3

Detection and Recognition[1,2]

R. Duncan Luce
University of Pennsylvania

1. This work was supported in part by grants NSF G-8864 and NSF G-17637 from the National Science Foundation to the University of Pennsylvania.
2. I wish to thank Robert R. Bush, Eugene Galanter, D. M. Green, Francis W. Irwin, M. V. Mathews, B. S. Rosner, Elizabeth F. Shipley, S. S. Stevens, and W. P. Tanner, Jr., for their careful reading and criticism of various drafts of this material. Numerous conversations with Dr. Galanter during the last four years have led to the view of psychophysics that is reflected in this as well as in Chapters 4 and 5.

Contents

Detection and Recognition

Modern psychophysics is concerned with asymptotic choice behavior when subjects respond to four general types of questions about simple physical stimuli:

1. Is a stimulus there? (The absolute threshold and, more generally, detection problems.)
2. Which of several possibilities is it? (The recognition problem.)
3. Is this stimulus different from that? (The discrimination problem.)
4. How different is this one from that? (The scaling problem.)

These, then, are the problems confronting the mathematical psychologist who attempts to create formal theories to summarize the empirical findings of psychophysics and to guide further research. A report of these mathematical studies for the first two questions is given in this chapter; for question 3, in the next; and for question 4, in Chapter 5.

In addition to this intuitive classification of the chapters, there is a more formal one based on the concepts defined in Chapter 2. Except for Sec. 9, the models in this chapter are for complete identification experiments, those in the next are for partial identification designs, and those in the scaling chapter are for choice experiments in which no identification function is defined.

It will be recalled from Chapter 2 that in complete identification experiments the experimenter establishes, and explains to the subject, a one-to-one correspondence ι between the response set R and the stimulus presentation set S. Because of this correspondence, each response designates a unique presentation both to the experimenter and to the subject.

Among the possible complete identification experiments, two substantively different groups of studies have been performed. Those called *detection experiments* have the set of stimuli $\mathscr{S} = \{\mathscr{s}, \emptyset\}$, where $\emptyset$ denotes the null and $\mathscr{s}$ a nonnull stimulus, a presentation set $S \subseteq \mathscr{S}^k$, and there may or may not be a background. In the simplest of these detection experiments the subject decides on each trial whether $\mathscr{s}$ or $\emptyset$ is added to the background; in more complicated ones he decides in which of several time intervals or spatial locations $\mathscr{s}$ has been added to the background.

In terms of the stimulation used, three types of detection experiments may be distinguished:

1. Absolute threshold studies—the classic procedure—in which no

background is introduced by the experimenter and $\mathscr{s}$ is so near the lower limit of perception that the subject is not perfectly certain when it is present.

2. Quantal studies in which there is a simple background, such as a pure tone, and $\mathscr{s}$ is simply an increment (or decrement) in one dimension of the background; for example, in energy or frequency.
3. Signal detection studies in which there is a complex background and $\mathscr{s}$ differs from it on more than one dimension; in an extreme but frequent case, the background is white noise and $\mathscr{s}$, a pure tone.

In *recognition experiments*, our other main category, the stimulus set includes at least two stimuli different from the null one and, with few exceptions, $S = \mathscr{S}$. In most recognition studies $\mathscr{S}$ does not include the null stimulus, but, as I shall argue in Sec. 7.1, this is a dubious practice when the stimuli are not perfectly detectable.

The data from complete identification experiments are often first summarized into what are called *confusion matrices*. The rows of such a matrix are identified with the stimulus presentations and the columns with the responses, ordered so that the ordinary correspondence between rows and columns is the same as that defined by the identification function ι. The entries are either the absolute frequencies f_{sr} or the relative frequencies

$$\hat{p}(r \mid s) = \frac{f_{sr}}{\sum_{r' \in R} f_{sr'}}$$

of response r to presentation s.

The relative frequency $\hat{p}(r \mid s)$ is generally interpreted as an estimate of a corresponding conditional response probability $p(r \mid s)$ which in most models is assumed to be a constant independent of the trial. In addition, the responses on different trials are usually assumed to be statistically independent. The first assumption is easily dropped by reinterpreting $\hat{p}(r \mid s)$ as an estimate of the expectation of $p_n(r \mid s)$ at asymptote; however, most of the models actually assume that the probabilities themselves, not just their expectations, are constant. The assumption of response independence is a good deal more troublesome, for there is considerable evidence (e.g., Neisser, 1955; Senders, 1953; Senders & Sowards, 1952; Speeth & Mathews, 1961) indicating that it is incorrect. We persist in making it, in spite of the evidence, because of the difficulty in constructing models that are tractable and have response dependencies. Although no one has formulated and proved any results to this effect, one suspects that there may be judicious ways to add dependencies to response-independent models so that certain of the asymptotic properties are unchanged. Such results are needed to justify many of our current practices of data analysis.

At one time the data from recognition experiments were either presented in uncondensed form as confusion matrices or they were summarized by one or another of the standard contingency-table statistics. Until quite recently, nothing like a coherent theory had evolved, and the empirical generalizations were few. Three may be mentioned. Let the presentations be labeled $s_1, s_2, \ldots, s_k$ and the responses, $1, 2, \ldots, k$ in such a way that $\iota(r) = s_r$. First, the largest entry in row s_r is generally the main diagonal one, $\hat{p}(r \mid s_r)$. Second, although the matrix is not strictly symmetric in the sense that $\hat{p}(r \mid s_{r'}) = \hat{p}(r' \mid s_r)$, there is a definite tendency in that direction. Third, when the presentations are physically ordered—for example, by intensity, size, frequency—so that $s_1 < s_2 < \ldots < s_k$, then the value of $\hat{p}(r \mid s_r)$ dips down rapidly from $\hat{p}(1 \mid s_1)$, reaches a plateau in the midrange, and then rises again rapidly to $\hat{p}(k \mid s_k)$. In other words, plots of $\hat{p}(r \mid s_r)$ versus r are usually U-shaped. Their exact nature depends upon the number and spacing of the presentations as well as upon the experimental conditions. Examples of size, color, and position confusion matrices can be found in Shepard (1958b).

In the early 1950's a number of psychologists began to analyze recognition confusion matrices in terms of Shannon's information measure, and several comparatively simple generalizations resulted. These we discuss in Sec. 7 on recognition experiments. What is still lacking is an adequate, detailed response theory to explain these somewhat gross results.

The study of detection has proceeded largely independently of the work on recognition, with, however, some fusion developing in the last several years. Detection research began early in the history of experimental psychology with determinations of absolute and difference thresholds. Theoretical contributions were scattered until the early 1950's when a program of theoretical and experimental research emerged at the University of Michigan. Somewhat later several related programs developed elsewhere in the United States.

Our study begins with the several analyses of detection experiments which are currently of interest, and in Secs. 7 and 8 some of the same ideas are applied to recognition experiments.

1. REPRESENTATIONS OF THE RESPONSE PROBABILITIES

Attempts to account for the behavioral relations among various types of identification experiments, both complete and partial, have so far resulted in three distinct response theories. In this section each is described in moderately general terms with little reference to specific experiments;

in the remainder of the chapter and in much of the next two they are applied to specific designs. The reader may well wonder why three different theories for the same behavior should be presented, when, after all, at most one can be correct. One reason is that there is no assurance that the same response theory is appropriate for all modalities or for different tasks within one modality, but more important, it has been impossible so far to choose among them on empirical grounds even for a single type of experiment within one modality. Their predictions tend to be similar, and where there are differences the experimental results have been either inconclusive or contradictory. This situation is hopefully transitory; in fact, considerable clarification can be expected in the next few years.

1.1 Signal Detectability Theory

The notions underlying signal detectability theory originally took root in psychology during the period bounded by Fechner and Thurstone. Later they reappeared in a slightly different guise explaining not just discrimination but also detection and recognition. W. P. Tanner, Jr., and his colleagues at the University of Michigan reinterpreted and modified analyses of optimal physical detection of electrical signals in noise (Peterson, Birdsall, & Fox, 1954; van Meter & Middleton, 1954) into a psychophysical theory. Some of the same ideas were also developed by Smith and Wilson (1953). In addition to the theory, a series of interrelated experiments have been performed (Birdsall, 1955, 1959; Clarke, Birdsall, & Tanner, 1959; Creelman, 1959, 1960; Egan, Schulman, & Greenberg, 1959; Green, 1958, 1960; Green, Birdsall, & Tanner, 1957; Speeth & Mathews, 1961; Swets, 1959, 1961b; Swets & Birdsall, 1956; Swets, Shipley, McKey, & Green, 1959; Swets, Tanner, & Birdsall, 1955, 1961; Tanner, 1955, 1956, 1960, 1961; Tanner & Birdsall, 1958; Tanner, Birdsall, & Clarke, 1960; Tanner & Norman, 1954; Tanner & Swets, 1954a,b; Tanner, Swets, & Green, 1956; Veniar, 1958a,b,c). These researches go under the name of *signal detection*, or, as Tanner prefers, *signal detectability theory*. Survey papers by Green (1960), Licklider (1959), and Swets (1961a) give summaries of the central ideas and experimental findings.

The main notion is that the pertinent information available to the subject as a result of the stimulation can be summarized by a number; however, repeated presentations of the same stimulus produce not the same number but a distribution of them. The subject is assumed to behave as if he knew these distributions. He evaluates the particular number arising on a trial in terms of the distributions from which it could have arisen, much as a statistician evaluates an observation to decide between a null and an alternative hypothesis (Wald, 1950). Indeed, the two models are formally the same.

The theory does not say where these distributions come from, although their usually assumed normality easily suggests a pseudoneurology in which many small independent neuronal errors accumulate to form the resultant error; nor does it tell how the subject comes to know the distributions, but a learning process during pretraining seems a likely candidate; nor does it suggest how the subject carries out the various needed transformations and calculations. Unexplained, "internal" numerical representations such as these are characteristic of almost all psychophysical theories and they simply indicate, I suspect, the relatively primitive state of the theory. Nonetheless, we are no more obliged to account for them in, say, physiological terms than were the authors of the first macroscopic physical theories required to explain planetary motions in terms of elementary particle properties.

Tanner and his colleagues arrived at the representation in this way. The effect of a presentation s is supposed to be a random vector $\mathbf{s}$ which assumes values in a k-dimensional Euclidean space E_k; that is, the effect of stimulation is assumed to be adequately described by a k-tuple of numbers. Not only does this seem moderately plausible, but for frequency bounded temporal signals the important sampling, or $2WT$, theorem (Shannon & Weaver, 1949) shows that in the limit as $T \to \infty$ such a representation of their physical properties is indeed possible.

If $\mathbf{x} \in E_k$, the probability density that stimulus s produces the effect $\mathbf{x}$, $p^{(k)}(\mathbf{x} \mid s)$ is assumed to exist. Suppose, for the moment, that one of two presentations, s or s', occurs on each trial and that on a particular trial an observation $\mathbf{x}$ occurs. The subject must use it and his assumed knowledge of the distributions $p^{(k)}(\cdot \mid s)$ and $p^{(k)}(\cdot \mid s')$ to decide which of the two presentations fathered it. In such matters there is an inherent uncertainty. It is plausible that he might decide by considering the relative likelihood of the two presentations generating $\mathbf{x}$. Specifically, let us suppose that he calculates the likelihood ratio

$$l(\mathbf{x}) = \frac{p^{(k)}(\mathbf{x} \mid s)}{p^{(k)}(\mathbf{x} \mid s')}. \tag{1}$$

If this number is large, it is only sensible to say that s was presented; if small, s'. This suggests that the subject should establish a *cut-point* (or *criterion*) c and use the *decision rule*:

$$\text{respond that } \begin{Bmatrix} s \\ s' \end{Bmatrix} \text{ was presented if } \log l(\mathbf{x}) \begin{Bmatrix} > \\ < \end{Bmatrix} c, \tag{2}$$

where the logarithmic transformation has been introduced for convenience later.

