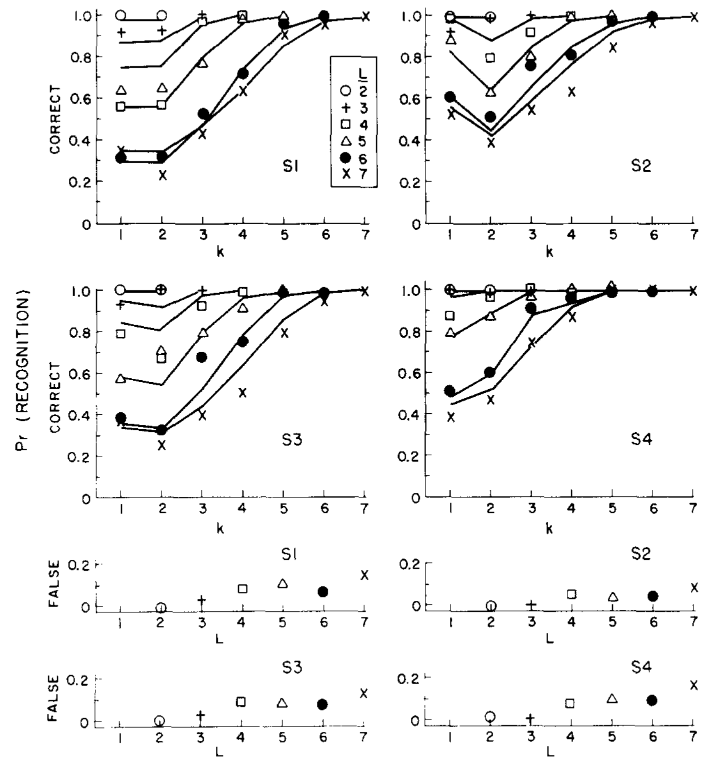

FIG. 3. Correct and false recognition rates for each subject as a function of serial position (k) with length of list (L) as the parameter. The lines are the theoretical predictions of the STM Acquisition-Primacy model.

The false recognition rate shows a general tendency to increase with list length, although a plateau shows up in each of the four subjects for list lengths four through six. A general increase in the false recognition rate can come about in two ways. First, if the criterion is fixed (independent of list length), false recognitions will increase if the mean of the new distribution increases with list length. This increase in the mean strength of the new distribution is predicted only by models that assume generalization. Second, the increase in false recognition can come about by a decrease in the criterion with list length, and this is compatible with any of the models.

Memory Operating Characteristics

The last item in each list and, for some subjects, the next to last item, were correctly recognized almost all of the time, so it was not possible to get reliable estimates of the operating characteristics for these conditions. Individual operating characteristics for conditions that did produce reliable estimates are shown in Figs. 4-8 on normal-normal plots. Points where either coordinate was below 0.02 or above 0.98 have not been included in these figures because the scale transformation makes the graphs extremely sensitive to slight variations in this region. Best-fitting straight lines were determined by eye for the points on each operating characteristic and values for *d* and slope were computed for each condition from the straight lines. The *d* values were estimated by entering the coordinates of the intersection of the straight lines with the negative diagonal in the published tables for d′(Elliott, 1964). This method of estimating *d* is often used in signal detection theory because it minimizes the effect of deviations

