Reprinted from:
HANDBOOK OF PERCEPTION, VOL. II
Psychophysical Judgment and Measurement

ACADEMIC PRESS, INC.
New York San Francisco London

Chapter 9

DETECTION, DISCRIMINATION, AND RECOGNITION*

R. DUNCAN LUCE and DAVID M. GREEN

* This work was supported in part by a grant from the Alfred P. Sloan Foundation to The Institute for Advanced Study and in part by grants from the National Institutes of Health, Public Health Service, U.S. Department of Health, Education, and Welfare, and from the National Science Foundation to the University of California, San Diego. Manuscript was completed January, 1971.

1. INTRODUCTION

During the period 1955–1970 many attempts were made to formulate precisely the basic mechanisms thought to underlie the processing of simple physical signals and to isolate some of the factors that affect subjects' responses to such signals. The main impetus for this work was the application of the theory of signal detectability to psychophysics by Tanner and Swets (1954). This work is well known and is adequately summarized elsewhere (Green & Swets, 1966; Luce, 1963a; Swets, 1964), and we will not review the basic ideas again. Rather, we will report some of the later developments sparked by these early studies.

Despite the diversity of extant models, all are able to predict certain basic features of the relevant data, and competing pairs of models tend to be about equally parsimonious as measured by the number of free parameters that one must estimate from data. Moreover, the predictions are often so similar that rather elaborate statistical tests are needed to ascertain which theory best accounts for a particular experiment. For those hoping to find a single, clearly superior theory, this chapter will prove disappointing. Our tentative conclusion is that among existing theories, different ones are appropriate to different situations. Or put another way, no current theory is really correct and which one better approximates the data varies with the experiment being analyzed. That state of affairs, though not ideal, would be acceptable if we were able to state clearly which experimental features serve to bound the region of reasonable application of a given theory. For example, we would like a rule such as: Threshold theories best handle discrimination of small changes in intensity when there is no noise background, whereas continuous theories are best when the signal is in noise. The sad fact is that no such rules have yet been formulated.

Another criterion for the worth of a theory, in addition to its accuracy, is the class of experimental designs for which it makes significant predictions. A minimum requirement of any theory is that it unify data from most of the simple (one and two stimulus) psychophysical tasks. Practically all current theories make predictions that can be cross validated in the simple yes–no (YN) and forced-choice (FC) experiments and in the single stimulus design with a rating response.* By "cross validate" we mean that certain parameters which, on theoretical grounds, should be the same in different experimental designs, appear to be so. It is our particular bias

* We abbreviate certain recurrent terms. At the first use of the term, we place the abbreviation in parentheses after it; thereafter we use the abbreviation.

that an adequate theory should apply to experimental situations in which the stimulus interval is not precisely marked as well as those in which it is. Experiments in which the signal can occur at any time mimic many practical detection and discrimination situations better than do the usual fixed interval (FI) designs. To generate models that account for behavior in these so-called free-response designs and to relate their parameters to those from the discrete FI tasks is a challenging and largely unsolved problem.

Applying the criterion of generality usually involves an element of individual taste, because one theory generalizes easily in one direction, another theory in a different direction. For example, choice theory is readily stated for any finite number of stimuli and responses, but the connection between its parameters and physical properties of the stimuli is exceedingly vague. The theory of signal detectability (TSD) is easily couched in very stimulus-oriented terms, but its extension to more than two responses produces so many free parameters that prediction becomes virtually impossible.

Let us consider further some of the problems that arise when we generalize designs and theories from $n = 2$ to $n > 2$ stimuli. Recall that the standard $n = 2$, YN design involves a noise background or null signal, $\varnothing$, a signal in noise, s, and a single observation interval in which one of these alternatives is presented. If we retain the single observation interval, then the natural way to generalize the YN design is to increase the number of signals that can be presented. We might do this by using several different frequencies or hues or by presenting several different levels of intensity. In contrast, we can keep only two signals, but increase the number of possible stimuli. If we allow all possible combinations of two signals in k observation intervals, there are 2^k possible stimuli. Thus, in the most general two-alternative forced-choice (2AFC) task there are four possible ordered pairs—stimuli—$\langle s, \varnothing\rangle$, $\langle \varnothing, s\rangle$, $\langle \varnothing, \varnothing\rangle$, and $\langle s, s\rangle$. Usually, the signal s is permitted to appear only once, which restricts the number of stimuli to the number of intervals. Under that restriction, a 3AFC design has the stimuli $\langle s, \varnothing, \varnothing\rangle$, $\langle \varnothing, s, \varnothing\rangle$, and $\langle \varnothing, \varnothing, s\rangle$. Most empirical studies have imposed this restriction; however, Markowitz (1966) explored the most general case of 2AFC.

The compass of such terms as detection, discrimination, and recognition for $n > 2$ is still confused. When there is one interval and n different nonnull signals in one-to-one relation with n responses, we speak of the design as a recognition task except when $n = 2$ and the stimuli differ only in intensity, in which case it is called an *intensity discrimination* task. However, if one of the n signals is replaced by the null signal $\varnothing$, we call it *simultaneous recognition and detection* except when $n = 2$ (one nonnull signal), in which case it is called a *yes–no detection* task. Any FC design

having a signal in one interval and the null signal in the other $n - 1$ intervals is also called a detection design. For more details, see Luce (1963a) and Green and Swets (1966). There is, of course, an obvious sense in which all of these designs are asking the subject to recognize which of n possible stimulus presentations has been presented. Were the terminology not so deeply ingrained in the psychophysical literature, we might be tempted to reserve the use of "detection" for FR designs concerned with the detection of a signal whose time of onset is uncertain, and refer to all of the FI designs as recognition experiments.

As the numbers n of stimuli and m of responses increase, programming problems and costs rise rapidly. Specifically, we must increase the number of trials by the factor $n(m - 1)/2$ in order to have, on the average, a constant number of observations per independent estimated probability. Thus, considerable experimental selectiveness is required. Most of the studies that have been performed have been motivated by specific considerations. In addition, there are theoretical difficulties of two types. First, several viable two-stimuli theories exist and each of them can be generalized in a variety of ways (many of the logical possibilities have not yet been worked out). Moreover, all of these generalizations proliferate free parameters at a spectacular rate. Often postulates not needed in the two-stimulus case must be added in order to test the general model. Even so, a great number of parameters must be estimated. This leads to the second difficulty: how to estimate optimally the many parameters from data and how to compare the adequacy of theories with different numbers of parameters? We have barely begun to face these problems.

Our review begins with a general classification scheme. It is first used to organize various general theories for situations with n signals and m responses. Following this, we turn to the special cases of $n = 2$ stimuli which have received the most careful attention in the literature. We first deal with operating characteristics and psychometric functions on the assumption of stationary mechanisms. We next turn to the limited literature dealing with sequential effects in the data. And, finally, we turn to the relatively few attempts to deal with FR data in which responses may occur at any time.

II. CLASSIFICATION SCHEME

Our classification of theories is based on guiding principles clearly evident in all of the theories so far proposed. First, they all postulate a sensory or perceptual stage followed by a memory process that stores a hypothetical representation of the stimulus. Second, following the sensory–memory process is a decision stage, which operates on the representation of the stimulus. This two-stage structure is characteristic of the successful two-stimulus theories, and it has been uniformly maintained in generalizing them. So we will class theories according to the two types of stages assumed.

A. Sensory–Memory Processes

The sensory–memory processes are of three general types, which provides the first dimension of our classification system:

1. Finite State

There are assumed to be k internal states of which one is activated with some probability when a stimulus is presented. These states may or may not be ordered. Two special cases are of importance.

(*a*) $k = n$, in which case the states are identified with the stimuli; these are *stimulus generalization* models.
(*b*) $k = m$, in which case the states are identified with the possible responses (*implicit responses*).

Note that, without further assumptions, the conditional probabilities relating states to stimuli constitute $n(k - 1)$ free parameters.

2. Continuous State

The internal states are assumed to form a continuum, usually in Euclidean space. In the one-dimensional case, the continuum is naturally ordered. Usually, each signal is represented by some distribution over the continuum. Such assumed distributions constitute, in a sense, a continuum of free parameters unless there are strong a priori arguments for restricting them to a particular family. Frequently, though not always, the family is assumed to be normal, in which case the parameters are reduced to two, namely the mean and variance of each normal distribution.

3. Stochastic Process

A signal is assumed to generate in time a train of pulses (neural, perhaps), which is described by some stochastic process. Most often, the process is assumed to be Poisson—in which case the interarrival times are independent and exponentially distributed, and so it can be viewed as a special type of continuous state process; however, the decision processes associated with continuous state and stochastic sensory models usually differ somewhat.

B. Response Processes

The response or decision processes are of three types; they provide the second dimension of our classification system.

1. Response Generalization

This is only compatible with the finite state representation in which $k = m$; it assumes that when the sensory process indicates one response, there is some tendency (possibly, probability) of another response occurring.

2. Response Bias

Response bias models postulate that among the possible responses or implicit responses, there are differential tendencies to employ one rather than another. The most widely used variant, called a choice model, assigns a weight $b(r) > 0$ to each response r and calculates the probability of choosing r out of a set R of possible responses as $b(r)/\sum b(x)$.

3. Response Criterion

Such models are compatible only with an ordered set of states; they are a natural generalization of TSD. Most often, the criteria are assumed to be fixed, but in some models they are assumed to be distributed in some manner and in others to be adjusted systematically, according to a fixed learning model.

III. THEORIES FOR FIXED-INTERVAL DESIGNS

In talking about the various theories that have actually been developed in any detail, we will identify them by the names used by their authors followed by a symbol i–j, where i identifies the sensory and j the decision process in the preceding classification. No model is described fully, and little more than mention is made of experiments that seem especially relevant to it. When a model has been tested for $n > 2$, we discuss the test in this section. All of the experimental work for $n = 2$ is dealt with in the following sections. Little by way of systematic confrontation of alternative models has yet taken place for $n > 2$.

A. Constant-Ratio Rule (CRR) (1a–)

Let $p(r|s)$ denote the typical entry of a confusion matrix when s is in the stimulus set S and r is in the response set R. The CRR asserts that if S' is a subset of S and R' is the response set corresponding to S', then the recognition experiment based on S' and R' has the confusion matrix

$$p(r'|s') \Big/ \sum_{x \text{ in } R'} p(x|s'),$$

where s' is in S and r' is in R' (Clarke, 1957). Supporting experimental evidence can be found in Anderson (1959), Clarke (1957, 1960), Clarke and Anderson (1957), Conrad (1964), Egan (1957a,b), Hodge (1962, 1967), Hodge, Crawford, and Piercy (1961); Hodge, Piercy, and Crawford (1961); Hodge and Pollack (1962), and Pollack and Decker (1960). Hodge and Pollack (1962) and Lee (1968) have suggested that the CRR may be a good empirical generalization when the stimuli are multidimensional but is less accurate when they are unidimensional. Clearly, this rule is of limited generality, since by varying the presentation probabilities, payoffs, or instructions, $p(r|s)$ can be affected [Nakatani (1968) and Shipley and Luce (1964) report such data]. Nonetheless, it suggests that the special case of a theory without response bias should predict the CRR or something close to it.

B. Stimulus and Response Generalization (1a–1)

Shepard (1957) proposed a recognition model ($m = n$) which has the matrix form $P_{SS}JP_{RR}$ where P_{SS} is a stochastic matrix representing stimulus generalization, J is an experimenter assigned permutation matrix mapping stimuli into responses, and P_{RR} is a response generalization matrix. He suggested how to estimate P_{SS} by running several experiments with different J's. In addition, he postulated that these probabilities of the stimulus matrix are exponentially related to an Euclidean distance, and he provided supporting experimental evidence (Shepard, 1958a,b). Applications are given in Shepard (1961a,b). These ideas grew ultimately into Shepard's very fruitful nonmetric scaling procedure (Kruskal, 1964a,b; Shepard, 1962a,b, 1963, 1966; see Carroll and Wish (Chapter 12) and Wish and Carroll (Chapter 13) of this volume), but they do not seem to have been pursued as a theory of recognition.