Assuming such a rule, we wish to calculate the response probabilities. To do so, we need expressions for the distributions of log l corresponding to the two presentations. Define the set

$$L(z) = \{\mathbf{x} \mid \log l(\mathbf{x}) = z\},$$

then if

$$\begin{aligned} p(z \mid s) &= \int_{L(z)} p^{(k)}(\mathbf{x} \mid s)\, d\mathbf{x} \\ p(z \mid s') &= \int_{L(z)} p^{(k)}(\mathbf{x} \mid s')\, d\mathbf{x} \end{aligned} \tag{3}$$

exist, they are the desired distributions. These are usually assumed to be normal distributions; one of the main reasons for this assumption is given in Sec. 6.1.

With the decision rule given in Eq. 2, the expressions for the response probabilities are easily seen to be

$$\begin{aligned} p(r \mid s) &= \int_c^\infty p(z \mid s)\, dz \\ p(r \mid s') &= \int_c^\infty p(z \mid s')\, dz, \end{aligned} \tag{4}$$

where r is the response such that $\iota(r) = s$.

Several comments and cautions. First, why, aside from plausibility, have we assumed that the decision axis is (the logarithm of) the likelihood ratio axis? The main reason, as we shall see in Sec. 5.1, is this. Had we begun with an uninterpreted decision axis, as Thurstone did in his analysis of discrimination problems, and were we to assume that the decision procedure is optimal in the sense of maximizing the expected payoff, then the decision axis must be the likelihood ratio axis or a monotonic function of it. The logarithm is, of course, a monotonic function.

Second, is a cut-point decision rule reasonable? It seems to be if the two distributions are like those of Fig. 1. Moreover, when they are distributions of likelihood ratios, it can be shown that cut-points lead to optimal behavior for various definitions of optimality. Were it possible, however, for the distributions to be multimodal, as, for example, in Fig. 2, then a cut-point rule would clearly be inappropriate. It is not difficult to show that distributions such as those in Fig. 2 cannot occur for a decision axis which is a monotonic function of likelihood ratio, For a given

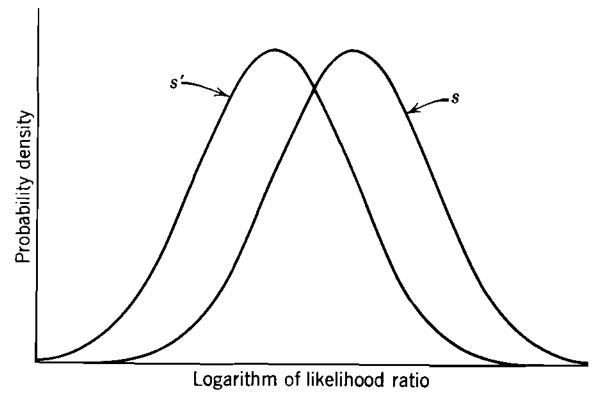

Fig. 1. Typical normal distributions of the logarithm of the likelihood ratio for two different stimulus presentations.

likelihood ratio l, we know by Eq. 1 that for any $\mathbf{x} \in L(\log l)$, then $p^{(k)}(\mathbf{x} \mid s) = lp^{(k)}(\mathbf{x} \mid s')$. Thus, by Eq. 2,

$$\begin{aligned} p(z \mid s) &= \int_{L(z)} p^{(k)}(\mathbf{x} \mid s)\, d\mathbf{x} \\ &= \int_{L(z)} lp^{(k)}(\mathbf{x} \mid s')\, d\mathbf{x} \\ &= lp(z \mid s'). \end{aligned}$$

Thus the two distributions must be closely related to one another; specifically, the ratio of the two values at $z = \log l$ must be just l. Therefore, because the ratio l is a monotonic function of $z = \log l$, distributions such as those shown in Fig. 2 are impossible.

Third, what happens when there are three or more presentations? This is a fairly subtle matter. In many of the experiments to which the

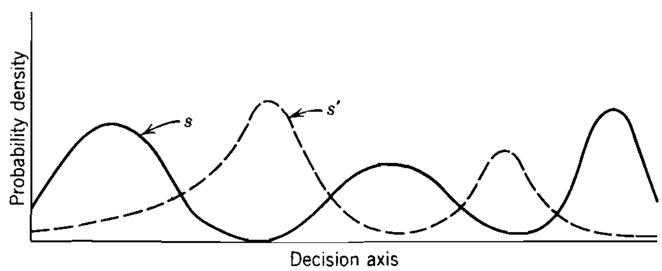

Fig. 2. Multimodal distributions for which a cut-point decision rule is not appropriate but which cannot occur if the decision axis is a monotonic function of likelihood ratio.

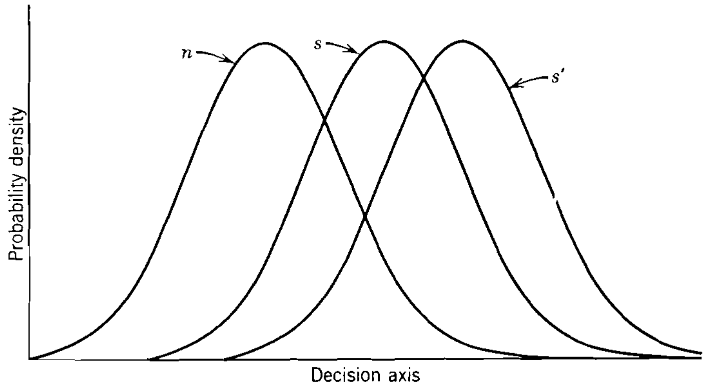

Fig. 3. Three normal distributions on a single decision axis, which cannot occur if the decision axis is a monotonic function of likelihood ratio.

theory has been applied one of the presentations, n, is a noise background that contains no stimulus and the others, s, s', . . . , consist of stimuli embedded in the noise. The several distributions of effects over the Euclidean k-space are assumed to be statistically independent of each other. If we suppose that the stimuli s and s' differ only on one dimension, say energy, it seems plausible to think of the three distributions as existing over a common decision axis, as in Fig. 3. Unfortunately, such a simple representation in terms of a likelihood ratio axis does not follow from the argument just given for two stimuli. The trouble is that when we compare n and s the likelihood ratio noise distribution, $p(z \mid n)$, depends not only upon $p^k(\cdot \mid n)$, which by assumption is independent of the other presentations, but also upon the set $L(z)$. Because $L(z)$ is defined in terms of the likelihood ratio, it depends upon $p^k(\cdot \mid s)$ as well as upon $p^k(\cdot \mid n)$. That is to say, in terms of likelihood ratio each stimulus has its own separate noise distribution. Or we may see it another way. In the two-stimulus case we have just shown that $p(\log l \mid s)/p(\log l \mid s') = l$, so where two distributions intersect, that is, where $p(\log l \mid s) = p(\log l \mid s')$, the likelihood ratio must be 1. But there are three such intersections in Fig. 3, all of which would have to correspond to the same likelihood ratio of 1, which is clearly impossible.

This means that complete identification experiments with more than two presentations cannot usually be reduced to a one-dimensional representation because each pair of presentations has its own likelihood ratio axis. It is customary in signal detectability theory to assume that the logarithms of these several axes can be embedded in an Euclidean space of appropriate dimension. Two serious problems result: when should the separate axes be assumed to be orthogonal and to what class of partitions should the decision rule be restricted? No class of rules seems nearly so compelling for a space as does the cut-point rule for a line.

1.2 Choice Theory

Choice theory has been discussed, in one variant or another, by Bradley (1954a,b, 1955), Bradley and Terry (1952), Clarke (1957), Luce, (1959), Restle (1961),[3] Shepard (1958a,b), and Shipley (1960, 1961) for either complete or partial identification experiments or both. No very systematic statement of the intuitions underlying these models has yet been given. Although I shall attempt some clarification, I am still far from a completely satisfactory axiomatic statement of all that is involved. First, the basic representation will be stated, and then I shall consider briefly some of the justifications that have been given for it.

Two ratio scales (i.e., scales unique up to multiplication by positive constants)

$$\eta\colon\ S \times S \rightarrow \text{positive real numbers}$$

$$b\colon\ R \rightarrow \text{positive real numbers}$$

are assumed to exist such that when ι is the identification function the response probabilities are of the form

$$p_\iota(r \mid s) = \frac{\eta[s, \iota(r)]\, b(r)}{\sum\limits_{r' \in R} \eta[s, \iota(r')]\, b(r')}. \tag{5}$$

The scale η is interpreted as a measure of the similarity between the presented stimulus s and the one, $\iota(r)$, for which r is the correct response. The scale b, which is associated only with responses, is interpreted as a measure of response bias.

At present, Eq. 5 is useful only if we make certain additional assumptions. Those we make, which are in large part suggested by Shipley's (1961) work, all arise from preconceived notions about the intuitive meaning of the η scale and from considerations of mathematical simplicity; they are neither obviously necessary nor clearly dictated by data, even though their consequences have received some empirical support.

The first three assumptions can be interpreted as formalizing our interpretation of η as a measure of the similarity between stimuli or, equally well, as postulating that the logarithm of η behaves like a measure of "psychological distance."

[3] Although I had an opportunity to read Restle's interesting book in manuscript form, it was not available to me when this chapter was being drafted and so no attempt has been made to incorporate his ideas directly.

Assumption 1. *For all* $s, s' \in S$, $\eta(s, s') = \eta(s', s)$.
Assumption 2. *For all* $s \in S$, $\eta(s, s) = 1$.
Assumption 3. *For all* $s, s', s'' \in S$, $\eta(s, s'') \geqslant \eta(s, s')\eta(s', s'')$.

The heart of Assumption 2 is that the number $\eta(s, s)$ is independent of s; setting it equal to 1 merely fixes the unit of the η scale.

It is easy to see that

$$d(s, s') = -\log \eta(s, s') \tag{6}$$

satisfies the usual distance axioms, namely:

1. $d(s, s') = d(s', s)$.
2. $d(s, s') \geqslant 0$ and $d(s, s) = 0$.
3. $d(s, s'') \leqslant d(s, s') + d(s', s'')$.

These three assumptions are used in all applications of the choice theory later. In addition, a fourth assumption, which is suggested by the interpretation of d as a distance measure, will sometimes be made. It plays exactly the same role in the choice theory as the orthogonal embeddings of the logarithm-of-likelihood-ratio axes into Euclidean spaces which are used in signal detectability theory.

Assumption 4. *If* $S = S_1 \times S_2 \times \ldots \times S_k$ *and if* η *is defined over* S *and over each of the* S_i, *then for* $s = \langle s^1, s^2, \ldots, s^k \rangle$ *and* $t = \langle t^1, t^2, \ldots, t^k \rangle \in s$,

$$d(s, t)^2 = \sum_{i=1}^{k} d(s^i, t^i)^2,$$

where d is defined by Eq. 6.

This states that if the stimuli can be viewed as having k distinct components—the S_i—then the several distance measures are interrelated as one would expect them to be, provided that the S_i were to correspond to the orthogonal dimensions of an Euclidean k-space and the distance, to the natural metric in that space. It is, of course, possible to write a weaker assumption in which the coordinates are not orthogonal, but this adds many more parameters to the model, namely the angles between the coordinates. Because we do not need this weaker version, it will not be described in detail.

The origins of this particular representation lie in two papers, a book, Shipley's thesis, and some unpublished work. Shepard (1957) suggested what amounts to Eq. 5 and Assumptions 1 to 3 to relate either stimuli to stimuli or responses to responses, and later Shepard (1958a) suggested the formula

$$\left[\frac{p_t(r \mid s')}{p_t(r' \mid s')} \frac{p_t(r' \mid s)}{p_t(r \mid s)}\right]^{1/2},$$

where $\iota(r) = s$ and $\iota(r') = s'$, as a "measure of stimulus generalization" between s and s'. By Eq. 5 and Assumptions 1 and 2, this equals

$$\left\{\frac{\eta[s', \iota(r)]b(r)}{\eta[s', \iota(r')]b(r')} \frac{\eta[s, \iota(r')]b(r')}{\eta[s, \iota(r)]b(r)}\right\}^{1/2} = \left\{\frac{\eta(s', s)}{\eta(s', s')} \frac{\eta(s, s')}{\eta(s, s)}\right\}^{1/2} = \eta(s, s'),$$

which is what was just termed a measure of stimulus similarity. Probably the same intuitions are involved, although the terminology differs.

