Psychological Review
1966, Vol. 73, No. 1, 44–58

SIGNAL-DETECTABILITY THEORY OF RECOGNITION-MEMORY PERFORMANCE[1]

THEODORE E. PARKS[2]

University of Wisconsin

Psychophysical signal-detection theory is applied to recognition-memory performance. Old items in the test list are considered to be analogous to "signals" while new items are analogous to "noise only." The resulting fundamental assumptions describe the covert responses which mediate recognition-memory performance as varying continuously in strength. Covert responses to both old and new items are normally distributed with the distance between distributions representing learning and retention. The overt response to an item depends on whether or not S's covert response exceeds an arbitrary criterion. The available evidence suggests that S's criterion will be set such that he will most probably choose a number of items approximately equal to the number of old items in the test.

A test of recognition memory typically involves presenting the subject with a list composed of old items (i.e., items that were in an inspection list presented previously) and new items. He is required to examine each item of the test list and to indicate whether or not it was a member of the inspection list. Two measures of performance are taken: (*a*) the proportion of the old items in the test list which the subject indicates are old and (*b*) the proportion of new items which the subject indicates (incorrectly) are old.

Customarily, these two scores have been combined in a "correction for guessing" which was thought to yield a single "true" measure of memory. Underlying this correction is the assumption that recognition memory is a sort of all-or-none process; that is, that correct recognitions consist of items recognized together with items chosen purely through random guessing. On the contrary, the present paper presents evidence that an item is not either recognized or not recognized, but, rather, the covert events which mediate overt recognition behavior vary in strength along a continuum.

Under this assumption, the subject is faced with a complex decision problem. The present theory attempts to describe this problem and the subject's solution in detail and in such a way as to make possible precise predictions concerning the effects of certain independent variables on recognition-memory performance. Following to some extent the suggestion of Egan (1958), many of the assumptions employed are quite similar to those found by Swets, Tanner, and Birdsall (1961) to be adequate for signal detection in a psychophysical situation.[3]

[1] The present paper is adapted from part of a dissertation submitted to the University of Wisconsin. The author wishes to express his sincere gratitude to W. J. Brogden, W. R. Thurlow, M. L. Posner, T. Natsoulas, and D. W. Bessemer for their encouragement and many helpful suggestions.

This research was supported in part by grants from the National Science Foundation and the Research Committee of the Graduate School from funds provided by the Wisconsin Alumni Research Foundation.

[2] Now with University of California, Davis.

[3] For those readers who are familiar with Swets et al. (1961), the application of their

GENERAL THEORY

Dependent Variables

In taking the recognition test, the subject may be required to choose a certain total number of test-list items (usually equal to the actual number of old items in the test), or he may be free to choose as many or few items as he desires. The present paper is concerned primarily with performance in the latter situation. Since the procedure employed in these experiments does not require the subject to pick some particular number of items, the number of old items that he calls old does not necessarily covary with the number of new items that he calls old. These two scores, expressed as the proportion of each type of item that the subject calls old, are the primary dependent variables of interest.

Fundamental Mechanisms

The present theory posits the existence of a psychological continuum according to which each decision to respond "old" or "new" is made. This dimension may be thought of as representing the degree of familiarity associated with test items and accordingly will be named the familiarity continuum (X). It represents that property of the covert response to the test stimuli which is effective in the determination of the subject's overt report.

theory to recognition memory may be made by considering each test item to be a "trial" with old items being "signal plus noise" trials and new items being "noise only" trials. However, the hypothesized mechanism which determines the position of the cutoff is not equivalent to the optimal strategy rule proposed by Swets et al. Egan in applying Swets et al. to recognition memory was not concerned with the precise determinants of the cutoff nor with the application of the theory to the multiple-choice test situation discussed later.

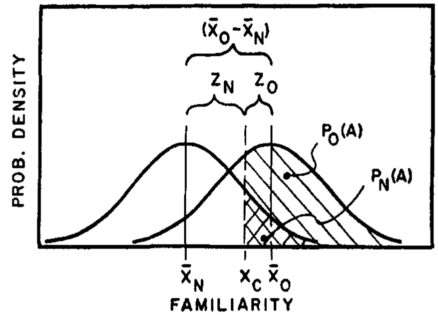

FIG. 1. Basic assumptions: $f_O(X)$, the curve to the right, represents the distribution of events which occurs on those occasions when the subject inspects an old item; $f_N(X)$, the curve to the left, represents the distribution of events which occurs when new items are inspected. (Assuming that the subject sets his cutoff at (X_C) and responds "old item" if and only if (X_i) exceeds (X_C), then $(P_O(A))$ represents the probability of the subject responding old when in fact the item is an old item. $(P_N(A))$ represents the probability of the subject reporting old when in fact the item is a new item.)

When the subject inspects an old item in the test list, it is assumed that a covert event occurs which has some value (X_i) on the familiarity continuum. Furthermore, (X_i) varies from item to item according to a normal distribution, $f_O(X)$. It is further assumed that an event on the same continuum occurs when the subject inspects any of the new items on the test list (as a result of generalization from the old items on transfer of prior learning). These events also form a normal distribution, $f_N(X)$, whose variance is equal to that of $f_O(X)$. The distance on the continuum between the means of the two distributions $(\bar{X}_O - \bar{X}_N)$ may be determined in Z units relative to the variance of the item distributions.[4] These assumptions are illustrated in Figure 1.

