Psychological Review
1967, Vol. 74, No. 6, 462–480

DECISION PROCESSES IN MEMORY

HARLEY A. BERNBACH

Cornell University

Several current theories of the response process in recognition memory, or in confidence rating in recall, have postulated a continuous decision process as in the theory of signal detectability. This assumption, however, does not rule out assuming that the process by which information is stored and retained in memory is a discrete one. In the finite-state decision theory presented here, the continuous decision process is combined with a representation of the memory trace as a tag that is either available or not available at the time of test. In its predictions for recognition memory experiments, this theory is at least the equal of theories based on traditional habit strength assumptions. Furthermore, data from studies of confidence ratings in recall provide considerable support for the finite-state assumption.

The recent growth of interest in the study of memory by recognition methods has led to a great deal of theoretical attention to the decision aspect of retrieval from memory. The subject (*S*) in a recognition memory experiment is faced with a decision problem very similar to that facing the *S* performing a psychophysical detection task. For example, in an experiment in which *S* must identify each of a series of stimuli as old or new, for example, Shepard and Teghtsoonian (1961), he can vary his proportion of correct identifications of old items by a criterion shift. That is, he can permit his proportion of false alarms (calling new stimuli old) to change in accordance with his perception of the relative cost of the two types of errors, false alarms and missed recognitions.

It was perhaps inevitable that theoretical developments in detection theory should be applied to the study of recognition memory, considering this similarity between these two tasks. With the advent of the theory of signal detectability (Peterson, Birdsall, & Fox, 1954; Tanner & Swets, 1954) and recent finite-state detection models (Atkinson, 1963; Luce, 1963), it has generally been agreed that it is appropriate in the study of signal detection to consider separately the sensory process(es), or input of information about the stimulus to the observer, and the decision process leading to the overt response. Similarly, it seems reasonable in the study of recognition memory to consider separately the process by which information is stored and retained in memory and the process by which a decision is made to respond "old" or "new."

A distinction may be made between two types of theories of the memory process, based upon whether their representation of the memory trace is continuous or discrete. Hull (1943), for example, assumed that an item in memory could be characterized by a continuous measure of habit strength. Some recent theories (Atkinson & Shiffrin, 1965; Bower, 1965) have adopted finite-state representations of the memory trace, with an item assumed either to be in the memory store (or one of several such stores) or to not be in memory at all. In similar fashion, a distinction may be made between theories about the decision process. In psychophysics, for example, discrete-state detection systems have been proposed in threshold theories such as those of Atkinson (1963) and Luce (1963). On the other hand, it is assumed in the theory of signal detectability that the response is based upon a continuous measure, the likelihood ratio.

In recognition memory, threshold assumptions have been quite common in theories that acknowledge the role of the response or decision process. Such assumptions underlie the standard procedure for correcting recognition scores for guessing (Woodworth & Schlosberg, 1954, p. 700), and a reaction threshold was an integral part of Hull's (1943) theory. Mathematical models combining finite-state memory representations and threshold decision processes were developed by Bernbach (1964, 1965) and by Kintsch (1966). However, since threshold models have been reviewed extensively elsewhere (e.g., Green & Swets, 1966), they will not be considered in this paper.

Egan (1958) used the operating characteristic method to question the appropriateness of threshold assumptions in recognition memory, and suggested applying the decision process of the theory of signal detectability to the recognition memory task. Since that time a number of writers have applied the theory of signal detectability to recognition memory (for a review see Green & Swets, 1966). Most of these papers have shared the characteristic that they have made no specific assumptions about the nature of the memory trace, simply assuming the existence of hypothesis distributions for new and old items. In fact, they have typically taken evidence for the continuous nature of these hypothesis distributions as evidence of a continuous representation for the memory trace. Norman and Wickelgren (1965) and Wickelgren and Norman (1966) have proposed a series of models combining the decision process of the theory of signal detectability with a specific continuous strength memory theory, and these models will be considered in detail here.

In this paper, a theory is presented that utilizes the decision process of the theory of signal detectability but assumes a finite-state representation of the memory trace. The predictions of this finite-state decision theory are compared with those of continuous strength theory by considering data from both recognition and recall studies.

It should be noted that this paper deals only with the process of retrieval from memory, and not with the storage or retention of information. For example, the only assumption about memory in the finite-state decision theory is that a trace may be represented in an all-or-none fashion, that is, as a stimulus tag that is either available or not available at the time of test. With this assumption, any of a large number of possible theories about the mechanisms of storage and forgetting could be combined with this retrieval theory to more completely model the processes involved in the memory task.

A FINITE-STATE DECISION THEORY

In its application to the detection of signals by human observers, the theory of signal detectability may be thought of as having two distinct parts. The first concerns transforming the characteristics of the signal and of the noise into a pair of likelihood distributions, $f_n(y)$ and $f_s(y)$. These are the probabilities, given the presentation of noise alone or signal plus noise, of an observation y on a unidimensional decision scale upon which the input is mapped. It is this part of the theory that was developed by Peterson, Birdsall, and Fox (1954) for the case of the ideal receiver. The second part of the theory consists of the application of statistical decision theory to determine an optimum response strategy, which is to respond "signal" whenever the likelihood ratio $f_s(y)/f_n(y)$ is greater than a criterion value determined by a priori probabilities and payoffs.

In order to apply the theory of signal detectability to recognition memory, we must construct a theoretical mechanism leading from the characteristics of the experiment and of the stimuli to two likelihood distributions, $f_n(x)$ and $f_o(x)$, that are the probabilities, given that the item presented is new or old, of some observation x on a unidimensional decision scale. Then, the decision feature of the theory of signal detectability can be applied directly to the task of responding "old" or "new." Such an application is the basis for the finite-state decision theory presented in this paper. The major part of the theory is concerned with the development of the likelihood distributions, $f_n(x)$ and $f_o(x)$; the decision process is the same as that proposed by the theory of signal detectability.

First, it is necessary to specify the assumed characteristics of the memory trace. We assume that the presentation of a stimulus item results in the storage of a stimulus tag containing identifying information about that stimulus. When such a tag already exists in memory at the time an item is presented for test, the item is said to be in State R. After the response is made, a stimulus tag is stored for later reference, so that every item is in State R immediately following its presentation. State N is defined as the condition in which some of the information needed to tell an item apart from the other stimuli is not available in memory at the time of test. Every item is assumed to be in one of these two states; therefore a new (not yet presented) item, as well as an old one that has been forgotten, will be in State N. The preceding statements are summarized by the conditional probabilities of an item being in each state at the time of test, as follows:

$$\begin{aligned} p(N|new) &= 1, \\ p(R|new) &= 0, \\ p(N|old) &= \delta, \\ p(R|old) &= 1-\delta, \end{aligned} \qquad [1]$$

where δ represents the probability that an item is forgotten, returning to State N, before test. While specifying a mechanism of forgetting is beyond the scope of this paper, δ will be, for example, an increasing function of the number of other stimuli intervening between an item's presentation and subsequent test.