Clarke (1957) proposed a model much like Eq. 5. Suppose that (S, R, ι) and (S', R', ι') are two complete identification experiments for which $s \in S' \subset S$, $r \in R' \subset R$, and ι' is the restriction of ι to R'; then he assumed what he called the *constant ratio rule*, namely,

$$p_{\iota'}(r \mid s) = \frac{p_\iota(r \mid s)}{\sum_{r' \in R'} p_\iota(r' \mid s)}.$$

In terms of Eq. 5, this assumption is equivalent to asserting that stimulus similarity $\eta(s, s')$ is independent of the particular S employed, with which I would agree, and that the response bias $b(r)$ is independent of the particular R employed, which I doubt. Both hypotheses must, of course, be subjected to experimental scrutiny, but in my view the model stands or falls on the first being correct; the second is not in the least crucial, given that b is interpreted as a response bias, and so it is not assumed here.

In *Individual Choice Behavior* (1959) I wrote down choice models for several specific detection and recognition experiments, but Eq. 5 was not stated in its full generality. These particular models, sometimes slightly modified to be consistent with Eq. 5, reappear below as applications of Eq. 5.

Recently, Bush, Luce, and Rose (1963) have shown that Eq. 5 arises as the asymptotic mean response probability of a simple experimenter-controlled (i.e., response-independent) linear learning model for complete identification experiments. (For a detailed discussion of stochastic learning models, see Chapter 9 of Vol. II). Specifically, they suppose that when stimulus s' is presented and r' is the correct response, that is, $\iota(r') = s'$, on trial i, then on trial $i + 1$

$$p_{i+1}(r \mid s) = p_i(r \mid s) + \eta(s, s')\theta(r')[\delta_{rr'} - p_i(r \mid s)],$$

where $\delta_{rr'}$ is the Kronecker delta (equal to 1 when $r = r'$ and 0 otherwise). They interpret $\theta(r')$ as a basic learning rate parameter associated with the response that the subject should have made and $\eta(s, s')$ as a similarity parameter representing the generalization from presentation s to presentation s'. It is clearly possible to define the learning rates so that $\eta(s, s) = 1$

for all $s \in S$. We see that whenever s' is presented the conditional probability of response r' occurring to any presentation is linearly increased, whereas the probabilities of all other responses are linearly decreased. The amount of the increase depends both upon the learning rate parameter and upon the generalization between s and s'.

Summing over the equation, it is easy to see that

$$\sum_{r \in R} p_{i+1}(r \mid s) = \sum_{r \in R} p_i(r \mid s) = 1,$$

so the model is consistent.

Let $P(s)$ denote the presentation probability of s, then

$$E[p_{i+1}(r \mid s) \mid p_i(r \mid s)] = p_i(r \mid s) \sum_{s' \in S} P(s') + P[\iota(r)]\eta[s, \iota(r)]\theta(r) - p_i(r \mid s) \sum_{r' \in R} P[\iota(r')]\eta[s, \iota(r')]\theta(r').$$

If we calculate expectations over $p_i(r \mid s)$, then take the limit as $i \to \infty$, and finally solve, we obtain

$$\lim_{i \to \infty} E[p_i(r \mid s)] = \frac{\eta[s, \iota(r)]b(r)}{\sum_{r' \in R} \eta[s, \iota(r')]b(r')},$$

where $b(r) = P[\iota(r)]\theta(r)$.

Thus the learning model not only leads to Eq. 5 as the asymptotic expected response probabilities, but it says that each response bias parameter is the product of the corresponding presentation probability and the learning rate parameter. The stimulus parameters are again interpreted as measures of similarity. In addition to accounting for Eq. 5, the learning model is of interest because it generates some sequential complexity in the trial-by-trial response patterns. As yet, however, little work has been done on the details of this stochastic model.

1.3 Threshold (or Neural Quantum) Theory

In the traditional psychological literature a chapter called "detection" is not found; however, what amount to detection experiments and some related theory are included, along with other designs, under the titles of "absolute" and "difference thresholds." The notion of a threshold, which is at least as old as Greek philosophy, is that some energy configurations, or differences, simply are not noted because of limitations imposed by the sensory and neural mechanisms. A characterization of this feature of the receptor system—when it exists—forms a partial description of the dynamics of that system, and a theory of its role in generating responses constitutes a partial description of the response mechanism of the organism.

Absolute and difference threshold experiments differ in this way. When the subject is asked to detect whether or not a stimulus has been presented, we are usually concerned with the value of the absolute threshold. But when he is asked to detect the difference between two temporal or spatial regions of stimulation, we are concerned with the difference threshold. Because the first can be viewed as detecting a difference between two regions of stimulation, one of which is null, no attempt is made in the formal theory to distinguish between the two problems.

The value of the absolute threshold is generally defined to be that level of stimulation that is detected 50 (or some other arbitrarily chosen) per cent of the time when the observing conditions are relatively ideal. The actual techniques used are various; some are rapid and probably yield biased or variable estimates, others are more painstaking. But, however the determinations may be made, two things are important to us. First, although the resulting numbers are called threshold values, there is nothing to prevent the procedures from yielding the numbers even if there are no thresholds. Both the detectability and choice theories, which postulate no thresholds, lead one to expect that "threshold" values can be determined. Second, the behavior of the subject is not really the object of study; rather, attention is paid to the limiting characteristics of his sensory system. When the behavior has been examined, it has usually been for "methodological" reasons—to improve the reliability or speed of the techniques. Examples of such research can be found in Blackwell's (1953) monograph on the determination of visual thresholds. Thus, although the threshold literature is large, it still remains for other psychologists to derive behavioral predictions from a threshold model, to use these to discover whether thresholds really exist, and to determine how such a psychophysical theory interacts with other factors affecting behavior.

In making threshold determinations, valued outcomes, or even information feedback, have rarely been used. Often the subject is instructed to minimize his "false-alarm" rate, and during pretraining he may be informed about his errors on catch trials. Some experimenters wait until the false alarm rate is sufficiently low—as they say, until the subject has established a "good criterion"—before proceeding to the main part of the experiment. Others simply estimate the false-alarm rate from the pretraining trials, and still others include catch trials during the experiment proper from which they estimate the rate. Sometimes these rates are simply reported; at other times they are used to "correct for guessing." I shall discuss the model for this correction presently (Sec. 2.3).

The existing threshold model was developed in two stages. The first, which in essence is a discrete analogue of the Thurstone-Tanner statistical model for stimulus effects, was initially stated along with supporting evidence by Békésy (1930); later Stevens, Morgan, and Volkmann (1941) refined the statement and added appreciably to the evidence. The second stage, which is concerned with the biases introduced by the subject, began with the model for correcting for guessing, was reformulated by Tanner and Swets (1954a), and was then extended by Luce (1963). For a general discussion of many of the issues involved and for some alternatives to the model we will discuss, see Swets (1961b).

The Békésy-Stevens model assumes that the effects of stimulation are discrete, not continuous as in the other two models. There is supposed to be a finite (or countable) sequence of "neural" quanta, which we may identify by the integers 1, 2, 3, ... A neural quantum is not identified with any particular neural configuration, although presumably it has some physiological correlate. At a given moment, stimulation is assumed to "excite" the first j of these quanta, but because of irregular fluctuations this number does not necessarily remain fixed over time, even when the stimulation is constant. The main feature of this quantal structure is that two stimuli, no matter how different they may be physically, cannot be different to the subject if they excite the same number of neural quanta. If this model is correct, the only changes that he can possibly notice are those producing a change in the number of excited quanta.

Let us suppose that just prior to the presentation of s on trial i, j quanta are excited by the residual environment plus the background, if any. When s is presented, suppose j' quanta are excited. The change, then, is $j' - j$, and so we can think of the effect of presenting s on trial i as $\theta(s, i) = j' - j$, that is, the presentations generate a function

$$\theta\colon\ S \times I_N \rightarrow I,$$

where I denotes the set consisting of zero and the positive and negative integers. Because the background is assumed to have a fluctuating effect, $\theta(\emptyset, i)$ is not necessarily 0, as one might first think.

The subject can detect a presentation only if $\theta(s, i)$ is not zero, but there are situations in which it is reasonable for a subject to require a change of more than one quantum before responding that a stimulus was presented. With that in mind, we have the following class of "internal" *detection rules:*

A presentation s on trial i is detected as $\begin{Bmatrix}\textit{the same as}\\ \textit{different from}\end{Bmatrix}$ *the background*

if $\begin{Bmatrix}-k \leqslant \theta(s, i) \leqslant k'\\ \theta(s, i) < -k \text{ or } > k'\end{Bmatrix}$, *where k and k′ are nonnegative integers.* (7)

Although the evidence makes one suspicious, it is generally assumed that Pr $[\theta(s, i) < -k \text{ or } > k']$ is independent of i but not of k and k'. We denote this probability by $q(s, k, k')$, or simply by $q(s)$ when k and k' are assumed fixed, and we speak of it as the *true detection probability*. This is not necessarily the same as the corresponding response probability which one estimates from experimental data—there may be response biases.

The second component of the model is the effect of the outcome structure. The obvious parallel to signal detectability theory is to suppose that biases are introduced by the selection of the cut-points k and k'. This model should be explored, but it has not been. Rather, it has been assumed that k and k' are fixed in a given situation and that the subject biases his responses in the light of the payoff structure simply by falsely converting some detection observations into negative responses or some no-detection observations into positive ones.

As in the other two models, there are two types of parameters. The true detection probabilities $q(s)$ are considered to be stimulus-determined, and the proportions of falsified responses are bias parameters which depend upon, among other things, the outcomes. The main body of the Békésy-Stevens work is concerned with the dependence of the stimulus parameters upon the stimulus (see Secs. 6.2 and 6.3), whereas other authors have focused more upon the response model and the dependence of the biasing parameters upon the outcome structure (see Secs. 5.2 and 6.3).

We turn now to two of the simplest detection experiments and show in turn how each of these three theories tries to account for the behavior.

2. SIMPLE DETECTION

Of the possible detection experiments, the simplest is the *Yes-No* design. At regular time intervals, which define the trials and which are marked off by, say, lights, a stimulus may or may not be added to a continuous background. Following each such interval, the subject responds, usually by pressing one of two buttons, to the effect that "Yes, a stimulus was present" or "No, none was there." Each presentation, stimulus plus background and background alone, is repeated some hundreds of times according to a random schedule, and the conditional choice probabilities are estimated by the relative frequencies of choices.

This is easily seen to be a complete identification experiment in which $\mathscr{S} = \{\mathscr{s}, \emptyset\}$, where $\emptyset$ is the null stimulus, $S = \mathscr{S}$, and $R = \{Y, N\}$, where Y means Yes and N, No. The identification function is, of course, $\iota(Y) = \mathscr{s}$ and $\iota(N) = \emptyset$. It is convenient to think of the background as noise and to change notation to the extent that s denotes "$\mathscr{s}$ plus noise" and n denotes "$\emptyset$ plus noise," that is, noise alone. So we treat $\{s, n\}$ as the presentation set. In summary, the confusion matrix is

$$\begin{array}{ccc} \text{Presentation Probability} & \text{Stimulus Presentation} & \text{Response} \\ & & \begin{array}{cc} Y & N \end{array} \\ P & s & \left[\begin{array}{cc} p(Y \mid s) & p(N \mid s) \\ p(Y \mid n) & p(N \mid n) \end{array}\right], \\ 1-P & n & \end{array}$$

where $p(Y \mid x) + p(N \mid x) = 1$, $x = s$ or n.