The quantity $(\bar{X}_O - \bar{X}_N)$ is in-

[4] The measure $(\bar{X}_O - \bar{X}_N)$ corresponds to (d') in Swets et al. (1961).

fluenced by learning and retention. These variables determine the position of the old-item distribution on (X). In addition, $(\bar{X}_O - \bar{X}_N)$ may be affected by generalization (which influences the final position of the new-item distribution) and by transfer of prior learning (which affects the initial positions of the new and old distributions). If, however, the same population of old and new items is used in all conditions of an experiment, then variation in $(\bar{X}_O - \bar{X}_N)$ represents variation in learning and retention.

It is further assumed that the subject's overt report regarding item (i) of the test list corresponds to the decision that the event (X_i) associated with item (i) is a member of $(f_O(X))$ or, contrariwise, that it is a member of $(f_N(X))$. We shall assume that this decision is made in terms of a cutoff (X_C) set on (X). That is, the subject will report that item (i) is an old item if, and only if, $(X_i) > (X_C)$ (see Figure 1).

For any given position of the cutoff, the probability that an old item will be reported as old (call this value $P_O(A)$ or the probability of a "*true* positive") is represented by the area under $(f_O(X))$ which lies to the right of (X_C). Similarly the probability that a new item will be called old (call this value $P_N(A)$ or the probability of a "*false* positive") in the area of $(f_N(X))$ to the right of the cutoff (see Figure 1). The proportions of true positives $(P_O(A))$ and of false positives $(P_N(A))$ actually obtained in an experiment are taken to be estimates of $(P_O(A))$ and $(P_N(A))$, respectively.

It follows that if the distance between the distributions is fixed (that is, if learning, retention, generalization, and transfer are constant), then the probability of a false positive may be reduced only by moving the cutoff to the right. Such a change would produce a simultaneous decrease in true positives.

Egan (1958) presented evidence in support of these basic assumptions in a recognition-memory test situation. His procedure involved noting the effects of encouraging or discouraging guessing on the probabilities of true positives $(P_O(A))$ and of false positives $(P_N(A))$. His method for testing these basic assumptions (i.e., plotting "ROC" curves) is equivalent to the following procedure. The proportions of true positives and that of false positives obtained in one of the test conditions are taken as estimates of the probability of true positives, $P_O(A)$, and false positives, $P_N(A)$, respectively. This pair of values is used to estimate the distance $(\bar{X}_O - \bar{X}_N)$ between the old- and new-item distributions (see Figure 1). That is, a table of normal curve will show what the position of the cutoff (in units of Z from the mean of the old-item distribution) *must have been* in order that the area of the curve above the cutoff would be equal to the theoretical $P_O(A)$. Similarly, the value $P_N(A)$ is used to find the distance between the cutoff and the mean of the new-item distribution. These two values are added together to give the estimate of $(\bar{X}_O - \bar{X}_N)$.

When this procedure was repeated for each of the test conditions, it was found that the resulting values for $(\bar{X}_O - \bar{X}_N)$ were essentially constant. Thus, changes in $P_O(A)$ produced by encouraging guessing were accompanied by appropriate changes in $P_N(A)$.

The Cutoff Rule: The Expected-Proportion Model

Beyond these basic assumptions, the present theory attempts to describe exactly *where* the cutoff will be set with respect to the old- and new-item distributions.

It is assumed that the subject is instructed as to the proportion of all the test items that are, in fact, old items. If the cutoff is set too low, the number of items that will be called old will probably be greater than the number of old items in the test. Similarly, if the cutoff is too high, the subject will probably choose fewer items than the actual number of old items. The unique assumption of the present model is that the subject's cutoff will be set at such a place on the familiarity continuum that he will be *expected to choose a proportion of items equal to the proportion of items which are, in fact, old.*[5]

For any given cutoff, the expected proportion of items which a subject will choose is given by the straightforward expression:

$$EP = (P(O))(P_O(A)) + (P(N)(P_N(A)) \quad [1]$$

where EP equals the proportion of test-list items the subject will be expected to check or, in other words, will be *most likely* to check; $P(O)$ is the proportion of all test items which are, in fact, old; and $P(N)$ is the proportion of all test items which are new.

The basic assumption of this model is that the subject's cutoff will be set such that the expected proportion of items checked (EP) will equal some proportion (K) of all test items which are, in fact, old. That is, it is assumed that:

$$EP = (K)(P(O)) \quad [2]$$

By substituting Equation 1 into Equation 2, we have the definition of the expected-proportion cutoff rule:

$$(P(O))(P_O(A)) + (P(N))(P_N(A)) = (K)(P(O)) \quad [3]$$

The only unknown of Equation 3, the value of (K), is assumed to be influenced in an unspecified manner by the costs and rewards of correct and incorrect responses. In experiments in which no deliberate attempt is made to vary these values, (K) is assumed to be constant. The fundamental assumptions and the expected-proportion rule may be tested separately or may be combined and tested simultaneously. This latter procedure will be illustrated shortly.

EVIDENCE FOR THE EXPECTED-PROPORTION MODEL

The Davis, Sutherland, and Judd Experiment

Davis, Sutherland, and Judd (1961) examined recognition-memory performance using both consonant-vowel nonsense syllables and two-digit numbers as stimuli.[6] The data for the nonsense-syllable condition will be examined first.

Each nonsense syllable consisted of 1 of 18 consonants (omitting q and z) followed by 1 of 5 vowels. A random list of 15 items was presented to the subject as the inspection list, each item being presented

[5] The word "expected" is used in the mathematical sense. The expected proportion of outcomes of a certain sort (in this case, of an item being called old) is the proportion which is *most likely* to occur over a large sample of occasions (in this case, a large number of test items). The expected proportion may be found simply by multiplying the probability of the outcome (the response) by the proportion of occasions (items).

