Psychological Bulletin
1974, Vol. 81, No. 12, 945–958

SIGNAL DETECTION THEORY: CONSIDERATIONS FOR GENERAL APPLICATION

R. E. PASTORE [1] AND C. J. SCHEIRER

State University of New York at Binghamton

While there exist a number of papers describing the theory of signal detection, it appears that many psychologists are not aware of the ease with which signal detection theory can be applied, the range of applications possible, or the limitations of signal detection theory. This paper briefly summarizes the assumptions of signal detection theory and describes the procedures, the limitations, and practical considerations relevant to its application. A worked example of an application of signal detection theory to the study of cognitive processes is included.

In recent years, researchers in many diverse areas of psychology have begun to employ the theory of signal detection to separate the ability of the subject to differentiate between classes of events from motivational effects or response biases. In addition to its extensive application in sensory psychophysics, signal detection theory has found application in such diverse areas as speech perception (Egan & Clarke, 1956), memory (Banks, 1970; Bernbach, 1967; Parks, 1966), animal learning (Rilling & McDiarmid, 1965; Suboski, 1967), audiology (Campbell & Moulin, 1968), attention (Moray, 1970; Sorkin, Pastore, & Pohlmann, 1972), clinical psychology (Sutton, 1972), and sensory-evoked potentials (Hillyard, Squires, Bauer, & Lindsay, 1971). The purpose of this article is to review and briefly summarize the more common models [2] of signal detection theory, describe the procedures required to apply each model, and discuss the limitations inherent in each. While there are a number of excellent theoretical papers describing signal detection theory and its various models (e.g., Egan & Clarke, 1966; Green & Swets, 1966; Licklider, 1959; Peterson, Birdsall, & Fox, 1954; Swets, Tanner, & Birdsall, 1961; Van Meter & Middleton, 1954), there is a clear need for a concise, unified explanation of how and when to use the various models of signal detection theory. This article attempts to fulfill that need. The first section of this article presents a brief summary of the models of signal detection theory on a general level. The second section presents practical considerations for the application of signal detection theory and the specific procedures used in these applications. The third section outlines the potential use of signal detection theory in several experimental situations and presents a worked example of an application to the study of cognitive processes.

OVERVIEW

The purpose of this section is to provide a brief overview and summary of the theoretical underpinnings of signal detection theory. For a more complete introduction and theoretical presentation, the reader should refer to the articles by Egan and Clarke (1966), by Swets et al. (1961), or others. Signal detection theory is an adaptation of statistical decision theory (e.g., Wald, 1950). A major aspect of both signal detection theory and statistical decision theory concerns the specification of a set of ideal processes or observers as a standard against which a subject's performance is compared. While this comparison is an important aspect of signal detection theory, the specification of an ideal observer depends on the exact area or modality under

[1] The authors wish to thank Crawford Clark, Don Ronken, John Swets, Douglas Creelman, William Lutz, and Charlotte MacLatchy for their helpful criticisms of earlier drafts of this paper.

Requests for reprints may be sent to either author, Department of Psychology, State University of New York, Binghamton, New York 13901.

[2] The term *model* is used in this article to connote a special case of the theory based on certain clearly defined assumptions and, therefore, is limited in scope.

study and the assumed capabilities of the observer. Therefore, this aspect of signal detection theory is not considered in this article; for a general discussion of the use of ideal observers, see Tanner (1961) or Tanner and Sorkin (1972).

General Case

Signal detection theory is applicable to those situations in which two classes of events are to be discriminated. It also can be generalized to situations involving more than two classes of events (Tanner, 1956; but also see Luce, 1963), although this generalization is not discussed in this article. The basic assumption of signal detection theory is that each decision made by the subject is based on a statistic that is derived from the many (i.e., M) characteristics of the event in question. This statistic reflects the relative probability that the observed characteristics of the event arose from one of a specific class of events. The optimum statistic for such a decision is the likelihood ratio or some monotonic transform of the likelihood ratio (Green & Swets, 1966). The likelihood ratio, $\lambda(u)$, is the relative likelihood that the event, u, arose from one as opposed to the other class of events. That is, $f_i(u)$, the likelihood that the given M-dimensional observation u arose from class i, is the product of the probability that each of the M observed characteristics arose from a class i event, and

$$\lambda(u) = f_1(u)/f_2(u). \qquad [1]$$

The theory assumes that the subject computes $\lambda(u)$, or some monotonic transform of $\lambda(u)$ [e.g., $\lambda'(u) = \log \lambda(u)$], for each event and makes a response decision based on that computed value. It is further assumed that the subject adopts a fixed criterion value of $\lambda(u)$, called β and that the decision corresponding to any event, u, is simply a statement of whether $\lambda(u)$ is greater than β. The capability of the subject to discriminate between the two classes of events is inversely proportional to the total area common to the two conditional probability density functions [$f_i(u)$; $i = 1, 2$]. This common area is assumed to be invariant during the measurement interval.[3]

Assumption of Normality

One specific set of signal detection theory models, the Gaussian models, assumes that the two conditional probability density functions [$f_i(u)$] are Gaussian (normal). One theoretical basis for this assumption is that there are a large number (i.e., M) of *independent* characteristics of the event sampled on each observation, and that the system performs a logarithmic transform of the computed likelihood ratio (Egan & Clarke, 1966). It should be noted that logarithmic transforms are common in perceptual and behavioral data (e.g., Fechner's and Stevens' laws). Because of this hypothetical logarithmic transformation, the separate likelihood statistics are the sums of a large number of independent factors. Then, according to the central limit theorem, the distribution of the likelihood statistics approximates a normal distribution.

