Psychological Bulletin
1970, Vol. 74, No. 2, 100–109

MEMORY AND THE THEORY OF SIGNAL DETECTION[1]

ROBERT S. LOCKHART AND BENNET B. MURDOCK, JR.[2]

University of Toronto

The use of signal detection theory as a model of the decision process in memory is discussed in terms of the problems of identification and testing. Identification refers to the matching of elements of the formal theory with constructs constituting a theory of memory. A number of issues are raised concerning the identification of the sensory continuum, the decision axis, and the distinction between signal and noise. The use of detection theory is not a theoretically neutral method of dealing with response bias; the validity of its use and the manner of its application depend heavily on certain assumptions about the memory system. The form of the operating characteristic can be used to test a number of these assumptions but such tests are relatively insensitive. The use of detection theory in the analysis of cued-recall data poses additional problems and is meaningful in the context of some, but not all, theories of recall.

A recurrent problem in the study of recognition memory has been that of combining hits and false alarms (or correct and incorrect responses) into a single index of performance. That the proportion of correct responses in a yes/no or multiple-choice recognition test may in some way be contaminated by factors other than the state of the memory system has never been seriously disputed, but only recently has the question received detailed and systematic attention. Traditionally the problem has been viewed in terms of "correcting for chance success"; in current terminology, and stated more generally, it is a problem of providing an adequate theory of the decision system which maps a given state of the memory system into an overt response.

The reasons for this increased attention to decision processes in memory are not difficult to trace. In the first place, the development of quantitative theories of recognition memory has necessitated a precise and explicit account of how the memory and decision systems interact to produce a given response. If a theory of memory is to be tested against data from a recognition-memory experiment, it is necessary to obtain performance measures that are independent of those parameters which, while they affect performance, do so only through the decision process. A second major influence has undoubtedly been the developments in psychophysics and choice behavior which have provided an extensive conceptual framework for dealing with decision processes, and it has not proved difficult to translate this framework into the area of memory research.

It is important to recognize at the outset that an adequate account of the decision process for a given test of memory performance cannot be made without a commitment to certain aspects of a theory of the memory system for that situation. This is so because, in order to state the decision rule, it is necessary to specify the nature of the mnemonic information on which the decision rule is to operate. The early methods of correcting for chance success were based on the assumption that the decision system operated on two or three possible values from the memory system, the states being characterized in terms of threshold or all-or-none principles. More recently, the theory of signal detection has been used as a basis for a decision model in recognition-memory performance. Its essential point of departure is to formalize the input to the decision system as the value of a random variable defined on a sensory continuum, rather than as one of a finite number of states, and to assume that the decision system operates on this input with a deterministic rather than a probabilistic decision rule.

[1] This work was supported by research grants APA 355 and 146 from the National Research Council and grant OMHF 164 from the Ontario Mental Health Foundation.

[2] Requests for reprints should be sent to Robert S. Lockhart, Department of Psychology, University of Toronto, Toronto 181, Canada.

The purpose of the present paper is to discuss what appear to be the more important problems associated with formalizing the decision process in memory within the framework of signal detection theory. The discussion is confined largely to recognition memory, although in a later section the question of applying detection theory to the analysis of cued recall is specifically considered.

IDENTIFICATION

Psychological applications of detection theory have evolved largely in the context of auditory and visual psychophysics, and the initial problem is therefore one of identifying the elements of the formal theory with constructs constituting a theory of memory, rather than of audition or vision. This problem may be termed one of identification. Three aspects are considered: the nature of the sensory (memory) continuum, the distinction between signal and noise, and the form of the decision axis.

The Sensory Continuum

As stated, an essential feature of the theory of signal detection is that the decision rule is applied to the value of a continuous variable. The most straightforward identification of this variable within a memory system is the continuous strength theory of Wickelgren and Norman (1966). The identification is straightforward in the sense that the sensory continuum is identified directly with a continuum of memory states (strength) and processes such as acquisition, forgetting, generalization, etc., serve simply to increase or decrease an item's value on the strength continuum. This assumption about how the memory system is to be represented is a very strong one; but it does have the advantage that given the assumption, little else is required by way of a detailed theory of the memory system in order for it to provide a suitable basis for incorporating a signal detection model of the decision process. It also serves to maintain a close analogy with traditional psychophysical detection theory and no doubt is the main reason why most "SDT analyses" of recognition-memory data have applied the theory in a way that is formally equivalent to strength theory.