Although the subject is not *required* to choose some fixed proportion of the items, we assume that he behaves in such a way as to make it *most likely* that he will choose a certain proportion of items.

[6] They also tested recall performance with these same stimuli, but only the recognition data is of interest here.

for 1.5 seconds. Immediately following the presentation of the inspection list, the subject was given a sheet of paper containing the 15 inspection (old) items together with 15, 45, or 75 new items (the test list). Thus, the proportion of old items, $P(O)$, in the test list was either .500, .250, or .167. There were 24 subjects in each condition. The instructions, which were the same in all conditions, required the subject to check as many old items as he could but not to guess wildly. The fundamental dependent variables, for our purposes, are the proportion of true positives ($p_O(A)$) and the proportion of false positives ($p_N(A)$) obtained in each condition. These values appear in Figure 2.

The test of applicability of the expected-proportion model to these data is relatively direct. Since the conditions of learning and recognition are constant (except for the values of $P(O)$ and $P(N)$), the difference between the means ($\bar{X}_O - \bar{X}_N$) and the value of (K) are presumably constant. The values of $p_O(A)$ and $p_N(A)$ for any one group can be used to estimate both of these parameters.

The data obtained under the $P(O) = .50$ condition were arbitrarily chosen for this purpose. In this group $P_O(A)$ was found to be (.694). A table of the normal curve shows that the probability of a true positive would be (.694) if the cutoff were located at (.51) Z units below the mean of the new-item distribution. Similarly, the probability of a false positive would be equal to the obtained value of (.186) if the cutoff were located at (.89) Z units above the mean of the new-item distribution. Adding these distances (see Figure 1) we estimate that the distance between the means ($\bar{X}_O - \bar{X}_N$) is equal to (1.40) Z units.

Now in order to estimate (K), these same values of $P_O(A)$ and $P_N(A)$ are placed in Equation 3 along with values of $P(O) = P(N) = .50$. Solving the equation, (K) equals .880.

Since both of these parameters are assumed to be constant, we assume that the subjects in each of the other conditions are faced with two normal distributions at a distance of (1.40 Z). Furthermore, we assume that their cutoff will be set such that they will be expected to check a proportion of items equal to (.880) times the value of ($P(O)$) for their conditions. Thus, for example, when $P(O) = .25$, they will attempt to check (.880), (.25), or (.220) of all the test-list items.

On the basis of these assumptions we may predict precisely the values of $P_O(A)$ and $P_N(A)$ which should be obtained in each of the other conditions. This procedure involves a trial and error search for the place where the cutoff *should* be set in order that Equation 3 will be satisfied.[7] The

[7] A cutoff point in terms of the distance (Z_O) between the cutoff and the mean of the old distribution is *tentatively* selected, and the resulting tentative value of $P_O(A)$ is found from a table of the normal curve. Since the distance between the means is assumed to be (1.40 Z), there will be a corresponding tentative value of Z_N and of $P_N(A)$. These tentative values of $P_O(A)$ and $P_N(A)$ are used in Equation 3 along with the values for $P(O)$ and $P(N)$ for the condition. This procedure is repeated until a cutoff position is found such that the expected proportion equals (.880) ($P(O)$). This procedure may be illustrated for the $P(O) = .250$ condition as follows:

Tentative Z_O	Tentative $P_O(A)$	Corresponding Z_N	Tentative $P_N(A)$	Expected proportion
−.20 Z	.579	1.20 Z	.115	.231 Reject
−.15 Z	.560	1.25 Z	.106	.220 *Accept*
−.10 Z	.540	1.30 Z	.097	.208 Reject

Since .220 equals (.880) (.250), we accept the prediction that the cutoff will be located .15 Z units below the mean of the old-item distribution and that, therefore, $P_O(A)$ will equal .560, and $P_N(A)$ will equal .106.

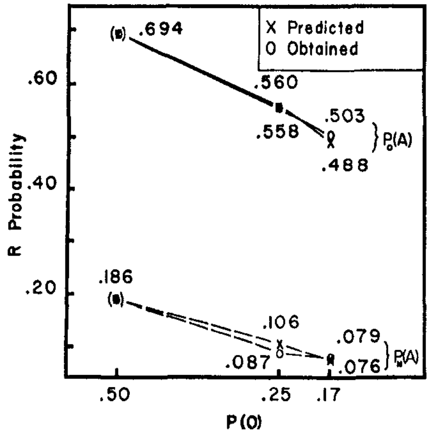

FIG. 2. Obtained and predicted recognition data when nonsense syllables are used as stimuli; $K = .880$; $(\bar{X}_O - \bar{X}_N) = 1.40Z$ (Davis, Sutherland, & Judd, 1961). (In this and the following figures, the data in parenthesis were used in making the predictions and, therefore, must coincide with the predictions.)

cutoff which yields values of $P_O(A)$ and $P_N(A)$ that satisfy this equation is taken as the predicted place of the cutoff, and these values of $P_O(A)$ and $P_N(A)$ are taken as the predictions of $p_O(A)$ and $p_N(A)$, respectively. The resulting predictions are compared to the obtained data in Figure 2.

In another experiment reported in the same paper Davis et al. (1961) used numbers as stimuli. These items consisted of all possible permutations of 10 digits taken 2 at a time. In all other aspects the experiment was identical to the one given above, and consequently the analysis is the same. Once again, the data obtained when $P(O) = .50$ is used to estimate (K) and $(\bar{X}_O - \bar{X}_N)$, and these values are used to predict the data under $P(O) = .250$ and $P(O) = .167$.