On a practical level, the important question is whether there exists evidence that in the given experimental situation, the assumption of normality is tenable. Such evidence might be in the form of reliable results published in the literature or in a test of the assumption by the experimenter (see following sections entitled "Assumptions of Normality and Equal Variance" and "Rating Procedure"). If the Gaussian assumption cannot be justified, then alternative statistics should be em-

[3] The computation of a likelihood ratio statistic assumes that the observer knows the probability distribution for each sampled characteristic conditional on each of the two classes of events. Obviously, the discriminability of the events from the two classes of events reflects the subject's knowledge of the actual differences between the two classes. The assumption of a decision statistic based on the likelihood ratio is simply an assumption that the subject's knowledge of the classes of events can be used in terms of the conditional probability density functions for each characteristic, and that this probability information is combined in an efficient, systematic manner. The theory further states that any factor (i.e., learning) that changes the subject's knowledge of these differences will alter the likelihood statistics, and therefore the discriminability of the two classes of events.

ployed. Such statistics might include simple response probabilities, estimates of "thresholds," statistics derived from a different parametric model of signal detection, or nonparametric indexes of signal detection theory. Some of these alternatives, including nonparametric indexes of signal detection theory, are also discussed in the section entitled "Nonparametric Model."

Since the Gaussian distribution is completely defined by its first two moments, the mean and variance, the area common to two Gaussian density functions, and thus the discriminability of the two classes of events giving rise to these functions, is a monotonic function of the distance between the distribution means scaled in terms of the pooled or average variance (a z transformation). The most common version of the Gaussian model is based on the further assumption that the variances of the probability density functions conditional on the two classes of events are equal. If this and the other assumptions are valid, then the differential weighting of the two distributions caused by any response bias does not affect the estimate of the pooled variance, and the scaled distance between the two means is called d'. This value is equal to the difference of the *signed* distances, in z-score units, from each mean to the subject's criterion. These distances may be estimated by converting into z-score units the probability (relative frequency) of the subject correctly identifying events from the separate classes. The value of d' also may be estimated from appropriate tables of d' (e.g., Elliott's tables in Swets, 1964) by arbitrarily calling one of the two classes of events (i.e., Class 1) the "signal," and the other class, "noise."

If one is interested in motivational or criterion factors, the criterion (β) employed by the subject should be examined. In the equal-variance Gaussian model this criterion is independent of d' and is defined by the following equation:

$$\beta = f(b|1)/f(b|2), \qquad [2]$$

where $f(b|i)$ is the height of the probability density function for class i at the criterion or boundary, b, between the two response classes. The values of $f(b|i)$ may be estimated from the values of P (response i|event i) with a table of ordinates of the normal curve. It should be noted that P (Response 2|Event 2) and β are monotonically related.

Receiver Operating Characteristics (ROC)

A receiver operating characteristic is the locus of points representing the performance of a subject across all criteria under a fixed experimental condition. The ordinate of the receiver-operating-characteristic function is P(response i|event i) and the abscissa is P(response i|event j), $i \neq j$, for each criterion. Since these probabilities are determined by the form of the underlying density functions, the shape of the receiver-operating-characteristic curve also is determined by the nature of these underlying distributions.[4]

The receiver-operating-characteristic curve is generated by sampling the performance of a subject under a given experimental condition while the subject uses various criteria. The subject's criterion can be manipulated through the use of instructions or by the use of a differential payoff schedule; in standard binary decision tasks (yes–no and two-alternative forced choice) the criterion is manipulated across, but not within (see section entitled "Criterion Stability") blocks of trials. A receiver operating characteristic can also be generated within blocks of trials by the use of the rating scale procedure discussed in a subsequent section entitled "Rating Procedure." Since it is assumed that the distributions on which the subject bases decisions are invariant under fixed experimental conditions independent of the criterion employed, the receiver-operating-characteristic curve is the

[4] Any factor that affects the density functions on which the subject bases his or her decision will affect the shape of the receiver-operating-characteristic curve. Some factors known to be important in these respects include the experimental paradigm employed by the experimenter (Markowitz & Swets, 1967), the strategy (i.e., decision rule) employed by the subject (Luce, 1963), the modality (auditory, visual, memory, etc.) in which the subject is operating (Green & Swets, 1966), the subject's criterion stability (Healy & Jones, 1973), and the intertrial intervals employed (Green & Swets, 1966).

map of data points for all possible criteria at a fixed level of sensitivity. Thus, the receiver-operating-characteristic curve is also called the Isosensitivity curve. In memory tasks, it is referred to as the memory operating characteristic (MOC).

Application Considerations

Assumptions of Normality and Equal Variance

If there is sufficient evidence in the literature to warrant the assumption of equal-variance Gaussian density functions, d' and β may be employed to describe, respectively, the subject's ability to discriminate the two classes of events and the subject's response bias, subject to the considerations described below. If there is insufficient evidence to support the assumptions of the model (that the density functions are Gaussian and of equal variance), these assumptions should be tested directly in applying the model. The most typical method for testing these assumptions is the use of a rating procedure (cf. Egan, Schulman, & Greenberg, 1959). This procedure should be used also if the density functions are suspected of being Gaussian, but of unequal variance, and might be employed by careful researchers even when both assumptions are supported by previous research.