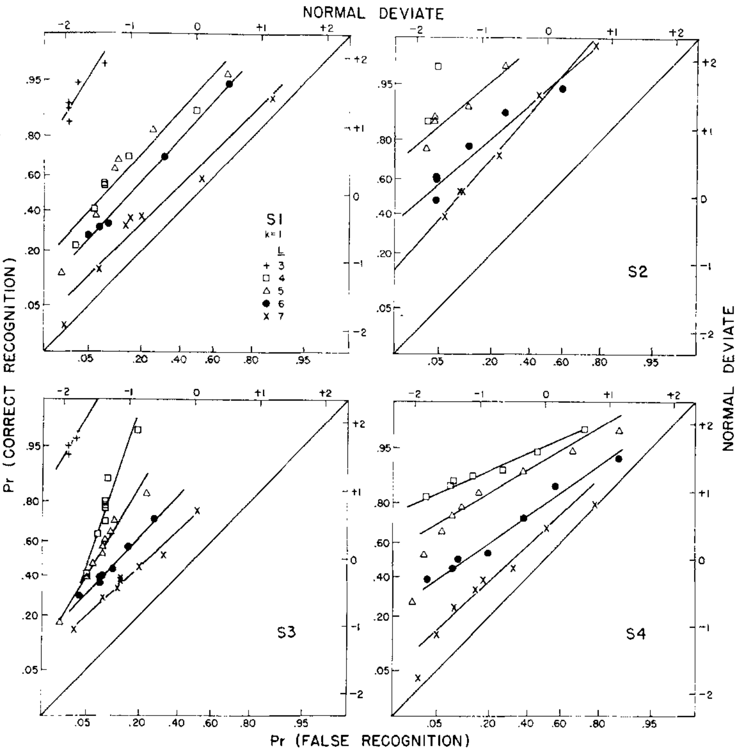

Fig. 4. Memory-operating characteristics for individual subjects for the first item in the list with length of list as the parameter. Points with a correct recognition rate greater than 0.98 or a false recognition rate less than 0.02 have not been included.

in the slope of the operating characteristics from unity. The unit of d, in this case, is the average of the standard deviations for the old and new distributions. We also computed the d value for each condition by using the point on the operating characteristic corresponding to the "yes-no" criterion. Both of these d values for each condition for each subject are shown in Table 2. Conditions where the probability of correct recognition was too high to produce reliable estimates of d are not included. The two ways of estimating d values give almost the same estimates. In all subsequent analyses we shall use the values estimated by the fitted lines because they are based on more data.

According to the criterion decision rule with normally distributed noise, the operating characteristics should be straight lines on normal-normal probability paper. The fit of straight lines to the data appears to be reasonably good. The slope of each line gives the ratio of the standard deviation of the new distribution to that of the old.

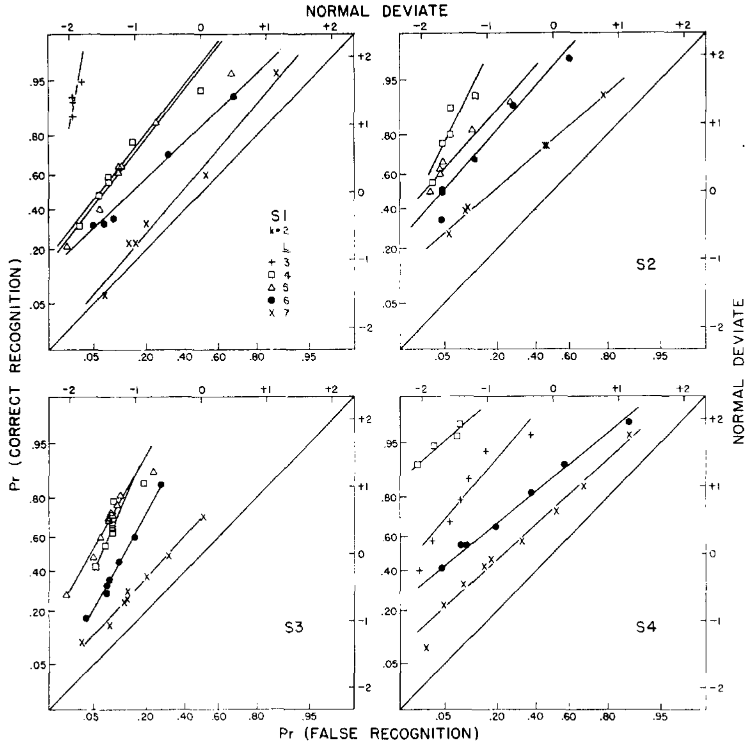

FIG. 5. Memory-operating characteristics for individual subjects for the second item in the list with length of list as the parameter. Points with a correct recognition rate greater than 0.98 or a false recognition rate less than 0.02 have not been included.

According to our theory, slope ought to be constant, independent of distance from the chance diagonal (*d* value), and always equal to unity. To test the hypothesis that slope is independent of *d* value, we computed Pearson product moment correlations between the slope and *d* value for each of the four subjects. The correlations were .16, .32, .31, and —.03, far from being significant in every case. To test the hypothesis that the slope is unity, we used both the binomial test and the *t* test. By the former, no difference was significant at the .05 level. By the latter, S3 had slopes significantly above unity at the .05 level, S4 had slopes significantly below unity at the .05 level, and the other two subjects demonstrated no significant deviation from the null hypothesis. No deviation reached significance at the .01 level, and the 99 % confidence limits for the true slope for each subject are: 1.12 ± .29, 1.16 ± .33, 1.41 ± .44, and .84 ± .17. Although some subjects may have slopes that tend to be above or below unity, the deviations are not great, not consistent over subjects, and not correlated with *d* value.