The Davis et al. experiments, while providing excellent support for the expected-proportion model, leave certain points unclarified.[8] The entire test list was exposed, and the subject was allowed to respond to the items in any order he wished. Under these conditions the subjects can produce data which are consistent with the expected-proportion model without, technically, using a cutoff at all (although the other, more fundamental assumptions would need to be valid). That is, the subject, faced with the two normal distributions, could start checking off the items with highest familiarity and work down to items with lower familiarity and simply stop when he had checked a number of items approximately equal to the number of old items on the test list. Such a strategy, on the other hand, would be impossible if the subject were allowed to view only a single test item at a time and was required to make his decision on that item before moving on to the next item. A demonstration that the expected-proportion model is adequate

[8] The goodness of fit of predicted to obtained data is ascertained by inspection throughout this paper except in Experiments I and II described later. The test used in the latter cases requires data for individual subjects. These data are not available for the other experiments that are discussed.

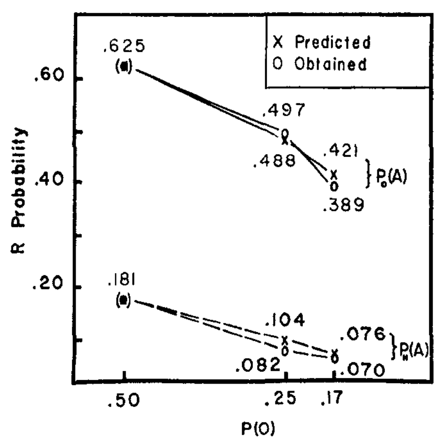

FIG. 3. The predicted and obtained recognition-memory data when two-digit numbers are used as stimuli; $K = .806$; $(\bar{X}_O - \bar{X}_N) = 1.23\ Z$.

under such a serial-test procedure would provide less ambiguous support for the notion of a cutoff. Two studies by the present author were designed to shed some light on this question. The method and results of these studies are described in detail in Parks (1964).

Experiment I

A total of 10 groups of 8 subjects each was presented with a 16-item list of consonant-vowel-consonant trigrams of low association value (Archer, 1960). Each subject viewed each item of the list only once, the exposure of each item being 1.67 seconds with a 1.33-second break between items. All subjects received the same inspection list, but eight different orders of presentation were used within each group. Immediately following the presentation of the last inspection item, the subject was presented with a test list composed of a random mixture of the 16 inspection items and 4, 8, 16, 32, or 64 new consonant-vowel-consonant trigrams. Thus, the proportion of test items that were old items was either $P(O)$ = .80, .67, .50, .33, or .20. In each case the subject was carefully instructed as to the number of old and number of new items which would appear in his test list. Care was taken to balance the effects of the particular new items used across the levels of $P(O)$.

In addition to differences in $P(O)$, the tests differed in response procedures. Half of the

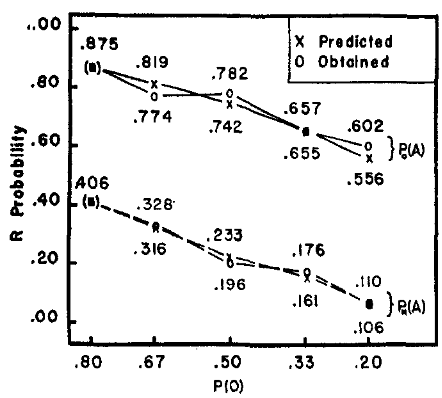

FIG. 4. Obtained and predicted data for the free-responding condition of Experiment I. (Predictions are made from the expected-proportion model and the $P(O)$ = .80 data; K = .977; $(\bar{X}_O - \bar{X}_N)$ = 1.39 Z.)

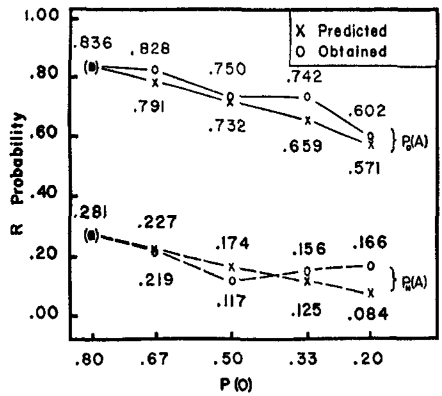

FIG. 5. Obtained and predicted data for the serial-responding condition of Experiment I. (Predictions are made for the expected-proportion model and the $P(O)$ = .80 data; K = .906; $(\bar{X}_O - \bar{X}_N)$ = 1.56 Z.)

groups were allowed to uncover the entire list at the beginning of the test and to respond in any order they desired ("free-responding" condition). The other five groups had a series of sliding masks over their test lists such that they could view only a single item at a time and were forbidden to return to an item already examined ("serial-responding" condition).

In both cases the subjects were instructed to underline any item which they thought had been on the original list. They were permitted to guess but encouraged not to guess wildly. The actual lists used in the two responding conditions were identical except for the masks placed over the lists in the serial condition. Finally, in both conditions, the subjects were free to take as much time as they desired but were encouraged to "keep moving along."

The treatment of the data is identical to that employed above with the data of Davis et al. The $P_O(A)$ and $P_N(A)$ scores obtained in $P(O)$ = .80 under the free-responding condition are used to estimate $(\bar{X}_O - \bar{X}_N)$ and (K) and, from these values, to predict the remaining free-responding data. The mean data and predictions appear in Figure 4. Similarly, the serial-responding $P(O)$ = .80 data are used to predict the performance of serial-responding groups (see Figure 5).