Rating Procedure

With the rating procedure, the subject is asked to use N different responses that reflect the subject's confidence that a Class 1 event has occurred. Typically, five to eight confidence ratings or responses are employed. It is assumed that the subject operates in a manner similar to that employed in binary decision tasks (yes–no or two-alternative forced choice) and adopts $N-1$ criteria separating each adjacent pair of the N responses rather than the single criterion employed in the binary task. The results generated by the use of these $N-1$ criteria are plotted as $N-1$ points (x_j, y_j), where $j = 1, 2, \ldots, N-1$, and where x_j and y_j are defined as

$$x_j = \sum_{i=1}^{j} P(\text{response } i \mid \text{Event 2})$$

and

$$y_j = \sum_{i=1}^{j} P(\text{response } i \mid \text{Event 1}). \qquad [3]$$

Thus x_j and y_j are the values of the distribution functions for Events 1 and 2 at criterion j. Obviously, if the assumptions of equal-variance Gaussian functions hold, the expected values of d' calculated for each point (x_j, y_j) are equal and therefore not correlated with the criterion employed by the subject.

The functional relationship between y_i (the probability of a "hit") and x_i (the probability of a "false alarm") describes the contour of criteria for a fixed set of density functions under a specific experimental condition. This function of equal sensitivity is the receiver-operating-characteristic curve described earlier. If the probabilities x_i and y_i are transformed to the equivalent z scores [$x'_j = z$ of $(x_j - .5)$ and $y'_j = z$ of $(y_j - .5)$], the normalized receiver-operating-characteristic curve can be used to test the validity of the assumptions of normality and equal variance. If the underlying density functions are Gaussian, the normalized receiver-operating-characteristic curve will describe a linear function:

$$y' = ax' + c. \qquad [4]$$

The slope of this function, a, is equal to the ratio of the standard deviations of the two density functions ($a = \sigma_2/\sigma_1$), and the intercept, c, is related to the distance between the distribution means.

If the receiver-operating-characteristic curve exhibits a systematic deviation from linearity, the Gaussian assumption may be invalid. If this deviation from linearity is large, but not systematic, there exists an actual deviation from normality and/or a large error factor that may be correlated with the criterion of the subject. Any criterion-correlated error factor will distort the form of the normalized receiver-operating-characteristic curve. However, an equally important concern is that any large error factor might mask an actual deviation from linearity. Such error factors may be due to a number of different problems, including a high degree of criterion variability and/or an insufficient number of trials employed in the experiment (see section entitled "Number of Trials"). If the error factor is large, or is suspected of being large, the interpretation of the results should reflect this fact.

The linear function best describing the data, expressed as standard normal deviates, should be estimated by standard curve-fitting operations.[5] If the data are adequately described by a linear function, the Gaussian assumption is supported. It should be noted, however, that this use of cumulative rating data to generate the receiver-operating-characteristic curve provides data points that are not independent and, at a minimum, imposes a monotonic relationship between successive data points.

If the Gaussian assumption is not rejected, the equal-variance assumption is tested with the slope, a, of the normalized receiver-operating-characteristic curve (Equation 4). If the slope is approximately equal to 1.0, the equal-variance assumption is not rejected and the equal-variance Gaussian model may be legitimately employed with d' and β as parameters. While the values of d' estimated from the $N - 1$ criteria are not totally independent estimates, the mean or median value of d' may be employed as the estimate of observer sensitivity. The measures derived for the unequal-variance Gaussian model, discussed in the next section, and the nonparametric model, discussed in the section entitled "Nonparametric Model," may be more desirable than those derived from the equal-variance model since fewer restrictive assumptions are involved.

General Gaussian Case

If the equal-variance assumption is violated, d' and β will be correlated to a degree that is related to the deviation from equality of variance. The general Gaussian (or unequal-variance Gaussian) model is applicable when the Gaussian assumption is justified, independent of the relative magnitude of the variances. Application of the general Gaussian model requires knowledge of the slope, a, of the normalized receiver-operating-characteristic curve (Equation 4) which may be estimated with the rating procedure (see section entitled "Rating Procedure"). The basic goal of the general Gaussian model is to develop a statistic that describes sensitivity, is independent of the subject's criterion, and reflects the average spread or variance of the two distributions. Several statistics use the fact that when $\beta = 1.0$, β is equidistant from the two distribution means in terms of standard normal deviates (z scores) for each of the given distributions. At $\beta = 1.0$, the "hit rate" for the two classes of events [P(response i| event i)] are equal. Thus d' computed at $\beta = 1.0$, the minimum total error criterion (given equal probability of presentation for the two classes of events), will be based on the average standard deviation with equal weighting given to the two distributions. This minimum error criterion is the negative diagonal ($y' = -x'$) of the receiver-operating-characteristic space (see Figure 1). The coordinates (x'_m, y'_m) of the intersection of the estimated receiver-operating-characteristic curve (Equation 4) and the negative diagonal define the value of d' for this minimum error criterion. The value of d' for this point is equal to the distance along the negative diagonal from the positive diagonal (chance line) scaled in terms of the difference between the coordinates ($y'_m - x'_m$), and is called d'_s

[5] Many researchers (e.g., Swets et al., 1961) have used simple visual fits to determine the "best-fitting" straight line. This crude method is probably sufficient for most proposed uses of the function. Conventional least squares curve-fitting procedures are theoretically inappropriate because both variables are dependent variables and subject to error. For rough approximations, this problem is of minor importance since the error introduced is likely to be small relative to the noise in the data. However, for researchers interested in precise estimates of the parameters of receiver-operating-characteristic curves and in a test for goodness-of-fit of the theoretical model, maximum-likelihood estimators giving exact fits have been developed by Ogilvie and Creelman (1968).

(Clarke, Birdsall, & Tanner, 1959). This value of d' (d'_s) is the distance between the distribution means scaled in terms of the average of the standard deviations for the two distributions.