The assumed mechanism for the response process is similar in many respects to that underlying the application of the theory of signal detectability to a *matching* task (Sorkin, 1962), in which *S*s must identify the members of a pair of stimuli as "same" or "different." In that case, Sorkin assumed that the decision is based upon an observation determined by some measure of the difference between the two stimuli. Thus, when this difference exceeds a specified criterion value, S responds "different"; otherwise the response "same" is made. In the recognition memory task, we assume that the item presented for test is compared with the available stimulus tags in memory, and that the observation is determined by the minimum difference found between the test stimulus and any tag. For convenience, we define a scale of *apparent oldness* such that a small minimum difference yields a high degree of apparent oldness, and very little apparent oldness is produced when the smallest difference obtainable from matching is relatively large.

If an item is in State N when presented for test, its apparent oldness is assumed to depend on its relation to the previously presented stimuli that make up the contextual background. Distinctive items, for example, having little in common with the background, will have little apparent oldness. We assume that the characteristics of the set of stimuli used in an experiment determine a likelihood function, $f_n(x)$, which is the probability of an observation x on the apparent oldness scale given that the item presented is in State N. Similarly, imperfection or noise in the matching process is assumed to give rise to a likelihood distribution, $f_r(x)$, that gives the probability of an observation x on the apparent oldness scale given that the item presented is in State R. From this likelihood function and $f_n(x)$, the distribution $f_o(x)$ giving the probability of an observation x for an old item is determined as follows:

$$\begin{aligned} f_o(x) &= p(N|old)f_n(x) \\ &\qquad + p(R|old)f_r(x) \qquad [2] \\ &= \delta f_n(x) + (1-\delta)f_r(x). \end{aligned}$$

The likelihood distribution for new items is simply $f_n(x)$, since we assume that all new items are in State N.

The decision to respond "old" or "new" is assumed to be made by comparing the observation obtained with a cut-off x_c on the apparent oldness scale, so that an observation $x \geq x_c$ gives rise to the response "old," and the response "new" is made otherwise. From the theory of signal detectability, the criterion value β of the likelihood ratio $f_o(x)/f_n(x)$ that maximizes the expected value of the response is given by:

$$\beta = \frac{p(new)}{p(old)} \cdot V, \qquad [3]$$

where V is the ratio of the relative importance of correct responses to new and old items, in terms of costs and values (Swets, Tanner, & Birdsall, 1961). If the distributions $f_n(x)$ and $f_o(x)$ are known, a value of the cut-off x_c corresponding to β is determined. Since the distribution $f_o(x)$ depends on δ, the location of the cut-off is dependent on the forgetting probability as well as on the odds, costs, and values of Equation 3.

The Operating Characteristic

In order to display graphically the effect of criterion changes and signal level on detection, it is common to plot the *operating characteristic,* that is, the function relating correct detections and false alarms. A point on the operating characteristic corresponds to a particular value of the cut-off, x_c. In recognition memory, the correct recognition rate, denoted $p(y|old)$, is given by:

$$p(y|old) = \int_{x_c}^{\infty} f_o(x)dx. \qquad [4]$$

Similarly, the false alarm rate is given by

$$p(y|new) = \int_{x_c}^{\infty} f_n(x)dx. \qquad [5]$$

Since $f_o(x)$ is a function of δ (from Equation 2), the operating characteristic depends on both the value of δ and the nature of the likelihood distributions $f_n(x)$ and $f_r(x)$. When there is no forgetting ($\delta = 0$), the operating characteristic, in this case denoted the *natural operating characteristic,* reflects these likelihood distributions only. Consider, for example, the three sets of likelihood functions shown in Figure 1. In Figure 1a, there is no overlap between $f_n(x)$ and $f_r(x)$;

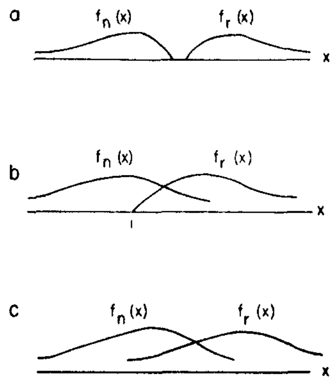

FIG. 1. Hypothetical likelihood distributions for items in States N and R.

therefore, for cut-off locations yielding all possible values of $p(y|new)$, the value of $p(y|old)$ will remain constant at one. The natural operating characteristic for this case is shown in Figure 2a.

In Figure 1b, there is no overlap of $f_n(x)$ and $f_r(x)$ to the left of the point labeled i. Therefore, as the cut-off is moved from the extreme left (where both $p(y|new)$ and $p(y|old)$ equal one) to Point i, the natural operating characteristic follows the horizontal line of that in Figure 2a. To the right of Point i, however, a decrease in $p(y|new)$ is accompanied by a corresponding decrease in $p(y|old)$, until both $p(y|new)$ and $p(y|old)$ equal zero as the cut-off reaches the far right. If the likelihood functions are continuous as in Figure 1c, the natural operating characteristic will be continuously curvilinear. Examples of

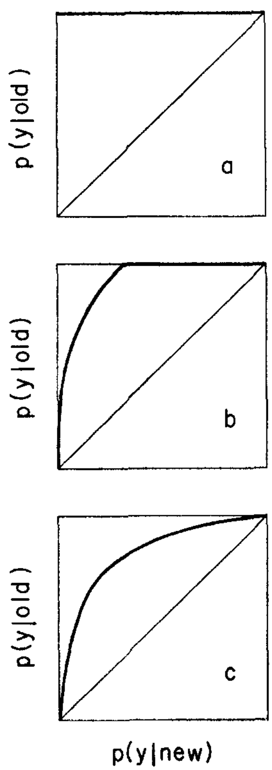

FIG. 2. Natural operating characteristics for the distributions of Figure 1.

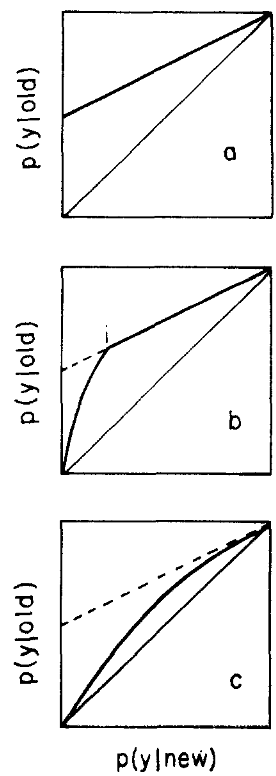

FIG. 3. Operating characteristics for the distributions of Figure 1, with $\delta = \frac{1}{2}$.