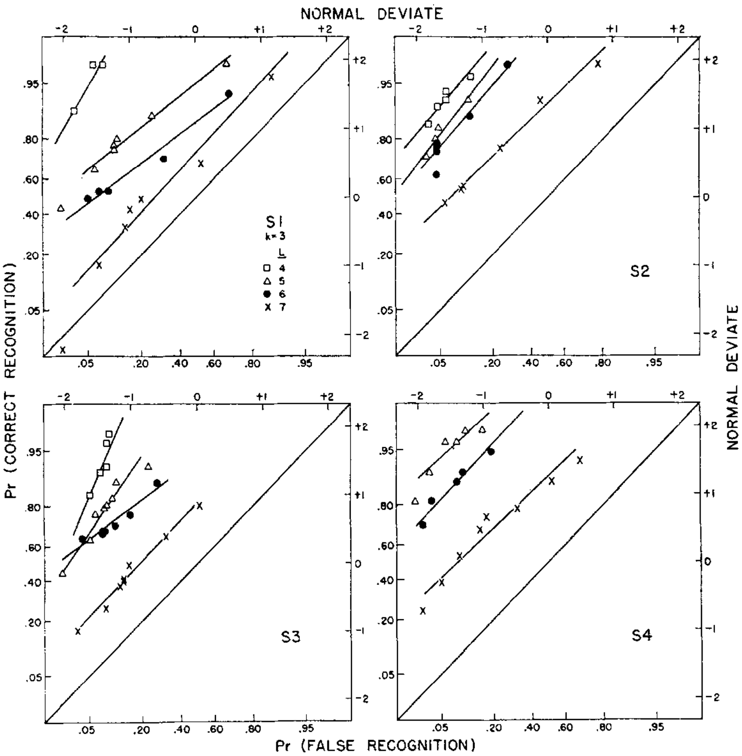

FIG. 6. Memory-operating characteristics for individual subjects for the third item in the list with length of list as the parameter. Points with a correct recognition rate greater than 0.98 or a false recognition rate less than 0.02 have not been included.

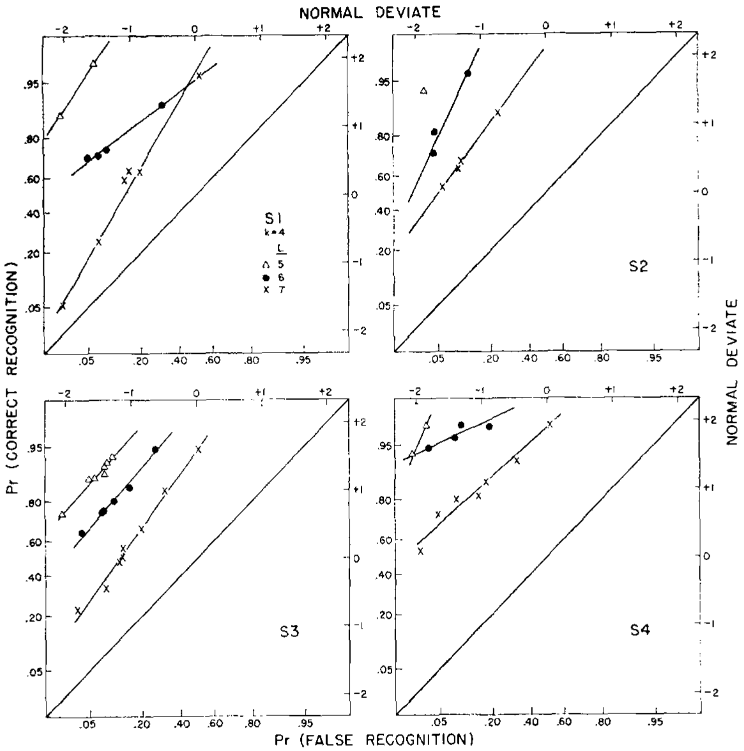

FIG. 7. Memory-operating characteristics for individual subjects for the fourth item in the list with length of list as the parameter. Points with a correct recognition rate greater than 0.98 or a false recognition rate less than 0.02 have not been included.