A less direct solution to the identification problem is found in theories postulating a finite state representation of the memory trace which treat the decision process within the signal detection framework by associating with each state a likelihood distribution defined on a continuous variable. Thus Bernbach's (1967) two-state model associated with each state (forgotten and remembered) a likelihood distribution defined on a scale of "apparent oldness," while Kintsch (1967) identified the three states of his Markov recognition-learning model with regions on a familiarity continuum.

While it is not the purpose of the present paper to consider the relative merits of these approaches as theories of recognition memory, it is important to stress the fact that they represent quite different solutions to the identification problem, and that the difference between them is not merely a question of the name used to label the continuum; the parameters of each model differ not only numerically but are subject to quite different interpretations. It is a mistake to suppose that there is just one, theoretically neutral, way in which signal detection theory can be applied to recognition-memory data.

The preceding comments represent one way to consider the identification problem for the sensory continuum. Yet there is a sense in which the distinction between detection theories and finite state theories is not quite so sharp. All theorists seem to agree that the decision process operates on the value of a continuous random variable, and in general a posteriori probabilities seem graded as they should. In signal detection theory there are two states; the signal is either present or absent, so both detection theory and a finite state model assume two possible "states of the world." The difference though, would be that detection theory (or strength theory) would permit different values of d' to be associated with different levels of signal strength. A finite state model would be restricted to differences in state probabilities, but could not tolerate different sensitivity values. Thus, the main issue between finite state and detection models is not whether the underlying distributions are discrete or continuous, but whether or not the parameters of the model change appropriately with specified experimental manipulations.

Signal and Noise

The problems associated with the distinction between signal and noise are best discussed in terms of the yes/no forced-choice task. In the case of auditory detection, these terms often have a quite literal meaning but in fact they can be used to denote any two levels of intensity on the sensory continuum. The designation of one as signal and the other as noise might be quite arbitrary, as is the case in Creelman's (1965) study of discrimination of linear extent.

In the case of the strength interpretation of recognition memory, the noise distribution (N) is identified with the distribution of strength values associated with new items, while the signal-plus-noise distribution (SN) is the distribution of these values for old items. A number of differences between the memory task and the usual psychophysical task are immediately apparent. The most obvious is the fact that in recognition memory a SN trial does not consist of two physically distinct components—signal added to noise—and one might be tempted to ask in what sense the noise of an SN trial is the same as the noise of an N trial. Clearly the identity cannot be specified in objective physical terms. Given the discussion of the previous section, this question becomes rather pointless. The validity of identifying the presentation of an old item as an SN trial depends not on the possibility of specifying the independent S and N components in physical terms but on the validity of certain assumptions about the memory system. Briefly, these assumptions are that all items (old and new) can be represented as having some value on the memory continuum and that the effects of acquisition and forgetting can be represented by increments in this value.

While the lack of any objective specification of the noise level poses no great conceptual problem, it does increase the difficulty of ensuring a *constant* noise level across conditions for which comparisons of d' are to be made. Since d' provides only a difference (discriminability) measure, meaningful comparisons will often necessitate a common base line. If the base line is clearly different across conditions, then care needs to be taken in making subsequent interpretations. A common example of this latter situation is the study of recognition memory as a function of item properties such as frequency, concreteness, etc. Suppose one were to compare the value of d' obtained from a recognition test in which both old and new items were low-frequency words with that obtained from a list in which both were high-frequency words. If a larger d' were obtained for the low-frequency condition, it would clearly be wrong to interpret this as indicating that low-frequency words had, on the average, greater values on the strength continuum. Rather, the greater d' indicates that the *difference* in strength between old and new low-frequency words is greater than the difference between old and new high-frequency words. This is an obvious point and is simply a restatement of the well-known fact that any measure of recognition memory depends on the kind of material used as lures.