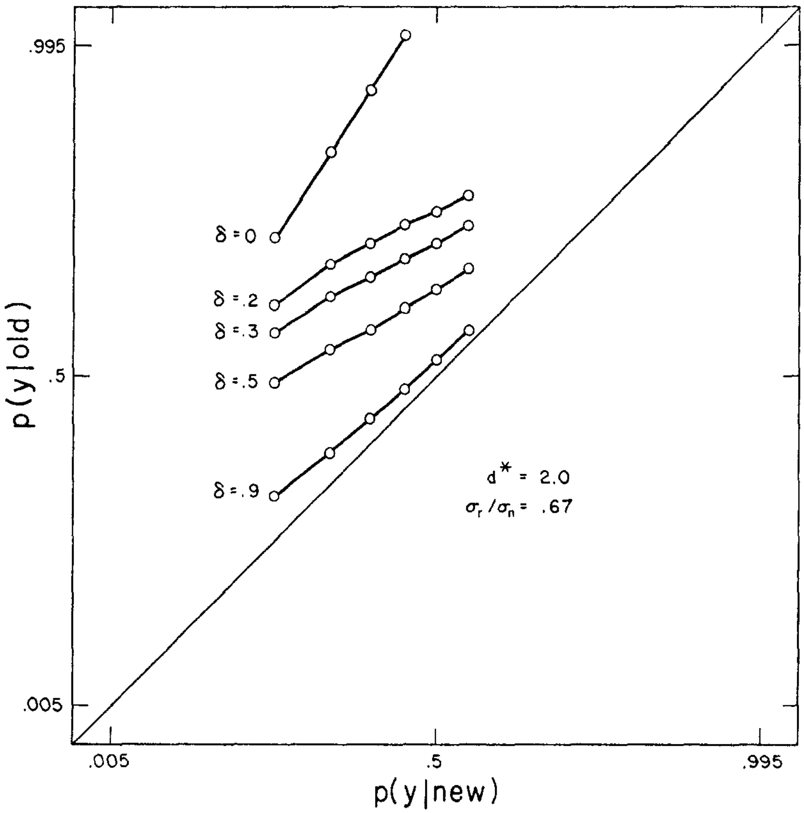

FIG. 4. Normal-deviate plot of operating characteristics assuming normal likelihood distributions for States N and R.

such functions are given in Figures 2b and 2c.

If there is complete forgetting ($\delta = 1$), $f_o(x)$ will equal $f_n(x)$, from Equation 2, and the operating characteristic will simply be the major diagonal, reflecting chance performance. Thus, the operating characteristic for other values of δ will be a weighted average of the natural operating characteristic and the major diagonal, with δ the weight assigned the diagonal. Operating characteristics for the likelihood functions of Figure 1 and $\delta = \frac{1}{2}$ are shown in Figure 3. It is interesting to note that the operating characteristic in Figure 3a for nonoverlapping likelihood functions is identical with that predicted by a high threshold theory (Green & Swets, 1966), while the function in Figure 3b is very similar to that predicted by Luce's (1963) low threshold theory.

If the distributions $f_n(x)$ and $f_r(x)$ are normal, the natural operating characteristic plotted on normal probability coordinates will be a straight line whose slope is the ratio of the standard deviations, σ_n/σ_r. Such a function is shown as the $\delta = 0$ line in Figure 4, with arbitrarily chosen parameter values, $d^* = 2.0$ and slope $= 1.5$. The effect of changing the value of δ is also shown in this figure. Since fairly linear operating characteristics are often obtained in memory experiments, it is important to note that the operating characteristics for the mid-range values of δ do not depart very much from straight lines.

The Measure of Sensitivity

In the theory of signal detectability, the sensitivity, d', is defined as the difference between the means of $f_n(y)$ and $f_s(y)$, expressed in units of the standard deviation of $f_n(y)$. The measure d' reflects the effect of characteristics of the stimulus (e.g., signal-to-noise ratio) and is independent of decision factors such as signal probability and payoffs (Swets, Tanner, & Birdsall, 1961). In recognition memory, the separation of the likelihood distributions $f_o(x)$ and $f_n(x)$ depends not only on the characteristics of the stimuli, but on factors influencing memory as well. Thus, d' depends both on δ, the probability of forgetting, and on d^*, the difference between the means of $f_n(x)$ and $f_r(x)$ in units of the standard deviation of $f_n(x)$. The parameter d^* is a measure of how much better a match is obtained when a stimulus tag is available in memory than when one is not. As such, it may be affected by certain stimulus characteristics, such as dimensionality and meaningfulness, but should be unaffected by memory factors such as retention interval.

Unfortunately, unless there is some independent means to estimate δ, a value of d^* cannot be obtained from recognition memory data. Furthermore, there is no straightforward way to estimate an accurate value of d' in this situation, as $f_o(x)$ is not a normal distribution (even if $f_r(x)$ is normal) and existing estimating procedures require such an assumption. If $f_o(x)$ is close to normal, however, as would be indicated by near linearity of the operating characteristics, such as shown in Figure 4, the measure of d' based on the normal assumption should not be far off.

It is a simple matter to express d' analytically in terms of δ and d^*. Taking expectations in Equation 2, we have

$$E[f_o(x)] = \delta E[f_n(x)] + (1 - \delta)E[f_r(x)].$$

If $E[f_n(x)]$ is set equal to 0, $E[f_o(x)]$ will equal d' and $E[f_r(x)]$ will equal d^*. Thus,

$$d' = (1 - \delta)d^*. \qquad [6]$$

Therefore, the value of d' which is obtained from recognition memory data can be transformed into the probability of remembering (i.e., that an item is in State R) by the simple operation of multiplication by a constant. Even if d^* is not known, so long as it remains unchanged d' may be an appropriate qualitative indicator of the effect on memory of many common variables, such as retention interval. The d' measure is not appropriate, however, for the study of the effect on memory of variables which may influence the value of d^*, for example, stimulus meaningfulness.

The Rating-Scale Method

In practice, in order to obtain an operating characteristic without manipulating the cut-off over a series of experiments, the rating-scale method (Egan, Schulman, & Greenberg, 1959; Pollack & Decker, 1958) is generally used. Here S, in addition to responding "old" or "new," rates his confidence that the response is correct. It is assumed that the confidence scale running from "highest confidence new" at one extreme to "highest confidence old" at the other reflects the apparent oldness scale directly; that is, S reports a monotonic transformation of the apparent oldness observation x as his rating response. This is equivalent to assuming that the confidence judgment reflects the subjective posterior probability that the item is old, since this posterior probability is monotonic with the apparent oldness scale. The procedure for generating an operating characteristic from rating data is discussed in detail by Green and Swets (1966).