Primacy and Recency Effects

Primacy and recency effects on trace strength can be assessed by comparing d values of operating characteristics for items with different numbers of prior items $(k - 1)$ and different numbers of subsequent items $(L - k)$. According to the STM Acquisition-Primacy model,

$$d(k, L) = \alpha(k)\, \varphi^{L-k}.$$

Hence,

$$\log d(k, L) = \log \alpha(k) + (L - k) \log \varphi.$$

This makes $d(k, L)$ a linear function of the number of subsequent items $(L - k)$ on a semi-logarithmic plot. Furthermore, the slope is the same for all serial positions, but the intercept should be a function of serial position. By the one-step primacy assumption, all items after the first should have the same intercept.

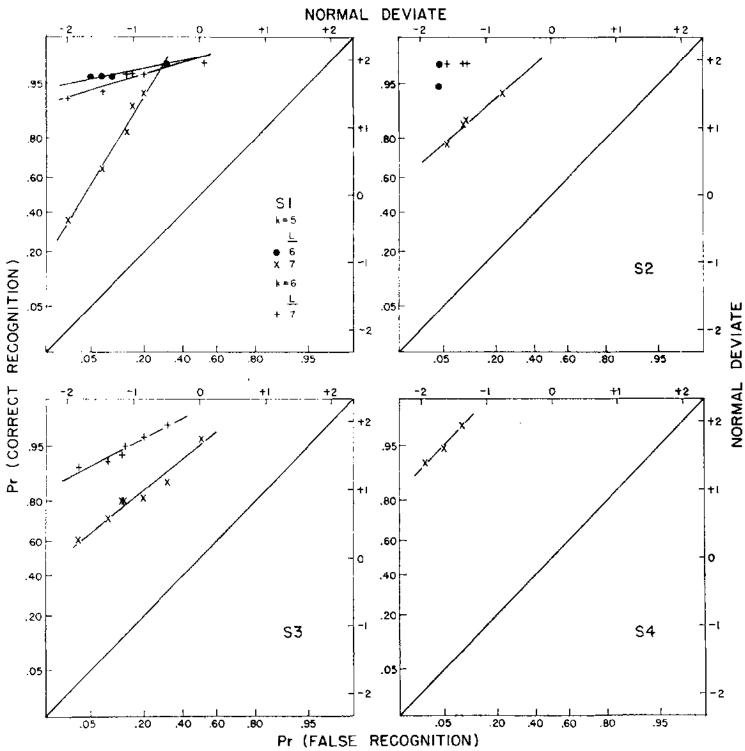

FIG. 8. Memory-operating characteristics for individual subjects for the fifth and sixth items in the list with length of list as the parameter. Points with a correct recognition rate greater than 0.98 or a false recognition rate less than 0.02 have not been included.

According to the STM Decay-Primacy model, it is the slope that is a function of serial position, rather than the intercept. Thus, the first item should be fit by a line with the same intercept as all other items, but the slope should differ.

Figure 9 presents semi-logarithmic plots of $d(k, L)$ against the number of subsequent items $(L - k)$ for each subject for all conditions in which reliable estimates of d were obtained. The two straight lines in each plot are the least-squares fits to the data. The upper line is the fit to the first items $(k = 1)$, and the lower line is the fit for all the other items $(k > 1)$.

The effects of proportional generalization, LTM, and primacy in LTM are not easily assessed by inspection of Fig. 9, and so consideration of models involving these effects will be deferred until the next section. However, it is possible to compare the STM Acquisition-Primacy model with the STM Decay-Primacy model by examining the dependence of the acquisition and decay parameters on list position. By using a least squares criterion, the acquisition and decay parameters are estimated by the

TABLE 2

$d(k, L)$ VALUES ESTIMATED FROM OPERATING CHARACTERISTICS (OC) AND FROM "YES-NO" PROBABILITY DATA (YN)