One possible virtue of the detection theory analysis applied, say, in the context of strength theory is that it provides a very simple method of scaling the potency of different sets of lures that might vary, for example, in the degree of their formal or semantic similarity with the old items. Conceptually this could be seen as keeping the SN distribution constant and systematically varying the location of the N distribution. One possible complication in this kind of analysis would be the existence of generalization or diffusion processes. If the strength of an old item is incremented by the presentation of a similar new item, then the location of the SN distribution would no longer be independent of the nature of the lures. Again, there would seem to be some danger in this kind of analysis if one is not working within the context of an explicit theory of the memory system.

The diffusion or generalization of the memory trace would cause difficulties in other situations as well. It might result in the location of the N distribution varying with list length or, in experimental designs such as the Shepard and Teghtsoonian (1961) continuous-memory task, it could vary over blocks of trials. The usual prediction would be that its location on the memory continuum should increase as the amount of material being stored increases. If the subject maintains a constant cutoff for his yes/no decision on the memory continuum, then an increase in the false-alarm rate would result. This in fact is what Donaldson and Murdock (1968) found in their study using the Shepard and Teghtsoonian procedure. Wickelgren and Norman (1966) found that the false-alarm rate increased with list length. Of course other interpretations are possible—indeed more probable—but it is nonetheless important to recognize the possibility that a change in experimental conditions might affect not only the location of the SN distribution, but also that of the N distribution.

Thus far, only questions related to the *location* of the N distribution have been considered. An equally important question concerns its variance and, in particular, the source of the variance. Traditionally, detection theory has assumed that noise variability is involved in the mapping of the physical stimulus into the sensory continuum so that a stimulus of given intensity corresponds to a distribution of sensory values. Wickelgren and Norman (1966), on the other hand, assumed that the random component is due entirely to variability in the decision process, an assumption dictated as much by convenience as by evidence. The assumption is convenient since it makes more plausible the further assumption that the variance is independent of location on the continuum; in particular, that the N and SN distributions will have equal variances. A further advantage is that it avoids adding stochastic components to acquisition or retention functions. Of course it seems unreasonable to suppose that subjects can maintain decision criteria with sufficient consistency (especially in a rating task) to make criterion variance negligible, but it seems even less reasonable to suppose that strengths operate in a strictly algebraic fashion. Wickelgren (1968a) considered the general case in which both criterion placement and strength are random variables, while Tennent (1968) used computer simulation to investigate the effects of nonconstant criterion variance.

The Decision Rule and the Criterion

In signal detection theory a decision rule involves the partitioning of a *decision axis*, and associating with each partition one and only one response. Traditionally this axis has been taken to be the likelihood ratio $\lambda(x)$ rather than the sensory continuum x. In the case of yes/no detection, the subject is assumed to set a cutoff value β, and respond "yes" if $\lambda(x) > \beta$ and "no" if $\lambda(x) < \beta$. In the case of an n-point rating-scale task, it is assumed that multiple criteria $\beta_1, \beta_2, \ldots, \beta_{n-1}$ are set, which partition the decision axis into n regions corresponding to the n points of the rating scale.

There has not been a great deal of attention paid to the problem of identifying the decision axis in the case of memory. An exception is the criterion cutoff rule of Wickelgren and Norman (1966). They identified the decision axis with the strength continuum itself, so that criteria were set in terms of actual strength values rather than values of likelihood ratio. Parks (1966) also discarded the likelihood-ratio identification in favor of a cutoff point set on a familiarity continuum.

If the memory and likelihood-ratio continua are strictly monotonically related, then from some points of view it is unimportant whether the decision axis is identified with one or the other; the decision made will be the same whether the criterion is thought of as being set at a point, β, on the likelihood-ratio scale, or a corresponding point on the memory continuum.

From other points of view, however, there are important differences, even when the two continua are monotonically related. Questions about systematic shifts in criteria resulting from changes only in signal-to-noise ratio will receive different answers depending upon whether the decision axis is identified as the sensory or the likelihood-ratio continuum. As an example, consider the hypothetical experiment in which a subject is run for two separate sessions in a yes/no test of recognition memory. Within each session he is tested at only one retention interval, but this interval differs substantially between sessions. The situation may therefore be represented by an

N distribution whose location on the memory continuum is constant across sessions and SN distributions with different locations depending on the session. Assume the standard detection theory model of normal distributions with equal variances and assume that payoffs, a priori probabilities, etc., are constant throughout. If the false-alarm rate remains constant across sessions, it might be argued that the cutoff is being set as a fixed value on the strength continuum. But it could also be argued that a likelihood ratio is being judged and a more lax criterion is set for the short retention interval condition. Similarly, a change in the false-alarm rate such that a constant cutoff on the likelihood continuum is observed could be interpreted as a shift in x_c criterion. Under these circumstances questions about criterion changes can be answered only relative to a particular identification of the decision axis.