CONTINUOUS STRENGTH THEORY

Wickelgren and Norman (1966) have proposed a series of models for recognition memory that are similar in their response rules to the finite-state decision theory, but that are based on the assumption of continuous habit strength rather than all-or-none storage. Their basic assumption is that the decision to respond "old" or "new" to an item is based upon whether or not the strength of that item is above some criterion value. The strength measure reflects directly the effect of memory factors such as retention interval, longer retention intervals leading to less strength. The response theory involves the introduction of a random component, *decision variability*, yielding distributions of strength for new and old items that are normal and have equal variance. The application of the theory of signal detectability to this process consists simply of equating the likelihood distributions $f_n(x)$ and $f_o(x)$ to these strength distributions; d' is then the measure of the difference in mean strength between new and old items. The operating characteristic predicted by this theory is symmetric about the minor diagonal, as shown in Figure 5; if plotted on normal coordinates this would be a straight line with unit slope and intercept d'.

Since Wickelgren and Norman assumed that the random variability is introduced in the decision process, thus being independent of the mean strength, equal variance of $f_n(x)$ and $f_r(x)$ is a feature of their model. However,

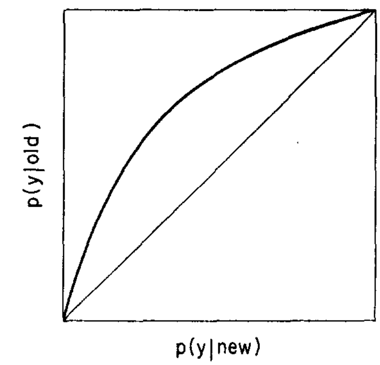

FIG. 5. Operating characteristic for continuous strength theory assuming equivariate normal likelihood distributions.

other conclusions about these variances could follow from assumptions about the introduction of random error that would be equally consistent with a continuous strength theory of the memory trace. Therefore, a more general strength model need not assume that the variance of the likelihood distributions are equal.

Norman and Wickelgren (1965) have proposed a variation of the continuous strength model that permits partial acquisition. First, they assumed that there is a distribution of strengths for new items. Next, they assumed that when an item is presented it receives an increment in strength with probability π; with probability $1 - \pi$ it remains in the distribution of new items. If π were set equal to $1 - \delta$, the predictions of this model would be identical to those of the finite-state decision theory. These models differ primarily in that the measure d^* in the finite-state decision theory is assumed to be a constant independent of memory factors. On the other hand, the size of the strength increment in the Norman and Wickelgren model would be expected to vary with conditions traditionally assumed to affect the strength of a memory trace. Furthermore, Norman and

Wickelgren state that their partial acquisition assumption is not a general one, but is meant to apply only to the special conditions of their experiment, as discussed in the next section of this paper.

RECOGNITION MEMORY EXPERIMENTS

In the traditional method for the measurement of recognition memory, S first learns a set of stimuli. He is then tested with a larger set consisting of new items as well as the old or learned items, and he must identify each as "old" or "new" (Woodworth & Schlosberg, 1954, p. 699). This procedure was followed by Egan (1958) in a study of recognition memory for words, with the addition that S made a confidence rating response instead of a simple binary choice. For the learning phase, 100 words were presented one at a time; the recognition test consisted of a list of 200 words, half of which were new items. The Ss gave each item on the recognition test a rating from "1" (positive the word was in the list to be learned) to "7" (positive the word is a new one).

Figure 6 shows an operating characteristic for Egan's Experiment 1, based on ratings averaged across Ss. Although the lack of symmetry of the operating characteristic about the minor diagonal is inconsistent with the equal variance assumption about $f_n(x)$ and $f_o(x)$ that was made by Wickelgren and Norman (1966), the data do not discriminate between continuous strength theory and the finite-state decision theory. In fact, since the form of the operating characteristics shown in Figure 4 are so close to the linear predictions of continuous strength theory, we will make no attempt to discriminate between theories on the basis of the shape of the operating characteristic.

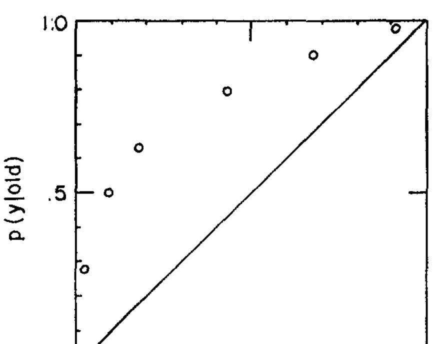

FIG. 6. Operating characteristic for recognition of words (data from Egan, 1958).

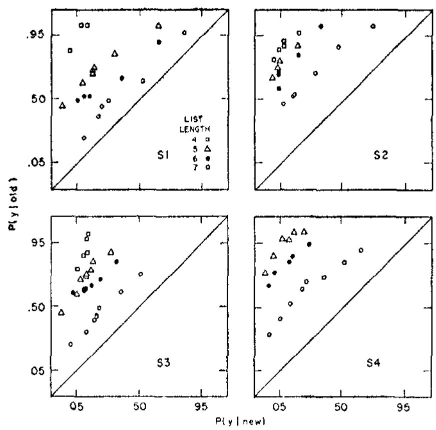

FIG. 7. Operating characteristics for individual subjects for the third item in a list of 3-digit numbers, with length of list as the parameter (data from Wickelgren & Norman, 1966).

The current emphasis on the study of short-term memory has led to the development of some new procedures for studying recognition memory. One of these is the probe experiment. A short list of stimuli is presented, one at a time, and after a short interval this is followed by a probe test consisting of a stimulus that has some fixed probability of being the same as one of the members of the list. The S must respond "old" if he thinks that the test item was in the study list, and "new" otherwise. Such a procedure was used in a study by Wickelgren and Norman (1966). They presented lists of from two to seven three-digit numbers, and followed each list with a single probe test. The *S*s indicated whether they thought the probe item had been in the study list and stated their confidence on a scale from 1 to 5.

A representative set of operating characteristics for each of the four individual *S*s is shown in Figure 7. These curves are for the third item in the list with length of list as the parameter. The basic finding that the distance of the operating characteristic from the major diagonal varies inversely with list length is consistent with both continuous strength theory and the finite-state decision theory. Wickelgren and Norman assume that an item's strength decays as intervening items are presented, thus predicting that the strength of a particular item in the list would decrease as list length increases. This increase is reflected in the value of d'. In the finite-state decision theory, δ is assumed to increase with the number of intervening items. Since $d' = (1 - \delta)d^*$, from Equation 6, the observed decrease in d' as list length increases corresponds to this increase in the value of δ.

As shown in Figure 4, the finite-state decision theory predicts an unusual pattern of slopes for normal-deviate operating characteristics as a function of δ. For small values of δ (and large d') the slope might be quite large. As δ increases, the slope decreases to a value less than one at mid-range values of δ, and then returns to unity as the line approaches the major diagonal. This pattern is borne out strikingly well in the data of Figure 7, despite the fact that Wickelgren and Norman concluded that there was no significant relationship between the values of slope and d'.