k	L	S1		S2		S3		S4	
		OC	YN	OC	YN	OC	YN	OC	YN
1	2	—	—	—	—	—	—	—	—
2		—	—	—	—	—	—	—	—
1	3	3.32	3.26	—	—	3.63	3.44	—	—
2		3.40	3.34	—	—	—	—	—	—
3		—	—	—	—	—	—	—	—
1	4	1.43	1.56	3.36	3.79	2.40	2.16	2.36	2.58
2		1.69	1.62	2.67	2.40	2.10	1.79	3.33	3.34
3		3.24	3.34	2.98	2.96	2.84	2.77	—	—
4		—	—	—	—	—	—	—	—
1	5	1.43	1.61	2.62	2.90	1.83	1.59	1.86	2.07
2		1.63	1.64	2.03	2.05	2.14	1.94	2.24	2.39
3		1.86	1.99	2.64	2.57	2.30	2.20	3.15	3.22
4		3.40	3.48	3.60	3.96	2.74	2.73	—	—
5		—	—	—	—	—	—	—	—
1	6	1.05	.96	1.72	1.99	1.16	1.14	1.01	1.33
2		1.03	1.00	1.76	1.72	1.49	.97	1.25	1.57
3		1.25	1.55	2.44	2.41	1.80	1.90	2.56	2.64
4		1.94	2.05	2.75	2.60	2.14	2.09	3.25	3.26
5		3.40	3.22	3.50	3.65	3.89	3.67	—	3.55
6		—	—	—	—	—	—	—	—
1	7	.38	.62	1.44	1.42	.66	.80	.48	.66
2		.33	.26	.82	1.05	.54	.46	.72	.87
3		.64	.83	1.43	1.46	.88	.87	1.32	1.62
4		1.35	1.37	1.86	1.70	1.35	1.14	2.06	2.06
5		2.16	2.33	2.34	2.39	1.86	1.97	3.29	—
6		3.10	2.78	3.58	3.30	2.92	2.76	—	—
7		—	—	—	—	—	—	—	—

intercepts and slopes of the best fitting straight lines to the data of Fig. 9 for each value of k. The resulting parameter estimates for each subject are shown in Fig. 10. The results are clear. The slope shows no tendency to change with k, but the intercept decreases with k, in accord with the STM Acquisition-Primacy model.

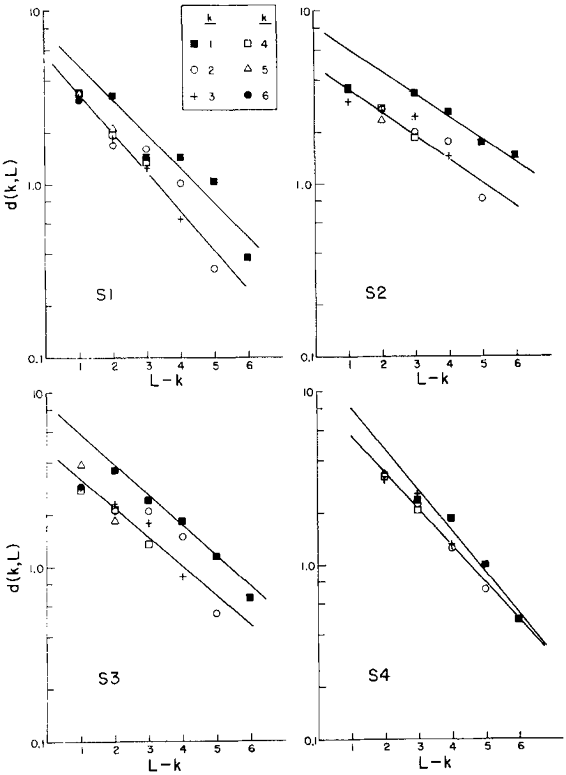

FIG. 9. The $d(k, L)$ values for each serial position (k) as a function of the number of subsequent items (L-k). The two straight lines in each plot are the least-squares fits to the data. The upper line is the fit to the first items ($k = 1$) and the lower line is the fit for all the other items ($k > 1$).

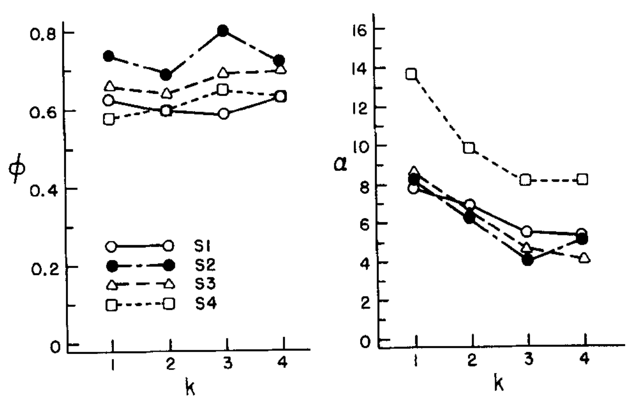

FIG. 10. Least-squares estimates for φ and α (slopes and intercepts) as a function of serial position (k) for each subject. There are not enough points in our data for $k = 5$, 6, and 7 to determine reliable estimates of φ and α for these conditions.