Next in complexity is the *two-alternative forced-choice design* in which two time intervals (or space locations) are marked off, and the subject is told that the stimulus is in one of the two intervals, but not both. The subject responds by indicating whether he believes the stimulus was in the first or the second interval. Otherwise, the experimental conditions are the same as in the Yes-No design. So, $\mathscr{S} = \{\mathscr{s}, \emptyset\}$, $S = \{\langle \mathscr{s}, \emptyset\rangle, \langle \emptyset, \mathscr{s}\rangle\}$, $R = \{1, 2\}$, and $\iota(1) = \langle \mathscr{s}, \emptyset\rangle$ and $\iota(2) = \langle \emptyset, \mathscr{s}\rangle$. Again, we write the presentations as $\langle s, n\rangle$ and $\langle n, s\rangle$ to emphasize the addition of the noise background. The confusion matrix is

$$\begin{array}{ccc} \text{Presentation Probability} & \text{Stimulus Presentation} & \text{Response} \\ & & \begin{array}{cc} 1 & 2 \end{array} \\ P & \langle s, n\rangle & \left[\begin{array}{cc} p(1 \mid \langle s, n\rangle) & p(2 \mid \langle s, n\rangle) \\ p(1 \mid \langle n, s\rangle) & p(2 \mid \langle n, s\rangle) \end{array}\right], \\ 1-P & \langle n, s\rangle & \end{array}$$

where $p(1 \mid x) + p(2 \mid x) = 1$, $x = \langle s, n\rangle$ or $\langle n, s\rangle$.

2.1 Signal Detectability Analysis

Because we have already discussed the signal detectability model for a general two-element presentation set, the model for the Yes-No experiment follows from Eq. 4 simply by making the appropriate notational changes:

$$\begin{aligned} p(Y \mid s) &= \int_c^\infty p(z \mid s)\, dz \\ p(Y \mid n) &= \int_c^\infty p(z \mid n)\, dz. \end{aligned} \tag{8}$$

Nothing has really been specified, however, until the forms of the density functions $p(z \mid s)$ and $p(z \mid n)$ are known. Throughout signal detectability theory they are assumed to be normal in z, the logarithm of the likelihood ratio. There is no loss of generality if we set the mean of the noise distribution at zero, for the location of the zero of the decision axis is arbitrary. Of course, the specific magnitude and sign of c depends upon that choice. Let the mean of the stimulus distribution be d. For the while, let us assume that the variances of the two distributions are equal, the common value being σ^2, even though later certain data force us to abandon this in favor of an assumption that $\sigma_s > \sigma_n$. Under these assumptions, Eq. 8 becomes

$$p(Y \mid s) = \frac{1}{\sqrt{2\pi}\sigma} \int_c^{\infty} \exp\left[-\frac{(z-d)^2}{2\sigma^2}\right] dz$$
$$p(Y \mid n) = \frac{1}{\sqrt{2\pi}\sigma} \int_c^{\infty} \exp\left[-\frac{z^2}{2\sigma^2}\right] dz. \tag{9}$$

In effect, the stimulus simply displaces the noise distribution to the right by the amount d. Because the unit of the decision axis is arbitrary, we can choose it to be σ, in which case the displacement is $d' = d/\sigma$. This normalized distance is a basic parameter of the signal detectability model. It is interpreted as a measure of the subject's sensitivity to the stimulus and it is thought to be independent of other experimental conditions such as the payoffs. The cut-point c is thought to be a bias that depends upon factors such as the presentation probability and the payoffs.

There is no question at this point that the model accounts for the observed response probabilities because there are two parameters, d' and c, and only two independent probabilities. (Of course, it may not account for sequential properties of the data). So we turn to the two-alternative forced-choice design to see whether the same stimulus parameter can be used to predict the data that are obtained there.

We make exactly the same assumptions as in the Yes-No experiment about the effect of a presentation of either s or n: the distributions are both normal, they have the same standard deviation, and the difference of the means is d. In the forced-choice design the subject makes two observations, $\mathbf{X}_1$ and $\mathbf{X}_2$ in the logarithm of likelihood ratio, corresponding to the two intervals. It is plausible that he uses the following decision rule:

$$\textit{For some value } c'\textit{, respond } \begin{Bmatrix}1\\2\end{Bmatrix} \textit{ if } \mathbf{X}_1 - \mathbf{X}_2 \begin{Bmatrix}\geqslant\\<\end{Bmatrix} c'. \tag{10}$$

For stimulus presentation $\langle s, n\rangle$, $\mathbf{X}_1 - \mathbf{X}_2 = \mathbf{S} - \mathbf{N}$, and for $\langle n, s\rangle$, $\mathbf{X}_1 - \mathbf{X}_2 = \mathbf{N} - \mathbf{S}$.

The question is: how are these two differences distributed? When the background and stimulus plus background differ simply in one dimension, such as energy or frequency in the quantal experiments (see Secs. 6.2 and

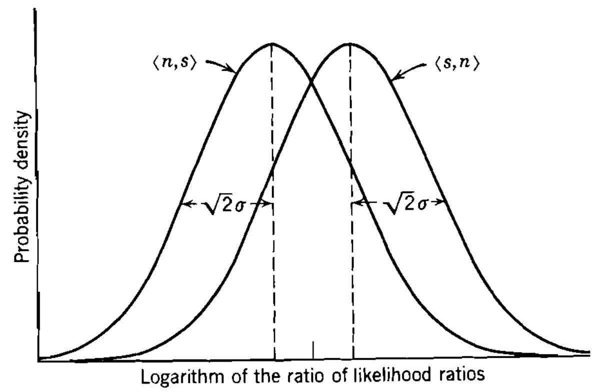

Fig. 4. Normal distributions of $\mathbf{X}_1 - \mathbf{X}_2$ for the two-alternative forced-choice design assuming **S** and **N** are normally distributed with standard deviation σ and separated by an amount d.

6.3), or when each interval includes a stimulus and they differ only in one dimension, as in the traditional discrimination experiments (see Sec. 3.1 of Chapter 4), then it is usually assumed that the effects of successive presentations are correlated. No very firm argument has been given why this should be, but the feeling seems to be something to the effect that the random errors introduced by the subject are due to comparatively slow changes in his reactions to the dimension being varied. When, however, the background is random noise and the stimulus is a tone, it is believed that it is more plausible to suppose that successive presentations have effects that are independent of one another. Under that assumption, it is easy to determine the two distributions we need by invoking the well-known fact that the distribution of the difference of two independent, normally distributed random variables is normal with mean equal to the difference of the means and variance equal to the sum of the variances. Thus the distribution of effects from $\langle s, n\rangle$ is normal with mean d and standard deviation $\sqrt{2}\sigma$, and that from $\langle n, s\rangle$, normal with mean $-d$ and standard deviation also $\sqrt{2}\sigma$. The situation is shown graphically in Fig. 4. From this and the assumed decision rule, Eq. 10, we obtain

$$p(1 \mid \langle s, n \rangle) = \frac{1}{2\sqrt{\pi}\sigma} \int_{c'}^{\infty} \exp\left[-\frac{(z-d)^2}{4\sigma^2}\right] dz$$
$$p(1 \mid \langle n, s \rangle) = \frac{1}{2\sqrt{\pi}\sigma} \int_{c'}^{\infty} \exp\left[-\frac{(z+d)^2}{4\sigma^2}\right] dz. \qquad (11)$$

Observe that $c' = 0$ corresponds to no bias in the sense that $p(1 \mid \langle s, n\rangle) = p(2 \mid \langle n, s\rangle)$. The data from two-alternative experiments suggest that there is little or no bias when the presentation probability is $\frac{1}{2}$ and the payoff matrix is symmetric, so we will assume that $c' = 0$ for such studies. The common response probability, the probability of a correct response, is denoted $p_2(C)$ —2 for the number of alternatives and C for "correct." Given an estimate of $p_2(C)$ from data, we can calculate the corresponding normal deviate $2d/\sqrt{2}\sigma = \sqrt{2}d'$, where d' is the sensitivity parameter used in the analysis of the corresponding Yes-No experiments. Thus d' can be independently estimated from both Yes-No and two-alternative forced-choice experiments. Later, in Sec. 2.4, we compare Yes-No and forced-choice estimates for two sets of data.

2.2 Choice Analysis

If we assume that the choice model, Eq. 5 and Assumptions 1 to 3, hold, and if we denote $\eta(s, n)$ by η, $b(N)/b(Y)$ by b, and recall that $\eta(s, s) = 1$, then the scale values for the Yes-No experiment are

$$\begin{array}{lc} & \text{Response} \\ & \begin{array}{cc} Y & N \end{array} \\ \begin{array}{l} \text{Stimulus} \\ \text{Presentation} \end{array} & \begin{array}{c} s \\ n \end{array} \begin{bmatrix} 1 & \eta b \\ \eta & b \end{bmatrix}. \end{array}$$

The corresponding confusion matrix is

$$\begin{array}{c} \\ s \\ \\ n \end{array} \begin{bmatrix} \dfrac{1}{1+\eta b} & \dfrac{\eta b}{1+\eta b} \\ \\ \dfrac{\eta}{\eta+b} & \dfrac{b}{\eta+b} \end{bmatrix} \quad (\text{columns } Y, N). \tag{12}$$

In what follows, we usually write down only the tables of scale values and not the corresponding probability tables, which are obtained by dividing each scale value by the sum of the scale values in its row.

As with signal detectability theory, this model describes the response frequencies perfectly: there are two parameters to account for two independent probabilities. The equations for the parameters are

$$\begin{aligned} \eta &= \left[\frac{p(N \mid s)\, p(Y \mid n)}{p(Y \mid s)\, p(N \mid n)}\right]^{1/2} \\ b &= \left[\frac{p(N \mid s)\, p(N \mid n)}{p(Y \mid s)\, p(Y \mid n)}\right]^{1/2}. \end{aligned} \tag{13}$$

Our interest in η and b is not as simple transformations of the Yes-No data but in the possibility that they can be used to predict other data. This possibility stems from our interpretation of the two scales: under otherwise fixed experimental conditions and for a given subject, η is supposed to depend only upon the stimuli and b, upon the payoffs, presentation probabilities, and instructions. It is believed that η is a measure of the subject's detection sensitivity, just as d' is in the first model, and that b is a bias which, like c, reflects the relative attractiveness to him of the two responses.

The matrix of scale values for the two-alternative, forced-choice design is

$$\begin{array}{c} \\ \langle s, n\rangle \\ \langle n, s\rangle \end{array} \begin{array}{c} \begin{array}{cc} 1 & \qquad\qquad\qquad 2 \end{array} \\ \begin{bmatrix} \eta(\langle s, n\rangle, \langle s, n\rangle)b(1) & \eta(\langle s, n\rangle, \langle n, s\rangle)b(2) \\ \eta(\langle n, s\rangle, \langle s, n\rangle)b(1) & \eta(\langle n, s\rangle, \langle n, s\rangle)b(2) \end{bmatrix} \end{array}.$$

By Assumption 2, we know that

$$\eta(\langle s, n\rangle, \langle s, n\rangle) = \eta(\langle n, s\rangle, \langle n, s\rangle) = 1,$$

and because $S \subset \mathscr{S} \times \mathscr{S}$, it is reasonable to invoke Assumption 4 under the same conditions as we did the independence assumption in the signal detectability model (e.g., a noise background) yielding

$$\begin{aligned} [-\log \eta(\langle n, s\rangle, \langle s, n\rangle)]^2 &= [-\log \eta(n, s)]^2 + [-\log \eta(s, n)]^2 \\ &= 2[-\log \eta(s, n)]^2 \\ &= [-\sqrt{2} \log \eta]^2 \\ &= [-\log \eta^{\sqrt{2}}]^2. \end{aligned}$$

Thus, by Assumption 1,

$$\eta(\langle n, s\rangle, \langle s, n\rangle) = \eta(\langle s, n\rangle, \langle n, s\rangle) = \eta^{\sqrt{2}},$$

and so, letting $b' = b(2)/b(1)$, the matrix of scale values reduces to

$$\begin{array}{c} \\ \langle s, n\rangle \\ \langle n, s\rangle \end{array} \begin{array}{c} \begin{array}{cc} 1 & \quad 2 \end{array} \\ \begin{bmatrix} 1 & \eta^{\sqrt{2}}b' \\ \eta^{\sqrt{2}} & b' \end{bmatrix} \end{array}, \tag{14}$$

where η is the Yes-No detection parameter. Note that $b' = 1$ corresponds to no response bias.