Melton, Bernbach, and Reicher (1964) studied the effect of repetition on recognition memory in the continuous task first developed by Shepard and Teghtsoonian (1961). Critical items that appeared four times each, at intervals of 2, 5, 20, or 40 intervening items, were imbedded in a long, continuous list of consonant trigrams. The *S*s were informed that each stimulus presented was equally likely, a priori, to be old or new, and responses to each item were made on a 10-point rating scale. The operating characteristics, pooled across intervals, are shown in Figure 8 for one, two, and three repetitions. The responses on the first presentation were used to calculate the false-alarm rates for all of these cases. It is apparent that the effect of repetition is to move the operating characteristic farther from the major diagonal, reflecting the increasing probability of the response "old" with repetition.

Like the effect of list length (Figure 7), this result is consistent with both strength theory and the finite-state decision theory. Traditional incremental strength theories of learning assume that the effect of repetition is to in-

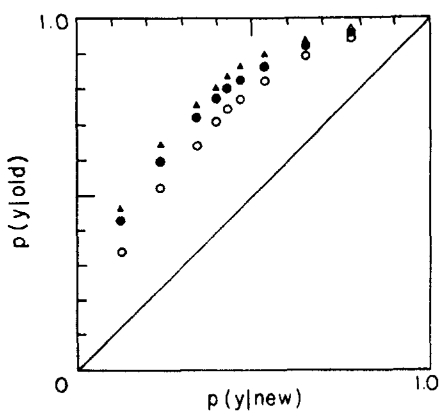

FIG. 8. Operating characteristics for 1, 2, and 3 repetitions of consonant trigrams in continuous recognition memory (data from Melton, Bernbach, & Reicher, 1964).

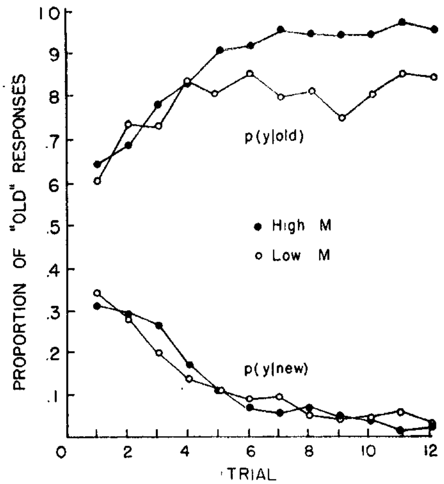

FIG. 9. Proportion of correct recognitions and false alarms for high M and low M consonant trigrams (data from Martin, 1967).

crease the strength of the repeated item, and this increase would be reflected in the value of d'. In finite-state learning models, on the other hand, it is assumed that the effect of repetition is to increase the proportion of items in the learned or remembered state. Thus, in the finite-state decision theory the value of δ decreases with repetition, and this is reflected by a corresponding increase in the observed value of d'.

Since the finite-state decision theory assumes that the effect of repetition is to decrease the probability of forgetting, the theory makes a strong asymptotic prediction in recognition learning. The smallest value that δ can attain is zero; thus, the maximum value of $p(y|old)$ is $\int_{x_0}^{\infty} f_r(x)$. The theory predicts then that the asymptotic value of $p(y|old)$ will be less than one in recognition learning experiments, though recall tests of learning generally yield asymptotes at a probability of one. This prediction is also made by one of the strength models suggested by Wickelgren and Norman (1966), in which a maximum value of strength is assumed such that d' has a finite maximum value.

Fortunately, this prediction is not inconsistent with experimental results. Lachman and Field (1965), for example, had *S*s learn a list of 50 words by a recognition method. On each trial, *S*s first studied the list and then were required to respond "old" or "new" to each of a series of 100 test items, containing 50 new filler items on each trial. Lachman and Field reported that the proportion of "old" responses to old items on the 128th trial was only .90, significantly less than one. If any further improvement in performance should be expected after 128 trials, it would probably be caused by phenomena outside of the scope of the present theory.

While in no sense discriminating between theories, this asymptotic prediction has important implications for the study of learning by recognition methods. Consider, for example, a study of the effect of meaningfulness (M) on recognition learning by Martin (1967). The *S*s learned a list consisting of four high M and four low M consonant trigrams; on each test trial, 16 new filler items were mixed with these eight, and *S*s responded "new" or "old" to each item. The proportions of correct recognitions and false alarms for each of the 12 trials of the experiment are shown in Figure 9. It appears in these data that the $p(y|old)$ curves for the two levels of M are approaching different asymptotes, while the false-alarm rate is not dependent on M. This would be predicted by the finite-state decision theory if the likelihood distribution for new items, $f_n(x)$, were independent of M and the value of d^* were greater for high M than for low M stimuli. Since one might expect high M trigrams to be more readily discriminable than low M items, this does not seem an unreasonable assumption. Moreover, if this interpretation is correct, the observed difference in performance between high and low M trigrams in this experiment can be attributed completely to the effect of item discriminability at the time of retrieval from memory, rather than to a difference in learning rate as a function of M.

In the final experiment considered in this section, Norman and Wickelgren (1965) used a slight variation of the probe procedure to study short-term recognition memory for digits. Their study list was a series of five digits, and this was followed by an interference set of eight digits that S copied. The probe test, for the experimental condition considered here, was either an adjacent and correctly ordered pair of digits from the study set (old) or a pair of new digits. The Ss indicated whether neither, one, or both probe digits had occurred in the study set and, if both, whether in the same order. For this analysis, responses were counted as "old" only if S responded that both digits were old and in the same order. The Ss also rated their confidence that they were correct on a scale from 1 to 5. The operating characteristic for this experiment is shown in Figure 10.

The shape of this function appears quite different from that of the smooth curves shown in Figures 6 and 8. This operating characteristic is very similar to the function shown in Figure 3b, which was predicted by the finite-state decision theory for the case in which the likelihood distributions, $f_n(x)$ and $f_r(x)$, do not completely overlap. There is, of course, no a priori reason to assume partially overlapping likelihood distributions for this experimental situation. The partial overlap condition is closely approximated with normal distributions, however, when the ratio of standard deviations, σ_r/σ_n, is very small. In this case, there exists a large region of $f_n(x)$ for which the value of $f_r(x)$ may be too small to be reflected in the data.

It is to these data that Norman and Wickelgren (1965) applied the partial acquisition version of the continuous strength model that we discussed previously. Noting that recognition memory for single digits in their experiment yielded a more typical rounded operating characteristic than obtained for pairs, they concluded that the strength for a single digit was incremented with probability, $\pi = 1$, while pair strengths were incremented with probability, $\pi < 1$. This conclusion was explained by assuming "that Ss code and rehearse the digits in nonoverlapping pairs. This would result in attending to every digit, but only about half of the pairs . . . [p. 487]." Thus, it is clear that the partial acquisition model of Norman and Wickelgren was not intended for ordinary probe experiments (e.g., Wickelgren & Norman, 1966).