C. Choice Theory (1a–2)

Shipley (1960) generalized to various experimental designs some of the psychophysical models in Luce (1959); a general statement of these models was given in Luce (1963a); and they were modified to account for experimental data by Shipley (1965). Consider an $n = m$ experiment where response r_i corresponds to stimulus s_i. Assume there is a stimulus generalization function η_{ij} (not necessarily a probability measure) and a response bias function b_j such that

$$p(r_j|s_i) = \frac{\eta_{ij}b_j}{\sum_{k=1}^{n}\eta_{ik}b_k}.$$

The scale values, though not the probabilities, can be written in matrix form as

$$\begin{bmatrix} & \cdot & \\ & \cdot & \\ & \cdot & \\ \cdots & \eta_{ij}b_j & \cdots \\ & \cdot & \\ & \cdot & \\ & \cdot & \end{bmatrix} = \begin{bmatrix} & \cdot & \\ & \cdot & \\ & \cdot & \\ \cdots & \eta_{ij} & \cdots \\ & \cdot & \\ & \cdot & \\ & \cdot & \end{bmatrix} \begin{bmatrix} b_1 & 0 & \cdots & 0 \\ 0 & b_2 & \cdots & 0 \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ 0 & 0 & \cdots & b_n \end{bmatrix}.$$

Note that in the unbiased case, i.e., b_j is independent of j, this model predicts the CRR. To reduce the number of free parameters, it is assumed, as in Shepard's work, that $-\log \eta_{ij}$ acts like a distance (in particular, it is symmetric, i.e., $\eta_{ij} = \eta_{ji}$); if the stimuli have components (as when there are several listening intervals), the squared distances add like orthogonal components of an Euclidean space; and if the stimuli vary in only one dimension, distances are assumed to be additive. For complex stimuli, such as words, which account for much of the recognition data, none of these assumptions has much bite. Only by holding the stimuli fixed and manipulating the response bias by, say, varying the presentation probabilities, do the number of free parameters increase more slowly than the number of independent data. However Broadbent (1967) has shown that the word frequency effect can be predicted simply by assuming η_{ij} is related to the frequency of the word in the language, and that theories based on pure response bias, $b_i \neq b_j (\eta_{ij} = 1)$ are inadequate. This paper is criticized by Catlin (1969) and Nakatani (1970), but defended by Treisman (1971).

An experiment on lifted weights was reported by Shipley and Luce (1964). Both $n = 2$ and $n = 3$ data were collected and fit reasonably well by the choice models using the same stimulus parameters where appropriate.

Using two tones and noise, Shipley (1965) investigated 12 different YN and FC detection, recognition, and simultaneous detection-and-recognition experiments. She was led to postulate that in the mixed case, a two-stage process is involved in which recognition of the stimulus precedes a detection decision; and when only detection is required in the uncertain frequency experiment the subject behaves as if he made covert recognition responses and simply ignored them later. These ideas given an adequate account of her data which, so far, have not been analyzed in detail in terms of any other theory. However, these choice models are surely wrong because they predict symmetric ROC curves for both YN and 2AFC experiments, whereas empirically the former is usually asymmetric and the latter appears to be linear (Atkinson & Kinchla, 1965; E. F. Shipley's data in Norman 1964a).

See also the discussion of the uncertain frequency experiments in Section III,F on discriminal dispersion theory.

D. Confusion-Choice Recognition Theory (1b–2)

Nakatani (1968, 1972) proposed a theory in which the internal states are implicit responses that correspond to the possible responses. When stimulus s_i is presented, implicit response j occurs with probability a_{ij}. Thus, any subset T of the set $M = (1, 2, \ldots, m)$ of implicit responses can occur. The probability that T occurs, called an equivocation probability, is

$$\prod_{k \text{ in } T} a_{ik} \prod_{k \text{ in } M-T} (1 - a_{ik}).$$

Given that a nonempty set T occurs, the decision process is assumed to select one response from T; If $T = \varnothing$, then the process treats it as if $T = M$. It is postulated that there is a response bias vector $(b_1, b_2, \ldots, b_m)$ such that j in T is selected with probability $b_j/\sum_{k \text{ in } T} b_k$. Thus, summing over all possible sets T such that T is a subset of M and j is in T yields

$$p(r_j|s_i) = \sum_{T} \prod_{k \text{ in } T} a_{ik} \prod_{k \text{ in } M-T} (1 - a_{ik}) \left(b_j \Big/ \sum_{k \text{ in } T} b_k\right).$$

When $m = n$, the data suggest making the auxiliary assumption that the matrix A is symmetric, $a_{ij} = a_{ji}$. With this, there are $\frac{1}{2}n(n+1) + n - 1 = \frac{1}{2}n(n+3) - 1$ parameters, and so only for $n \geq 5$ do the number of independent probabilities exceed the number of parameters.

In his model, Nakatani views the probabilities a_{ij} as areas under the tails of normal distributions. By itself, this does not affect the fit to the data. However, in his 1971 paper, he used this as a way to estimate the similarity between the stimuli and responses and then used Kruskal's (1964a,b) nonmetric multidimensional scaling technique to locate stimuli and responses as points in a Euclidean space. If that space has k dimensions, this reduces the number of parameters to $(n-2)k + 1$ stimulus ones (the arbitrariness of the origin and rotations of the space drops $2k - 1$) and $n - 1$ bias ones.

Considering word-confusion data, Nakatani (1968) had considerable success with his model. Using the special case where the matrix A has identical main diagonal entries and identical off-diagonal ones and no response bias, he accounted nicely for plots of the probability of a correct

response versus the number n of words, with S/N ratio as a parameter, from data of Miller, Heise and Lichten (1951), Pollack (1959) and Rubenstein and Pollack (1963). Again, assuming no response bias, he gave as good an account as the CRR of the data in Clarke (1957) and Egan (1957a). Dorfman (1967) reported a series of 2×2 recognition experiments using four presentation probabilities and three tachistoscopic times. The key prediction of Nakatani's theory in this case is that, for each time, the data points should lie on a linear ROC curve. The data do not reject this, although a symmetric curve (as in the choice theory) appears equally adequate. In a study of his own, Nakatani used six nonsense syllables at two S/N levels and three presentation distributions. One A matrix and three b vectors were estimated and they gave a satisfactory account of the data. In a second experiment, he chose three pairs of words with little confusion between pairs and a great deal within. Presentation distributions were chosen so that, in one case, the absolute probability of a word remained about constant whereas its relative probability to its mate varied, and in the other the relative probability remained constant whereas the absolute one varied. The model predicts that articulation scores should change in the former case and not in the latter, and this was confirmed.

Using Kruskal's nonmetric multidimensional scaling technique, Nakatani (1971) analyzed confusion data for Munsell color chips (Shepard, 1958b) and for pure tones (Hodge & Pollack, 1962). The fits are impressive: The model accounts for 99.8% of the variance with 49 *df* and 99.2% with 30 *df*, respectively.

A direct confrontation of these models with the choice model, which with analogous assumptions has about the same number of parameters, has not been carried out.

E. Threshold Theory (1–2, 3)

Basically, all threshold models involve a discrete set of states, some probabilistic rule for their activation, and a response bias or a criterion response rule, or both. Many can be viewed as a discrete version of Thurstone's (1959) discriminal dispersion model (see Section III,F).

One body of literature, called neural quantum theory, is concerned with the shape of the distribution over states induced by the signal (Békésy, 1930; Stevens, Morgan, & Volkman, 1941; Corso, 1956); relevant data are discussed in Section IV,F. The other body of literature, much influenced by TSD, concerns the nature of the response criterion and the resulting ROC curve in the two-stimulus case (for a summary, see Luce, 1963a, Krantz, 1969, and Sections IV,A,D, and F).

The only attempt to use the theory for $n > 2$ was in connection with

Shipley's (1965) simultaneous detection-and-recognition data. Luce (1963a) pointed out that when one of two signals of different frequency was presented in a YN design and the subject was required to state which signal it was (independent of his detection response), there was no evidence of any residual recognition when he responded "no." Lindner (1968) carried out a systematic study in which the response bias was varied; his data suggested the opposite conclusion and rejected, at least, a two-state threshold model.

F. Discriminal Dispersion Theory (2–3)

The natural generalization of TSD to $n > 2$ was presented by Tanner (1956). Each signal relative to noise defines a likelihood ratio axis. These axes are thought of as directions in some multidimensional Euclidean space. Each signal is represented as a multinormal density in such a way that the likelihood ratio of one signal to another can be mapped to the straight line connecting the two signal points. Not only do the means and covariance matrices required to define the normal distributions introduce many parameters, but the response criteria of the $n = 2$ case generalize, even in the simplest generalization, to $\frac{1}{2}n(n-1)$ likelihood ratio criteria. As a result of this multiplicity of free parameters, this model has not been seriously pursued, and no useful generalization of TSD has been proposed. This remains the gravest weakness of TSD.

Much earlier, the successive interval or categorical judgment generalization of Thurstone's (1959) discriminal dispersion model was proposed (see Torgerson, 1958). One continued to assume an abstract one-dimensional representation (definitely not likelihood ratio) of the stimuli in which each signal is described by a normal distribution, and the m responses are represented as a partition of the continuum into m intervals. This model may be suitable for acoustic signals that vary only in intensity, but it is doubtful for word-recognition studies. If the normal distributions are correlated, the model has $\frac{1}{2}n(n-1) + n - 1 + n - 1 + m - 1 = \frac{1}{2}(n-1)(n+4) + m - 1$ free parameters. For references to the relevant literature see Torgerson (1958, p. 208) and Bock and Jones (1968, p. 213).

A careful, very general study of the discriminal dispersion model for both comparative and categorical judgments—called unidimensional strength theory—is given by Wicklegren (1968a).

A special dispersion model was used by Green and Birdsall (1964) to reanalyze the data of Miller *et al.* (1951). Essentially they assumed that the representation of a stimulus (word) is cross correlated with templates corresponding to the possible responses, that each correlation is a normally distributed random variable with unit variance and mean 0 except for the

correct one, which has mean d. It is assumed that the response is determined by the largest of the correlations. It appeared that a single function related signal-to-noise ratio to d independent of the number of words on the list, except for the 1000 word list. The distinctive feature of this theory is that each stimulus is represented as a random point in an n-dimensional space, and the decision rule does not involve any free parameters.

A key model-free phenomenon of the YN design when either of two signals ($n = 3$, $m = 2$) may be presented is that an appreciable decrement in detectability results as compared with the single signal case (Gundy, 1961; Shipley, 1965; Swets, Shipley, McKey, & Green, 1959; Tanner & Norman, 1954; Tanner, Swets, & Green, 1956; Veniar, 1958). Tanner *et al.* (1956) suggested that the hearing mechanism acts like a single filter that has to be shifted from one frequency to another. Later Green (1958) proposed a multiple filter model. Some subjects seem in accord with the one model, others with the other model. Shipley (1960) pointed out that a single stage, covert, choice model admits two solutions, which correspond closely to the two filter models. Swets and Sewall (1961) argued that a decision between the perceptual and response explanations can be made by comparing performance with a presignal presentation cue identifying the signal (if any) to that with a postsignal presentation cue. If the phenomenon is perceptual, the precue should result in the same performance as for a single signal, whereas the postcue should leave the decrement unaffected; if it is a response matter, both cues should return the performance to the level of a single signal. Swets and Sewall interpreted their data as supporting the perceptual hypothesis; actually neither cue was very effective and the support to either theory is most marginal. Their conclusion is at variance with Pollack (1959) in another context. Shipley's (1965) later two-stage, recognition-then-detection model does not allow for this differential prediction.

Greenberg and Larkin (1968) and Greenberg (1969a,b) have explored a design in which the detectabilities of various probe signals are measured when they are introduced without prior warning to the subject who is expecting only a single signal frequency. Detection is best at the expected signal frequency and decreases systematically at frequencies different from it. They failed to observe any decrement in detection when either of two signals might be used on a YN trial, despite a highly asymmetric payoff matrix favoring one of the two signals (Larkin & Greenberg, 1970). Thus, they found no evidence for a "listening strategy." However, Penner (1970), using a 2AFC design with asymmetric reward over the various signal frequencies, found detection best at the most highly rewarded frequency. If all frequencies were equally rewarded, then detection was the same at all frequencies.

G. Counter Theory (3–3)

McGill (1967) and Siebert (1965) have proposed that the sensory system converts signal energy into a "neural" pulse train with a rate that increases with intensity. Moreover, as an idealization they assume that the process is Poisson, i.e., the times between successive pulses are independent and exponentially distributed. Kiang's (1965) data on periphal auditory fibers supports the independence assumption; however, refractoriness and equilibration make the full Poisson assumption a considerable idealization. Ultimately, the exponential assumption will probably have to be replaced by a more descriptive one, but the mathematical complications will be considerable and, in all likelihood, more precise assumptions will have to be simulated on a computer.

The basic decision process assumed in the FI design is that the pulses occurring during the observation interval are counted and this number is compared with a criterion (in YN design) or with another count (in 2AFC designs). Under the Poisson assumption, explicit expressions can be calculated for all the relevant probabilities and, in general, it provides a good account of the data (McGill, 1967; McGill & Goldberg, 1968). Because the sum of a number of identically distributed random variables is approximately normally distributed, the predictions of the counter models tend to be similar to those of TSD, but without requiring the likelihood interpretation.

A modification of this theory to FR designs is outlined in Section VI,D.

IV. OPERATING CHARACTERISTICS AND PSYCHOMETRIC FUNCTIONS FOR FIXED-INTERVAL DESIGNS WITH TWO STIMULI

As we have already noted, little has been done to compare one model against another for $n > 2$. For $n = 2$, the main comparison has been between discrete and continuous sensory–memory states—this distinction served as the first component of our classificatory scheme. The issue is, basically, whether a sensory threshold exists in any measurable form. Some investigators have attempted to assess this question by examining the shape of the ROC curve.