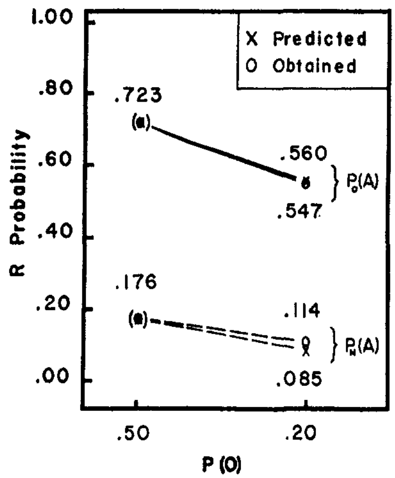

FIG. 6. Obtained and predicted data of Experiment II in which the $P(O)$ of the first few items was equal to the overall $P(O)$ for the condition. (Predictions are made from the expected-proportion model and the $P(O) = .50$ data; $K = .899$; $(\bar{X}_O - \bar{X}_N) = 1.52$.)

The goodness of fit of each of the four predicted curves was evaluated using a test suggested by Grant (1962). Essentially this procedure tests the correlation of predicted and obtained means against the variance within conditions. The null hypothesis is that the correlation is no greater than that which would be expected on the basis of chance. The results are as follows: free responding, $P_O(A)$: $F = 24.068$; free responding, $P_N(A)$: $F = 17.716$; serial responding, $P_O(A)$: $F = 19.826$; and serial responding, $P_N(A)$: $F = 5.030$. In each case $df = 1/35$. All of these are significant at the .05 level of confidence. In fact only the serial-responding $P_N(A)$ measure fails to achieve significance at .01.

Inspection reveals that under the serial-responding condition the obtained value of $p_N(A)$ is lower than predicted at $P(O) = .50$ and higher than predicted at $P(O) = .20$. Reexamination of the test lists employed in these conditions suggested that this effect might have been due to the proportion of old items among the first few items of the test list. It was found that the mean proportion of old items in the first 10 items for the $P(O) = .50$ lists was .60. The mean proportion among the initial 10 items for the $P(O) = .20$ lists was .10. These differences between overall $P(O)$ and $P(O)$ within the initial test items might well have made the subjects more conservative in the former case and more liberal in the latter. This variable would not systematically affect the free-responding subjects since they viewed the items in any order desired. A second experiment supports this hypothesis.

Experiment II

Two groups of eight subjects each were tested under identical conditions to those of the serial-responding $P(O) = .50$ and $P(O) = .20$ groups of Experiment I. In this case, however, the test lists employed were randomly constructed *but* with the restriction that the $P(O) = .50$ lists each have 5 old items among the first 10 test items and lists for $P(O) = .20$ each have 2 old items among the first 10 items.

The results of this study appear in Figure 6 along with the theoretical curve generated from the $P(O) = .50$ data. In this case the F value for correspondence of obtained and predicted means is equal to 12.598 for $P_O(A)$ and 5.000 for $P_N(A)$ ($df = 1/14$, $p < .01$, and $p < .05$), respectively.

In these experiments, as in those of Davis et al., when $P(O)$ is varied, the subjects' performances varied as though they were using the expected-proportion cutoff rule. Furthermore, the fit of the model to the serial-responding data indicates that the cutoff hypothesis may indeed be valid and lends support, in the interests of parsimony, to the cutoff type of description of free-responding performance.

Finally, it is interesting to note that the theoretical curves which fit these data are based, in part, on the assumption that $(\bar{X}_O - \bar{X}_N)$ is constant. According to the theory, this fact may be interpreted as indication that retention was constant across the various levels of $P(O)$ even though the subjects in the lower levels of $P(O)$ took more time to complete the test. In the $P(O) = .80$ condition, the subjects required an average of 112 seconds and 121 seconds to complete the test in the free-responding and serial-responding groups, respectively. When $P(O) = .20$, a mean of 394 and 273 seconds was required. A 2×5 analysis of variance on the time measures revealed a significant effect of $P(O)$ level ($F = 12.901$, $df = 4/70$, $p < .01$).

From a methodological standpoint, this finding illustrates the value and even necessity of precise theory since it could not have been ascertained by simple examination of the data. Perhaps, in fact, the main value of the present theory lies in its being able to provide a means of evaluating retention which allows for, rather than being confounded with, the particular conditions under which retention is measured.

Before proceeding, one possible source of confusion regarding these tests of the expected-proportion model must be discussed. If the subjects tend to choose a certain fixed number of items, the resulting pair of $p_O(A)$ and $p_N(A)$ values will necessarily satisfy Equation 3. But there are a variety of such pairs that may occur. The accuracy of the predictions given above demonstrates that the particular pair that occurs in each condition is that *unique* pair of areas that *both* (*a*) will satisfy Equation 3 and (*b*) *may occur* when a cutoff is moved along two normal curves of equal variance at a fixed distance.

Other Applications of the Expected-Proportion Model

The expected-proportion model may also be tested by holding $P(O)$ and $P(N)$ constant while some other variable of learning or retention is manipulated. In this case, however, as we shall see, only the ideas expressed in Equation 3 are tested (i.e., no test is made of the fundamental assumption of normality).

One such experiment has been reported by Strong (1912). The number of stimuli in the presentation list was the independent variable. Either 5, 10, 25, 50, 100, or 150 magazine advertisements were presented one at a time at a rate of one per second. Immediately following the presentation of the inspection list, the subject was given a stack of stimulus cards containing a random mixture of the inspection items (i.e., "old" items) together with an equal number of new advertisements. Thus, the test list contained either 10, 20, 50, 100, 200, or 300 items. The subject was required to sort the test items into four stacks according to his degree of conviction that each item was an old item. For our purposes, however, we will lump together the three classes that included items which the subject "had at least a faint idea he had seen before . . . i.e., those of which he was about 25% sure [p. 449]." The proportion of the old items in the test list which the subject placed into any one of these three classes shall be considered to be the proportion of true positives ($p_O(A)$). The proportion of all new items in the test list which the subject placed into any of these three classes is taken as the proportion of false positives ($P_N(A)$).