For obvious reasons the use of a single retention interval for each session is not a very satisfactory way to design a short-term memory experiment, so there is an important difference between memory experiments and the usual auditory detection experiment. In the latter, observers are typically tested at only one signal-to-noise ratio in any one session or block of trials. In memory experiments, on the other hand, various retention intervals are tested, usually in a random order, within a single session. This situation is represented by a single N distribution (and thus a single false-alarm rate) and as many SN distributions as there are retention intervals. From the point of view of the subject, then, an item may represent "noise only" or a signal from any one of a number of distributions. Under these circumstances, a number of difficulties are raised if the decision axis is identified as the likelihood-ratio continuum. Clearly it makes no sense to estimate a β criterion for each retention interval. Since a constant false-alarm rate is built into the design of the experiment, it would have to be said that a different criterion is set for each retention interval, suggesting that a subject must know the retention interval before he knows where to place his criterion.

A possible way around this difficulty would be to suppose that all SN distributions are treated as a single likelihood function and that the likelihood-ratio criterion is set in terms of this average distribution. This solution raises further difficulties of its own. If each SN distribution is taken to be normal with variance equal to that of the N distribution, then the average distribution formed by sampling randomly from each of the SN distributions will not be normal (in general, it will not even be unimodal) and its variance will certainly be greater than that of the N distribution. The immediate difficulty then is that the likelihood-ratio continuum is no longer monotonically related to the strength continuum.

Taking these considerations into account it is difficult to see how, in the typical memory experiment, a subject could respond in a fashion consistent with the assumption that he is judging a likelihood ratio. It seems more reasonable, therefore, that in applications of detection theory to memory the decision axis should be identified as the memory continuum itself. While this definition avoids the difficulties inherent in a nonmonotonic relationship between the sensory and likelihood-ratio continuum, it raises a new problem concerning the assumption of a *unidimensional* memory continuum. Swets, Tanner, and Birdsall (1961) pointed out that a multidimensional sensory continuum offered no difficulty for detection theory provided the *decision axis* was unidimensional. Since likelihood ratios, even if from a multivariate space, can always be represented on a line, the assumption that the decision axis is likelihood ratio guarantees that the unidimensionality assumption is satisfied. If the decision axis is the memory continuum, however, it becomes necessary to assume that strength itself can be represented unidimensionally. While this point could be argued at length, it does seem to be the case that in the typical recognition-memory task, subjects respond in a fashion consistent with this assumption. Figure 1 gives a typical plot of a posteriori probabilities (the probability of an item being old, given a particular confidence rating) against confidence level. The monotonically decreasing function is what would be expected on the assumption that subjects are ordering their judgments on a single axis.

It is possible to argue that at some higher level of abstraction the memory trace is more appropriately represented in a multidimensional space, and that confidence ratings on a single axis correspond to some weighted sum of the item's values on the various dimensions. Justifying the unidimensional assumption on these grounds would seem no less reasonable than the assumption that the subject can operate consistently in a multivariate likelihood space.

Testing Assumptions

The Continuity Assumption

The most common method of testing the adequacy of the signal detection model is to examine the form of the operating characteristic (e.g., Green & Moses, 1966). Since the predicted form of the operating characteristic depends on the particular assumptions made about both the memory and the decision systems, departures from expectation may reflect incorrect assumptions about either or both systems. For example, in testing the assumption that the input to the decision system is the value of a continuous variable against the alternative that it is one of two possible states,[3] it has been commonly assumed that a two-state theory predicts an operating characteristic consisting of two straight-line segments, while detection theory predicts a smooth curve. But as Broadbent (1966) and Wickelgren (1968b) pointed out, the former prediction holds only when certain additional constraints are placed on the decision rules. Unless, therefore, one has independent evidence about the decision rule being used, an operating characteristic which is a smooth curve rather than two straight-line segments does not constitute sufficient evidence to reject the two-state assumption. A detailed analysis of the difficulties of testing these assumptions by the use of operating characteristics has been given by Krantz (1969) and Wickelgren (1968b) and is not repeated here.