Of course, an assumption of probabilistic acquisition could easily be in-

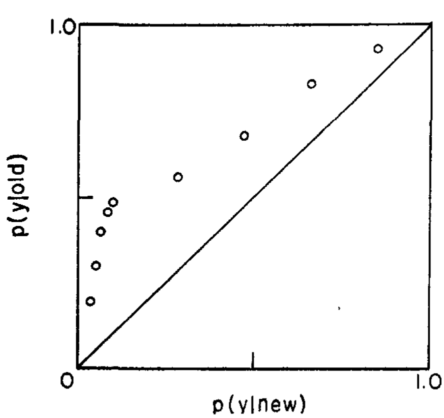

Fig. 10. Operating characteristic for recognition of digit pairs (data from Norman & Wickelgren, 1965).

corporated into a more general continuous strength model which could not be discriminated from the finite-state decision theory on the basis of data of this sort. Therefore, the remainder of this paper will concentrate on the most important distinction between the finite-state and strength theories, the assumption by the finite-state theory that the likelihood distributions $f_n(x)$ and $f_r(x)$ will not be affected by factors influencing memory.

CONFIDENCE JUDGMENTS IN RECALL

So far, the application of decision theory to memory has been discussed only with regard to recognition memory experiments. Though the finite-state decision theory was developed in that context, it also has an important application to recall tasks in which *S*s are asked to rate their confidence that they are correct.

First, we make the important assumption that *S*, after having made a recall response, makes an independent judgment as to whether that response was old or new (correct or incorrect). This implies that recall probability is independent of the apparent oldness scale. The probability of a correct response, $p(c)$, is given by the usual finite-state formula,

$$p(c) = p(R) + g[1 - p(R)], \quad [7]$$

where g is the probability of a correct response by guessing. The process of giving a confidence rating to the response made is assumed to be the same as that involved in recognition memory, and we expect imperfection or noise in the matching process to lead to likelihood distributions, $f_n(x)$ and $f_r(x)$, where x is an observation on the apparent oldness scale.

Before proceeding, we must distinguish between two types of operating characteristics. In the usual, or Type 1 operating characteristic, the proportion of "yes" responses given that the item is old is plotted against the proportion of "yes" responses given that the item is new. If a confidence rating method is used, "yes" is defined as a response above each successive criterion. In analyzing recall data involving more than two response alternatives, however, the proportion of confidence judgments above a criterion for responses that are correct are plotted against this proportion for incorrect responses. Such a response-conditional curve is a Type 2 operating characteristic. In general, as shown by Clarke, Birdsall, and Tanner (1959), the Type 2 operating characteristic and the d' value obtained from it will not be equivalent to those obtained from a Type 1 analysis.

Now consider a hypothetical experiment in which the underlying memory states are perfectly observable, so that the occurrence of an error (e) indicates that the item is in State N and a correct response (c) that the item is in State R. This implies that $p(y|c)$ and $p(y|e)$ are equal, respectively, to $p(y|R)$ and $p(y|N)$, so that the Type 2 operating characteristic is equivalent to the natural operating characteristic. Therefore, the value of d' obtained from a Type 2 analysis of an experiment with observable states is equivalent to the value of d^* for that experiment.

The above conditions are approximated fairly closely in a recall experiment involving an extremely large set of response alternatives, for example, a paired-associate learning experiment with common English words as responses. Since in this case the probability of a correct guess is close to zero, the occurrence of a correct response can be assumed to imply that the item is in State R. Similarly, since we assume that the probability of a correct response in State R is one, the occurrence of an error implies that the item is in State N. One such study has been reported by Murdock (1966). The Ss were presented lists of five A-B pairs of words; each list was followed by a probe test for one of the pairs in which one of the members (A or B) was presented. The Ss gave a response and rated their confidence on a 6-point scale running from + + + ("positive the word is correct") to − − − ("positive the word is wrong").

Figure 11 shows the Type 2 operating characteristic for each of the five serial positions, with the four central response categories decreased to two (− −, − and +, + +) because so many of the responses were at the extremes of the confidence scale. There is clearly no consistent effect of serial position on d' in these data. Murdock did find, however, a large effect of serial position on memory, the observed proportions of correct recall responses being .29, .29, .44, .59, and .85 for serial positions 1–5 respectively. If, as assumed, the Type 2 operating characteristics for this study are equivalent to the natural operating characteristic, this lack of dependence of the Type 2 d' on memory is predicted by the finite-state decision theory.

It is generally assumed by continuous strength theories that a confidence judgment directly reflects the strength of an item. For example, such an assumption is the basis for the application of the Wickelgren and Norman (1966) models to rating data. If this is the case, however, then the findings of the Murdock (1966) study cannot be accounted for by a traditional continuous strength theory. Such a theory assumes that an increase in response probability is caused by an increase in the strength of the correct response relative to the set of possible incorrect responses. This increase should have been reflected in a correlation between

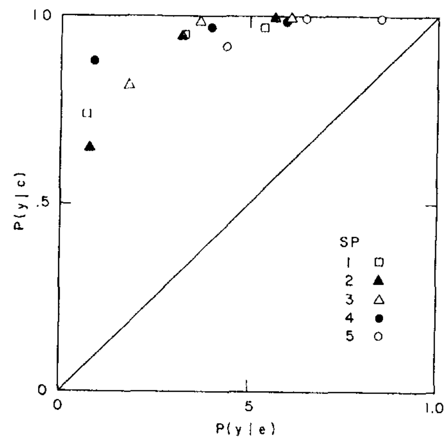

FIG. 11. Type 2 operating characteristics for short-term recall of paired-associates (data from Murdock, 1966).

the Type 2 d', which is the mean difference in strength between correct and incorrect responses, and the serial position of the item in Murdock's experiment. There is no evidence for such a relationship in the data shown in Figure 11.