A. The Shape of the ROC Curve

Given the same decision process—usually, a response criterion one—different theories about the sensory process predict qualitative differences

in the ROC curve [$p(\mathrm{Y}|\mathrm{s})$ versus $P(\mathrm{Y}|\varnothing)$ as the response criterion is varied]. Thus, one is led to collect such data. More often than not, the shape exhibited by the data is found largely in the eye of the beholder. Any data point estimates a point on the ROC curve with variability appearing in both dimensions, and even binomial variability, which is at best a lower bound on experimental variability, is sufficient to obscure somewhat the shape of any empirically determined curve (see Green & Swets, 1966, pp. 402–403). The only sure generalization from these earlier studies is that the high threshold model is definitely wrong (Swets, 1961; Tanner & Swets, 1954). A large amount of data appear to be fit adequately by the two straight-line segments of low threshold theory, and there are also considerable data that appear to be better fit by continuous curves. Krantz (1969) has provided a careful critique of the attempts to discriminate these hypotheses.

For data consistent with the Gaussian assumptions of TSD, the slope of the ROC curve on double probability paper is often interpreted as the ratio of the noise to the signal-plus-noise standard deviations. Wickelgren (1968b) and Nachmias (1968) have emphasized that this slope is also affected by criterion variability. Shipley (1970) found an apparent increase in criterion variance when the subject was asked to increase the number of response criteria. Markowitz and Swets (1967), who found systematic differences in the shape of binary and rating ROC curves (see Section IV,D) although the detection indices (d's) were nearly the same, suggested that the slope is related to the a priori probability of the signal's occurrence. Schulman and Greenberg (1970) found similar effects.

One safe generalization is that one seldom obtains data with a slope less than one (an exception being Shipley, 1970).* Thus, either the signal distribution is seldom less variable than the noise distribution or the criterion variability is so large compared with the variance of the sensory distributions that such changes cannot be observed.

B. Fitting a Theoretical ROC Curve to Data Points

A persistent problem in evaluating the shape of the ROC curve is the estimation of parameters. Recent work has improved this somewhat.

* Dr. Angus Craig (personal communication, 1974) has reanalyzed vigilance data reported by Colquhoun and Baddeley (1967) and by Colquhoun, Blake, and Edwards (1968) and has found that somewhat more than half of the individual ROC curves have slopes greater than one. Of course, the procedure for these experiments is somewhat different from the usual YN one, from which we made our "safe" generalization; in particular, the probability of a signal recurring is low and the subject does not make a response of no signal.

When we have only two data points, we can reformulate the question as a test of the following null hypothesis: Given a particular theory, the two points lie on the same theoretical curve. Gourevitch and Galanter (1967) gave a parametric analysis for large samples from normal distributions with equal variances. They presented an approximate estimate for the asymptotic variance of d'. Smith (1969) provided an approximate estimate for the variance of the criterion β associated with a single data point.

With three or more data points, the null hypothesis approach is awkward and is replaced by attempts to estimate the best fitting member of some family of ROC curves and to provide a measure of goodness of fit. The most general case so far studied (Abrahamson & Levitt, 1969) is the location-scale family where, for some probability density g, the noise and signal densities are given by

$$g_{\varnothing}(x) = \frac{1}{\sigma^g}\left(\frac{x}{\sigma}\right) \qquad \text{and} \qquad g_s(x) = \frac{1}{\sigma_s}\, g\left(\frac{x-d}{\sigma_s}\right).$$

The hit and false alarm probabilities are given by

$$p_H(\beta) = p(Y|s) = \int_\beta^\infty g_s(x)\,dx \quad \text{and} \quad p_F(\beta) = P(Y/\varnothing) = \int_\beta^\infty g_{\varnothing}(x)\,dx.$$

If G is the distribution function of g and if we define

$$\xi = G^{-1}(1 - p_F) \qquad \text{and} \qquad \eta = G^{-1}(1 - p_H),$$

then the (G-transformed) ROC curve is given by the linear relation

$$\xi = (\sigma_s/\sigma)\eta + d.$$

The problem is to estimate σ_s/σ and d. This is not a conventional linear regression because both ξ and η are estimated, and the joint distribution varies with (ξ, η). Madansky (1959) has given one general approximate treatment of such problems. Within the context of YN and rating-scale designs, Abrahamson and Levitt (1969) used numerical iteration to solve the resulting maximum likelihood (ML) equations. By Monte Carlo methods, they showed that Madansky's approximation is virtually identical to the ML method. They defined a goodness-of-fit statistic which is, asymptotically, related to a χ^2 variable. These results generalize earlier ones for the logistic and normal distributions applied to YN and rating methods (Ogilvie & Creelman, 1968; Dorfman & Alf, 1968, 1969; the latter is a specialization of Schonemann & Tucker's 1967 analysis of the Thurstonian model for successive intervals with unequal variances).

Various threshold models (Luce, 1963b; Norman, 1963, 1964a; Krantz, 1969) suggest that the ROC curve may be composed of several linear pieces. No statistical treatment has been given for these cases. The heart of the difficulty is in deciding which data points belong with each linear piece.

C. Sampling Variability of the Area under the ROC Curve

Green (1964; Green & Swets, 1966, p. 47) showed that the area under the YN ROC curve (as a proportion A of the unit square) equals the percentage of correct responses in the unbiased 2AFC model, and that this result is independent of the form of the assumed distributions. Pollack and Norman (1964) have provided a reasonable scheme for estimating A given only a single point on the ROC curve; see also Norman (1964b). Pollack, Norman, and Galanter (1964) have illustrated its use in a recognition memory experiment. Because of the nonparametric (although not model-free) character of Green's result, much interest is shown in A as a general unidimensional measure of sensitivity. Therefore, it is important to know something about its sampling variability. Green and Moses (1966) assumed a binomial sampling distribution, in which case the variance is $A(1 - A)/N$. This is not obviously true, and no one has yet worked out the actual sampling distribution for, say, the ML member of a location-scale family or even of a logistic or normal family. The only detailed study (Pollack & Hsieh, 1969) is a Monte Carlo exploration of A under step functions fitted to "data" points generated from normal, uniform, and exponential families. The relation between A and its standard deviation is roughly independent of the family and is slightly less than the binomial standard deviation. Among other things, they showed that correlations in the samples had little effect on this relationship.

D. Rating Operating Characteristics

Egan, Schulman, and Greenberg (1959) were the first to compare the operating characteristics determined from binary (YN) responses and ratings responses in which the subject indicated his confidence about the presence of the signal by choosing among a limited ($\sim$6) number of categories. They found good agreement between the two methods, and their results were replicated by Emmerich (1968). Markowitz and Swets (1967), however, found systematic differences between the methods as the a priori probability of the signals was varied.

Impatient with the slow accumulation of information in both of these methods, Watson, Rilling, and Bourbon (1964) pioneered the use of the rating method with a very large number of response categories. Using an auditory single-interval detection design in which the signal was a pure tone partially masked by noise, they asked subjects to position a movable rod to indicate their degree of confidence about the signal's presence. The possible positions of the rod were categorized into 31 distinct "ratings," and the operating characteristics produced by the procedure have been widely cited. The curves are remarkably smooth (as they must be, since they are cumulative in character) and the slope appears to change continuously.

Unfortunately, these plots have tended to be somewhat overinterpreted in attempts to decide whether the internal sensory states are continuous or discrete. The crux of the problem is that specific assumptions about the response process must be made in order to draw unambiguous inferences about the sensory mechanism from the data. The central issue is the assumption one makes about the decision or response process, in particular about the mapping relating responses to the internal sensory states.

It is generally admitted that the same internal state may lead to different responses—in effect, this is an assumption that some generalization takes place. There is considerable disagreement about the extent of this generalization. Clearly, the more responses that a single sensory state elicits, the less certain are the inferences made from responses to internal states. Sensory psychologists have tended to assume that extreme rating responses indicate, unambiguously, extreme detection states, whereas only the medium rating responses may be caused by either sensory state.

If such a view is adopted, a two-state threshold model predicts a rating operating characteristic that begins and ends with straight-line segments and is curved only in the middle portion, where medium ratings are encountered. This mapping between detection states and responses is clearly what Watson *et al.* (1964) had in mind in their original paper, and Nachmias and Steinman (1963) and Green and Moses (1966) assumed it explicitly. Broadbent (1966) and Larkin (1965) challenged this assumption, and the alternative hypothesis was defended by Wickelgren (1968b). Krantz (1969) has presented the most complete statement of the alternative view.

The key assumption of that alternative view is, of course, that any sensory state can lead to any response. The rationale for this position is simple and compelling. Suppose an observer has, in fact, only two sensory states, but that other, nonsensory, internal states affect his disposition toward a certain response. Then these other internal states can bias the rating responses and produce high confidence responses, even when the signal is undetected, or low ones when it is detected. So long as these other variables are partially independent of the sensory states, any sensory state may, with some probability, elicit one of the possible responses. We might think of these nonsensory variables as attentional factors.

If R_i, $i = 1, \ldots, r$, denote the possible ratings and D and $\bar{D}$ the two states, we let $\sigma_D(i) = p(R_i|D)$ and $\sigma_{\bar{D}}(i) = p(R_i|\bar{D})$. Let us make the

following two assumptions about these parameters. First, they are stochastic, i.e., $\sigma_D(i), \sigma_{\bar{D}}(i) \geq 0$ and $\sum_i \sigma_D(i) = \sum \sigma_{\bar{D}}(i) = 1$. Second, $\sigma_D(i)/\sigma_D(i)$ is nonincreasing as i moves from 1 to r, i.e., there is more likelihood that a high rating response is chosen in the presence of a detect state than a nondetect state. Thus the effect of these nonsensory states coupled with only two sensory states determines the relation between the ratio $\sigma_D(i)/\sigma_{\bar{D}}(i)$ and the response emitted.

From these assumptions and letting $q(s) = p(D|s)$ and $q(\varnothing) = p(D|\varnothing)$, the kth point on the ROC has coordinates

$$\begin{aligned} P(R \leq k|s) &= q(s) \sum_i \sigma_D(i) + [1 - q(s)] \sum_i \sigma_{\bar{D}}(i) \\ &= \sum_i \sigma_{\bar{D}}(i) + q(s) \sum_i [\sigma_D(i) - \sigma_{\bar{D}}(i)]; \\ P(R \leq k|\varnothing) &= q(\varnothing) \sum_i \sigma_D(i) + [1 - q(\varnothing)] \sum_i \sigma_{\bar{D}}(i) \\ &= \sum_i \sigma_{\bar{D}}(i) + q(\varnothing) \sum_i [\sigma_D(i) - \sigma_{\bar{D}}(i)]. \end{aligned}$$

Thus the kth point can be considered as the sum of two vectors, lying along the major diagonal one with coordinates $(\sum_i \sigma_{\bar{D}}(i), \sum_i \sigma_{\bar{D}}(i))$ and the other with coordinates $[q(s) \sum_i [\sigma_D(i) - \sigma_{\bar{D}}(i)], q(\varnothing) \sum_i [\sigma_D(i) - \sigma_{\bar{D}}(i)]]$. The term $\sum_i [\sigma_D(i) - \sigma_{\bar{D}}(i)]$ is a scaler which ranges between 0 and 1 (because of the stochastic assumptions).

Observe that the slope between successive points is simply

$$\frac{1 + q(s)\{[\sigma_D(i)/\sigma_{\bar{D}}(i)] - 1\}}{1 + q(\varnothing)\{[\sigma_D(i)/\sigma_{\bar{D}}(i)] - 1\}},$$

and so it is monotonic decreasing because $\sigma_D(i)/\sigma_{\bar{D}}(i)$ never increases. The limiting slope is $[1 - q(s)]/[1 - q(\varnothing)]$.

There is no intention of fitting data using all of these parameters, but it does suggest a plausible account of how a two-state theory can predict the smooth, apparently continuous, data of rating experiments. Krantz's assumptions (especially as expressed here) are, in effect, a multistate theory if one considers the two sensory states and the several nonsensory states needed to provide the different values for the ratio $\sigma_D(i)/\sigma_{\bar{D}}(i)$.

There is, in fact, little difference between the way continuous theory and multistate theory handle the comparison between YN and rating data. Note that if one starts by assuming two equal variance Gaussian distributions and a number of fixed response criteria, then replacing D by s and $\bar{D}$ by $\varnothing$ means that $\sigma_D(i) = p(R_i|s)$ and $\sigma_{\bar{D}}(i) = p(R_i|\varnothing)$. The parameters are clearly stochastic and the assumption that $\sigma_D(i)/\sigma_{\bar{D}}(i) =$

$p(R_i|s)/p(R_i|\varnothing)$ is nonincreasing follows simply from the likelihood ratio being monotonic in the equal variance Gaussian case. About the only observation damaging to either theory would be rating operating characteristics that lie above (i.e., include more area) than the YN ROC curves.

E. Latency Operating Characteristics

Another recent development has been the attempt to use latency measures as a means of determining the observer's confidence in his response, and from these data to construct curves claimed to be analogous to the rating operating characteristics. In this method, the subject is seldom informed that his response time is being measured, and no emphasis on speed is ever given—latencies greater than 1 sec are common. Moreover, the results of these experiments are not used, as are the reaction-time data discussed in Section VI,E, to make inferences concerning stochastic delays within the sensory processing. Rather, the idea is that latency of a response is an index of the confidence of the subject concerning his response. The basic argument, as expressed by Norman and Wickelgren (1969), is that if the response is quick, then the subject is probably sure; if it is slow, he is probably uncertain.