With knowledge of the slope, a, of the normalized receiver-operating-characteristic function, one can compute d'_s from a single data point (x'_k, y'_k) by the formula

$$d'_s = 2(y'_k - ax'_k)/(1 + a). \quad [5]$$

Other measures of sensitivity that may be derived for the general Gaussian model include d'_e, which equals the square of d'_s (Egan, 1958; Egan & Clarke, 1966) and Δm, the x intercept of the normalized receiver-operating-characteristic curve (Green & Swets, 1966).

Whenever d' or d'_s is computed from replicated sets of binary decisions, the various estimates of sensitivity for a given subject should be examined for any relationship between the sensitivity measure and the probability of a false alarm, P(Response 1|Event 2). Any such relationship indicates a deviation from the assumed slope of the receiver-operating-characteristic curve. This simple check should be followed whenever repeated estimates of sensitivity are obtained for each subject, whether or not the variances are assumed to be equal.

Nonparametric Model

The Gaussian models of signal detection theory are members of a class of models in which the parameters describing the ability of the subject to perform the given task and the subject's decision rule are dependent on the assumption of certain specific underlying density functions. Other parametric models of signal detection theory based on different assumed density functions (e.g., negative exponential, rectangular, Raleigh, and Rice) can be derived (Green & Swets, 1966; Pollack & Hsieh, 1969), but such models have not been fully developed for general use and, in any case, would be of questionable utility since the user must be able to justify the assumption of the given underlying density functions. Since it is unlikely that, in a particular situation, data that are sufficiently deviant to lead to rejection of the Gaussian model would be sufficiently regular to support an alternative model, a nonparametric model of signal detection theory whose measures are independent of the exact nature of the underlying density function would be of general utility.

Green (1964) proposed the use of the area under the receiver-operating-characteristic curve, A_g, as a measure of observer sensitivity. Assuming only that the subject bases his decision on two continuous probability density functions that are identical under the various experimental procedures, this measure can be shown to be identical to the expected percentage correct in a two-alternative forced choice experiment (Green & Moses, 1966). Since this relationship is independent of the form of the underlying probability density functions, it may be employed with no prior assumptions concerning the shape of these density functions. This area measure of sensitivity may be employed for both rating data (Pollack, Norman, & Galanter, 1964) and single data points (Pollack & Norman, 1964).

The derivation of the area measure and the corresponding nonparametric measure of criterion is based on the unit square (see Grier, 1971). This square has as its abscissa and ordinate, x_i and y_i, as defined in Equation 3. When only a single point relating x_i and y_i is available, the area under the curve joining the points (0,0) to (x,y) to (1,1) is determined. This area, A_g, is then taken to be the index of observer sensitivity; A_g is the average of the maximum and minimum possible areas under the receiver-operating-characteristic curve and is given by the following formula:

$$A_g = .5 + (y-x)(1+y-x)/4y(1-x). \quad [6]$$

The A_g measure has also been extended to cases where more complete receiving-operating-characteristic curves are available (Pollack, Norman, & Galanter, 1964) for example, when data have been obtained through the rating procedure (see previous section entitled "Rating Procedure"). In this case the area under the curve A_g, can be estimated by the formula:

$$A_g = (1/2) \sum_{j=1}^{N} (x_{j+1} - x_j)(y_{j+1} + y_j), \quad [7]$$

where x_j and y_j are defined by Equation 3.

Pollack and Hsieh (1969) have used Monte Carlo methods to sample from various density functions in order to investigate the sampling distributions of A_g and d'_e (discussed in a previous section entitled "General Gaussian Case"). They determined that when the normality assumptions of signal detection theory are satisfied, d'_e and a Gaussian transform of A_g, $N(A_g)$, are related by the formula

$$N(A_g) = [.707 - .234 \log_2 (\sigma_2^2/\sigma_1^2)]d_e'. \quad [8]$$

They found that the empirically determined values of $N(A_g)$ tended to overestimate the values of d'_e by 1%–6%.

Hodos (1970) developed a nonparametric measure of criterion or bias. This measure was based on the fact that the negative diagonal of the unit square represents the locus of points where the subject would be equally likely to respond "*i*" or "*j*" given ambiguous stimulus conditions. The measure reflects the degree to which a data point deviates from the negative diagonal relative to the maximum possible deviation. A computational formula for the nonparametric measure of criterion, β', based on the Hodos measure, has been developed by Grier (1971). The formula is:

$$\beta' = 1 - x_i(1 - x_i)/y_i(1 - y_i), \quad [9]$$

where x_i and y_i are defined in Equation 3.

Criterion Stability

If the criterion adopted by the subject is not stable during any given session, the variability of the criterion will affect the results. The presence of criterion variability cannot be detected easily, and will have the same effect on the results as an increase in the variance of both likelihood density functions. Criterion variability therefore decreases the estimate of d' by an amount that is related to the size of the criterion variance without actually affecting the *true* discriminability of the two classes of events. Obviously, any experimental manipulation that affects criterion variability will alter the estimate of d'. Safeguards against criterion variability include the use of trained subjects, strict instructions to the subjects about maintaining a stable criterion, and strict definitions of the subject's response classes. Since the subject's criterion is partially determined by the expectation of the probability of presentation of the two classes of events, the subject should be made aware of the absence of sequential dependencies across trials.

While it may be reasonable to assume that the criterion employed by a single subject during any measurement session (block of trials) is stable, it is less reasonable to assume that the subject will employ the same criterion across sessions, or even across separate blocks of trials within a session. Therefore, only the data for a single block of trials should be used to estimate a value of d'. The estimates of d' from the various blocks of trials may then be averaged.

Malingering

The positive diagonal of the receiver-operating-characteristic space ($x' = y'$) defines chance performance. Under the equal-variance Gaussian model, the receiver-operating-characteristic curve that corresponds to the positive diagonal is generated under the condition of exact equality for the two density functions. Data points below this chance line can be generated only by (*a*) measurement error or (*b*) the subject performing the discrimination and then emitting a response that is inconsistent with the computed decision statistic $[\lambda'(y)]$. If a subject consistently produces data that fall below the chance line, there is justification to assume that the subject can perform the discrimination, but is malingering.