Again, there is no question about the model reproducing the data. What is not automatic is that the estimate of η from the Yes-No data will be the same as that from the forced-choice data, as is alleged by the theory.

Before turning to the threshold analysis of these two detection experiments, a possible objection to our analysis must be examined. It may be quite misleading to treat the Yes-No experiment as one in which an absolute identification of s or n is made, for in many studies the subject hears the noise background both before and after the marked interval in which the stimulus may occur. That being so, it seems more realistic to say that the two stimulus presentations are $\langle n_b, n, n_a\rangle$ and $\langle n_b, s, n_a\rangle$, where n_b denotes the noise before and n_a, the noise after the interval. If these three intervals are uncorrelated, we may invoke Assumption 4 as follows:

$$\begin{aligned}[-\log \eta(\langle n_b, s, n_a\rangle, \langle n_b, n, n_a\rangle)]^2 \\ &= d(\langle n_b, s, n_a\rangle, \langle n_b, n, n_a\rangle)^2 \\ &= d(n_b, n_b)^2 + d(s, n)^2 + d(n_a, n_a)^2 \\ &= [-\log \eta(n_b, n_b)]^2 + [-\log \eta(s, n)]^2 + [-\log \eta(n_a, n_a)]^2.\end{aligned}$$

But $\eta(n_b, n_b) = \eta(n_a, n_a) = 1$, hence $\log \eta(n_b, n_b) = \log \eta(n_a, n_a) = 0$, and so

$$\eta(\langle n_b, s, n_a\rangle, \langle n_b, n, n_a\rangle) = \eta(s, n) = \eta.$$

Thus the more precise analysis leads to the same result as the simpler one, provided that the effects in successive intervals are uncorrelated.

2.3 Threshold Analysis

Let us suppose that when the stimulating conditions are held constant the two cut-offs k and k' of the threshold model are fixed quantities independent of the presentation probabilities, the payoffs, and the experimental design. So the true detection probabilities can be written simply as $q(s)$ and $q(n)$. For the Yes-No design, we may summarize the "internal" detection observations as

$$\begin{array}{llcc} & & \multicolumn{2}{c}{\text{Observation}} \\ & & D & \bar{D} \\ \text{Stimulus} & s & q(s) & 1 - q(s) \\ \text{Presentation} & n & q(n) & 1 - q(n) \end{array}$$

These, however, are not the response probabilities.

Suppose that the subject wishes to reduce his false-alarm rate—Yes responses to noise—below the true rate $q(n)$; then we assume that in addition to saying No to all $\bar{D}$ observations he also falsely responds No to some proportion $1 - t$ of his D observations. This means, therefore, that he responds Yes to only a proportion t of his D observations, that is, if $p(Y \mid n) \leqslant q(n)$,

$$\begin{aligned} p(Y \mid s) &= tq(s) \\ p(Y \mid n) &= tq(n), \end{aligned} \tag{15}$$

where $0 \leqslant t \leqslant 1$. Similarily, when he wishes to increase his rate of correct Yes responses above his true rate $q(s)$, albeit at the same time increasing his false-alarm rate above $q(n)$, he is assumed to say Yes to all D observations and to some proportion u of his $\bar{D}$ observations, that is, if $p(Y \mid n) \geqslant q(n)$

$$\begin{aligned} p(Y \mid s) &= q(s) + u[1 - q(s)] \\ p(Y \mid n) &= q(n) + u[1 - q(n)], \end{aligned} \tag{16}$$

where $0 \leqslant u \leqslant 1$.

It is clear that even if we knew which of these processes, Eq. 15 or 16, a subject had used, we would still have no way of testing the model because there are three parameters, $q(s)$, $q(n)$, and t or u, to account for two independent probabilities. Moreover, going to the two-alternative forced-choice design as we did for the other two models does not provide us with a test; however, in the next section testable conclusions are derived.

Before describing one possible threshold analysis of the two-alternative forced-choice experiment, we examine the familiar technique to correct for guessing (see Blackwell, 1953; Swets, 1961b; and Tanner & Swets, 1954a). The so-called high-threshold model underlying this technique assumes that the true probability of a false alarm, $q(n)$, is zero but that the observed rate is positive because the subject inflates the number of Yes responses to s beyond its true value $q(s)$. Thus Eq. 16 represents the situation, and, with $q(n) = 0$, it follows that $u = p(Y \mid n)$. Substituting this into the expression for $p(Y \mid s)$ and solving yields

$$q(s) = \frac{p(Y \mid s) - p(Y \mid n)}{1 - p(Y \mid n)}. \tag{17}$$

This formula is frequently recommended to "correct" threshold data for guessing. Because this correction involves the rather strong assumption that $q(n) = 0$, which, as we shall see in the next section, is surely incorrect, I very much doubt that this "correction" should be made.

On each trial of a two-alternative forced-choice experiment, the subject is confronted with two Yes-No determinations—one for each time interval. This means that there are four possible observation states, $\langle D, \bar{D} \rangle$, $\langle \bar{D}, D \rangle$, $\langle D, D \rangle$, or $\langle \bar{D}, \bar{D} \rangle$, where D denotes a detection observation in the Yes-No situation and $\bar{D}$, a nondetection observation. The first, $\langle D, \bar{D} \rangle$, surely suggests making response 1; the second, response 2; but the last two give him no hint how to respond. Presumably, these are the observations he should bias, at least when there is nothing to drive him to extreme biases. We assume this.

To write the equations for the response probabilities, we need to know the probabilities of each of these observation outcomes for each of the presentations. As with the other two models, we assume that successive stimulus effects are independent, so that, for example, the probability of a $\langle D, \bar{D}\rangle$ observation when $\langle s, n\rangle$ is presented is simply $q(s)[1 - q(n)]$. The other cases are similar:

$$\begin{aligned} p(1 \mid \langle s, n\rangle) &= q(s)[1 - q(n)] + vq(s)\,q(n) + w[1 - q(s)][1 - q(n)] \\ p(1 \mid \langle n, s\rangle) &= q(n)[1 - q(s)] + vq(n)\,q(s) + w[1 - q(n)][1 - q(s)], \end{aligned} \tag{18}$$

where v and w are biasing parameters such that $0 \leqslant v, w \leqslant 1$. Again, the model has too many parameters to permit any check on it with just forced-choice data. In Sec. 3 we discuss experiments in which it, as well as the other models, can be tested.

2.4 Comparison of Models with Data

Swets (1959) reported data for three subjects run in both Yes-No and two-alternative forced-choice designs. The stimuli were 1000-cps tones of 100-ms duration at several different energy levels in a background of white noise. Five hundred observations were obtained from each subject in each energy-design condition. The presentation probabilities were approximately $\frac{1}{2}$ and a symmetric payoff matrix was used. The data, which Professor Swets has kindly provided me, and the estimates of d' for two different designs are shown in Table 1.

Shipley (1961) also ran three subjects in both designs, using a background of white noise and 500- and 1000-cps stimuli at one energy level each. A total of 1600 observations were obtained in each condition for each subject using presentation probabilities of $\frac{1}{2}$ and a symmetric payoff matrix. The data and d' estimates are shown in Table 2. With the exception of subject 3 on the 1000-cps stimulus, all pairs of estimates are within 10 per cent of each other.

Shipley's estimates seem considerably more consistent than Swets's, but in large part this is due to the increased number of observations. To see this, suppose $d' = 1.2$ and $p(Y \mid n) = 0.2$, then $p(Y \mid s) = 0.640$. An increase of 10 per cent in d' and $p(Y \mid n) = 0.2$ yields $p(Y \mid s) = 0.683$. With 250 s presentations, a difference of $0.683 - 0.640 = 0.043$ corresponds to about 1.4 standard deviations, whereas with 800 presentations it corresponds to about 2.5 standard deviations.

These same data are reanalyzed in terms of the choice model in Tables 3 and 4, and much the same pattern is exhibited as for the detectability model. For example, the largest difference in Table 4 is in the same place

Table 1 Yes-No and Two-Alternative Forced-Choice Acoustic Data (Swets, 1959) and the Corresponding Estimates of d'

Subject	S/N in db	$p(Y \mid s)$	$p(Y \mid n)$	$p(1 \mid \langle s, n \rangle)$	$p(1 \mid \langle n, s \rangle)$	Yes-No d'	Two-Alternative Forced-Choice d'
1	9.4	0.793	0.226	0.824	0.187	1.57	1.29
	14.5	0.872	0.180	0.931	0.060	2.05	2.15
	16.6	0.902	0.120	0.963	0.071	2.45	2.29
2	9.4	0.753	0.288	0.670	0.149	1.24	1.03
	11.7	0.771	0.254	0.777	0.194	1.40	1.14
	14.5	0.833	0.295	0.854	0.145	1.51	1.50
	16.6	0.867	0.232	0.855	0.078	1.83	1.83
3	9.4	0.731	0.195	0.835	0.149	1.48	1.43
	11.7	0.836	0.254	0.870	0.142	1.65	1.56
	14.5	0.816	0.169	0.959	0.125	1.85	1.96
	16.6	0.895	0.149	0.953	0.037	2.29	2.45

The stimuli were 1000-cps tones of 100-ms duration in noise of 50 db re 0.0002 d/cm^2; 500 observations were made on each subject at each energy level for each condition. Presentation probabilities of about $\frac{1}{2}$ and symmetric payoff matrices were used.

Table 2 Yes-No and Two-Alternative Forced-Choice Acoustic Data (Shipley, 1961) and the Corresponding d'

Subject	$p(Y \mid s)$	$p(Y \mid n)$	$p(1 \mid \langle s, n \rangle)$	$p(2 \mid \langle n, s \rangle)$	Yes-No d'	Forced-Choice d'
	Stimulus 1					
1	0.768	0.148	0.895	0.880	1.78	1.72
2	0.712	0.258	0.798	0.796	1.20	1.18
3	0.746	0.216	0.836	0.866	1.44	1.48
	Stimulus 2					
1	0.695	0.201	0.835	0.838	1.35	1.38
2	0.675	0.199	0.795	0.812	1.30	1.20
3	0.693	0.287	0.791	0.832	1.07	1.25

Each stimulus lasted for 100 ms in a 500-ms interval and was imbedded in wide band noise at 0.0435 volt across the terminals of the ear phones. Stimulus 1 was 500 cps at 0.0023 volt and Stimulus 2 was 1000 cps at 0.0026 volt. Each presentation of each condition occurred approximately 800 times. Presentation probabilities of $\frac{1}{2}$ and symmetric payoff matrices were used.

and is of comparable magnitude. In both analyses the Yes-No parameters do not seem to be consistently larger or smaller than the forced-choice ones.

Thus, although the two theories differ in their approach, it is evident that they do not differ appreciably in their predictions from one simple detection design to another. To see this more vividly, we make the following calculation. For each of several different values of d', determine from Eq. 9 the values of $p(Y \mid s)$ and $p(Y \mid n)$ corresponding to different

Table 3 Estimates of η for Swets's (1959) Data

Subject	S/N in db	η Yes-No	η Forced-Choice
1	9.4	0.276	0.344
	14.5	0.180	0.151
	16.6	0.122	0.100
2	9.4	0.365	0.419
	11.7	0.319	0.394
	14.5	0.290	0.285
	16.6	0.216	0.201
3	9.4	0.299	0.304
	11.7	0.258	0.270
	14.5	0.214	0.164
	16.6	0.144	0.109

See Table 1 for a description of the experimental conditions.

Table 4 Estimates of η for Shipley's (1961) Data

	Stimulus 1		Stimulus 2	
Subject	Yes-No	Forced-Choice	Yes-No	Forced-Choice
1	0.238	0.242	0.332	0.317
2	0.375	0.369	0.347	0.370
3	0.305	0.290	0.421	0.356

See Table 2 for a description of the experimental conditions.

choices of c. Elliot's (1959) tables of d' are handy for this. For these pairs of probabilities, determine η from Eq. 13. A plot of the logarithm of η versus d' is shown in Fig. 5. It is evident that the relation is approximately linear and that the correlation is high. The points that differ the most from the main trend are those for which at least one of the probabilities is near 0 or 1.