Several features of the data should be noted. First, for all subjects, trace strength as a function of the number of subsequent items is quite well described by an exponential decay function. This is true for the first as well as later items, and the rate of decay appears to be the same for all items. Moreover, the rate of decay appears to be approximately the same for each of the four subjects. Although there are some deviations from straight lines on semi-logarithmic plots, the deviations do not suggest any reasonable alternative function. Second, for all subjects the initial item in a list has greater trace strength than later items, when the number of subsequent items is the same. Third, for all subjects the differences among other serial positions are small when the number of subsequent items is controlled. The results are in striking agreement with the STM Acquisition-Primacy model, although the STM Decay-Primacy model provides a reasonable first approximation to the data. For S4 the fit of the STM Decay model is extremely good without a primacy assumption.

Estimation of Parameters

To assess the goodness-of-fit of each of our models and to compare the fit of different models, we must have a fair method of estimating parameters. Our models do not predict behavior directly, but rather the underlying memory strengths and, through them, the probability of responses. In the present instance, it seemed difficult to use standard maximum likelihood techniques for parameter estimation. Instead, we estimated parameters for each subject by minimizing the chi-square value of the fit of each model to the "yes-no" probability data, using the six empirical false recognition probabilities to estimate the retrieval criterion for "yes" responses.

Because reasonable parameter estimates yield expected frequencies less than one in a number of instances, we could not use a standard chi-square goodness of fit test. Instead, we determined the probability of obtaining each observed frequency of correct recognition (or a more discrepant frequency) by using the binomial test. We found the exact probability using the binomial distribution, not the probability obtained from the normal approximation to the binomial. The probability of getting a frequency as extreme as the observed frequency in a particular condition (for a particular model with a particular set of parameter estimates) was transformed to χ^2, using the *exact* transformation proposed by Fisher (1938) for combining independent tests of significance. The transformation is, $\chi^2 = -2 \log_e p$, where χ^2 is on 2 *df*. These χ^2 were then summed for all 27 conditions to yield an overall test of the goodness-of-fit of a particular model with a particular set of parameter estimates. The number of degrees of freedom is found by taking the number of independent conditions, subtracting the number of parameters estimated from the data and then doubling this difference. Thus, with 27 observed frequencies of correct recognition and three parameter estimates, there are $2(27 - 3) = 48$ *df*.

We found the minimum chis-quare values by systematically varying parameters and testing goodness of fit at many points in the n-space determined by the parameters.[3] If done in one step, this gridding technique is very inefficient because a fine grid must be used to get precise estimates of parameters and hundreds of thousands of chi-squares calculated. A more efficient procedure is to examine a rough grid of the parameter space and the examine the (few) areas of interest in more detail, repeating this procedure as many times as is necessary to get the desired precision of the estimates. This hierarchical procedure is usually inefficient because each analysis depends upon the results of the previous analysis, and on the usual university computer facilities the computations might take months. Fortunately, we were able to use a remote console (located at Harvard) of the Compatible Time Sharing System (CTSS) of the Massachusetts Institute of Technology (Project MAC), so that the hierarchical gridding procedure was fast and efficient. We continued our analysis until we were sure that we were not examining a local minimum and we had reached a precision of slightly less than two significant figures.

Goodness-of-Fit of the Models

The parameter estimates and minimum χ^2 values for each of the models tested are shown in Table 3. The simple STM Decay model appears to be inadequate without some primacy assumption for at least three of the four subjects. The Proportional model fits much better than the Decay model, but is consistently less adequate than the STM Acquisition-Primacy model. The Proportional model gives some primacy effect, especially for low values of $L - k$, so it is reasonable to guess that this is responsible for the improvement in fit of the Proportional model over the Decay model. Since the more straightforward primacy models provide an even better fit to the data, it is reasonable to prefer them to the Proporional model. The LTM parameters appear to be completely unnecessary.

The fit of the STM Acquisition-Primacy model is uniformly better than the fit of the STM Decay-Primacy model for each of the four subjects. In addition, there are three other considerations that favor the STM Acquisition-Primacy model. First, the least squares results summarized in Fig. 10 lead us to conclude that primacy affects acquisition, rather than decay. Second, it is intuitively more plausible to assume that subjects spend more time or effort learning the first items compared to the later items than it is to assume that the decay process proceeds at a different rate for the first items than for later items. Third, the STM Acquisition-Primacy model yields estimates for the rate of decay in STM that are almost invariant over the four sub-

[3] We would like to thank Richard C. Atkinson and Edward J. Crothers for sending us a copy of their computer program for parameter estimation by the minimum chi-square method and discussing the estimation problem with us. We did not use their iterative hill-climbing program to obtain our estimates, however, but instead used the hierarchical grid procedure described in this section.

jects, while the STM Decay-Primacy model does not yield estimates for the initial learning that are comparably invariant over the four subjects.