Number of Trials

In applying signal detection theory, the experimenter is assuming that there are two fixed internal probability density functions, and the subject has established a fixed criterion along the dimension (decision axis) on which these functions lie. The purpose of the experiment is to estimate the area [P (response "i"|event j)] in the tail of each of the two distributions from the relative frequencies of the responses. The expected standard error in estimating the probablity [P("i"|j)] as a function of both the sample size and the expected value of this probability is $p \cdot q/s$, where p is the expected value of the probability, $q = 1 - p$, and s is the sample size. The expected error of estimation for d' can be obtained by applying a z transformation to $p \pm (p \cdot q/s)$. Green and Moses (1966) found that the actual error involved in estimating this parameter is slightly larger than predicted by this assumption of binomial variability. Pollack and Hsieh (1969), in a computer simulation, found that the error variance in A_g was slightly smaller than predicted by the assumption of binomial variability. Therefore, the expected binomial variability would seem to reflect the magnitude of error to be expected in a given measure. Obviously, the use of only a small number of trials for one or both of the event classes results in a large expected error in estimates of the parameters of the model. Furthermore, reliable estimates of sensitivity require a large number of trials when based upon extreme values of P("i"|j).

Applications of Signal Detection Theory

Since signal detection theory provides the researcher with a means of evaluating independently both the ability of an organism to discriminate between classes of events and motivational or other response effects, it can be a powerful research tool having application in a variety of different experimental settings. It is the purpose of this section to outline some potential applications of signal detection theory in areas of psychology where this method has not been widely used. The applications discussed include the evaluation of (*a*) the state of the organism or environment, (*b*) the relationship between stimuli and potential or actual responses, and (*c*) the independence of "channels" for processing stimulus information. Finally, a more traditional example from memory work is presented in some detail to provide a worked example for persons unfamiliar with the computational procedures involved in a signal detection theory analysis.

Evaluating the Condition of Subjects or Environment

The ability of a subject to perform detection, discrimination, or recognition tasks can be altered by a number of conditions including the psychological or physiological condition of the subject (e.g., behavioral or organic dysfunction), the existence of a drug state, or the imposition of an external stimulus. In many cases, however, it is unclear whether the performance difference is due to changes in the ability of the subject to perform the task or changes in the response tendencies of the subject. Signal detection theory may be used in a between-groups design to evaluate the cause of the observed differences between an altered and a control population. Similarly, signal detection theory could be used in a pretest–posttest design to investigate the locus of performance differences as a result of pharmacological or surgical interventions. A somewhat less obvious potential application occurs in the area of motivation. If an experimenter discovers that rats initially exhibit a preference for a given solution over water, but after two months of continuous ad libium intake of the solution exhibit no differential preference, the experimenter does not know whether the motivation of the rats or their ability to discriminate between the two solutions has been altered. However, by using the given solution and water as the discriminative stimuli with either an appetitive or avoidance conditioning technique, the researcher could apply nonparametric measures of signal detection theory to the performance data to evaluate the nature of this change.

Evaluating Response Factors

Performance in any given task is determined by two main classes of variables: those that affect the discriminability of stimulus events or conditions and those that affect the motivational state and response tendencies of the given subject. These two classes of variables are completely analogous to the sensitivity and criterion factors considered by signal detection theory. Therefore, differences between response or motivational effects may be evaluated by signal detection theory analysis techniques. For example, the responses of infrahuman subjects may be evaluated in terms of the "hit" and "false alarm" rates with nonparametric measures of signal detection theory applied to separate the two classes of effects. In addition, if one is willing to assume that the magnitude of the response (e.g., galvanic skin response or reaction time) is directly (or inversely) proportional to the likelihood ratio, one can generate a receiver-operating-characteristic curve and employ parametric measures of signal detection theory (cf. Pike, 1973).

An assumption with greater face validity and some empirical support (Emmerich, Gray, Watson, & Tanis, 1972) is that the magnitude of a response parameter reflects the absolute value of the difference between the given likelihood ratio and the subject's response criterion. Partitioning the data on the basis of a subject's binary (yes–no, etc.) response, the strength or speed of a "yes" (or "signal") response may be assumed to be monotonically related to the likelihood ratio, with "no" responses reflecting likelihood ratios that are below those for "yes" responses and are an inverse monotonic function of the strength or speed of response. Once a receiver-operating-characteristic curve is generated, the assumptions for the parametric measures (d', d'_s, etc.) can be tested, and the appropriate measure employed. For an example of this type of analysis applied to response latency, see Murdock (1966).

The use of modern data-acquisition and analysis procedures have opened the area of the neural coding of stimuli to study. Many analysis techniques (e.g., average evoked potential, poststimulus time, and interpulse interval histograms) pool the data across trials to extract statistically the average response from the data. While these techniques have statistical validity, the "average" response they extract may be typical of none of the actual responses. Signal detection theory offers another statistical means of extracting a "characteristic response." Assume that in a study of multicellular activity in a given nucleus, it is found that vibratory stimuli presented simultaneously with a light flash cause a characteristic "averaged" response that differs in certain aspects from the response to only a light flash. An analysis of the individual trial data using an arbitrarily defined response can estimate a complete outcome matrix (hit, false alarm, etc.) for the given response from which a signal detection theory measure of the adequacy of that "response" in distinguishing "stimulus-plus-noise" events from "noise-only" events. Systematic modification of the definition of a "response" can then be used to determine the response pattern that best discriminates the stimulus from the noise. Thus, the averaging may have indicated the importance of a burst of responding 10 milliseconds after the stimulus, while the iterative signal detection theory technique may indicate the requirement of two distinct bursts that must begin between 7 and 9 milliseconds, and must be separated by a 2-millisecond lull in firing.