Consider a simple YN task of the type described earlier. Associated with each cell of the response matrix is a corresponding latency distribution. For example, let l_{sY} be the latency associated with a "yes" response given that the signal was present on that trial. Similarly, $l_{\varnothing Y}$ is the latency associated with a false alarm, $l_{\varnothing N}$ the latency of a correct rejection, and l_{sN} the latency of a false "no." Note that there is no obvious dependency among the latencies as there was among the probabilities in the stimulus–response matrix. Over many trials, the four different latency distributions can be estimated. Two treatments of these data have been suggested, one by Carterette, Friedman, and Cosmides (1965) (CFC) and the other by Norman and Wickelgren (1969) (NW). A more recent example of such curves is in Katz (1970). Because they are different, yet related, and because the techniques are new, we explain both in some detail.

The CFC method constructs two curves from the data. One is constructed from the latencies of the "yes" responses by passing a temporal criterion through the two Y distributions:

$$y_1(k) = P(l_{sY} \geq k) \cong N(l_{sY} \geq k)/N_{sY}, \qquad 0 \leq k < \infty;$$
$$x_1(k) = P(l_{\varnothing Y} \geq k) \cong N(l_{\varnothing Y} \geq k)/N_{\varnothing Y}, \qquad 0 \leq k < \infty;$$

where $N(l_{sY} \geq k)$ is the number of latencies equal to or exceeding k and N_{sY} is the total number of such latencies. Thus, both y and x range from 0 to 1 and the curve is monotonic increasing. There is no mathematical necessity that $y > x$ and, in fact, some empirical data have shown the reverse order over the entire range of k.

Another curve is constructed in an analogous fashion from the negative responses, namely,

$$y_2(k) = P(l_{\varnothing N} > k) \cong N(l_{\varnothing N} \geq k)/N_{\varnothing N}, \qquad 0 \leq k < \infty;$$
$$x_2(k) = P(l_{sN} > k) \cong N(l_{sN} \geq k)/N_{sN}, \qquad 0 \leq k < \infty.$$

In the NW scheme, these two curves are composed into a single graph by rescaling the two CFC curves as follows:

$$\beta_1(k) = P(l_{sY} \geq k)p(Y|s) = y_1(k)p(Y|s)$$

versus

$$\alpha_1(k) = P(l_{\varnothing Y} \geq k)p(Y|n) = x_1(k)p(Y|s)$$

and

$$\beta_2(c) = p(Y|s) + [1 - p(Y|s)]P(l_{nY} \geq c) = p(Y|s) + [1 - p(Y|s)]y_2(c)$$

versus

$$\alpha_2(c) = p(Y|n) + [1 - p(Y|n)]P(l_{nN} \geq c) = p(Y|n) + [1 - p(Y|n)]x_2(c).$$

Although the two curves are related and are designed to analyze similar kinds of data, neither has any great theoretical rationale. They are both simply ways of presenting data and it is as pointless to argue whether one treatment is superior to the other as it is to argue that the median is a better measure of central tendency than the mean. Either curve may be useful in certain circumstances. It is clear that theoretical work is needed to reveal their exact relation to other, more traditional ROC curves. On the face of it, there is no apparent relation.

F. Psychometric Functions

By a "psychometric function" we mean a plot of some measure of detectability against physical signal intensity, $I(s)$. The examination of such functions, as a source of information about the nature of the sensory system, has a long and venerable history (Boring, 1942). A popular approach is to deduce the shape of the function from a consideration of physical fluctuations of the stimulus. In these theories the observer is treated as having either no threshold or one that is small compared with the fluctuations in the stimulus. The physical quantum theory in vision (Hecht, Schlaer, & Pirenne, 1942; Cornsweet, 1970) or the "ideal" detector theories in audition (Green & McGill, 1970; Green & Swets, 1966; Jeffress, 1964; Pfafflin & Mathews, 1962) are examples of this approach.

We will not consider this approach further at this time, but rather we will emphasize that potentially the psychometric function can be used to try to determine whether or not the observer possesses a threshold and to estimate its size. The problem is what dependent variable one should use in determining the detectability of the signal. The classic one is $p(\mathrm{Y}|\mathrm{s})$. Since TSD and our awareness of ROC curves, there has been a temptation to plot d' versus signal-to-noise ratio, especially since a special case of TSD predicts that d' is proportional to $(E/N_0)^{1/2}$. The objection to the first measure is that it is highly affected by the subject's location on the ROC curve and that location may very well not be independent of $I(\mathrm{s})$. The objection to the latter measure is that it is special to a particular theory.

An alternative proposal is to use A, the area under the YN ROC curve, which obviously is independent of response criterion and which, as was already noted, is equal in a broad class of models to the percentage correct $p(C)$ in the (symmetric) 2AFC design. Closely related is the proposal to use

$$\Delta = p(1|\langle \mathrm{s}, \varnothing\rangle) - p(1|\langle \varnothing, \mathrm{s}\rangle),$$

which has the following advantage. If the presentation probabilities are equal in the 2AFC design, $\Delta = 2p(C) - 1 = 2A - 1$, and if they are not equal, all of the threshold theories predict that Δ should be independent of the response bias or criterion (which is equivalent to saying that the 2AFC ROC curve is a straight line with slope 1). The surprising fact is that even though the advantages of using $p(C) = A$ have been known since at least 1964 and of Δ since 1963, we do not know of any plots of these functions against signal intensity.

It was early recognized that some appropriate psychometric function should be able to decide between a threshold and a continuous theory. In particular, as Krantz (1969) later emphasized, such a test is best made at low signal levels, where threshold theories imply that there is no detection whatsoever. In practice, the attempts to see this (as well as other features expected in the psychometric function) have been based upon plots of $p(\mathrm{Y}|\mathrm{s})$ versus $I(\mathrm{s})$ for pure tones, with no artificial noise in the background, and for light flashes. Békésy (1930) and later Stevens, Morgan, and Volkmann (1941) reported such functions as favoring a particular threshold theory known as neural quantum theory. Blackwell (1963) drew the opposite conclusion for light. A survey article by Corso (1956), summarizing many studies, was inconclusive. Luce (1963b) pointed out that the ROC literature suggests that there can be serious and complex biasing

of such functions; Larkin and Norman (1964) and Norman (1963) demonstrated this empirically. These studies made clear that threshold theories could account for a wide range of shapes of psychometric functions arising from YN designs, and so these functions could not possibly decide between the two classes of theories. The conclusion does not, however, apply to Δ, and so there is much to recommend a careful 2AFC study in which Δ is determined as a function of $I(\mathrm{s})$.

G. Efficient Estimates of Single Psychometric Points

With the advent of sequential analyses in statistics (Wald, 1947; Wetherill, 1966), the early introduction of sequential or tracking methods into psychophysics (Békésy, 1947) and the widespread availability of digital computers for on-line control of psychophysical experiments, adaptive control of stimulus presentation schedules has become increasingly common. The goal is a procedure to estimate the physical stimulus required to achieve a preassigned response probability. The procedure is to be efficient (not necessarily optimal), not too complex, not too biased, and robust (insensitive to the exact underlying model). The initially widely used method of limits has come under severe theoretical criticism (Brown & Cane, 1959; Herrick, 1967, 1969, 1970; Pollack, 1968), and it has been largely replaced by one or another variant of the up-and-down (or staircase) method. The original method, apparently due to Békésy (1947) and Dixon and Mood (1948), was designed to use fixed step size to ascertain the 50% point on a distribution function (Brownlee, Hodges, & Rosenblatt, 1953; Cornsweet, 1962; Wetherill, 1963; Wetherill & Levitt, 1965, Wetherill, Chen, & Vasudeva, 1966). Variable step size was early proposed and studied (Chung, 1954; Robbins & Munro, 1951) and is now commonly used. Cornsweet (1962) and Smith (1961) suggested that by randomly interleaving two up-and-down procedures the subject will not detect the strategy of presentation, and Levitt (1968) suggested a nonrandom interleaving to test for the existence of sequential dependencies. To estimate points other than 50% on a distribution function, transformed up-and-down methods have been proposed (Campbell, 1963; Cardozo, 1966; Heinemann, 1961; Levitt & Bock, 1967; Levitt & Treisman, 1969; Wetherill & Levitt, 1965; Zwislocki, Maire, Feldman, & Rubin, 1958).

Applications of these methods include, among other studies, Adler and Dalland (1959); Békésy (1947); Blough (1955, 1958); Blough and Schrier (1963); Elliott, Frazier, and Riach (1962); Gourevitch, Hack, and Hawkins (1960); Levitt (1968); Levitt and Rabiner (1967); Symmes (1962); and Zwislocki *et al.* (1958).

When on-line computer control is available, much more complicated adaptive procedures can be entertained: Hall (1968) suggested placing the next signal at the current ML estimate, and Smith (1966) proposed a strategy that maximizes the information gained on each trial. A somewhat different adaptive procedure, based on the concepts of sequential testing, is Taylor and Creelman's (1967) PEST procedure, for which a computer program is available.

For an excellent general survey of these methods see Levitt (undated, but not earlier than 1969), and for a complementary survey of their use in animal psychophysics see Blough (1966).

V. NONSTATIONARY RESPONSE PROCESSES

Differences in assumptions about the nature of the sensory–memory process and the resulting internal representation of the stimulus has been a major focus of the previous section and will be taken up again in Section VI. Whereas the impact of the response process on the observed behavior is generally acknowledged, it is often treated as a necessary evil—something that merely clouds our view of the sensory systems. Others find the decision processes inherently interesting. Over the past 15 years a number of studies have attempted to clarify their nature and to develop an organized theory for them comparable in scope to those of the sensory processes.

The corresponding empirical investigations have centered around sequential dependencies and the effect of stimulus presentation probability on the hit and false-alarm rate. The latter is closely related to questions about ROC curves. The emphasis on sequential effects is easily understood. With only one exception (Atkinson, 1963), all theories assume that the sensory processes are statistically stationary; thus, any evidence of nonstationary behavior is interpreted as trial-by-trial changes in the decision–response process. The emphasis on how the response criterion is affected by a priori probability was stimulated in part by the glaring failure in both TSD and threshold models of the postulate that the criterion is set so as to maximize the expected payoff.

Despite the rapid development of response theories, a number of empirical factors are still not completely understood. We attempt to summarize the better-established empirical generalizations. Often, some of the earlier studies used conditions that, in retrospect, were far from optimum and hence the effects demonstrated in these studies are not impressive. It is to be hoped that the bootstrap phase is nearly over and that future work, utilizing more judicious experimental conditions, will obtain more sizable effects.

A. Studies of Sequential Effects

Although judgments in psychophysical tasks are often treated as if they arise from a Bernoulli process, it has long been known that this is, at best, an idealization. Preston (1936a,b) clearly established response dependencies in psychophysical data, and a number of later papers have repeated his basic finding. Senders and Soward (1952) present a good history of these studies as well as some data of their own. In the early 1950s, a number of papers on this topic were published, reporting studies that used mainly visual detection tasks. Verplanck, Collier, and Cotton (1952) reported a tendency to repeat the last response. Correlations among the responses extended over a period of about 1 min in their study. Similar effects were also reported by Verplanck, Cotton, and Collier (1953) and Howarth and Bulmer (1956). Sequential dependencies using auditory detection were also demonstrated by Day (1956), Shipley (1961) and Speeth and Mathews (1961). Day also showed that the size of the sequential effect diminishes as a function of the length of the interstimulus interval. Since about 1960, a number of studies have investigated various determinants of sequential effects and have contrasted experimental results with theoretical models of the response process.

Sequential dependencies arise naturally in models that postulate a decision criterion that is updated from trial to trial. Among the models of this type are those of Atkinson (1963), Atkinson, Carterette, and Kinchla (1962), Dorfman and Biderman (1971), Kac (1962), Luce (1963b, 1964), Bush, Luce, and Rose (1964), Norman (1962, 1964a), and Schoeffler (1965). The advantage of these models over purely empirical estimates of the sequential probability is that they require much less data in order to estimate the magnitude of the effect.

B. Size of the Sequential Effects

Stating the size of a sequential effect is still largely a subjective matter. There simply is no standard way to record the results or for assessing the magnitude of the effects. The null hypothesis is that the occurrence of some response is a Bernoulli process with an unknown probability p. By computing various conditional probabilities, one may show that the occurrence of some response deviates significantly from the value p. A subjective element arises both in determining how large a deviation one believes is really important and in evaluating how much of the past history is needed in order to achieve the given deviation. For example, if the overall probability of a "yes" response is .5, and the conditional probability of a "yes" response given a previous "yes" response is .6, then clearly a fairly interesting sequential dependency has been uncovered. On the other hand, if one needs a sequence of five or six previous responses in order to achieve deviations of 10% from the overall mean, then because such sequences occur so infrequently, the deviation is less important. No one has yet suggested a method for combining the size of the deviation and the length of the conditioning sequence into a single measure. Thus, it is frequently impossible to compare the magnitudes of the sequential effects in two different experiments, despite the fact that the authors used the same number of responses and conditionalized on very similar events.