[3] This is not necessarily equivalent to testing the assumption that the form of the memory trace is continuous or discrete. It is possible to have a finite state representation of memory but continuous input to the decision system as in Bernbach's (1967) model.

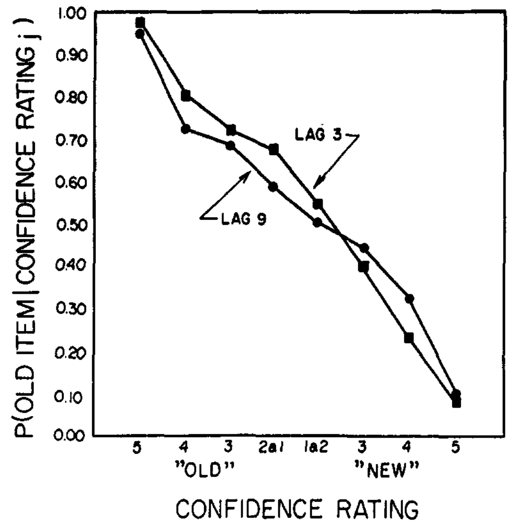

FIG. 1. A posteriori probabilities for an individual subject in a continuous-recognition task. (The two curves correspond to different retention intervals; data from Donaldson, 1968.)

It is sufficient to note that routine examination of the operating characteristic is, in general, inconclusive, especially if generated by the use of confidence ratings.

For similar reasons a smooth monotonic relationship between a posteriori probabilities and confidence ratings (such as in Figure 1) while predicted by signal detection theory, does not disprove the finite state assumption. On the other hand, Krantz (1969) derived a possible method for rejecting general threshold theories in terms of a lower bound on the ratio of the posterior odds of a signal given some response (e.g., confidence rating) to the prior odds of the signal. However, as Krantz pointed out, failure to reject threshold theories on these grounds does not provide evidence for the acceptance of threshold theories rather than signal detection theory.

Distribution Assumptions

The usual (though not essential) assumption of SDT is that the N and SN distributions are normal with equal variance. Again, these assumptions can be tested in terms of the operating characteristic and large sample likelihood-ratio tests of both assumptions are now available for rating and for yes/no designs (Dorfman & Alf, 1968, 1969; Ogilvie &

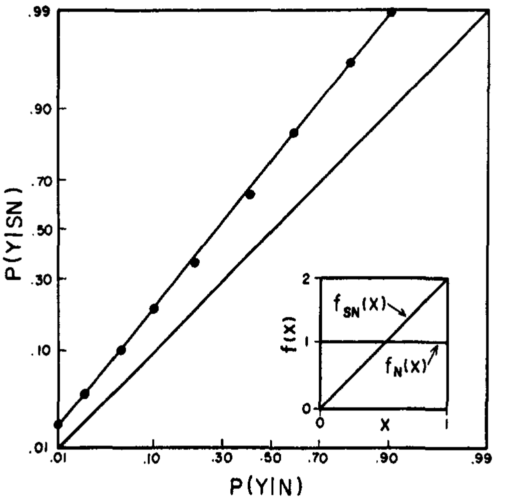

FIG. 2. Operating characteristic plotted on double normal-probability paper for the density functions depicted in the lower right corner.

Creelman, 1968). The normal distribution assumption implies a linear operating characteristic when plotted on double-normal probability paper and the equal-variance assumption implies unit slope. However, virtually any unimodal distribution will produce curves that are indistinguishable from straight lines in normal probability paper. Such a finding is perhaps not unexpected, with distributions such as the gamma distribution and the Poisson, since in the limit both approach the normal distribution. However, if a ramp and rectangular distribution are taken as SN and N, respectively, a reasonable linear operating characteristic can still result. In Figure 2, the probability-density functions $f_{SN}(x)$ and $f_N(x)$ are depicted in the lower right corner, and the resulting operating characteristic (defined as in Green & Swets, 1966, p. 35) is shown as plotted on double-normal probability coordinates. It would seem that linear operating characteristics do not provide a sensitive test of the normality assumption against even unreasonable alternatives. While this fact no doubt introduces a certain desirable robustness into the statistics of the situation, it does have the disadvantage of limiting the use of operating characteristics in distinguishing between theories that differ in terms of their predictions about the form of the underlying distributions.