Evaluating Channel Independence

Recently there has been considerable interest in the ability of subjects to process independently the information presented to different sensory or perceptual "channels." One factor contaminating much of the early work in this area is the change in the criterion of the subject with changes in the requirements of the task. Eijkman and Vendrik (1965) studied the independence of processing of auditory and visual signals. Signal detection theory measures of the detectability of the auditory and visual stimuli were estimated for a set of subjects in separate experiments. Then the same subjects were asked to perform simultaneously the same *independent* auditory and visual detection tasks with separate sets of responses for each type of stimulus; signal detection theory measures of sensitivity were computed for each of these two stimulus channels. A comparison of the signal detection theory measure under the simple and simultaneous condition provides an indication of the degree of interaction between the two tasks, which may be evaluated in quantitative terms by methods developed by Taylor, Lindsay, and Forbes (1967). Because there is independence across stimuli and across most stimulus–response categories in the response matrix, it is possible to calculate a signal detection theory measure of sensitivity in one channel conditional on the simultaneous stimulus, response, or outcome event in the other channel. With modern data-handling techniques, partitioning of the data in this manner has become a simple matter. Pastore and Sorkin (1972) examined the effects on sensitivity in a single sensory channel as a function of the various possible stimulus events and outcomes in the second channel in a simultaneous two-channel detection paradigm. This technique of analysis has also been successfully employed by Harvey and Treisman (1973) in a simultaneous task and by Sorkin, Pohlmann, and Gilliom (1973) in a successive two-channel task.

Evaluating Memory Processes

Lutz and Scheirer (1974) investigated differences in the processes involved in the encoding of visually presented verbal and pictorial stimuli. Each subject was presented with a series of 190 stimulus items; each item was presented for either .25, .50, 1.00, or 2.00 seconds with a fixed interstimulus interval of either .25, 1.00, or 2.00 seconds. All conditions in the $4 \times 3 \times 2$ (Presentation Interval $\times$ Interstimulus Interval $\times$ Stimulus Type) factorial arrangement were presented to independent groups of 12 subjects.

At the beginning of the session the subjects were told that a series of items would appear on the screen in front of them. The subjects were instructed to "pay careful attention to each item . . ." since the subjects would "later. . . be given a test based on these items." Following the presentation of the 190 stimulus items, the subjects were told that a second series of items would be presented, each for a few seconds. They were told that some of the items had been presented previously and some had not, and were given the following instructions:

> When an item appears, look at it and decide whether the item appeared in the first series. You should record your decision on the answer sheet. There are six categories you can respond with:
>
> +++ if you are *definite* that you have seen the item before;
> ++ if you *believe* that you have seen the item before;
> + if you *guess* that you have seen the item before;
> – if you *guess* that you have *not* seen the item before;
> –– if you *believe* that you have *not* seen the item before;
> ––– if you are *definite* that you have *not* seen the item before.
>
> For each stimulus you should respond with one and only one of the above categories. For each item that appears, circle the appropriate symbol on the answer sheet as soon as you have made your decision. Try to use all six categories but only where appropriate . . . In summary, when the first item appears look at it, decide whether the item appeared in the first series . . . Then wait for the next item to appear [p. 317].

All instructions were read aloud to the subjects, with the category definitions typed on a card given to the subject for reference during the experiment. A series of 120 test stimuli were presented to the subject, 60 of these test stimuli were randomly chosen from the original 190 items. These "old" items were randomly mixed with 60 "new" items that were not in the original set.

The relative frequency of responding with each of the six categories to each of the two classes of events is shown in Table 1 for two of the subjects. These rating response data were converted to cumulative response probabilities as described by Equation 3 and then plotted as receiver-operating-characteristic curves. Figure 1 is the normalized receiver-operating-characteristic curve for the two subjects reported in Table 1. The upper and right margins of the figure are delineated in z-score units. The lower and left margins are delineated in terms of probabilities. The data points are labeled according to the limits

TABLE 1

Relative and Cumulative Frequencies with Each Category for the Subjects Described in Text[a] and in Figure 1

Subject	response i	$P(i\|\text{new})$	$P(i\|\text{old})$	$\Sigma P(i\|\text{new})$	$\Sigma P(i\|\text{old})$
1	+++	.033	.533	.033	.533
	++	.000	.050	.033	.583
	+	.067	.117	.100	.700
	−	.067	.133	.167	.833
	−−	.200	.100	.367	.933
	−−−	.633	.067	1.000	1.000
Total		1.000	1.000		
2	+++	.033	.517	.033	.517
	++	.250	.200	.283	.717
	+	.167	.050	.450	.767
	−	.183	.100	.633	.867
	−−	.234	.117	.867	.984
	−−−	.133	.016	1.000	1.000
Total		1.000	1.000		

[a] See section entitled "Rating Procedure."

of summation indicated in Equation 3 and Table 1. The linear function describing each set of data is plotted in Figure 1. The data for the second subject appear to be curvilinear. Had the data for a majority of the other subjects been curvilinear, the Gaussian assumption would have to be rejected. However, a small and approximately equal number of receiver-operating-characteristic curves were curvilinear in each direction, and most functions were linear (e.g., see the curve for Subject 1 in Figure 1). Therefore, the Gaussian assumption was held to be supported by the data and the few deviations from linearity were assumed to be due to error.