TABLE 3

PARAMETER ESTIMATES AND MINIMUM CHI-SQUARE VALUES

Model	S	Parameters: STM			LTM		χ^2	*df*	*p*
STM Decay		α		φ					
$d(k, L) = \alpha\varphi^{L-k}$	1	(6.0)		(.66)			>300	50	.0001
	2	(5.0)		(.80)			>300	50	.0001
	3	5.0		.70			224	50	.0001
	4	8.5		.66			57	50	.23
STM Proportional		α	γ	φ					
$d(k, L) = \alpha\varphi^{L-k} + \gamma\varphi^{L}$	1	4.0	3.0	.67			87	48	.001
	2	2.7	5.5	.75			132	48	.0001
	3	3.3	5.5	.67			95	48	.0001
	4	6.0	6.0	.66			25	48	.99
STM Acquisition-Primacy		α	α_1	φ					
$d(k, L) = \alpha(k)\,\varphi^{L-k}$	1	5.0	7.5	.65			49	48	.44
	2	4.9	8.5	.74			71	48	.015
	3	4.6	7.5	.68			86	48	.001
	4	7.5	10.0	.67			20	48	.99
STM Decay-Primacy		α	φ_1	φ					
$d(k, L) = \alpha[\varphi(k)]^{L-k}$	1	5.5	.71	.62			69	48	.02
	2	5.3	.83	.72			87	48	.001
	3	5.0	.76	.65			122	48	.0001
	4	8.2	.70	.65			33	48	.95
LTM Acquisition-Primacy		α		φ	λ	λ_1			
$d(k, L) = \alpha\varphi^{L-k} + \lambda(k)$	1	5.3		.62	.1	.6	74	46	.01
	2	5.2		.72	.0	1.1	83	46	.001
	3	4.8		.65	.1	.8	124	46	.0001
	4	7.8		.66	.1	.4	39	46	.76
STM-LTM Primacy		α	α_1	φ	λ	λ_1			
$d(k, L) =$	1	5.0	7.0	.65	.0	.1	49	44	.28
$\alpha(k)\,\varphi^{L-k} + \lambda(k)$	2	4.9	8.5	.74	.0	.0	71	44	.01
	3	4.6	7.5	.68	.0	.0	86	44	.0001
	4	7.5	10.0	.67	.0	.0	20	44	.99

The fit for the LTM Acquisition-Primacy model is also uniformly poorer than the fit of the STM Acquisition-Primacy model for each of the four subjects. Again, the parameter estimates for the combined STM-LTM Primacy model confirm the choice of the STM Acquisition-Primacy model over the LTM Acquisition-Primacy model. The best estimates for the STM-LTM Primacy model indicate a large difference in the acquisition parameter for STM and effectively no difference in the acquisition parameter for LTM. In fact, LTM appears to be playing a negligibly small role in our experiment, so it is quite reasonable to assign the largest part of the primacy effect to STM.

DISCUSSION

The principal finding of the present study is that STM trace strength decays exponentially with the number of subsequently presented items when the criterion decision rule is used to transform strengths into probabilities. This conclusion is based on the satisfactory fit of the STM Acquisition-Primacy model to the data. In our view, this provides strong support for both the criterion rule and the exponential decay assumption. This conclusion is based as much on a criterion of simplicity as on a criterion of adequacy, because if we applied an exponential transformation to strengths, the ratio decision rule would apply equally well to our data.

In the present study the decay in recognition probability is not exponential. We find an exponential decay only when we transform probability measures into strength measures using the criterion rule. Other studies of recognition memory have also shown that the probability of recognition departs from an exponential decay function (see Shepard, 1961, p. 186, and Shepard and Teghtsoonian, 1961).

The fact that the trace strengths estimated with the criterion decision rule are best described by a simple exponential decay function provides strong support for the validity of the criterion rule and the choice of strength over probability as the variable to which the laws of forgetting apply.

The simple exponential decay model accounts for a good deal of our data. However, of the models we studied, the best was the STM Acquisition-Primacy model. It seems that the first item in the list is remembered better than subsequent items because it reaches a higher initial strength in STM than other items. In order to account for these results it does not seem necessary to assume that any items acquire substantial strength in LTM or that strength in STM decays at a different rate for items in different serial positions.