On the other hand, alternatives to the normal distribution assumption will usually predict nonunit slopes that may provide a more powerful test. For example, if noise is assumed to result from a Poisson process, then the SN distribution will necessarily have a variance greater than that of the N distribution, and the same is true for any distribution in which the variance is a direct function of the mean (e.g., the exponential and gamma distributions). However, with such distributions the relationship between variance and slope on double-normal paper is *not* always the same as with normal distributions. Also, in such cases the largest discrepancies may be predicted when discrimination is high (the slope decreases from unity as the operating characteristic moves away from the chance diagonal) and it is in these regions that estimation is least reliable.

If operating characteristics have been estimated on the basis of the normal distribution assumption, and nonunit slopes result, there are several courses of action that may be taken. The first possibility that should be considered is that the departure from unit slope is the consequence of extraneous variance being introduced through inappropriate averaging of data. Averaging confidence ratings before estimating the operating characteristic can very easily produce slopes of less than unity because the effect of averaging is usually to inflate the variance of the SN distribution. This is especially true if one averages the confidence ratings across different retention intervals, since one may be adding systematic (strength) variance to the SN but not to the N distribution. This averaging may, at least in part, account for why slopes of less than unity are often obtained when recognition memory is tested using a study-test procedure (e.g., Egan, 1958). Similar (though, in general, less predictable) consequences can result from averaging confidence ratings across subjects. This form of averaging has two difficulties: individual differences in criterion location and in sensitivity. It is difficult to specify the precise effect of the former either on the estimate of d' or on the slope. What can be said is that if the assumption of normal distributions and equal variances is true for individual subjects, it will not in general be true for the average rating data. The effect of differences across subjects in d' will be the same as that resulting from averaging across different retention intervals for a single subject; namely, to inflate the SN variance. In general the use of group data should be avoided, or if average data are to be reported then d' and slope should be estimated for individual subjects and these estimates averaged.

The direct consequence of nonunit slope is that two parameters rather than one are required to specify the operating characteristic. One way of reporting such results is, therefore, to specify both the difference between the means of the two distributions (with the standard deviation of the N distribution as unit) and the slope (Green & Swets, 1966). Alternatively, one can ignore certain information and use a single index such as d_e' which gives equal weight to the standard deviations of the N and SN distributions. A further alternative is to use the nonparametric area analysis of Pollack and Norman (1964).

A more radical solution to the problem is to discard the normal distribution assumption in favor of distributions such as the Poisson or exponential, which predict particular patterns of nonunit slope relative to the sensitivity index and which avoid the introduction of an extra parameter. This solution is hardly a satisfactory strategy if employed on a post hoc basis. However, if a given task produces a consistent pattern of nonunit slopes, one might consider reformulating the theory of the memory or decision system such that distributions other than the normal might be expected. For example, Murdock (1970) suggested several different mechanisms that might underlie a subject's confidence rating in a cued-recall task, all of which predict nonnormal distributions.

Applications to Cued Recall

Operating characteristics can be constructed for a cued- or probe-recall experiment by forcing the subject to make some response for every probe and obtaining ratings on a scale ranging from "certain right" to "certain wrong" (Murdock, 1966). The curve is then a plot of *prob* (correct | rating $\geq j$) against *prob* (incorrect | rating $\geq j$).

Such curves have been termed Type 2 or response-conditional operating characteristics, and their essential difference from the usual Type 1 curves is that the "stimulus" to be judged is a subject-generated response. The a priori probabilities are not under the experimenter's control but are a function of the recall level. Interpreting the results of such an analysis raises a number of difficulties both in terms of the d' value and the nature of the criterion. In a general qualitative sense, the d' indicates the subject's ability to discriminate correct from incorrect responses; a large value of d' reflects a situation in which "certain correct" ratings tend to be assigned to all (and only) correct responses and "certain wrong" ratings are assigned to responses that are in fact wrong. It can be seen that there is no obvious reason why there should be any relationship between the value of d' and the probability of correct recall. Furthermore, there is no reason to suppose that an experimenter should be able to increase the number of items correctly recalled simply by instructing subjects to use a more lax criterion. Such a change in the payoff matrix could affect the confidence ratings given, but not the recall probabilities themselves.