The values of the parameters ob-

TABLE 1

PROPORTION OF TRUE POSITIVES AND FALSE POSITIVES OBTAINED BY STRONG (1912) AS A FUNCTION OF THE NUMBER OF ADVERTISEMENTS IN THE INSPECTION LIST

Condition	No. *Ss*	No. old list items	No. old list items	Obtained $P_O(A)$	Obtained $P_N(A)$
Rec 10	80	5	5	.891	.036
Rec 20	40	10	10	.885	.063
Rec 50	40	25	25	.823	.074
Rec 100	40	50	50	.726	.114
Rec 200	40	100	100	.678	.135
Rec 300	40	150	150	.530	.129

Note.—Test list consisted of the inspection-list items together with an equal number of new items.

tained in each condition are given in Table 1 along with the number of subjects for each condition.

Ordinarily, the expected-proportion model might be tested against these data by simply placing the estimates of $P(O)$ and $P(N)$ obtained in each group along with values $P(O) = P(N) = .50$ into Equation 3 and solving for (K). The value (K) should, of course, be constant. However Equation 3 is based on the assumption that the subject has been informed of the values, $P(O)$ and $P(N)$. According to his report, Strong's subjects were apparently not so instructed. Thus, the model must be modified as follows:

$$(K)(P(O))' = (P(O))'(P_O(A)) + (P(N))'(P_N(A)) \quad [4]$$

where these prime marks indicate values assumed by the subject rather than the actual values of the parameters. That is, the prime values are fitting parameters.

By one interpretation this equation simply states that the subject will set his cutoff such that the expected proportion of items he *believes* he will check (based on his estimates of $P(O)$ and $P(N)$) will be equal to some constant fraction of what he *believes* to be the proportion of test-list items which are old items.

In order to test this model against Strong's data, Equation 4 must be rewritten as follows:

$$(K) = (P_O(A)) + \frac{P(N)'}{P(O)'}(P_N(A)) \quad [5]$$

Now if it is assumed that $P(O)'$ and $P(N)'$ are constant across groups, then $P(N)'/P(O)'$ will be constant and may be treated as a single fitting parameter. As usual, (K) is also assumed to be constant. Since $P_O(A)$ and $P_N(A)$ are known for each group, an equation in two unknowns may be written from the data of each group. Solving two such equations (from the 10 and 20 conditions) simultaneously, it is found that $P(N)'/P(O)' = (2.22)$ and $(K) = (.971)$.

It is now possible to predict the value of $P_N(A)$ for any group by placing these values together with the estimate of $P_O(A)$ for that group in Equation 5. These predictions are compared to the obtained values in Figure 7. Except for data obtained with the 300-item list, the fit appears to be satisfactory.

The expected-proportion model may also be applied to data gathered using what has been called a multiple-choice or forced-choice testing procedure. In this case, the test list is divided into subsets consisting of one old item and

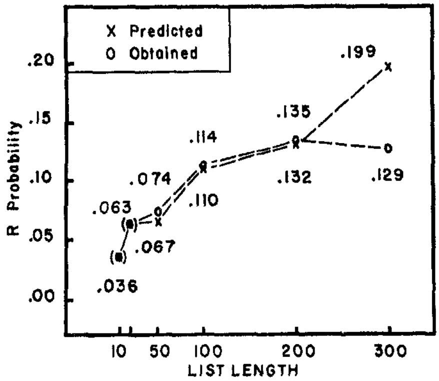

FIG. 7. Obtained and predicted values of $P_N(A)$ using the modified expected-proportion model and obtained $P_O(A)$ in each condition. (This experiment differed from Davis et al. in that (*a*) $P(O)$ and $P(N)$ were not varied; and (*b*) the subject did not know the value of $P(O)$—Strong, 1912.)

(n) new items. The instructions employed along with this procedure usually require the subject to choose no more nor less than one item out of each subset. This requirement ordinarily would place the experiment beyond the scope of the present theory. That is, if the subject is required to examine a subset *and* to pick one item *before proceeding* to the next set, he cannot use a cutoff at all since it may be that no item exceeds his cutoff, and yet he must respond.[9] On the other hand, the cutoff rule may be employed in a multiple-choice situation if the subject is free to move back and forth among the sublists. In this case, if no item exceeds the cutoff in one subset, the subject may go on to another subset and return later.

In order to apply the expected-proportion model to this sort of data, all of the assumptions of the expected-proportion model are made. That is, it is assumed that the subject is faced with two normal distributions at fixed distance on a familiarity continuum and that the cutoff is set according to the expected-proportion model. Furthermore, the additional assumptions are made that (*a*) in each subset the subject will choose the first item that he examines which exceeds the cutoff, and (*b*) that (K) = 1.[10] According to Assumption *a* the probability of a true positive (estimated by $p_O(A)$) will be given by:

$$p_O(A) \simeq \frac{\text{expected number of old items that exceed cutoff}}{\text{expected total number of items that exceed cutoff}} \qquad [6]$$

The expected number of old items that exceed the cutoff is equal simply to the probability that the single old item in each subset exceeds the cutoff—that is, it is equal to $P_O(A)$. The expected number of new items that exceed the cutoff is given by the probability that any particular new item exceeds the cutoff—that is, by $P_N(A)$—multiplied by the number of new items (n) in the subset. Therefore, Equation 6 may be rewritten as:

$$p_O(A) \simeq \frac{(1)\ (P_O(A))}{(1)\ (P_O(A)) + (n)(P_N(A))} \qquad [7]$$

[9] In this case the subject would probably simply pick the item with the greatest familiarity response. In support of this hypothesis, it can be shown that the raw (uncorrected) data reported by Flores (1958) may be fit by the equation $P_O(A) = P(X_O > X_N)^n$ where (n) is the number of new items in each subset, and $P(X_O > X_N)$ is the probability that one old item will exceed any one new item. This latter quantity is treated as a fitting parameter and is assumed to be constant if the conditions of learning, retention, and test (except n) are constant.