C. Generalizations about Sequential Dependencies

Although the preceding remarks make clear why it is difficult to compare different studies, we nonetheless offer the following generalizations concerning response dependencies. Each conclusion is stated in terms of a single variable because there have been very few studies that provide information about potential interactions among variables. Sequential effects are as follows:

1. small in simple auditory detection situations (Atkinson, 1963; Atkinson & Kinchla, 1965; Carterette, Friedman, & Wyman, 1966; Friedman, Carterette, Nakatani, & Ahumada, 1968a,b);
2. possibly large in simple visual detection situations (Kinchla, 1964); however, the design of this study may mean it was really a probability prediction rather than a detection design, and it is well known that sequential effects are large in probability prediction designs;
3. relatively large in recognition situations (at least of intensity) (Kinchla, 1966; Tanner, Haller, & Atkinson, 1967; Tanner, Rauk, & Atkinson,1970);
4. smaller when feedback is given then when it is not (Atkinson & Kinchla, 1965; Kinchla, 1966; Tanner, Haller, & Atkinson, 1967; Tanner, Rauk, & Atkinson, 1970).

Based in part on these generalizations, Tanner, Rauk, and Atkinson (1970) have attempted to state a fairly complete model dealing with these various factors. The following summary illustrates the general nature of models in this area. Given a stimulus presentation, the observer is assumed to compare it with whatever standard is available to him. For a sine wave in noise, the noise itself plays the role of the standard. In this case, very little memory is involved in the decision making, and, hence, the sequential effects are small. In experiments in which the task is to recognize the more intense of two 1000-Hz tones, the basic decision process is treated as a comparison of the stimulus presentation with a memory trace, which is

influenced by the past presentations and so changes from trial to trial. The observer's response is assumed to result from a weighted combination of two tendencies: One is to repeat the response he made on the previous trial and the other is to report the signal labeled correct by the feedback that occurred on the preceding trial. These tendencies are combined according to a linear weight. If there is no feedback, the tendency to repeat the previous response is given all of the weight. Thus the model can account for the smaller sequential effects observed with feedback.

This model exhibits several interesting effects. One is sequential dependencies (because the pattern of previous stimuli and responses influence the present decision). A second is a very interesting prediction about the effect of presentation probabilities on the response probabilities, a topic we take up in the next section. A third implication of this general view concerns the role played by a standard in a psychophysical judgment task. Although relatively large sequential effects have been found in the intensive discrimination task, Parducci and Sandusky (1970) report relatively small ones in an experiment having, as an explicit standard, a constant intensity signal preceding each trial. The view of memory introduced by Tanner *et al.* is reminiscent of adaptation level theory, Helson (1964), a similarity which has been pursued in papers by Parducci and Marshall (1962) and Parducci and Sandusky (1965).

D. The Effect of a Priori Probability on Response Bias

As we noted previously, the original impetus for many of the adaptive models of response criteria was the attempt to specify how the subject's response changes as a function of values and costs and the a priori probability of the stimulus alternatives. One of the early dramatic findings was that the change in the criterion as a function of a priori probability could be opposite to that predicted by the expected-value model. Kinchla (1966) and Tanner *et al.* (1967) reported data in which the observer decreased his false-alarm rate when the a priori probability of the signal was increased. This finding is at odds with considerable previous data obtained in simple detection experiments. The crucial difference appears to be whether or not the subject is given feedback and his degree of experience in these experiments. Naive subjects with no feedback invariably reduce their false-alarm rates as the signal probability is increased. With feedback, the opposite tendency is observed (Tanner *et al.*, 1967).

Although this finding is accounted for by the general model proposed by Tanner *et al.*, another quite simple explanation of it can be given. Suppose that the observer is trying to maintain an equal number of yes and no responses in the experimental situation. A tendency to equalize the frequency of all the available responses is a well-known property of many judgment tasks, as was first demonstrated by Arons and Irwin (1932). If one attempts to achieve this goal by dividing a sensory scale in such a way that half the responses will be yes and the other half no, then the criterion must be increased when the signal is presented more frequently (since more drawings are presented from the signal distribution) and so the false-alarm rate will decrease. Note that this view is consistent with a stationary observer, one whose response dependencies do not change from trial to trial. For more complete discussions of the matching hypothesis to account for the change in the hit and false-alarm rate caused by altering the a priori probabilities, see Creelman and Donaldson (1968), Dorfman (1969), Parks (1966), and Thomas and Legge (1970).

Tanner *et al.* argued that both the change in false-alarm rate with a priori probability and the large sequential effects support their view that a memory process is heavily involved in these kinds of tasks. Parducci and Sandusky (1970), however, exhibit this change in false-alarm rate with a priori probability without finding any sequential effects. Thus, presumably, the presence of the standard in the Parducci and Sandusky experiment minimizes the changes in criterion on a trial by trial basis, but nonetheless the criterion did change systematically in response to changes in a priori probability. Further work is needed to clarify this general area. Unfortunately, the next series of studies must concentrate on various interactions among the major variables.

VI. THEORIES FOR FREE RESPONSE DATA

Although we do not wish to contend that the setting for all laboratory measurements must simulate natural ones, one should realize that the study of detection and discrimination by means of fixed interval (FI) designs is highly artificial. Most signals in nature occur at unpredictable times, and one is seldom quizzed in an interval just following a potential signal about whether or not one seemed to be present. The reason for using a FI procedure is that it allows one to avoid several thorny questions concerning the temporal properties of the detection process. It is exactly these questions that one must face in trying to extend the analysis of a detection or discrimination mechanism to more practical and realistic situations.

A. A Methodological Note

In all free response (FR) experiments, an attempt is made to schedule signal presentations so that the subject is unable to predict their arrival.

The complete achievement of this goal is, perhaps, more subtle than is often realized. In any event, the schedules employed in some of the studies do not achieve the desired ends and, in fact, should strongly reinforce behavior that is highly nonuniform over time. Because this problem has not been discussed explicitly in many of the papers and because it is important, the following remarks seem in order.

Probably because of ease of programming, the most common presentation schedule is the discrete approximation to the rectangular (uniform) distribution over a fixed interval. For example, with an interval of 100 sec, signals might be presented with equal probability at each of the second marks of the 100 sec after the preceding signal. The mean time to the next signal is approximately 50 sec, and there is considerable randomness in the time of presentation. The difficulty with this schedule is that the conditional density of the signal is not uniform throughout the interval: Given that it has failed to occur up to some point in time, the probability that it will occur in the next second is a monotonic increasing function of the duration of the wait. In particular, the probability that it will occur in the first interval is 1/100, whereas the probability that it will occur in the last interval, given that it has not occurred in any of the previous ones, is 100 times as great. The potential bias that can result from this depends upon the degree to which the subject can estimate when the last signal occurred, and so it will interact with signal level. Even if no bias is evident at low intensities, this may not be true for higher intensities. It is clear that, in some circumstances, subjects make use of this information and, in fact, in a reaction time experiment by Nickerson (1967), the number of false anticipations increased monotonically with the wait from the warning light.

To avoid this problem and so the possibility of encouraging subjects to adopt nonhomogeneous response strategies, one should use the correct temporal analogue to uniform uncertainty, namely the Poisson process in which the times between successive events are independent and have a common exponential distribution. A crucial property of this distribution is the fact that the probability of an event occurring in the next instant in time is independent of how long it has been since the last event. Put another way, the conditional density of a signal occurring is constant. Experimentally, the major difficulty in using this distribution is that successive events are often quite close together in time, and so it becomes difficult to distinguish which event caused a later response (see Luce & Green, 1970). With a low signal rate, this is rarely a problem, and so it is unfortunate that the exponential schedule (or its discrete analogue, the geometric distribution) is seldom used in vigilance tasks, for the bias introduced by the uniform distribution may be considerable.

B. Temporal Partitioning

Although FI designs are special cases of FR ones, most analyses have attempted to reduce FR data to FI models. In practice, the motivation has been largely expediency, not a belief that the organism actually quantizes time and so is a kind of sample data system. If the latter were actually true and if we could divide time so as to agree with the natural quanta, then it would be perfectly reasonable to treat the FR situation as a sequence of yes–no experiments. Although the hypothesis that organisms quantize time has existed for some time (e.g., Stroud, 1955; and White, 1963), Kristofferson (1965, 1966, 1967a,b, 1969) is the only contemporary psychophysicist who argues seriously for it. His evidence consists of the numerical agreement of three parameters estimated from successive discrimination and two reaction experiments. From these data, he estimates the time quantum to be about 50 msec, which it should be noted is about half the alpha rhythm period. Although both of the present authors are skeptical, as are others (for example, Carterette, 1969), it is important to realize how crucial this unresolved point is for finding the appropriate generalization of present detection models.

The simplest practical procedure for partitioning time, one often followed in the analysis of vigilance data and first adopted by Broadbent and Gregory (1963) and Mackworth and Taylor (1963), divides time into nonoverlapping intervals of equal duration, usually about 1 sec, although other durations have been used both by the original authors and by subsequent investigators. They estimate the hit probability as the relative frequency of positive responses in the interval following the signal and estimate the false-alarm probability as the relative frequency of positive responses in all other intervals. The interval is chosen to be sufficiently large so that even the slowest responses are not misclassified as false alarms. Because the density of false alarms is usually quite low, there is little chance of making the opposite misclassification. These probabilities are then used in a TSD yes–no analysis to estimate d' and β.

Egan, Greenberg, and Shulman (1961) and Watson and Nichols (1966) carried out a somewhat more subtle analysis. In their experiments the number of false positive responses was at least an order of magnitude higher than usually is encountered in vigilance experiments; hence, the problem of misclassifying false alarms as hits becomes much more serious, and so requires more care. They constructed empirical histograms showing the frequency of response following the onset of a signal. Naturally, there is a high response rate immediately following signal onset; this is especially so with loud signals. By 1 or 2 sec after the signal interval, this rate returns to a fairly stable value, which they treated as a quantity similar to the

false-alarm probability in the FI design. Their analogue of the hit probability in the FI design is the proportion of responses occurring between the signal onset and the return to the base rate. They then proceeded as follows: Suppose the observer "divides time into a succession of subjective intervals, each of duration T_σ. It will be considered that each of these subjective intervals implicitly defines a trial for the listener, and that he makes a decision after each interval." Obviously the observed rate is then proportional to the probability of each response, the constant of proportionality being T_σ. The ROC data generalized by varying the subject's criterion yields an estimate of d' which is close to the FI one, e.g., 1.29 and 1.55. Egan *et al.* pointed out two major defects with this procedure. As the subject relaxes his criterion, the data suggest an apparent increase in the detection index, and the method does not lead to an independent estimate of the subject's criterion.

Another important approach involving temporal partitioning is based on sequential decision making. One class of empirical studies allows the observer to determine how many intervals he observes before making a response (Swets & Green, 1961; Swets & Birdsall, 1967). And one class of theoretical studies of reaction time assumes that the subject partitions time, observes a random variable in each interval, and arrives at a response decision by means of a sequential decision procedure (Carterette, 1966; Edwards, 1965; La Berge, 1962; Laming, 1968; Stone, 1960).

C. Vigilance

A vigilance task is, by definition, a free response one in which the signal rate is low (perhaps one per minute) and the total observation period is long (an hour or two); a general reference is Buckner and McGrath (1963). A major empirical generalization from a long history of research in this area is that there is a very marked decrement in performance as the period of the watch increases (a good summary of these data is Jerison and Pickett, 1963).

Broadbent and Gregory (1963) applied TSD to the data simply by dividing the FR situation into 1-sec intervals, as already explained. Assuming the signal and noise distributions are both Gaussian, with equal variance, they estimated both d' and β. The surprising finding was that although β changed during the course of the watch, d' appeared to remain constant. They argued that the decline in performance was, in fact, simply a shift on the part of the observer to a more conservative response criterion. They argued from the apparent constancy of d' that sensitivity is independent of the period on the watch. At about the same time, however, Mackworth and Taylor (1963) applied TSD to a continuous visual display and found a systematic decline in d' as a function of watch time. It appears that the modality of the display plays an important role in whether or not a decline in sensitivity is observed. Apparently the decline is related to some sort of attentional variable and may be related to observing responses, as argued by Loeb and Binford (1964), Jerison, Pickett, and Stenson (1965), and Jerison (1967). In auditory experiments, where no orienting or observing responses are required, the data show little change in sensitivity but large changes in criterion as a function of the time in the watch (Broadbent & Gregory, 1965; Davenport, 1968; Levine, 1966; Loeb & Binford, 1964; Mackworth, 1968; Hatfield & Soderquist, 1970).

Mackworth (1965) tried a variety of different temporal intervals to estimate the hit and false alarm rate. She concluded that a 30-fold change in the size of the interval has little effect upon the conclusion one draws. Although such stability is impressive, it must be remembered that in vigilance situations the false-alarm rate is extremely low. Indeed, in many of the experiments, when something like 10^4 responses are counted as hits, the number of false alarms is of the order of 10. With such tiny probabilities, it is hardly surprising that the size of the interval used to estimate them has little effect on the estimate. A more serious problem, one recognized by practically all investigators, is the stability of the estimates. Given the paucity of false alarms, the confidence intervals on both d' and β are enormous. Also, as Mackworth and Taylor (1963) pointed out, the estimates depend heavily on the assumed form of the distribution. This is not too serious as long as we only wish to compare parameter estimates within similar conditions of an experiment, but the extrapolation of these parameter estimates to other situations, especially to FI ones, is indeed hazardous. This point, along with some other cautions about the overacceptance of TSD, is well expressed in Jerison's paper (1967).