The linear functions for the two subjects plotted in Figure 1 have estimated slopes of .91 and .85, respectively. While these slopes do not differ substantially from the slope of unity required by the assumption of equal-variance Gaussian distributions, larger deviations were found for a number of subjects. Since acceptance of the equal-variance assumption is therefore tenuous and since the rating procedure allows the use of the general Gaussian model (see sections "Rating Procedure" and "General Gaussian Case"), d'_s was used as the measure of discriminability rather than d'. The negative diagonal (minimum error criterion) of the receiver-operating-characteristic space in Figure 1 is delineated in units of d'_s. The intersection of this diagonal with the linear regression lines for the obtained data yield d'_s estimates of 1.94 and 1.14 for the two subjects. The corresponding values of the nonparametric area measure of sensitivity, A_g, computed with the use of Equation 7, are .892 and .785. These are estimates of the ability of the subjects to discriminate the two classes of events independent of the criteria employed and any differences in variability within the classes of events.

Using any positive (+, ++, +++) response as a response indicating an "old" stimulus and any negative response (−, −−, −−−) as a response indicating a "new" stimulus, the probability of a correct response was computed for each subject. A within-cell product-moment correlation of .91 was obtained between d'_s and the probability of being correct. This high correlation appears to be at least partially due to the use of a strict set of criterion categories. While this procedure was intended to minimize within-subject variability, it also appears to have caused most subjects to adopt a set of criteria whose

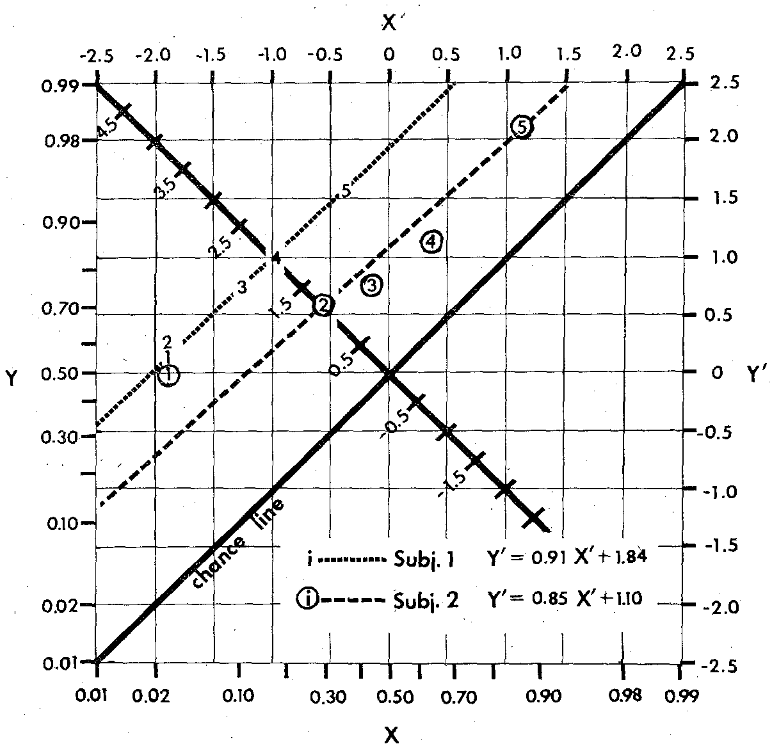

FIGURE 1. Normalized receiver-operating-characteristic for two subjects, which is the probability of a "hit" [$y = P$(response i|stimulus i)] plotted as a function of the probability of a "false alarm" [$x = P$(response i|stimulus j), $i \neq j$] in terms of z scores. (The right and upper edges of the figure are delineated in terms of z scores [x', y'], while the left and lower edges are delineated in terms of the equivalent probabilities [x, y]. The procedures used to generate the receiver-operating-characteristic curve are described in the last part of this article. Units of d'_s are delineated along the negative diagonal and are explained in the section entitled "General Gaussian Case.")

centroid was consistent. This improved the validity of the probability-of-correct response measure by limiting the criterion variability between subjects.

FINAL CONSIDERATIONS

While this article is intended to provide the reader with an overview of the general theory and requisite procedures to use signal detection theory, it is recommended that this article also be used as a guide to reading the more detailed statements of signal detection theory, most of which deal only superficially with many of the application considerations described in the second section. This article has treated only the simple, more common models of signal detection theory. Models of signal detection theory have been generalized to multicomponent recognition tasks (Pastore & Sorkin, 1971; Tanner, 1956), and to a broad spectrum of research applications in modern psychology (i.e., Swets, 1973). In addition, statistical tests for various signal detection theory parameters have been developed (i.e., Gourevitch & Galanter, 1967; Ogilvie & Creelman, 1968). Critiques of signal detection theory may be found in Luce (1963), and Abrahamson and Levitt (1969).

REFERENCES

Abrahamson, I. G., & Levitt, H. Statistical analysis of data from experiments in human signal detection. *Journal of Mathematical Psychology*, 1969, 6, 391–417.