Further evidence for the independence of serial position and the Type 2 d' has been obtained in two replications of Murdock's (1966) study, both run as parts of senior honors theses at Cornell University in 1967. In the first experiment, which was run by Patricia Geer, the study lists consisted of five paired-associates. The stimuli were the digits 1–5, and the responses were consonant trigrams. After each study list, one of the digits was presented, and Ss made a trigram response and rated their confidence on a 4-point scale running from + + to − −. The second study was run by Jules Tanenbaum, who used five-item lists with two-digit numbers as stimuli and consonant trigrams as responses. Confidence was rated on the 6-point scale used by Murdock (1966), ranging from + + + to − − −. Type 2 operating characteristics for each serial position in these experiments are shown in

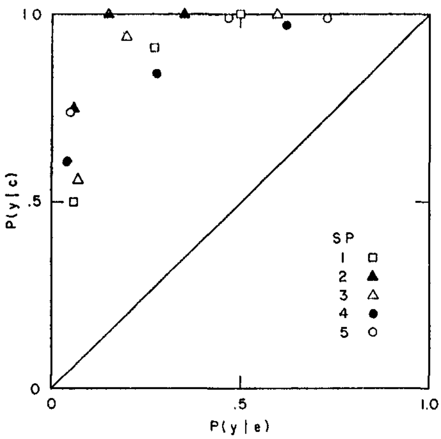

FIG. 12. Type 2 operating characteristics for short-term recall of paired-associates (data from Geer, personal communication).

Figures 12 and 13. As predicted by the finite-state decision theory, there is no evidence for an increase in d' with serial position in either of these sets of data.

An obvious objection to the above analysis is that the assumption of a zero guessing probability may not be very realistic. For example, *S*s might remember some of the responses even though the specific association to the test stimulus is forgotten. This would limit the population from which a guess is made and produce a value of the guessing rate, $g > 0$. Under this condition, the set of correct responses at any serial position would contain a proportion of correct guesses that would be a function of the values of both $p(c)$ and g, and $p(y|c)$ could not be assumed equal to $p(y|R)$.

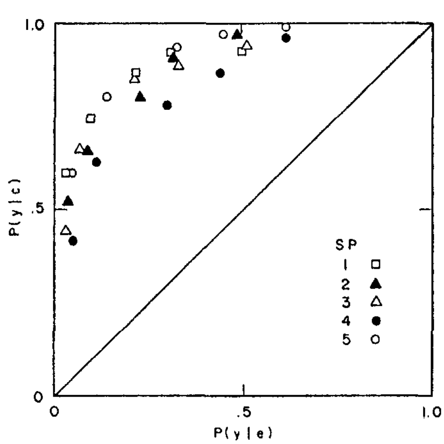

FIG. 13. Type 2 operating characteristics for short-term recall of paired-associates (data from Tanenbaum, personal communication).

If the value of g is very small (as we assumed it was in the experiments just discussed), the effect of guessing on the Type 2 operating characteristic should be very small also. This effect cannot be quantified unless the precise value of g is known. However, if a value for g can be estimated, it is possible to convert the observed values of $p(y|c)$ to estimates of $p(y|R)$, in the following manner.

From probability theory, we have the relationship,

$$p(y|c) = p(y|R)p(R|c) + p(y|N)p(N|c), \quad [8]$$

since States N and R form mutually exclusive and exhaustive categories. Using the relationship in Equation 7 and the fact that $p(y|N) = p(y|e)$, the above can be rearranged and solved for $p(y|R)$ as follows:

$$p(y|R) = \frac{p(y|c) - p(y|e)[1 - p(R|c)]}{p(R|c)}, \quad [9]$$

where

$$p(R|c) = \frac{p(c) - g}{(1 - g)p(c)}. \quad [10]$$

If the proportion of correct recall responses is used as $p(c)$, Equations 9 and 10 provide a means to "correct" Type 2 operating characteristics for guessing. The corrected function is a plot of $p(y|R)$ against $p(y|e)$, which is equivalent to the natural operating characteristic of the finite-state decision theory.

In a paired-associate learning experiment using a small and well-defined set of responses, a reasonable estimate of the guessing rate is simply $g = 1/n$, where n is the number of response alternatives. Therefore, if confidence judgments are collected in such an experiment, the above procedure can be used to convert the Type 2 operating characteristics to natural operating characteristics. According to the finite-state decision theory, these "corrected-for-guessing" curves should be independent of trial number or any other factors influencing the probability that an item is in State R.

The data we will consider are from an unpublished study by Douglas Hintzman. He had Ss learn lists of 16 paired-associate items. The stimuli were 16 nonsense syllables; the response set in the three experimental conditions consisted of either 4, 8, or 16 numbers. In addition to giving a paired-associate response to each stimulus, S rated his confidence that he was correct on a scale from 1 to 11. The natural operating characteristics, that is, the Type 2 operating characteristics modified according to Equations 9 and 10, are shown in Figure 14. As predicted, the trial number has no consistent effect on the distance between the natural operating characteristics and the major diagonal.

Like the serial-position data, these results are not consistent with continuous strength theory. Strength theory assumes that the effect of repeated trials in paired-associate learning is to increment the strength of the correct response. If this were so, the mean difference in strength between correct and incorrect responses, as measured by the Type 2 d', would increase as a function of trials. Of course, the operating characteristics shown in Figure

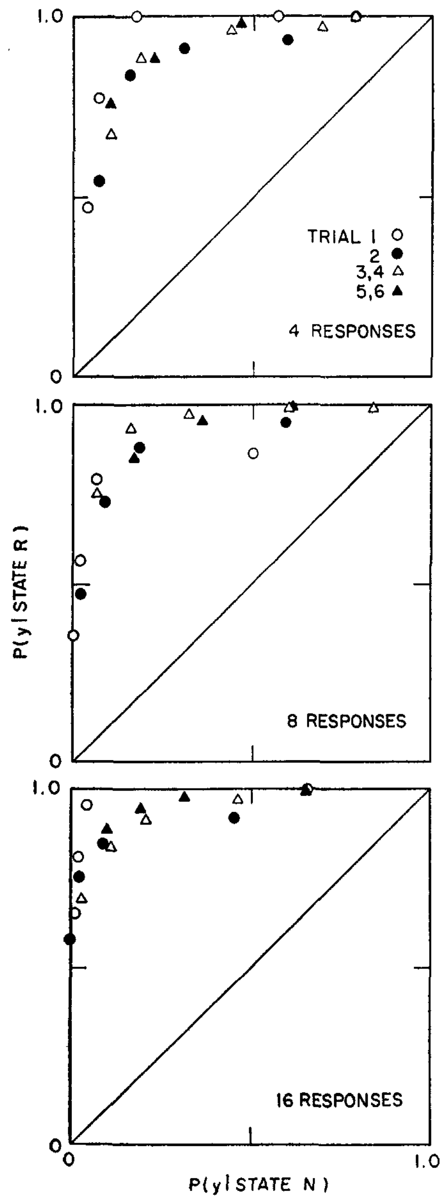

FIG. 14. Type 2 operating characteristics for paired-associate learning with 4, 8, or 16 response alternatives (data from Hintzman, personal communication).

14 depend on the specific correction-for-guessing procedure that was used. It is not unreasonable, however, to conclude that *some* correction is required

when the probability of a correct guess is quite significant. The procedure outlined in Equations 9 and 10 may not be intuitively obvious; however, it is based only on the assumption of Equation 7, which is the basis for the traditional correction-for-guessing procedure.