As noted earlier, the threshold model has too many parameters to be tested with these data.

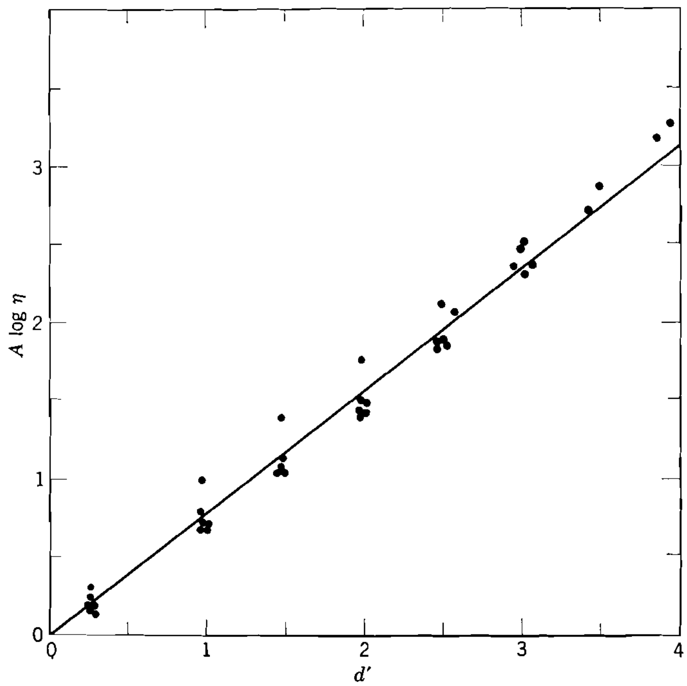

Fig. 5. Relationship between the stimulus parameters d' and η of the signal detectability and choice models for the Yes-No design.

3. ISOSENSITIVITY CURVES

It has been suggested that the parameters d', η, $q(s)$, and $q(n)$ of the three theories represent the subject's sensitivity to stimuli and that c, b, t, and u are biases more or less under his control. Among the experimental conditions that are thought to affect the biases, but not the stimulus parameters, are the presentation probabilities, the payoffs, and the names given to the response alternatives. It is, of course, necessary to show empirically that there is some justice to these interpretations and, if there is, to examine how the parameters are related to objective features of the experiment. The justification we consider here; the dependence of model parameters upon experimental parameters is treated in Secs. 5 and 6.

Suppose that we have a physiologically stable subject (no drugs, minimal fatigue, etc.), and suppose that d', η, or $q(s)$ and $q(n)$, as the case may be, depend only upon the stimulus and noise, then up to errors due to binomial variability our estimates of them should be the same no matter what instructions, presentation probabilities, and payoffs are used. This should be true in spite of the fact that varying these factors produces changes in the response probabilities. For example, if in a Yes-No design one of the two types of errors is very costly, we anticipate that subjects will bias their responses away from the one that makes this expensive error possible. In terms of the models, the payoffs affect either their choice of c, of b, or of t or u. Nevertheless, if one of these models is correct, then as the bias parameters change the response probabilities are constrained by one of the Eqs. 9, 12, 15, or 16. That is to say, each model establishes an exchange relation between the two probabilities so that if one probability is altered the other is also and in a predetermined way. To find the equation for this relation, one merely eliminates the biasing parameter from the pair of equations determining $p(Y \mid s)$ and $p(Y \mid n)$. The resulting function depends upon the stimulus parameters but not upon the biases. Plots of these relations I shall call *isosensitivity* curves.[4] In the literature they are commonly called R.O.C. curves, which stands for *receiver operating characteristic curves*, a term used in the original engineering publications. Because this seems an unfortunate psychological phrase, I suggest that it be changed.

The isosensitivity curves for the choice model can easily be written down. From Eq. 12, we know

$$p(Y \mid s) = \frac{1}{1 + \eta b}$$

$$p(Y \mid n) = \frac{\eta}{\eta + b},$$

and, eliminating b, we obtain

$$\left[\frac{1 - p(Y \mid s)}{p(Y \mid s)}\right]\left[\frac{p(Y \mid n)}{1 - p(Y \mid n)}\right] = \eta^2. \tag{19}$$

One cannot write an explicit function for the isosensitivity curves of the signal detectability model because Eq. 9 involves integrals of normal distributions, but it is easy to calculate them numerically. The curves for the two theories, along with detection data for a pure tone in noise (Tanner, Swets, & Green, 1956), are shown in Fig. 6. These data were generated by varying P from 0.1 to 0.9 in steps of 0.2; each point is based upon a total of 300 observations. It is clear that the theories produce substantially the same curves and that both are in reasonable accord with the data.

[4] Because "sensitivity" derives from Latin, one should use the term "equisensitivity," but the common use of "iso" in similar scientific contexts makes the term of mixed origin seem more natural to nonpurists.

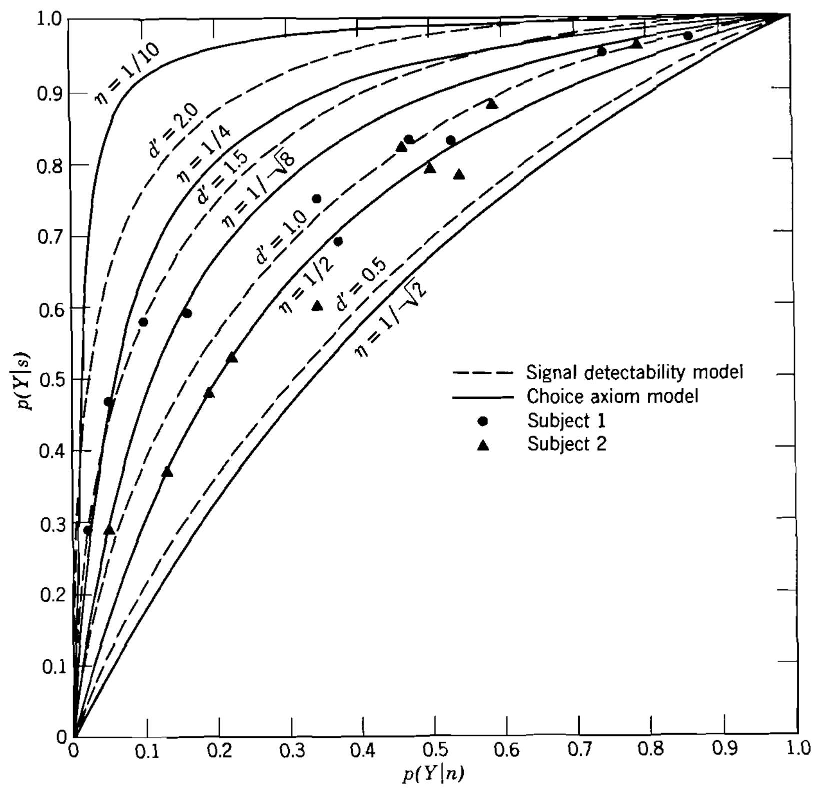

Fig. 6. Typical signal detectability and choice model isosensitivity curves for Yes-No design. The data points, reported in Tanner, Swets, & Green (1956), were obtained by presenting pure tones in noise, with P varied from 0.1 to 0.9 in steps of 0.2 and a fixed symmetric payoff matrix.

Note that the theoretical curves in Fig. 6 are symmetric about the main diagonal that runs from $\langle 0, 1 \rangle$ to $\langle 1, 0 \rangle$. Not all data are symmetric, however, as the visual ones shown in Fig. 7 for one of four subjects studied by Swets, Tanner, and Birdsall (1955, 1961) indicate. The empirical isosensitivity curve was swept out by varying the payoffs and holding $P = \frac{1}{2}$. The data for the other subjects are similar. It is evident that these data reject both the choice and the equal-variance signal detectability models. The theoretical curves of Fig. 7, one of which corresponds reasonably well with the data, were obtained from the detectability model by assuming that the stimulus plus noise standard deviation is $1 + \frac{1}{4}d'$ times the noise standard deviation. For these data, d' is in the range of 2 to 4, so the factor is 1.5 to 2. Thus a second stimulus parameter allows detectability theory to account for these data, but I am at a loss to understand why adding a faint tone to the noise should have such major repercussions on the variance of the distribution of effects.

In all likelihood, there is some plausible way to add a second stimulus parameter to the choice model so that it does just about as well, but none has yet been suggested.

The threshold theory isosensitivity curves are obtained by eliminating t from Eq. 15 and u from Eq. 16:

$$p(Y \mid s) = \begin{cases} p(Y \mid n)\left[\dfrac{q(s)}{q(n)}\right], & \text{if } p(Y \mid n) \leqslant q(n) \\ p(Y \mid n)\left[\dfrac{1 - q(s)}{1 - q(n)}\right] + \dfrac{q(s) - q(n)}{1 - q(n)}, & \text{if } p(Y \mid n) \geqslant q(n). \end{cases} \tag{20}$$

This equation represents two line segments: one from $\langle 0, 0\rangle$ to $\langle q(n), q(s)\rangle$, which is referred to as the *lower limb*, and the other, the *upper limb*, from $\langle q(n), q(s)\rangle$ to $\langle 1, 1\rangle$.

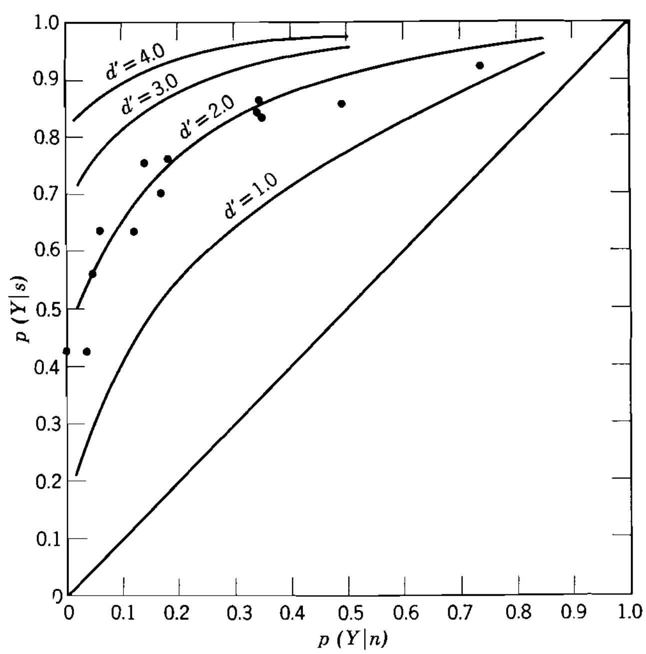

Fig. 7. Nonsymmetric signal detectability isosensitivity curves for the Yes-No design. The data points were obtained by presenting local increases in light intensity, with $P = \frac{1}{2}$ and different payoff matrices. See text for an explanation of the theoretical curves. Adapted with permission from Swets, Tanner, & Birdsall (1961, p. 319).

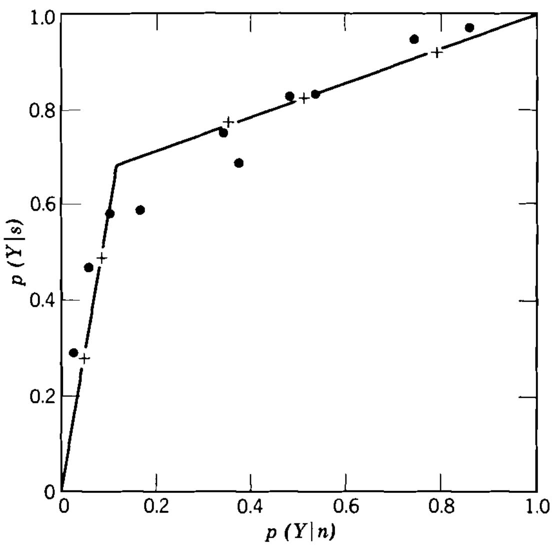

Fig. 8. Threshold isosensitivity curve fitted by eye to subject 1 acoustic data previously plotted in Fig. 6. The theoretical crosses are explained in Sec. 5.2.