While the d' value may be interpreted in this qualitative way it can be given no quantitatively precise meaning without an explicit theory of the recall process itself. Even with such a theory, no simple interpretation may be possible. The latter is true for continuous strength theory which provides such a straightforward analysis in the case of recognition. In particular, it is a gross oversimplification to say that the Type 2 d' is a measure of the strength of the correct association. It might be said to indicate the difference in strength between the correct and incorrect associations, but this formulation adds little to the qualitative interpretation given earlier. If the number of response alternatives is small and known to the subject then it is possible to handle the situation as one would an m-alternative–forced-choice recognition test. But for the more usual (and more interesting) case—in which there is no guarantee that the correct response will be accessed—it is impossible, within the context of existing strength theory, to give any precise interpretation of the value of d' obtained from Type 2 analyses.

The interpretation of d' from the point of view of a finite state theory of recall is rather more straightforward. The reason is because, in such theories, the processes determining recall probabilities are distinct from those underlying discrimination. In the two-state theories such as Bernbach (1967, 1969) and Murdock (1970) the analysis is particularly simple. If the set of response alternatives is large, then correct and incorrect responses can be identified directly with the remembered and forgotten states, respectively. The d' can be interpreted as the subject's ability to discriminate from which of the two states the response was recalled, and the only remaining problem is to identify the continuum on which this discrimination is made. This latter question has been considered in some detail by Murdock (1970). Whether the same memory system is used in recall and recognition and the interpretation of confidence judgments and response latencies has recently been considered by Norman and Wickelgren (1969).

Care must also be taken in interpreting criterion changes in relation to Type 2 analyses. In this connection it is important to distinguish between a forced-response procedure (which is necessary to generate the isomnemonic curve) and the more usual one in which omissions are allowed. In this latter case, the subject must set some criterion for overt responding, a criterion such that if a potential response falls below a given confidence level, it will not be given. Since the resulting omissions may contain "misses" as well as correct identifications of "noise only," a shift in this criterion may result in changes in recall probabilities. Forcing a response to every probe effectively eliminates this source of variance. The criterion problem in Type 2 analyses concerns the amount of evidence the subject requires about the correctness of the response before he will assign a confidence rating at a given level; there is no sense in which a Type 2 d' "corrects" recall probabilities for response biases.

Summary

By way of summary, the main points in the paper are recapitulated. To test theories of recognition memory by experimental data, it is necessary to obtain performance measures which separate memory and decision processes. Signal detection theory represents the input to the decision stage as a continuous random variable defined on a sensory continuum. Two major problems are (*a*) to identify the constructs in signal detection theory with constructs of a theory of memory, and (*b*) to test the concepts of signal detection theory as applied to a theory of memory.

Three major problems of identification are (*a*) the nature of the sensory (memory) continuum, (*b*) the distinction between signal and noise, and (*c*) the form of the decision axis. Strength theories and finite state theories both suggest that the decision process operates on the value of a continuous random variable, but differ in their identification of the underlying memory continuum, and in the amount of variation possible for the sensitivity parameter (d'). The identification of signal and noise in the memory task depends on assumptions about the memory system, and it would be unwise to uncritically assume constancy in the noise distribution. The decision axis may be the memory dimension itself or a likelihood-ratio transformation, but if it is the latter it is hard to see how the subject could respond consistently in the typical memory task with different retention intervals.

To test assumptions, the finding of an operating characteristic that is a smooth curve, rather than two discontinuous straight-line segments, does not necessarily constitute negative evidence for a finite state model. Also, distribution assumptions in general cannot be tested by the criterion of linearity on double-normal probability paper, since many different underlying distributions will yield essentially linear plots. Nonunit slopes may be indicative of averaging data, but also may be predicted when underlying distributions other than the normal are assumed.