SUMMARY AND CONCLUSIONS

A number of recent investigators have demonstrated the usefulness of a decision-theory model, as in the theory of signal detectability, in accounting for recognition-memory data. Such a model follows very naturally from a continuous strength theory of the memory trace, since the decision scale can simply be equated to the strength scale. For example, one might assume a distribution, $f_n(x)$, giving the probability of a strength value x for a new item, and another distribution, $f_o(x)$, giving this probability for an old item. The difference between the means of these distributions, d', would be a measure of the mean difference in strength between new and old items. A recognition memory experiment might yield a family of $f_o(x)$ distributions, with values of d' reflecting memory factors such as list length, number of repetitions, and the like. Such a strength theory (e.g., Wickelgren & Norman, 1966) performs quite adequately in relating these memory factors to recognition memory data analyzed by operating characteristic methods.

Despite the general adequacy of strength models in this regard, it is a mistake to take evidence for continuous decision processes in recognition memory as evidence for the continuous nature of the memory trace. We have presented in this paper a finite-state decision theory that assumes only two states of memory, N and R, and therefore only two likelihood distributions, $f_n(x)$ and $f_r(x)$, on the decision scale. These distributions and d^*, which is the measure of the distance between their means, are assumed independent of memory factors; forgetting is reflected in δ, the probability that an item is in State N rather than R at the time of test.

In order to relate memory factors to the operating characteristic in recognition memory, we showed that the measure d' for the finite-state decision theory is equal to the product $(1 - \delta)d^*$. This yields the same general relationship between memory and d' as does continuous strength theory, that is, a lower value of d' under conditions in which less retention is expected. Furthermore, the predictions of the detailed shape of the operating characteristic in recognition memory were shown not to be different enough to permit distinguishing between the theories on this basis.

In their application to the analysis of confidence ratings in recall experiments, the continuous strength and finite-state decision theories make quite different predictions, however. According to strength theory, an increase in the probability of a correct response is caused by an increase in its strength relative to the set of possible incorrect responses. Therefore, the value of d' obtained from the Type 2 (response-conditional) operating characteristic should increase with the probability of a correct response. On the other hand, the finite-state decision theory predicts that the Type 2 d' (corrected for guessing if necessary) is equal to d^*, and therefore independent of response probability in recall. Data from several experiments were shown to provide considerable evidence for the independence of the Type 2 d' from memory factors, and thus for the finite-state assumption.

The finite-state decision theory has an important implication for the study of memory by recognition procedures. Memory, as represented in the theory by the probability of forgetting, δ, is but one of three factors influencing a recognition or confidence rating response. The response also depends upon the decision criterion, β, and upon the asymptotic discriminability of the stimulus as represented by d^*. The criterion is reflected in the false alarm rate, and correcting for false alarms by the use of the d' measure has become quite common in the study of recognition memory. It is equally important, however, to take account of the asymptotic stimulus discriminability, particularly when manipulating factors such as meaningfulness that may be expected to affect this discriminability. Thus, a value of d' from recognition memory data should be "corrected" for the projected maximum (d^*) in order to provide a pure measure of memory. When, as is usually the case, no reasonable estimate of d^* is available, it is necessary at least to give cognizance to this factor when drawing conclusions from studies of recognition memory.

REFERENCES

ATKINSON, R. C. A variable sensitivity theory of signal detection. *Psychological Review,* 1963, **70**, 91–106.

ATKINSON, R. C., & SHIFFRIN, R. M. Mathematical models for memory and learning. Technical Report No. 79, 1965, Institute for Mathematical Studies in the Social Sciences, Stanford University.

BERNBACH, H. A. A decision and forgetting model for recognition memory. Technical Report No. 64-4, 1964. Michigan Mathematical Psychology Program.

BERNBACH, H. A. Stimulus learning and recognition in paired-associate learning. Unpublished doctoral dissertation, University of Michigan, 1965.

BOWER, G. H. Notes on a descriptive theory of memory. Paper presented at the Second Conference on Learning, Remembering and Forgetting, New York Academy of Sciences, Princeton, September 1964.

CLARKE, F. R., BIRDSALL, T. G., & TANNER, W. P., JR. Two types of ROC curves and definition of parameters. *Journal of the Acoustical Society of America,* 1959, 31, 629–630.

EGAN, J. P. Recognition memory and the operating characteristic. Technical Note AFCRC-TN-58-51, 1958, Indiana University, Hearing and Communication Laboratory.

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

GREEN, D. M., & SWETS, J. A. *Signal detection theory and psychophysics.* New York: Wiley, 1966.

HULL, C. L. *Principles of behavior.* New York: Appleton-Century, 1943.

KINTSCH, W. Recognition learning as a function of the length of the retention interval and changes in the retention interval. *Journal of Mathematical Psychology,* 1966, **3**, 412–433.

LACHMAN, R., & FIELD, W. H. Recognition and recall of verbal material as a function of degree of training. *Psychonomic Science,* 1965, **2**, 225–226.

LUCE, R. D. A threshold theory for simple detection experiments. *Psychological Review,* 1963, **70**, 61–79.

MARTIN, E. Stimulus recognition in aural paired-associate learning. *Journal of Verbal Learning and Verbal Behavior,* 1967, **6**, 272–276.

MELTON, A. W., BERNBACH, H. A., & REICHER, G. M. Stimulus frequency in continuous short-term recognition memory. Paper presented at the meeting of the Psychonomic Society, Niagara Falls, Ontario, October 1964.

MURDOCK, B. B., JR. The criterion problem in short-term memory. *Journal of Experimental Psychology,* 1966, **72**, 317–324.

NORMAN, D. A., & WICKELGREN, W. A. Short-term recognition memory for single digits and pairs of digits. *Journal of Experimental Psychology,* 1965, **70**, 479–489.

PETERSON, W. W., BIRDSALL, T. G., & FOX, W. C. The theory of signal detectability. *IRE Transactions,* 1954, **PGIT-4**, 171–212.

POLLACK, I., & DECKER, L. R. Confidence ratings, message reception, and the receiver operating characteristic. *Journal of the Acoustical Society of America,* 1958, 30, 286–292.

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

SORKIN, R. D. Extension of the theory of signal detectability to matching procedures in psychoacoustics. *Journal of the Acoustical Society of America,* 1962, **34**, 1745-1751.

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

TANNER, W. P., JR., & SWETS, J. A. A decision making theory of visual detection. *Psychological Review,* 1954, **61**, 401-409.

WICKELGREN, W. A., & NORMAN, D. A. Strength models and serial position in short-term recognition memory. *Journal of Mathematical Psychology,* 1966, **3**, 316-347.

WOODWORTH, R. S., & SCHLOSBERG, H. *Experimental psychology.* New York: Holt, 1954.

(Received July 14, 1966)