Banks, W. P. Signal detection theory and human memory. *Psychological Bulletin,* 1970, **74**, 81–99.
Bernbach, H. A. Decision processes in memory. *Psychological Review,* 1967, **74**, 462–480.
Campbell, R. A., & Moulin, L. K. Signal detection audiometry: An exploratory study. *Journal of Speech and Hearing Research,* 1968, **11**, 402–410.
Clarke, F. R., Birdsall, T. G., & Tanner, W. P., Jr. Two types of ROC curves and definitions of parameters. *Journal of the Acoustical Society of America,* 1959, **31**, 629–630.
Egan, J. P. *Recognition memory and the operating characteristic.* (Tech. Note AFCRC-TN-58-51) Indiana University: Hearing and Communication Laboratory, 1958.
Egan, J. P., & Clarke, F. R. Source and receiver behavior in the use of a criterion. *Journal of the Acoustical Society of America,* 1956, **28**, 1267–1269.
Egan, J. P., & Clarke, F. R. Psychophysics and signal detection. In J. B. Sidowski (Ed.), *Experimental methods and instrumentation in psychology.* New York: McGraw-Hill, 1966.
Egan, J. P., Schulman, A. I., & Greenberg, G. A. Operating characteristics determined by binary decisions and by ratings. *Journal of the Acoustical Society of America,* 1959, **31**, 768–773.
Eijkman, E., & Vendrik, A. J. H. Can a sensory system be specified by its internal noise? *Journal of the Acoustical Society of America,* 1965, **37**, 1102–1109.
Emmerich, D. S., Gray, J. L., Watson, C. S., & Tanis, D. C. Response latency, confidence, and ROCs in auditory signal detection. *Perception and Psychophysics,* 1972, **11**, 65–72.
Green, D. M. General prediction relating yes–no and forced choice results. *Journal of the Acoustical Society of America,* 1964, **36**, 1042.
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.
Gourevitch, V., & Galanter, E. A significance test for one parameter isosensitivity functions. *Psychometrika,* 1967, **32**, 25–33.
Grier, J. B. Nonparametric indexes for sensitivity and bias: Computing formulas. *Psychological Bulletin,* 1971, **75**, 424–429.
Harvey, N., & Treisman, A. M. Switching attention between the ears to monitor tones. *Perception and Psychophysics,* 1973, **14**, 51–59.
Healy, A. F., & Jones, C. Criterion shifts in recall. *Psychological Bulletin,* 1973, **79**, 335–340.
Hillyard, S. A., Squires, K. C., Bauer, J. W., & Lindsay, P. H. Evoked potential correlates of auditory signal detection. *Science,* 1971, **172**, 1357–1360.
Hodos, W. Nonparametric index of response bias for use in detection and recognition experiments. *Psychological Bulletin,* 1970, **74**, 351–354.
Licklider, J. C. R. Three auditory theories. In S. Koch (Ed.) *Psychology: A study of a science. Vol. 1.* New York: McGraw-Hill, 1959.
Luce, R. D. Detection and recognition. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology.* New York: Wiley, 1963.
Lutz, W. J., & Scheirer, C. J. Encoding processes in pictures and words. *Journal of Verbal Learning and Verbal Behavior,* 1974, **13**, 316–320.
Markowitz, J., & Swets, J. A. Factors affecting the slope of empirical ROC curves: Comparison of binary and rating responses. *Perception and Psychophysics,* 1967, **2**, 91–100.
Moray, N. Time sharing in auditory perception: Effect of stimulus duration. *Journal of the Acoustical Society of America,* 1970, **47**, 660–661.
Murdock, B. B., Jr. The criterion problem in short-term memory. *Journal of Experimental Psychology,* 1966, **72**, 317–324.
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.
Pastore, R. E., & Sorkin, R. D. Simultaneous two-channel signal detection. I. Simple binaural stimuli. *Journal of the Acoustical Society of America,* 1972, **51**, 544–551.
Peterson, W. W., Birdsall, T. G., & Fox, W. C. The theory of signal detectability. *Transactions IRE Professional Group on Information Theory,* 1954, **4**, 171–212.
Pike, R. Response latency models for signal detection. *Psychological Review,* 1973, **80**, 53–68.
Pollack, I., & Hsieh, R. Sampling variability of the area under the ROC-curve and of d'_e. *Psychological Bulletin,* 1969, **71**, 161–173.
Pollack, I., & Norman, D. A. A non-parametric analysis of recognition experiments. *Psychonomic Science,* 1964, **1**, 125–126.
Pollack, I., Norman, D. A., & Galanter, E. An efficient non-parametric analysis of recognition memory. *Psychonomic Science,* 1964, **1**, 327–328.
Rilling, M., & McDiarmid, C. Signal detection in fixed-ratio schedules. *Science,* 1965, **148**, 526–527.
Sorkin, R. D., Pastore, R. E., & Pohlmann, L. D. Simultaneous two-channel signal detection. II. Correlated and uncorrelated signals. *Journal of the Acoustical Society of America,* 1972, **51**, 1960–1965.
Suboski, M. D. Signal detection methods in the analysis of classical and instrumental discrimination conditioning experiments. *Proceedings of the 75th Annual Convention of the American Psychological Association,* 1967, **2**, 37–38.
Sutton, S. Fact and artifact in the psychology of schizophrenia. In M. Hammer, K. Salzinger, & S. Sutton (Eds.), *Psychopathology.* New York: Wiley, 1972.
Swets, J. A. (Ed.). *Signal detection and recognition by human observers.* New York: Wiley, 1964.

Swets, J. A. The relative operating characteristic in psychology. *Science,* 1973, **182**, 990–1000.

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

Tanner, W. P., Jr. Theory of recognition. *Journal of the Acoustical Society of America,* 1956, **28**, 882–888.

Tanner, W. P., Jr. Physiological implications of psychophysical data. *Annals of the New York Academy of Sciences,* 1961, **89**, 752–765.

Taylor, M. M., Lindsay, P. H., & Forbes, S. M. Quantification of shared capacity processing in auditory and visual discrimination. *Acta Psychologica,* 1967, **27**, 223–229.

Van Meter, D., & Middleton, D. Modern statistical approaches to reception in communication theory. *Transactions IRE Professional Group on Information Theory,* 1954, **4**, 119–141.

Wald, A. *Statistical decision functions.* New York: Wiley, 1950.

(Received January 16, 1974)