D. A Continuous Free-Response Model

Luce (1966) proposed an approach to the analysis of free response data which does not involve any arbitrary division of time into decision intervals. The most striking feature of this approach is that it redefines the basic data of a free response experiment as a family of temporal distributions: the distribution of times from a signal to the next response, the distribution of times between successive responses given that no signal has intervened, and so forth. What a theory of free-response behavior must account for are these temporal distributions, rather than some artificially constructed quantities analogous to the hit and false-alarm rate of an FI experiment.

Such notions as the hit and false-alarm probabilities are, at best, uncertain concepts because, when false alarms can occur and when the time between a stimulus initiating a response and its response is not constant, it is impossible to be sure what preceding event caused a given response.

The original model has undergone considerable experimental testing and modification in Green and Luce (1967, 1971) and Luce and Green (1970, 1971). The present version can be formulated in terms of three main ideas. First, as in the counter models (Section III,G), the intensity of a signal is assumed to be transduced and ultimately represented at the observer's decision center as a temporal pulse train of (perhaps, neural) events whose average rate is assumed to increase with signal intensity. A more specific assumption, which ultimately may have to be abandoned, is that the transduction of a signal of constant intensity is a Poisson process. In this case, the pulse rate, which is the reciprocal of the expected interpulse time, completely characterizes the process. The pulse rate due to noise ν and that due to signal-plus-noise μ play roles somewhat similar to β and d' in TSD.

The second assumption is that this pulse train is subjected to a decision process which determines when and which response is to be activated. The time taken to reach such a decision, which depends upon the pulse rate, is called the *sensory-decision latency*. A variety of decision rules in addition to counter models are possible. Initially, Green and Luce investigated the simplest possible rule, namely that the arrival of each pulse activates a response. (If so, these theoretical pulses are surely not the same as the peripheral ones which, even in the absence of an auditory signal, often have rates of about 10 per second.) Ultimately, they showed that this rule is untenable (see the next subsection). The next simplest model, still not involving a time partition, assumes that the momentary pulse rate (and hence, by the first assumption, the intensity of the signal) is estimated from the reciprocal of the interarrival times (IAT) of the pulses and that certain sums of IATs are compared with a criterion or with another sum of IATs, much as likelihood ratios are compared with a criterion or with each other in TSD. Such rules can be applied to a variety of psychophysical designs, including all the usual FI ones as well as free response types (Luce & Green, 1972).

The third and last feature of the theory concerns delays introduced by the afferent and motor systems. The sum of all these is called the *residual latency*. This latency is assumed to convolve independently with the decision latency to produce the observed response time. The only assumption made about the residual latency is that it is a bounded random variable. According to the theory, reaction times to intense signals are approximately the residual latencies, so the bound is probably about 200–300 msec. Aside from being bounded, the distribution of residual times is not otherwise constrained in the theory; undoubtedly, it depends on the input channel and on the exact response required.

The key role of the boundedness assumption is this: any observable density that is a convolution of a decision latency whose tail is exponential with a residual latency has a tail (beyond the bound) that is entirely dictated by the tail of the decision density. This is of great interest because the tail of the decision latency reflects something of the decision rule and the interpulse density. With weak signals, these tails represent an appreciable fraction of all the data, and so they can be used to test hypotheses about the nature of the decision process and to estimate mean pulse rates. For example, in Green and Luce (1967) and Luce and Green (1970) certain signal–response and response–response densities were shown to have approximately exponential tails; from these the pulse rates ν and μ were estimated, μ/ν was shown to increase smoothly with S/N in decibels, and criterion changes altered both in a way such that $\mu-\nu$ is approximately constant.

A basic difficulty in applying this type of analysis to FR data arises when two signals occur close together, in which case the response to the second may be initiated before the response to the first is completed. This is especially an issue when the signal presentation schedule of the signals is Poisson, since short intersignal times are common. Using schedules other than the Poisson for the signal presentation greatly complicates the mathematical analysis and is bound to invite nonhomogeneous strategies on the part of the observer, as we already discussed.) Some simple assumptions were explored and rejected, and more complex ones have so far proved mathematically intractable. This led to consideration of experiments in which freedom not to respond was retained, but in which multiple responses were avoided; we turn to these next.

E. Related Reaction Time Experiments

For the stochastic models of the detection process just outlined, a natural source of information about them is a reaction-time (RT) experiment with exponentially distributed foreperiods and weak signals. Although the literature on simple RT is enormous, that having to do with very weak signals is quite small. Yet such signals are admirably suited to study the decision process since they generate processing delays that are appreciable with respect to other lags in the system. In a sense, weak signals serve as a microscope to provide a more detailed view of the sensory part of the perceptual system. The nature of these delays is predicted in detail by the stochastic model. Green and Luce (1971) and Luce and Green (1970, 1972) have

pursued this approach. They show an increase in RT of over an order of magnitude as signal-to-noise ratio is decreased. In addition, sizable effects were observed when the criterion of the subject was altered and some dependence on mean signal wait was also noted.

Using the assumptions stated in the preceding section and assuming that each pulse initiates a response, they employed Fourier transforms of the observed RT distribution to try to deconvolve the sensory decision process from the remaining residual distribution. They found that the calculated residual distribution was bounded, as predicted by the theory, but was negative for a short period after the bound (which, of course, is impossible). Moreover, the mean of this inferred residual distribution was about 100 msec longer than the response times observed in simple RT to very strong signals. This is inconsistent with their model since if a bounded number of IATs enter into any response decision, as would be the case with a finite buffer store, the time for processing the pulse train vanishes as $\mu \to \infty$; hence the RT to a very strong signal should be identical to the residual distribution. Although some aspect of that model, most likely the assumption that single pulses activate responses, is surely incorrect, the authors believe that the approach has considerable promise.

Since the observed RT distribution for strong signals approximates the residual latency distribution, and the RT distribution for weak and moderate signals is the convolution of the corresponding decision latency with this RT distribution for strong ones, there is no reason to expect this convolution to have any simple mathematical form. The extraction of information from RT distributions is therefore a matter of considerable delicacy and probably requires more complicated techniques of analysis than has been traditionally assumed.

References

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

Adler, H. E., & Dalland, J. I. Spectral thresholds in the starling. *Journal of Comparative and Physiological Psychology*, 1959, **52**, 438–445.

Anderson, C. D. The constant-ratio rule as a predictor of confusions among visual stimuli of brief exposure duration. Indiana University, Hearing and Communication Lab., TN-58-60, 1959.

Arons, L., & Irwin, F. W. Equal weights and psychophysical judgments. *Journal of Experimental Psychology*, 1932, **15**, 733–751.

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

Atkinson, R. C., Carterette, E. C., & Kinchla, R. A. Sequential phenomena in psychophysical judgments: a theoretical analysis. *I.R.E. Transactions of Information Theory*, 1962, **8**, 155–162.

Atkinson, R. C., & Kinchla, R. A. A learning model for forced-choice detection experiments. *British Journal of Mathematical and Statistical Psychology*, 1965, **18**, 183–206.

Békésy, G. von. Über das Fechneresche Gesetz und seine Bedeutung für die Theorie der akustishe Beobachtungsfehler und die Theorie des Hörens. *Annalen der Physik*, 1930, **7**, 329–359.

Békésy, G. von. A new audiometer. *Acta Oto-laryn*, 1947, **35**, 411–422.

Blackwell, H. R. Neural theories of simple visual detection. *Journal of the Optical Society of America*, 1963, **53**, 129–160.

Blough, D. S. A method for tracing dark adaptation in the pigeon. *Science*, 1955, **121**, 703–704.

Blough, D. S. A method for obtaining psychophysical thresholds from the pigeon. *Journal of the Experimental Analysis of Behavior*, 1958, **1**, 31–43.

Blough, D. S. The study of animal sensory processes by operant methods. In W. K. Honig (Ed.), *Operant behavior*. New York: Appleton 1966. Pp. 345–379.

Blough, D. S., & Schrier, A. M. Scotopic spectral sensitivity in the rhesus monkey. *Science*, 1963, **139**, 493–494.

Bock, R. D., & Jones, L. V. *The measurement and prediction of judgment and choice*. San Francisco: Holden-Day, 1968.

Boring, E. *Sensation and perception in the history of experimental psychology*. New York-London: D. Appleton-Century, 1942.

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

Broadbent, D. E. Word-frequency effect and response bias. *Psychology Review*, 1967, **74**, 1–15.

Broadbent, D. E., & Gregory, Margaret. Vigilance considered as a statistical decision. *British Journal of Psychology*, 1963, **54**, 309–323.

Broadbent, D. E., & Gregory, M. Effects of noise and of signal rate upon vigilance analyzed by means of decision theory. *Human Factors* 1965, **7**, 155–162.

Brown, J., & Cane, Violet An analysis of the limiting method. *British Journal of Statistical Psychology*, 1959, **12**, 119–126.

Brownlee, K. A., Hodges, J. L., & Rosenblatt, M. The up-and-down method with small samples. *Journal of the American Statistical Association*, 1953, **48**, 262–277.

Buckner, D. N., & McGrath, J. J. (Eds.), *Vigilance: A symposium*. New York: McGraw-Hill, 1963.

Bush, R. R., Luce, R. D., & Rose, R. M. Learning models for psychophysics. In R. C. Atkinson (Ed.), *Studies in mathematical psychology*. Stanford: Stanford University Press, 1964. Pp. 201–217.

Campbell, R. A. Detection of a noise signal if varying duration. *Journal of the Acoustical Society of America*, 1963, **35**, 1732–1737.

Cardozo, B. L. A sequential up-and-down method. *Institut voor Perceptie Onderzock, Annual Progress Report*, 1966, **1**, 110–114.

Carterette, E. C. Random walk models for reaction times in signal detection and recognition. *Proceedings of the XVIII International Congress on Psychology*, 1966, **16**, 84–95.

Carterette, E. C. Symposium on attention. *Proceedings of the XIX International Congress on Psychology*, 1969, London: British Psychological Society, 1971. Pp. 226–227.

Carterette, E. C., Friedman, M. P., & Cosmides, R. Reaction-time distributions in the detection of weak signals in noise. *Journal of the Acoustical Society of America*, 1965, **38**, 531–542.

Carterette, E. C., Friedman, M. P., & Wyman, M. J. Feedback and psychophysical variables in signal detection. *Journal of the Acoustical Society of America*, 1966, **39**, 1051–1055.

Catlin, J. On the word-frequency effect. *Psychology Review*, 1969, **76**, 504–506.

Chung, K. L. On a stochastic approximation method. *Annals of Mathematical Statistics*, 1954, **25**, 463–483.

Clarke, F. R. Constant-ratio rule for confusion matrices in speech communication. *Journal of the Acoustical Society of America*, 1957, **29**, 715–720.

Clarke, F. R. Confidence ratings, second-choice responses, and confusion matrices in intelligibility tests. *Journal of the Acoustical Society of America*, 1960, **32**, 35–46. Reprinted in J. A. Swets (Ed.), *Signal detection and recognition by human observers*. New York: Wiley, 1964. Pp. 620–648.

Clarke, F. R., & Anderson, C. D. Further test of the constant-ratio rule in speech communication. *Journal of the Acoustical Society of America*, 1957, **29**, 1318–1320.

Colquhoun, W. P. & Baddeley, A. D. Influence of signal probability during pretraining on vigilance decrement. *Journal of Experimental Psychology*, 1967, **73**, 153–155.

Colquhoun, W. P., Blake, M. J., & Edwards, R. S. Experimental studies of shift-work I: a comparison of "rotating" and "stabilized" 4-hour shift systems. *Ergonomics*, 1968, **11**, 437–453.

Conrad, R. Acoustic confusions in immediate memory. *British Journal of Psychology*, 1964, **55**, 75–84.

Cornsweet, T. N. The staircase-method in psychophysics. *American Journal of Psychology*, 1962, **75**, 485–491.

Cornsweet, T. N. *Visual perception*. New York: Academic Press, 1970.

Corso, J. F. The neural quantum theory of sensory discrimination. *Psychology Bulletin*, 1956, **53**, 371–393.

Creelman, C. D., & Donaldson, W. ROC curves for discrimination of linear extent. *Journal of Experimental Psychology*, 1968, **77**, 514–516.

Davenport, W. G. Auditory vigilance: the effects of costs and values on signals. *Australian Journal of Psychology*, 1968, **20**, 213–218.

Day, W. F. Serial non-randomness in auditory differential thresholds as a function of interstimulus interval. *American Journal of Psychology*, 1956, **69**, 387–394.

Dixon, W. J., & Mood, A. M. A method for obtaining and analyzing sensitivity data. *Journal of the American Statistical Association*, 1948, **43**, 109–126.