In other experiments with other types of material, multiple presentation, slower rates of presentation, or greater opportunity for rehearsal it will probably by necessary for models to include an LTM component in the total memory trace, even when there is a very brief retention interval. Such a combination of short- and long-term traces was found necessary by Waugh and Norman (1965) in order to account for the forgetting curves in a variety of recall experiments. The almost complete absence of LTM effects in the present experiment can probably be obtained only under controlled conditions. In our view, the most important experimental conditions leading to this result are control over rate of presentation, the nature of the materials, and the rehearsal instructions to the subject. Large LTM effects, especially for the first item(s), can probably be obtained rather easily if subjects are permitted to rehearse in any way they please, if the items are simple and highly differentiated in their associations, and if the rate of presentation is not extremely fast. In the present experiment we repeatedly instructed our subjects to rehearse only the currently presented item. Furthermore, we used a fast presentation rate and relatively complex and undifferentiated items.

If the assumption that STM and LTM traces are combined additively is true, one need not be concerned about the direct effect of the initial strength of LTM traces, provided they are approximately equal for all items in the population being used. As demonstrated in the theoretical section, initial strength in LTM (λ_0) does not occur in the final equation for $d(k, L)$, in every model. Strong initial LTM associations do not interfere with immediate memory.

Finally, proportional generalization does not account for the primacy effect as well as the more straightforward assumption that the first item(s) are learned better. This does not mean that there was no generalization in our exepriment. Evidence for or against generalization should come from an experiment in which the similarity of items is varied in a systematic way.

CONCLUSION

The theoretical assumptions that receive strongest support are: (a) strength in STM decays exponentially with the number of subsequently presented items and (b), subjects respond that they recognize an item, if and only if its strength in memory exceeds a criterion. Primacy in our experiment was best explained by assuming that the first item(s) received a greater increment in STM when they were presented than did later items, with no differences in their LTM values or in the rate of their decay in STM.

REFERENCES

ADAMS, E. W., AND MESSICK, S. An axiomization of Thurstone's successive intervals and paired comparison scaling models. *Technical Report*, No. 12, Contr. Nonr. 225 (17), Appl. Math. and Statist. Lab., Stanford Univer., 1957.

BURKE, C. J., AND ZINNES, J. L. A paired comparison of pair comparisons. *Journal of Mathematical Psychology*, 1965, **2**, 53-76.

BUSH, R. R. Estimation and evaluation. In R. D. Luce, R. R. Bush, and E. Galanter (Eds.), *Handbook of mathematical psychology*, Vol. 1. New York: Wiley, 1963. Pp. 429-469.

EGAN, J. P. Recognition memory and the operating characteristic. Indiana Univer., Hearing and Communication Lab., AFCRC-TN-58-51, AD-152650, 1958.

EGAN, J. P., SCHULMAN, A. I., AND GREENBERG, G. Z. Operating characteristics determined by binary decisions and by ratings. *Journal of the Acoustical Society of America*, 1959, **31**, 768-773.

ELLIOTT, P. B. Tables of d′. In J. A. Swets (Ed.), *Signal detection and recognition by human observers*. New York: Wiley, 1964. Pp. 651-684.

FISHER, R. A. *Statistical methods for research workers.* (7th ed.) Edinburgh: Oliver & Boyd, 1938. Pp. 103-106.

LUCE, R. D. *Individual choice behavior*· New York: Wiley, 1959.

LUCE, R. D. Detection and recognition. In R. D. Luce, R. R. Bush, and E. Galanter (Eds.), *Handbook of mathematical psychology*, Vol. 1. New York: Wiley, 1963. Pp. 103-189.

SHEPARD, R. N. Application of a trace model to the retention of information in a recognition task. *Psychometrika*, 1961. **26**, 185-203.

SHEPARD, R. N., AND TEGHTSOONIAN, MARTHA. Retention of information under conditions approaching a steady state. *Journal of Experimental Psychology*, 1961, **62**, 302-309.

SWETS, J. A., TANNER, W. P., JR., AND BIRDSALL, T. G. Decision processes in perception. *Psychological Review*, 1961, **68**, 301-340.

THURSTONE, L. L. A law of comparative judgment. *Psychological Review*, 1927, **34**, 273-286.

TORGERSON, W. S. *Theory and methods of scaling*. New York: Wiley, 1958.

WAUGH, N. C., AND NORMAN, D. A. Primary memory. *Psychological Review*, 1965, **72**, 89-104.

RECEIVED June 21, 1965