The traditional "correction-for-guessing" procedure assumes that $q(n) = 0$, in which case the lower limb runs along the ordinate to $q(s)$ at which point the upper limb departs for $\langle 1, 1\rangle$. It is abundantly clear that the $q(n) = 0$ model does not describe the data of Figs. 6 or 7; as a result, Tanner and Swets (1954a) concluded that these detection data reject the high-threshold hypothesis. Sometimes their conclusion has been interpreted as a rejection of all sensory thresholds, but the more general threshold model appears to be quite adequate. In Fig. 8, threshold curves are fitted to Tanner, Swets, and Green's (1956) subject 1 data and in Fig. 9 to those of Swets, Tanner, and Birdsall's (1955) subject 4. These theoretical curves are comparable to those of Fig. 7, not to the symmetric ones of Fig. 6, because the threshold model, like the unequal-variance signal-detectability model, has two estimated parameters. The threshold curves appear to be just as satisfactory as the signal-detectability ones.

I shall not carry out the parallel development of isosensitivity curves for the two-alternative forced-choice design using the detectability or choice models. Suffice it to say that the same equations result, except that $d'/\sqrt{2}$ replaces d' and $\eta^{\sqrt{2}}$ replaces η. The isosensitivity curve for the threshold model is obtained by subtracting the second expression in Eq. 18 from the first and rewriting the result as

$$p(1 \mid \langle s, n\rangle) = p(1 \mid \langle n, s\rangle) + q(s) - q(n). \tag{21}$$

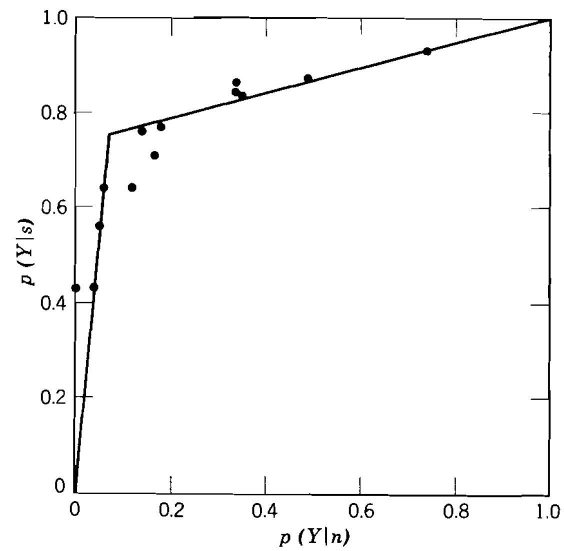

Fig. 9. Threshold isosensitivity curve fitted by eye to subject 4 visual data previously plotted in Fig. 7.

This represents a straight line with slope 1. By setting

$$v = w = \frac{q(s)\,q(n)}{q(s)\,q(n) + [1 - q(s)][1 - q(n)]}$$

in Eq. 18, we see that the isosensitivity curve passes through $\langle q(n), q(s)\rangle$. Thus, when the stimulating conditions are the same, the Yes-No and forced-choice isosensitivity curves must be related, as shown in Fig. 10; so,

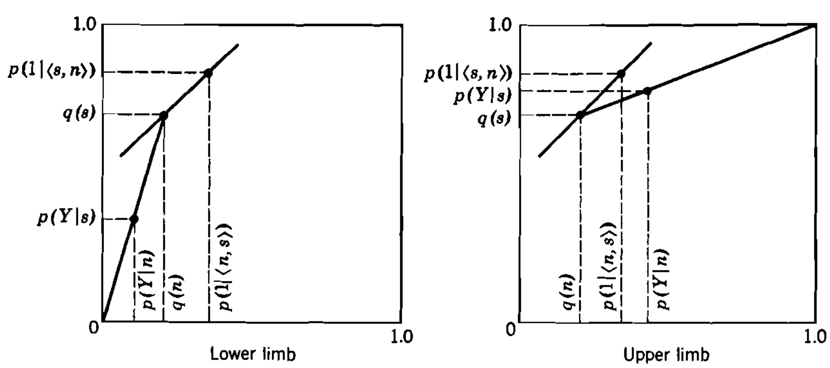

Fig. 10. The geometry relating the threshold Yes-No and two alternative forced-choice isosensitivity curves.

aside from the ambiguity about whether the Yes-No data point lies on the upper or lower limb of its isosensitivity curve, the data from the two experiments provide an estimate of the true detection probabilities. This method seems a suitable replacement for the incorrect correction-for-guessing procedure. Examples of such estimates can be found in Luce (1963), or they can be reconstructed from the raw data presented in Tables 1 and 2.

For some reason not apparent to me, no one has yet reported two-alternative forced-choice data where the presentation probabilities or payoff function are varied, and so we do not know what the empirical isosensitivity curves look like.

In summary, the following conclusions seem justified.

1. The models with but one stimulus parameter, including the threshold model that underlies the correction-for-guessing equation (17), are inadequate to account for existing visual data. In addition, the $q(n) = 0$ threshold model is incorrect for acoustic data.
2. The detectability and threshold models that have two estimated stimulus parameters both handle the visual data quite well.
3. The stimulus parameters for either of these models can be estimated from empirical Yes-No isosensitivity curves, which can be generated by varying the presentation probability, the payoffs, or both. In addition, a simple scheme exists to find the "true" threshold probabilities, which uses data from just one run in each of the Yes-No and two-alternative forced-choice designs.

4. COMPLEX DETECTION

Three somewhat more complex detection designs are our next concern. The first is the *multiple-look Yes-No design* in which there are m distinct intervals, all or none of which contain the stimulus. Thus $S = \{\langle s, s, \ldots, s\rangle, \langle n, n, \ldots, n\rangle\}$, $R = \{Y, N\}$, and $\iota(Y) = \langle s, s, \ldots, s\rangle$ and $\iota(N) = \langle n, n, \ldots, n\rangle$.

The second is the *k-alternative forced-choice design* in which there are k intervals, exactly one of which contains the stimulus. If we let s_i denote the presentation in which s occurs in the ith interval and n in all others, $S = \{s_1, s_2, \ldots, s_k\}$, $R = \{1, 2, \ldots, k\}$, and $\iota(r) = s_r$ for $r \in R$.

The third is the *multiple-look k-alternative forced-choice* design in which the k-alternative forced-choice presentation s_i is repeated a total of m times. Thus $S = \{\langle s_1, s_1, \ldots, s_1\rangle, \langle s_2, s_2, \ldots, s_2\rangle, \ldots, \langle s_k, s_k, \ldots, s_k\rangle\}$, $R = \{1, 2, \ldots, k\}$, and $\iota(r) = \langle s_r, s_r, \ldots, s_r\rangle$ for $r \in R$.

For each of these designs, we assume that the noise is uncorrelated on successive presentations, except when explicitly stated otherwise. In particular, this means that successive presentations cannot simply be tape recordings of the initial one. The reason for imposing this experimental limitation is to permit us to assume independence of effects in the analysis, as was done earlier in the discussions of the two-alternative forced-choice design.

4.1 Signal Detectability Analysis

Let $\mathbf{X}_i$ denote the logarithm of the likelihood ratio of the observation on the ith look of the multiple-look Yes-No experiment; then it is assumed that $\mathbf{X} = \sum_{i=1}^{m} \mathbf{X}_i$ is the random variable used to arrive at a decision. First, suppose noise is presented m times. If each presentation is independent and normally distributed with mean 0 and standard deviation σ, then $\mathbf{X}$ is normally distributed with mean 0 and standard deviation $\sigma_m = \sqrt{m}\sigma$. Similarly, if the stimulus is presented each time and the presentations are independent, then the mean is $d_m = md$ and the standard deviation is $\sigma_m = \sqrt{m}\sigma$. Thus the effective detection parameter is

$$d_m' = \frac{d_m}{\sigma_m} = \frac{md}{\sqrt{m}\sigma} = \sqrt{m}d', \tag{22}$$

and so we can predict multiple-look data from simple Yes-No data (Swets, Shipley, McKey, & Green, 1959).

The generalization of the two-alternative signal detectability model to the k-alternative forced-choice design is comparatively complicated if response biases are included and very simple if they are not. I shall sketch the general idea of the former and carry out the latter in detail.

As presented in terms of differences, it is not easy to see how to generalize the two-alternative analysis; however, if we view it in a different but equivalent way, the outlines of the generalization become clear (Swets & Birdsall, 1956). As before, suppose that the observations in the two intervals are independent, in which case it is plausible to represent the two decision axes as orthogonal coordinates in the plane. Joint normal distributions for $\langle s, n \rangle$ and for $\langle n, s \rangle$ are assumed to exist and to have equal variances. When projected on either axis, these distributions generate the usual one-dimensional noise and signal distributions, the means of the noise distributions being at the origin. This is diagrammed in Fig. 11. The observational random variable is the pair $(\mathbf{X}_1, \mathbf{X}_2)$, and the decision rule is no longer characterized by a point but by a division of the plane

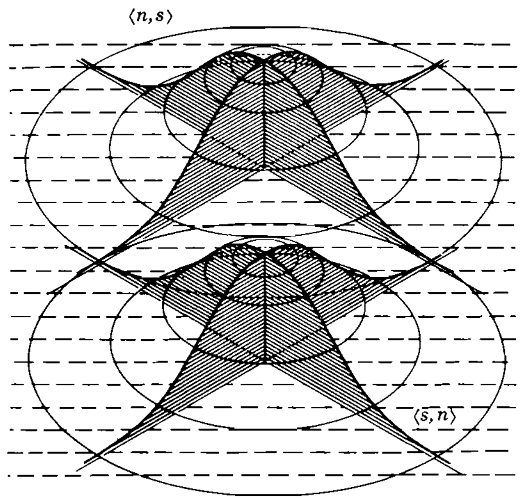

Fig. 11. The two-dimensional signal detectability representation of the two-alternative forced-choice experiment. The dotted lines represent the family of decision rules that correspond to the cutpoints in the decision axis representation.

into two nonoverlapping regions. Under reasonable assumptions about the subject's goals, it can be shown that the division of the plane must be by a line located at 45° between the two decision axes; typical ones are shown dotted. It is not difficult to see that our original representation in terms of differences is simply the projection of the present model onto a plane orthogonal both to this family of 45° lines and to the decision plane. The intersection of each 45° line with this plane corresponds to a possible cut-point c'.

The generalization to k-alternatives is now clear. There are k random variables, $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_k$, corresponding to observations in each of the intervals. For each stimulus presentation, s_i, where the stimulus is in the ith interval and noise is in the others, there is a density function having the value $p(x_1, x_2, \ldots, x_k \mid s_i)$ at $\mathbf{X}_1 = x_1$, $\mathbf{X}_2 = x_2, \ldots,$ and $\mathbf{X}_k = x_k$. These are assumed to be independent multivariate normal distributions with equal variances. The decision rule is a partition of the k-dimensional Euclidean space into k response regions; the simplest rule involves a division of the space by hyperplanes. The mathematics required for specific numerical calculations is, of course, rather clumsy, and, so far as I know, no actual work has used this general form of the model.

If, however, we assume that the payoffs are symmetric and that the subject introduces no biases, matters are very much simpler. The subject makes an observation, $\mathbf{X}_i$, in each interval, and he is assumed simply to say that the stimulus is located in the interval having the largest observation. Because there is no bias, it does not matter which interval actually contains the stimulus—the probability of a correct response, $p_k(C)$, is the same for all. The probability density that s generates an effect x that is the largest is simply $p(x \mid s)$ times the probability that all $k - 1$ of the noise observations are less than x, that is, $p(x \mid s)\, P(x \mid n)^{k-1}$, where $P(x \mid n) = \int_{-\infty}^{x} p(z \mid n)\, dz$. Because the particular value of the largest value is immaterial to the response made, we integrate over all x to obtain

$$p_k(C) = \int_{-\infty}^{\infty} P(x \mid n)^{k-1}\, p(x \mid s)\, dx. \tag{23}$$

Of course, we assume that $p(x \mid s)$ and $p(x \mid n)$ are normal, have the same variance, and are separated by an amount d, just as in the Yes-No model. That being so, an estimate of d' from either the Yes-No or two-alternative forced-choice data is sufficient to predict $p_k(C)$ from Eq. 23 (Tanner & Swets, 1954a).