Dorfman, D. D. Recognition of taboo words as a function of a priori probability. *Journal of Personality and Social Psychology*, 1967, **7**, 1–10.

Dorfman, D. D. Probability matching in signal detection. *Psychonomic Science*, 1969, **17**, 103.

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 and determination of confidence intervals—rating-method data. *Journal of Mathematical Psychology*, 1969, **6**, 487–496.

Dorfman, D. D., & Biderman, M. A learning model for a continuum of sensory states. *Journal of Mathematical Psychology*, 1971, **8**, 264–284.

Edwards, W. Optimal strategies for seeking information: models for statistics, choice reaction times and human information processing. *Journal of Mathematical Psychology*, 1965, **2**, 312–329.

Egan, J. P. Message repetition, operating characteristics, and confusion matrices in speech communication. Indiana Univ., Hearing and Communication Lab., Tech. Rep. 57-50, 1957. (a)

Egan, J. P. Monitoring task in speech communication. *Journal of the Acoustical Society of America,* 1957, **29,** 482–489. (b)

Egan, J. P., Greenberg, G. Z., & Schulman, A. I. Operating characteristics, signal detectability, and the method of free response. *Journal of the Acoustical Society of America,* 1961, **33,** 993–1007. [Reprinted in J. A. Swets (Ed.), *Signal detection and recognition by human observers.* New York: Wiley, 1964. Pp. 316–348.]

Egan, J. P., Schulman, A. I., & Greenberg, G. Z. Operating characteristics determined by binary decisions and by ratings. *Journal of the Acoustical Society of America,* 1959, **31,** 768–773. [Reprinted in J. A. Swets (Ed.), *Signal detection and recognition by human observers.* New York: Wiley, 1964. Pp. 172–186.]

Emmerich, D. S. ROCs obtained with two signal intensities presented in random order, and a comparison between yes–no and rating ROCs. *Perception & Psychophysics,* 1968, **3,** 35–40.

Elliott, D. N., Frazier, L., & Riach, W. A tracking procedure for determining cat's frequency discrimination. *Journal of the Experimental Analysis of Behavior,* 1962, **5,** 323–328.

Friedman, M. P., Carterette, E. C., Nakatani, L. H., & Ahumada, A. Comparison of some learning models for response bias in signal detection. *Perception & Psychophysics,* 1968, **3,** 5–11. (a)

Friedman, M. P., Carterette, E. C., Nakatani, L., & Ahumada, A. Feedback and frequency variables in signal detection. Tech. Rpt. #30, Human Communication Laboratory, Dept. Psychol., UCLA, 1968. (b).

Gourevitch, G., Hack, M. H., & Hawkins, J. E., Jr. Auditory thresholds in the rat measured by an operant technique. *Science,* 1960, **131,** 1046–1047.

Gourevitch, V. & Galanter, E. A significance test for one parameter isosensitivity functions. *Psychometrika,* 1967, **32,** 25–33.

Green, D. M. Detection of multiple component signals in noise. *Journal of the Acoustical Society of America,* 1958, **30,** 904–911. [Reprinted in J. A. Swets (Ed.), *Signal detection and recognition by human observers,* New York: Wiley, 1964. Pp. 491–507.]

Green, D. M. General prediction relating yes–no and forced choice results. *Journal of the Acoustical Society of America,* 1964, **36,** 1042 (abstract).

Green, D. M., & Birdsall, T. G. The effect of vocabulary size on articulation score. In J. A. Swets (Ed.), *Signal detection and recognition by human observers: contemporary readings.* New York: Wiley, 1964. Pp. 609–619.

Green, D. M., & Luce, R. D. Detection of auditory signals presented at random times. *Perception & Psychophysics,* 1967, **2,** 441–450.

Green, D. M., & Luce, R. D. Detection of auditory signals presented at random times III. *Perception & Psychophysics,* 1971, **9,** 257–268.

Green, D. M., & McGill, W. J. On the equivalence of detection probabilities and well-known statistical quantities. *Psychological Review,* 1970, **77,** 294–301.

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. *Signal detection theory and psychophysics.* New York: Wiley, 1966.

Greenberg, G. Z. Frequency selectivity during amplitude discrimination of signals in noise. *Journal of the Acoustical Society of America,* 1969, **45,** 1438–1442. (a)

Greenberg, G. Z. Frequency-selective detection at three signal amplitudes. *Perception & Psychophysics,* 1969, **6,** 297–301. (b)

Greenberg, G. Z., & Larkin, W. D. Frequency response characteristics of auditory observers detecting signals of a single frequency in noise: The probe-signal method. *Journal of the Acoustical Society of America,* 1968, **44,** 1513–1523.

Gundy, R. F. Auditory detection of an unspecified signal. *Journal of the Acoustical Society of America,* 1961, **33,** 1008–1012. [Reprinted in J. A. Swets (Ed.), *Signal detection and recognition by human observers.* New York: Wiley, 1964. Pp. 187–198.]

Hall, J. L. Maximum-likelihood sequential procedure for estimation of psychometric functions. *Journal of the Acoustical Society of America,* 1968, **44,** 370 (abs.).

Hatfield, J. L., & Soderquist, D. R. Coupling effects and performance in vigilance tasks. *Human Factors,* 1970, **12,** 351–361.

Hecht, S., Schlaer, S., & Pirenne, M. H. Energy, quanta, and vision. *Journal of General Physiology,* 1942, **25,** 819–840.

Heinemann, E. G. The relation of apparent brightness to the threshold for differences in luminance. *Journal of Experimental Psychology,* 1961, **61,** 389–399.

Helson, H. *Adaptation-level theory.* New York: Harper, 1964.

Herrick, R. M. Psychophysical methodology: Comparison of thresholds of the method of limits and of the method of constant stimuli. *Perceptual and Motor Skills,* 1967, **24,** 915–922.

Herrick, R. M. Psychophysical methodology: comparisons within the method of limits. *Perceptual and Motor Skills,* 1969, **28,** 503–514.

Herrick, R. M. Psychophysical methodology: Deductions from the phi-gamma hypothesis and related hypotheses. *Perception & Psychophysics,* 1970, **7,** 73–78.

Hodge, M. H. The constant-ratio rule and identification tasks. Univ. of Georgia, Dept. of Psychology, Final report, 1962.

Hodge, M. H. Some further tests of the constant-ratio rule. *Perception & Psychophysics,* 1967, **2,** 429–437.

Hodge, M. H., Crawford, M. J., & Piercy, M. L. The constant-ratio rule and visual displays. Univ. of Georgia, Dept. of Psychology, Report 2, 1961.

Hodge, M. H., Piercy, M. L., & Crawford, M. J. The constant-ratio rule and lifted weights. Univ. of Georgia, Dept. of Psychology, Report 1, 1961.

Hodge, M. H., & Pollack, I. Confusion matrix analysis of single and multidimensional auditory displays. *Journal of Experimental Psychology,* 1962, **63,** 129–142.

Howarth, C. I., & Bulmer, M. G. Non-random sequences in visual threshold experiments. *Quarterly Journal of Experimental Psychology,* 1956, **8,** 163–171.

Jeffress, L. A. Stimulus oriented approach to detection. *Journal of the Acoustical Society of America,* 1964, **36,** 766–774.

Jerison, H. J. Signal detection theory in the analysis of human vigilance. *Human Factors,* 1967, **9,** 285–288.

Jerison, H. J., & Pickett, R. M. Vigilance: a review and re-evaluation. *Human Factors,* 1963, *5,* 211–237.

Jerison, H. J., Pickett, R. M., & Stenson, H. H. The elicited observing rate and decision process in vigilance. *Human Factors,* 1965, **7,** 107–128.

Kac, M. A note on learning signal detection. *I.R.E. Transactions on Information Theory,* 1962, **IT-8,** 126–128.

Katz, L. A comparison of type II operating characteristics derived from confidence ratings and from latencies. *Perception & Psychophysics,* 1970, **8,** 65–68.

Kiang, N. Y-S. *Discharge patterns of single fibers in the cat's auditory nerve.* Cambridge, Massachusetts: MIT Press, 1965.

Kinchla, R. A. A learning factor in visual discrimination. In R. C. Atkinson (Ed.), *Studies in mathematical psychology*. Stanford: Stanford Univ. Press, 1964. Pp. 233–249.

Kinchla, R. A. A comparison of sequential effects in detection and recognition. Experimental Psychology Series, Psychology Dept. of New York University, Tech. Rept. 1, 1966.

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

Kristofferson, A. B. Attention in time discrimination and reaction time. Washington, D.C.: NASA CR-194, 1965.

Kristofferson, A. B. A time constant involved in attention and neural information processing. Washington, D.C.: NASA CR-427, 1966.

Kristofferson, A. B. Attention and psychophysical time. *Acta Psychologica,* 1967, **27,** 93–100. (a)

Kristofferson, A. B. Successiveness discrimination as a two-state quantal process. *Science,* 1967, **158,** 1337–1339. (b)

Kristofferson, A. B. Sensory attention. Hamilton, Ontario: McMaster Univ. Tech. Rept. 36, 1969.

Kruskal, J. B. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika,* 1964, **29,** 1–27. (a)

Kruskal, J. B. Nonmetric multidimensional scaling: a numerical method. *Psychometrika,* 1964, **29,** 115–129. (b)

LaBerge, D. L. A recruitment theory of simple behavior. *Psychometrika,* 1962, **27,** 375–396.

Laming, D. R. J. *Information theory of choice-reaction times.* New York: Academic Press, 1968.

Larkin, W. D. Rating scales in detection experiments. *Journal of the Acoustical Society of America,* 1965, **37,** 748–749.

Larkin, W. D., & Norman, D. A. An extension and experimental analysis of the neural quantum theory. In R. C. Atkinson (Ed.), *Studies in mathematical psychology.* Stanford: Stanford Univ. Press, 1964. Pp. 188–200.

Larkin, W., & Greenberg, G. Z. Selective attention in uncertain frequency detection. *Perception & Psychophysics,* 1970, **8,** 179–185.

Lee, W. Detection theory, micromatching, and the constant-ratio rule. *Perception & Psychophysics,* 1968, **4,** 217–219.

Levine, J. M. The effects of values and costs on detection and identification of signals in auditory vigilance. *Human Factors,* 1966, **8,** 525–537.

Levitt, H. Testing for sequential dependencies. *Journal of the Acoustical Society of America,* 1968, **43,** 65–69.

Levitt, H. Transformed up–down methods in psychoacoustics. City Univ. of New York Graduate Center, mimeographed, undated.

Levitt, H., & Bock, D. Sequential programmer for psychophysical testing. *Journal of the Acoustical Society of America,* 1967, **42,** 911–913.

Levitt, H., & Rabiner, L. R. Use of a sequential strategy in intelligibility testing. *Journal of the Acoustical Society of America,* 1967, **42,** 609–612.

Levitt, H., & Treisman, M. Control charts for sequential testing. *Psychometrika,* 1969, **34,** 509–518.

Lindner, W. A. Recognition performance as a function of detection criterion in a simultaneous detection-recognition task. *Journal of the Acoustical Society of America,* 1968, **44,** 204–211.

Loeb, M., & Binford, J. R. Vigilance for auditory intensity changes as a function of preliminary feedback and confidence level. *Human Factors,* 1964, **6,** 445–458.

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

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

Luce, R. D. A threshold theory for simple detection experiments. *Psychological Review,* 1963, **70,** 61–79. (b)

Luce, R. D. Asymptotic learning in psychophysical theories. *British Journal of Statistical Psychology* 1964, **17,** 1–14.

Luce, R. D. A model for detection in temporally unstructured experiments with a Poisson distribution of signal presentations. *Journal of Mathematical Psychology,* 1966, **3,** 48–64.

Luce, R. D., & Green, D. M. Detection of auditory signals presented at random times, II. *Perception & Psychophysics,* 1970, **7,** 1–14.

Luce, R. D., & Green, D. M. A neural timing theory for response times and the psychophysics of intensity. *Psychological Review,* 1972, **79,** 14–57.

Mackworth, J. F. Decision interval and signal detectability in a vigilance task. *Canadian Journal of Psychology,* 1965, **19,** 111–117.

Mackworth, J. F. The effects of signal rate on performance in two kinds of vigilance tasks. *Human Factors,* 1968, **10,** 11–18.

Mackworth, J. F., & Taylor, M. M. The d' measure of signal detectability in vigilance-like situations. *Canadian Journal of Psychology,* 1963, **17,** 307–325.

Madansky, A. The fitting of straight lines when both variables are subject to error. *Journal of the American Statistical Association,* 1959, **54,** 173–205.

Markowitz, J. *Investigation of the dynamic properties of the neural quantal model.* Unpublished doctoral dissertation, Univ. of Pennsylvania, 1966.

Markowitz, J., & Swets, J. A. Factors affecting the slope of empirical ROC curves: comparison of binary and rating responses. *Perception & Psychophysics,* 1967, **2,** 91–100.

McGill, W. J. Neural counting mechanisms and energy detection in audition. *Journal of Mathematical Psychology,* 1967, **4,** 351–376.