[10] For our purposes this assumption is equivalent to the assumption that the subject rejects some members of each subset (those that do not exceed the cutoff) and makes a random guess among the remaining items. Murdock (1963) has suggested and defended

Now if $(K) = 1$, Equation 3 given earlier may be rewritten as:

$$P_N(A) = \frac{1 - P_O(A)}{n}$$

Substituting this expression into Equation 7 and simplifying:

$$p_O(A) \simeq P_O(A) \qquad [8]$$

In other words, it follows from the assumptions, that the obtained proportion of true positives will be, as usual, an estimate of the probability that an old item will exceed the cutoff. Similarly, it can be shown that $p_N(A)$ estimates $P_N(A)$.[11] Thus, the model may be tested in exactly the same manner in which it was tested against the data of Davis et al. In this case, of course, (K) is not a fitting parameter but is assumed to equal one.

It should be noted that in testing the theory against multiple-choice data, only the fundamental assumptions of normality and a movable cutoff are being tested. That is, it is known a priori that the data will satisfy Equation 3 since the subject is forced by the instructions to choose a constant number of items (i.e., one per subset). On the other hand, although a variety of pairs of $p_O(A)$ and $p_N(A)$ values will satisfy this requirement, there is a unique pair which will also satisfy the requirement of representing corresponding areas under two normal distributions at a fixed distance.

a hypothesis which is very similar to this latter assumption.

[11] The score $p_N(A)$ is the proportion of new items that are called old. Therefore, it is equal to the proportion of subsets on which a new item is chosen divided by the number of new items in each subset.

The Postman Experiment

An experiment reported by Postman (1950) used this paradigm. He presented a list of 36 nonsense syl-

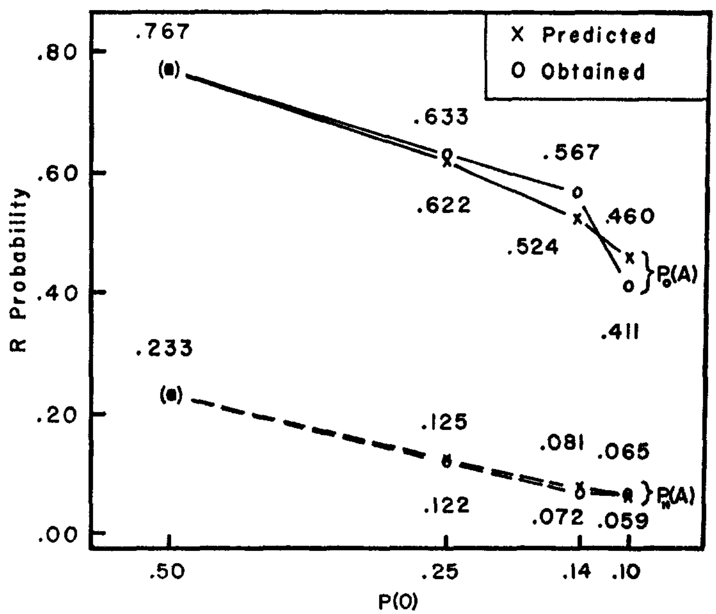

FIG. 8. Obtained and predicted data using a modified expected-proportion model; $(\bar{X}_O - \bar{X}_N) = 1.46$ (Postman, 1950).

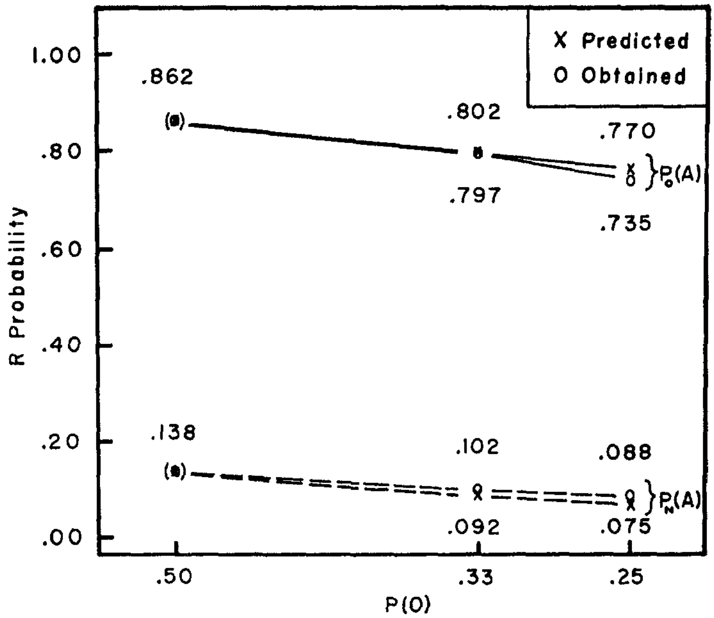

FIG. 9. Obtained and predicted data using a modified expected-proportion model; $(\bar{X}_O - \bar{X}_N) = 2.18$ (Murdock, 1963).

lables, the list being presented five times in five different orders. Following these learning presentations, a written multiple-choice recognition-memory test was administered. This test consisted of a series of 36 subsets, each subset containing one old item and either one, three, six, or nine new items. Thus, there are four recall conditions; $P(O) = .50, .25, .14$, and $.10$, respectively. These 36 multiple-choice subsets were presented on a single page, and the subject was asked to indicate which member of each list was an old item.