In Type 2 or response-conditional procedures, d' is a measure of discriminability and not, necessarily, of memory. A Type 2 procedure is not recommended for a continuous strength model, but may be quite reasonable for certain finite state models. In a forced-responding procedure, the criterion in a Type 2 analysis reflects the amount of evidence required for a given confidence judgment; it does not "correct" recall probabilities for response bias.

REFERENCES

Bernbach, H. A. Decision processes in memory. *Psychological Review,* 1967, **74**, 462–480.

Bernbach, H. A. Replication processes in human memory and learning. In G. H. Bower & J. T. Spence (Eds.), *The psychology of learning and motivation: advances in research and theory.* Vol. III. New York: Academic Press, 1969.

Broadbent, D. E. Two-state threshold model and rating-scale experiments. *Journal of the Acoustical Society of America,* 1966, **40**, 244–245.

Creelman, C. D. Discriminability and scaling of linear extent. *Journal of Experimental Psychology,* 1965, **70**, 192–200.

Donaldson, W. An examination of false-positives in short-term recognition memory. Unpublished doctoral dissertation, University of Toronto, 1967.

Donaldson, W., & Murdock, B. B. Jr. Criterion change in continuous recognition memory. *Journal of Experimental Psychology,* 1968, **76**, 325–330.

Dorfman, D. D., & Alf, E., Jr. Maximum likelihood estimation of parameters of signal-detection theory—a direct solution. *Psychometrika,* 1968, **33**, 117–124.

Dorfman, D. D., & Alf, E., Jr. Maximum-likelihood estimation of parameters of signal-detection theory—rating method data. *Journal of Mathematical Psychology,* 1969, **6**, 487–496.

Egan, J. P. *Recognition memory and the operating characteristic.* (Tech. Rep. No. AFCRC-TN-58-51, AD-152650) Indiana University, Hearing and Communication Laboratory, 1958.

Green, D. M., & Moses, F. L. On the equivalence of two recognition measures of short-term memory. *Psychological Bulletin,* 1966, **66**, 228–234.

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

Kintsch, W. Memory and decision aspects of recognition learning. *Psychological Review,* 1967, **74**, 496–504.

Krantz, D. H. Threshold theories of signal detection. *Psychological Review,* 1969, **76**, 308–324.

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

Murdock, B. B., Jr. Short-term memory for associations. In D. A. Norman (Ed.), *Models of human memory.* New York: Academic Press, 1970, in press.

Norman, D. A. A comparison of data obtained with different false-alarm rates. *Psychological Review,* 1964, **71**, 243–246.

Norman, D. A., & Wickelgren, W. A. Strength theory of decision rules and latency in retrieval from short-term memory. *Journal of Mathematical Psychology,* 1969, **6**, 192–208.

Ogilvie, J. C., & Creelman, C. D. Maximum likelihood estimation of receiver operating characteristic curve parameters. *Journal of Mathematical Psychology,* 1968, **5**, 377–391.

Parks, T. E. Signal detectability theory of recognition-memory performance. *Psychological Review,* 1966, **73**, 44–58.

Pollack, I., & Norman, D. A. A non-parametric analysis of recognition experiments. *Psychonomic Science,* 1964, **1**, 125–126.

Shepard, R. N., & Teghtsoonian, M. Retention of information under conditions approaching a steady state. *Journal of Experimental Psychology,* 1961, **62**, 302–309.

Swets, J. A., Tanner, W. P., Jr., & Birdsall, T. G. Decision processes in perception. *Psychological Review,* 1961, **68**, 301–340.

Tennent, R. D. The effects of Gaussian criterion noise in signal detection theory. Unpublished master's thesis, University of Toronto, 1968.

Wickelgren, W. A. Unidimensional strength theory and component analysis of noise in absolute and comparative judgments. *Journal of Mathematical Psychology,* 1968, **5**, 102–122. (a)

Wickelgren, W. A. Testing two-state theories with operating characteristics and a posteriori probabilities. *Psychological Bulletin,* 1968, **69**, 126–131. (b)

Wickelgren, W. A., & Norman, D. A. Strength models and serial position in short-term recognition memory. *Journal of Mathematical Psychology,* 1966, **2**, 316–347.

(Received July 7, 1969)