McGill, W. J., & Goldberg, J. P. Pure-tone intensity discrimination and energy detection. *Journal of the Acoustical Society of America,* 1968, **44,** 576–581.

Miller, G. A., Heise, G. A., & Lichten, W. The intelligibility of speech as a function of the context of the test materials. *Journal of Experimental Psychology,* 1951, **41,** 329–335.

Nachmias, J. Effects of presentation probability and number of response alternatives on simple visual detection. *Perception & Psychophysics,* 1968, **3,** 151–155.

Nachmias, J., & Steinman, R. M. Study of absolute visual detection by the rating scale method. *Journal of the Optical Society of America,* 1963, **53,** 1206–1213.

Nakatani, L. H. *A confusion-choice stimulus recognition model applied to word recognition.* Univ. of California, Los Angeles, Dept. of Psychology, Ph.D. dissertation (TR 31), 1968.

Nakatani, L. H. Comments on Broadbent's response bias model for stimulus recognition. *Psychological Review,* 1970, **77,** 574–576.

Nakatani, L. H. Confusion-choice model for multi-dimensional psychophysics. *Journal of Mathematical Psychology,* 1972, **9,** 104–127.

Nickerson, R. J. Expectancy waiting time and the psychological refractory period. *Acta Psychologica,* 1967, **27,** 23–34.

Norman, D. A. Stochastic learning and a quantal model of signal detection. Proc. Joint Automatic Control Conference Symposium on discrete adaptive systems. *American Institute of Electrical Engineers,* 1962, 35–40. Reprinted in *Transactions of the I.E.E.E. on Applications and Industry,* 1964, **83,** 292–296.

Norman, D. A. Sensory thresholds and response bias. *Journal of the Acoustical Society of America,* 1963, **35,** 1432–1441.

Norman, D. A. Sensory thresholds, response bias, and the neural quantum theory. *Journal of Mathematical Psychology,* 1964, **1,** 88–120. (a)

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

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.

Parducci, A., & Marshall, Louise M. Assimilation vs. contrast in the anchoring of perceptual judgments of weight. *Journal of Experimental Psychology,* 1962, **63,** 426–437.

Parducci, A., & Sandusky, A. Distribution and sequence effects in judgment. *Journal of Experimental Psychology,* 1965, **69,** 450–459.

Parducci, A., & Sandusky, A. J. Limits on the applicability of signal detection theories. *Perception & Psychophysics,* 1970, **1,** 63–64.

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

Penner, M. J. Detection of sinusoids of uncertain frequency. Univ. of California, San Diego, Unpublished doctoral dissertation, 1970.

Pfafflin, S. M., & Mathews, M. Energy detection model for monaural auditory detection. *Journal of the Acoustical Society of America,* 1962, **34,** 1842–1852.

Pollack, I. Message uncertainty and message reception. *Journal of the Acoustical Society of America,* 1959, **31,** 1500–1508.

Pollack, I. Computer simulation of threshold observations by method of limits. *Perceptual and Motor Skills,* 1968, **26,** 583–586.

Pollack, I., & Decker, L. Consonant confusions and the constant ratio rule. *Language and Speech,* 1960, **3,** 1–6.

Pollack, I., & Hsieh, R. Sampling variability of the area under the ROC-curve and of *d'*. *Psychological Bulletin,* 1969, **71,** 161–173.

Pollack, I., & Norman, D. A. A nonparametric 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.

Preston, M. G. Contrast effects and the psychometric judgments. *American Journal of Psychology,* 1936, **48,** 297–303. (a)

Preston, M. G. Contrast effects and the psychometric functions. *American Journal of Psychology,* 1936, **48,** 625–631. (b)

Robbins, H., & Munro, S. A stochastic approximation method. *Annals of Mathematical Statistics,* 1951, **22,** 400–407.

Rubenstein, H., & Pollack, I. Word predictability and intelligibility. *Journal of Verbal Learning and Verbal Behavior,* 1963, **2,** 147–158.

Schoeffler, M. S. Theory for psychophysical learning. *Journal of the Acoustical Society of America,* 1965, **37,** 1124–1133.

Schönemann, P. H., & Tucker, L. R. A maximum likelihood solution for the method of successive intervals allowing for unequal stimulus dispersions. *Psychometrika,* 1967, **32,** 403–417.

Schulman, A. I., & Greenberg, G. Z. Operating characteristics and a priori probability of the signal. *Perception & Psychophysics,* 1970, **8,** 317–320.

Senders, V., & Soward, A. Analysis of response sequences in the settings of a psychophysical experiment. *American Journal of Psychology,* 1952, **65,** 358–374.

Shepard, R. N. Stimulus and response generalization: A stochastic model relating generalization to distance in psychological space. *Psychometrika,* 1957, **22,** 325–345.

Shepard, R. N. Stimulus and response generalization: Deductions of the generalization gradient from a trace model. *Psychological Review,* 1958, **65,** 242–256. (a)

Shepard, R. N. Stimulus and response generalization: Tests of a model relating generalization to distance in psychological space. *Journal of Experimental Psychology,* 1958, **55,** 509–523. (b)

Shepard, R. N. Application of a trace model to the retention of information in a recognition task. *Psychometrika,* 1961, **26,** 185–203. (a)

Shepard, R. N. Role of generalization in stimulus-response compatibility. *Perceptual and Motor Skills,* 1961, **13,** 59–62. (b)

Shepard, R. N. The analysis of proximities: Multidimensional scaling with an unknown distance function. I. *Psychometrika,* 1962, **27,** 125–140. (a)

Shepard, R. N. The analysis of proximities: multidimensional scaling with an unknown distance function. II. *Psychometrika,* 1962, **27,** 219–246. (b)

Shepard, R. N. Analysis of proximities as a technique for the study of information processing in man. *Human Factors,* 1963, **5,** 33–48.

Shepard, R. N. Metric structure in ordinal data. *Journal of Mathematical Psychology,* 1966, **3,** 287–315.

Shipley, E. F. A model for detection and recognition with signal uncertainty. *Psychometrika,* 1960, **25,** 273–289.

Shipley, E. F. Dependence of successive judgments in detection tasks: correctness of the response. *Journal of the Acoustical Society of America,* 1961, **33,** 1142–1143.

Shipley, E. F. Detection and recognition: Experiments and choice models. *Journal of Mathematical Psychology,* 1965, **2,** 277–311.

Shipley, E. F. A signal detection analysis of a category judgment experiment. *Perception & Psychophysics,* 1970, **7,** 38–42.

Shipley, E. F., & Luce, R. D. Discrimination among two- and three-element sets of weights. In R. C. Atkinson (Ed.), *Studies in mathematical psychology.* Stanford: Stanford Univ. Press, 1964. Pp. 218–232.

Siebert, W. M. Some implications of the stochastic behavior of primary auditory neurons. *Kybernetik,* 1965, **2,** 206–215.

Smith, J. E. K. Stimulus programming in psychophysics. *Psychometrika,* 1961, **26,** 27–33.

Smith, J. E. K. Discussion on psychophysical testing. In *Conference on computer assisted testing.* Princeton, New Jersey: Educational Testing Service, 1966.

Smith, J. E. K. A note on the standard error of the estimates of TSD parameters from "Yes–No" experiments. University of Michigan, MMPP Tech. Rep. 69–8, 1969.

Speeth, S. D., & Mathews, M. V. Sequential effects in the signal-detection situation. *Journal of the Acoustical Society of America,* 1961, **33,** 1046–1054.

Stevens, S. S., Morgan, C. T., & Volkmann, J. Theory of the neural quantum in the discrimination of loudness and pitch. *American Journal of Psychology,* 1941, **54,** 315–335.

Stone, M. Models for choice-reaction time. *Psychometrika,* 1960, **25,** 251–260.

Stroud, J. M. The fine structure of psychological time. In H. Quastler (Ed.), *Information theory in psychology.* Glencoe, Illinois: Free Press, 1955.

Swets, J. A. Is there a sensory threshold? *Science,* 1961, **134,** 168–177. [Reprinted in J. A. Swets (Ed.), *Signal detection and recognition by human observers.* New York: Wiley, 1964. Pp. 122–144.

Swets, J. A. (Ed.). *Signal detection and recognition by human observers.* New York: Wiley, 1964.

Swets, J. A., & Birdsall, T. G. Deferred decision in human signal detection: a preliminary experiment. *Perception & Psychophysics,* 1967, **2,** 15–28.

Swets, J. A., & Green, D. M. Sequential observations by human observers of signals in noise. In C. Cherry (Ed.), *Information theory.* London: Butterworths, 1961. [Reprinted in J. A. Swets (Ed.), *Signal detection and recognition by human observers.* New York: Wiley, 1964. Pp. 221–242.]

Swets, J. A., Shipley, E. F., McKey, M. J., & Green, D. M. Multiple observations of signals in noise. *Journal of the Acoustical Society of America,* 1959, **31,** 514–521. [Reprinted in J. A. Swets (Ed.), *Signal detection and recognition by human observers.* New York: Wiley, 1964. Pp. 201–220.]

Swets, J. A., & Sewall, S. T. Stimulus versus response uncertainty in recognition. *Journal of the Acoustical Society of America,* 1961, **33,** 1589–1592. [Reprinted in J. A. Swets (Ed.), *Signal detection and recognition by human observers.* New York: Wiley, 1964. Pp. 431–446.]

Symmes, D. Self-determination of critical flicker frequencies in monkeys. *Science,* 1962, **136,** 714–715.

Tanner, T. A., Haller, R. W., & Atkinson, R. C. Signal recognition as influenced by presentation schedules. *Perception & Psychophysics,* 1967, **2,** 349–358.

Tanner, T. A., Jr., Rauk, J. A., & Atkinson, R. C. Signal recognition as influenced by information feedback. *Journal of Mathematical Psychology,* 1970, **7,** 259–274.

Tanner, W. P., Jr. Theory of recognition. *Journal of the Acoustical Society of America,* 1956, 28, 882–888. [Reprinted in J. A. Swets (Ed.), *Signal detection and recognition by human observers.* New York: Wiley, 1964. Pp. 413–430.]

Tanner, W. P. Jr., & Norman, R. Z. The human use of information, II. Signal detection for the case of an unknown signal parameter. *Transactions of the I.R.E. Professional Group on Information Theory,* 1954, 222–226.

Tanner, W. P., Jr., & Swets, J. A. A decision-making theory of visual detection. *Psychological Review,* 1954, **61,** 401–409.

Tanner, W. P., Jr., Swets, J. A., & Green, D. M. Some general properties of the hearing mechanism. Tech. Rep. No. 30, Electronic Defense Group, Univ. of Michigan, 1956.

Taylor, M. M., & Creelman, C. D. PEST: efficient estimates on probability functions. *Journal of the Acoustical Society of America,* 1967, **41,** 782–787.

Thomas, E. A. C., & Legge, D. Probability matching as a basis for detection and recognition decisions. *Psychological Review,* 1970, **77,** 65–72.

Thurstone, L. L. *The measurement of values.* Chicago: Univ. of Chicago Press, 1959.

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

Treisman, M. On the word frequency effect: Comments on papers by J. Catlin and L. H. Nakatani. *Psychological Review,* 1971, **78,** 420–425.

Verplanck, W. S., Collier, G. H., & Cotton, J. W. Nonindependence of successive responses in measurements of the visual threshold. *Journal of Experimental Psychology,* 1952, **44,** 273–282.

Verplanck, W. S., Cotton, J. W., & Collier, G. H. Previous training as a determinant of response dependency at the threshold. *Journal of Experimental Psychology,* 1953, **46,** 10–14.

Veniar, F. A. Signal detection as a function of frequency ensemble. I. *Journal of the Acoustical Society of America,* 1958, **30,** 1020–1024.

Wald, A. *Sequential analysis.* New York: Wiley, 1947.

Watson, C. S., & Nichols, T. L. Replication and revisions of Egan's method of free response. *Journal of the Acoustical Society of America,* 1966, **36,** 1247 (abstract).

Watson, C. S., Rilling, M. E. & Bourbon, W. T. Receiver-operating characteristics determined by a mechanical analog to the rating scale. *Journal of the Acoustical Society of America,* 1964, **36,** 283–288.

Wetherill, G. B. Sequential estimation of quantal response curves. *Journal of the Royal Statistical Society,* 1963, **25,** 1–48.

Wetherill, G. B. *Sequential methods in statistics.* London: Methuen, 1966.

Wetherill, G. B., Chen, H., & Vasudeva, R. B. Sequential estimation of quantal response curves: A new method of estimation. *Biometrika,* 1966, **53,** 439–454.

Wetherill, G. B., & Levitt, H. Sequential estimation of points on a psychometric function. *British Journal of Mathematical and Statistical Psychology,* 1965, **18,** 1–10.

White, C. T. Temporal numerosity and the psychological unit of duration. *Psychological Monographs,* 1963, **77** (Whole No. 575).

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 theory. *Psychological Bulletin,* 1968, **69,** 126–130. (b)

Zwislocki, J., Maire, F., Feldman, A. S., & Rubin, H. On the effect of practice and motivation on the threshold of audibility. *Journal of the Acoustical Society of America,* 1958, **30,** 254–262.