The data obtained under the $P(O) = .50$ condition are used to estimate $(\bar{X}_O - \bar{X}_N)$. This value along with $(K = 1)$ and level of $P(O)$ are used to predict the rest of the data in the manner described for the Davis et al. experiment. The obtained and predicted values of $p_O(A)$ are given in Figure 8. The fit appears to be satisfactory.

The Murdock Experiment

Murdock (1963, p. 18) reports a very similar experiment. Using general knowledge statements as stimuli, a list of 200 multiple-choice subsets was administered to 97 subjects. Of these 200 subsets, 67 contained one correct statement ("old" item) and one incorrect statement ("new" item), 67 contained one old and two new items, and 66 contained one old and three new items. Thus, $P(O) = .50$, $.33$, or $.25$. The instructions required the subject to select one item of each subset. The obtained values of $p_O(A)$ and $p_N(A)$ after Murdock's correction for guessing is removed are given in Figure 9 along with the predicted values. The predictions are made from the $P(O) = .50$ data (where (K) is assumed to equal 1). The fit appears to be satisfactory.

The studies reviewed in this section are especially important in distinguishing the movable-cutoff model underlying the present theory from the sophisticated fixed-threshold model proposed by Luce (1963) for psychophysical detection. Luce's model, when suitably translated for application to the recognition-memory situation, may be shown to be consistent with the data obtained by Davis et al. (1961), Parks (1964), and Strong (1912).[12] However, this fixed-threshold model is apparently incapable of explaining the precise effect of $P(O)$ on performance under the multiple-choice procedure of Postman (1950) and Murdock (1963). Although the data of these latter experiments are consistent with the very broad predictions yielded by Luce's "K alternative" model, there is no ready explanation of the fact that the obtained data lie in each case on a single ROC function.

[12] Rather than assuming that (t) and (u) are determined by an asymptotic learning process as described by Luce, it is assumed that these biasing factors result from the subject's attempt to match the expected proportion of old responses to $P(O)$. Thus, for example, (t) is set such that:

$$(t)\ (q(O))\ (P(O)) + (t)\ (q(N))\ (P(N)) = K\ (P(O)).$$

That is:

$$t = \frac{K}{q(O) + \dfrac{P(N)\,q(N)}{P(O)}}$$

where K is determined, in an unspecified manner, by the cost and values involved. It follows that:

$$p_O(A) + \frac{P(N)}{P(O)}\,(p_N(A)) = K$$

If, as in Strong's procedure, the subject employs estimates $P(O)'$ and $P(N)'$, then:

$$p_O(A) + \frac{P(N)'}{P(O)'}\,(p_N(A)) = K$$

These predictions are, as far as they go, identical to those made from the expected-proportion, movable-cutoff theory.

The two theories differ in that the sophisticated fixed-threshold theory predicts that, if the conditions of learning and retention are constant, the obtained values $p_O(A)$ and $p_N(A)$ will lie on an ROC curve composed of two intersecting straight-line segments. In practice such a function has been indistinguishable from the Gaussian ROC function of the movable-cutoff theory.

CONCLUDING COMMENTS

Of the independent variables which occur in the studies reviewed, only those concerned with test procedure are specifically handled by the theory. The other variables which have been mentioned (length of list, type of stimuli, and procedure) are indifferent with respect to the theory. That is, they may or may not be effective variables of $(\bar{X}_O - \bar{X}_N)$ and (K). In this sense, the present theory is concerned with recognition-memory performance rather than learning and retention per se. Nevertheless, in describing the end product of learning and retention, the theory has implications for any description of that learning upon which recognition responses are based. On a broader scale, the great similarity of the present theoretical assumptions to those which have been found to be adequate in psychophysical situations suggests that the essential features of this theory might, in time, be shown to be only a special application of a very broad theory of decision behavior.

REFERENCES

ARCHER, E. J. Re-evaluation of the meaningfulness of all possible CVC trigrams. *Psychological Monographs*, 1960, **17**(10, Whole No. 497).

DAVIS, R., SUTHERLAND, N. S., & JUDD, B. R. Information content in recognition and recall. *Journal of Experimental Psychology*, 1961, **61**, 422–429.

EGAN, J. P. Recognition memory and the operating characteristic. Technical Report No. AFCRC TN-58-51, AD-152650, 1958, Hearing and Communication Laboratory, Indiana University.

FLORES, C. Etude sur les relations entre le rappel et la reconnaissance. *L'Annee Psychologie*, 1958, **58**, 365–375.

GRANT, D. A. Testing the null hypothesis and the strategy and tactics of investigating theoretical models. *Psychological Review*, 1962, **69**, 54–61.

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

MURDOCK, B. B. An analysis of the recognition process. In C. N. Cofer & B. S. Musgrave (Eds.), *Verbal behavior and learning.* New York: McGraw-Hill, 1963. Pp. 10–21.

PARKS, T. E. *An experimental and theoretical analysis of recognition memory performance.* (Doctoral dissertation, University of Wisconsin) Ann Arbor, Mich.: University Microfilms, 1964, No. 64–12741.

POSTMAN, L. Choice behavior and the process of recognition. *American Journal of Psychology*, 1950, **63**, 576–583.

STRONG, E. K. The effect of length of series upon recognition memory. *Psychological Review*, 1912, **19**, 447–462.

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

(Received September 16, 1964)