The analysis of the multiple-look k-alternative forced-choice design is analogous to that for the multiple-look Yes-No design (Swets, Shipley, McKey, & Green, 1959). The subject is assumed to make observation $\mathbf{X}_{ij}$ in interval i on the jth observation. The sums $\sum_{j=1}^{m} \mathbf{X}_{ij}$ are calculated, and then the subject chooses the interval having the largest sum. Because the $\mathbf{X}_{ij}$ are assumed to be independent and normally distributed, $\sum_{j=1}^{m} \mathbf{X}_{ij}$ is normally distributed with mean 0 and standard deviation $\sqrt{m}\sigma$ when n is presented in interval i and with mean md and standard deviation $\sqrt{m}\sigma$ when s is presented. Thus Eq. 23 can be used to calculate the probability of a correct detection. Note that if $d'(k)$ denotes the value of d' estimated from the simple k-alternative forced-choice design and $d_m'(k)$ denotes that corresponding to the m-look design, they are related by

$$d_m'(k) = \sqrt{m}\, d'(k). \tag{24}$$

4.2 Choice Analysis

By repeated use of the independence Assumption 4, it is easy to see that the choice model matrix of scale values for the multiple look Yes-No design is

$$\begin{array}{c} \\ \langle s, s, \ldots, s\rangle \\ \langle n\ n, \ldots, n\rangle \end{array} \begin{array}{c} \begin{array}{cc} Y & \quad N \end{array} \\ \begin{bmatrix} 1 & \eta^{\sqrt{m}} b \\ \eta^{\sqrt{m}} & b \end{bmatrix}. \end{array} \tag{25}$$

The argument is much the same as that used for the two-alternative forced-choice design.

Letting s_i denote the presentation in which the stimulus is in interval i, then by a similar argument we obtain as the matrix for the k-alternative forced-choice design

$$\begin{array}{c} \\ s_1 \\ s_2 \\ \cdot \\ \cdot \\ \cdot \\ s_k \end{array} \begin{array}{c} \begin{array}{cccc} 1 & 2 & \ldots & k \end{array} \\ \left[\begin{array}{cccc} 1 & \eta^{\sqrt{2}} & \ldots & \eta^{\sqrt{2}} \\ \eta^{\sqrt{2}} & 1 & \ldots & \eta^{\sqrt{2}} \\ & & & \\ \ldots & \ldots & \ldots & \ldots \\ & & & \\ \eta^{\sqrt{2}} & \eta^{\sqrt{2}} & \ldots & 1 \end{array}\right] \end{array}, \tag{26}$$

where I have omitted writing the response biases. When the biases are equal—the assumption we shall make in analyzing data—the equation for the probability of a correct response is seen to be

$$p_k(C) = \frac{1}{1 + (k - 1)\eta^{\sqrt{2}}}. \tag{27}$$

When the biases are not assumed equal, then it is easy to see from Eq. 26 that

$$\prod_{i=1}^{k} \prod_{j=1}^{k} \frac{p(j \mid s_i)}{p(i \mid s_i)} = \eta^{\sqrt{2}\,k(k-1)}$$

and

$$\prod_{l=1}^{k} \frac{p(i \mid s_l)}{p(j \mid s_l)} = \left[\frac{b(i)}{b(j)}\right]^k;$$

these equations can be used to estimate the parameters η and $b(i)$.

Because $(\eta^{\sqrt{2}})^{\sqrt{m}} = \eta^{\sqrt{2m}}$, the matrix for the multiple-look k-alternative forced-choice design is

$$\begin{array}{c} \\ \langle s_1, s_1, \ldots, s_1 \rangle \\ \langle s_2, s_2, \ldots, s_2 \rangle \\ \ldots\ldots\ldots \\ \langle s_k, s_k, \ldots, s_k \rangle \end{array} \begin{array}{c} \begin{array}{cccc} 1 & 2 & \cdots & k \end{array} \\ \left[\begin{array}{cccc} 1 & \eta^{\sqrt{2m}} & \cdots & \eta^{\sqrt{2m}} \\ \eta^{\sqrt{2m}} & 1 & \cdots & \eta^{\sqrt{2m}} \\ \ldots & \ldots & \ldots & \ldots \\ \eta^{\sqrt{2m}} & \eta^{\sqrt{2m}} & \cdots & 1 \end{array}\right] \end{array}, \tag{28}$$

where again I have not written the biases explicitly.

4.3 Threshold Analysis

In the multiple look Yes-No experiment the threshold model says that the subject will observe some sequence of D's and $\bar{D}$'s on the basis of which he must say Yes or No. We assume that these observations are independent and that he bases his response upon the number of D's that occur; specifically, that when there are m looks he says Yes if and only if the number of D's is k_m or greater. It is easy to see that

$$
\begin{aligned}
p(Y \mid s) &= \sum_{i=k_m}^{m} q(s)^i[1 - q(s)]^{m-i} \binom{m}{i} \\
p(Y \mid n) &= \sum_{i=k_m}^{m} q(n)^i[1 - q(n)]^{m-i} \binom{m}{i}.
\end{aligned} \tag{29}
$$

The two most extreme cases are when $k_m = 1$, that is, when the subject says Yes if at least one D observation occurs, in which case

$$p(Y \mid s) = 1 - [1 - q(s)]^m \quad \text{and} \quad p(Y \mid n) = 1 - [1 - q(n)]^m,$$

and when $k_m = m$, that is, when the subject says Yes only if the D observation occurs for all presentations, in which case

$$p(Y \mid s) = q(s)^m \quad \text{and} \quad p(Y \mid n) = q(n)^m.$$

A more plausible assumption is something like majority rule (we suppose that he says Yes 50 per cent of the time when there is an equal number of D's and $\bar{D}$'s). No simple equation can be written for this case, but it is easy to calculate specific values from Eq. 29. Note that the response probabilities for successive odd-even m's are identical.

For the k-alternative forced-choice design, we assume that the responses are unbiased as in the other two models. If the subject obtains D observations in m of the k intervals, we assume that he chooses one of these intervals at random. Luce (1963) has shown that this implies

$$p_k(C) = \frac{1}{kq(n)} \{q(s) - [1 - q(n)]^{k-1}[q(s) - q(n)]\}. \tag{30}$$

Note that

$$p_2(C) = \tfrac{1}{2}[1 + q(s) - q(n)],$$

and so two-alternative forced-choice data determine $q(s) - q(n)$, but not $q(s)$ and $q(n)$. However, for $k > 2$, the value of $p_k(C)$ depends upon $q(s)$ and $q(n)$ separately, not just upon their difference. This is to be contrasted with the other two models in which the $p_2(C)$ data uniquely determine $p_k(C)$.

No threshold analysis for the multiple look k-alternative forced-choice design has yet been suggested.

4.4 Comparison of Models with Data

There do not appear to be any published raw data for the multiple-look Yes-No design, and so all we can do is attempt to compare the several theories. In all three cases there is a free parameter which gives one a good deal of freedom: neither the bias parameters in the detectability and choice models nor the value of k_m in the threshold model need be the same as the number of looks is changed. If we assume no bias in the first two models and suppose that, for $m = 1$, $p(Y \mid s) = \frac{2}{3}$ and $p(Y \mid n) = \frac{1}{3}$, then we get the solid points on the diagonal of Fig. 12. The two models differ so little that the separate points cannot be shown on a graph of this size. The majority rule for the threshold model also yields points on the diagonal, but they do not approach the corner quite so rapidly. The threshold $m = 7$ and 8 point is nearly the same as the detectability and choice $m = 5$. The other two sets of threshold points are for the extremes $k_m = 1$ and $k_m = m$, and it is clear that by other choices for k_m almost

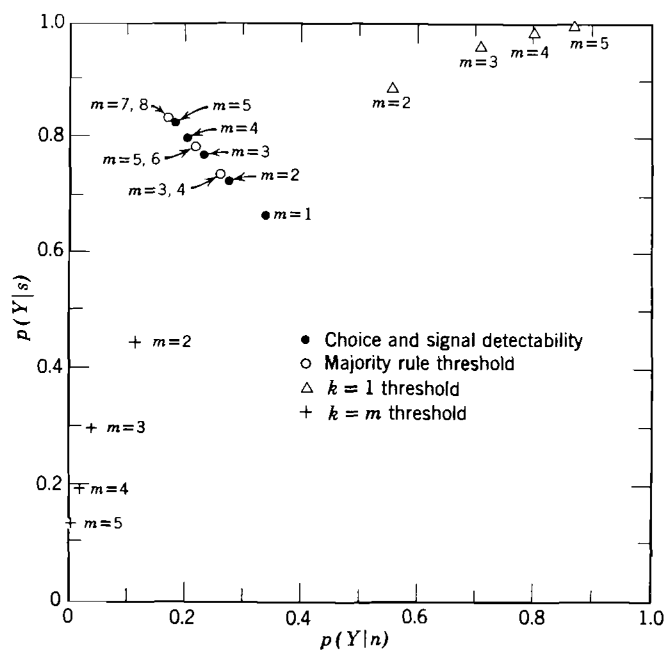

Fig. 12. Typical predictions of the several models for the multiple look Yes-No design. The parameter m denotes the number of independent repetitions of the stimulus plus noise or of the noise alone.

any other curve to the left and above these extremes can be generated. Much the same is true for the other two models because of the freedom in choosing the bias parameter.

In addition to the Yes-No and two-alternative forced-choice data given in Table 1, Swets (1959) collected four-alternative forced-choice data on the same subjects. Using Eq. 23, or Elliott's (1959) tables, $p_4(C)$ for the detectability model can be predicted from the observed values of $p_2(C)$ and

Table 5 Observed and Predicted Values of $p_4(C)$ for Swets's (1959) data

Subject	E/N_0 in db	Threshold Parameters: Upper Limb q(n)	Upper Limb q(s)	Lower Limb q(n)	Lower Limb q(s)	Observed	Predicted: Signal Detectability	Choice	Threshold: Upper Limb	Threshold: Lower Limb
1	9.4	0.13	0.77	0.25	0.89	0.623	0.66	0.60	0.67	0.62
	14.5	0.00	0.87	0.13	1.00	0.823	0.87	0.83	0.90	0.82
	16.6	0.00	0.89	0.11	1.00	0.883	0.89	0.86	0.92	0.85
2	9.4	0.19	0.72	0.33	0.86	0.524	0.57	0.52	0.58	0.53
	11.7	0.17	0.74	0.28	0.86	0.626	0.61	0.55	0.62	0.57
	14.5	0.07	0.78	0.29	1.00	0.750	0.71	0.66	0.75	0.64
	16.6	0.02	0.83	0.19	1.00	0.792	0.80	0.75	0.84	0.75
3	9.4	0.00	0.69	0.25	0.94	0.677	0.70	0.64	0.76	0.65
	11.7	0.07	0.80	0.27	1.00	0.734	0.73	0.68	0.77	0.66
	14.5	0.00	0.83	0.17	1.00	0.847	0.83	0.79	0.88	0.80
	16.6	0.00	0.92	0.08	1.00	0.895	0.91	0.88	0.94	0.88

See Table 1 for a description of experimental conditions.

for the choice model from Eq. 27. The predictions for the threshold model, Eq. 30, depend upon knowing both $q(n)$ and $q(s)$. These may be estimated from the Yes-No and two-alternative forced-choice data, with, however, the upper limb-lower limb ambiguity inherent in the Yes-No model. The details about how this was done can be found in Luce (1963). Both sets of estimates and the predictions for all three models are shown in Table 5. It is clear that there is little to choose between the detectability and choice models. The threshold model is adequate only if we admit the possibility that the subjects did not all operate on the same limb and that some may have shifted from one limb to the other as the stimulus energy was increased. Both seem like reasonable possibilities.

Swets (1959) also reported $p_k(C)$ estimates for three subjects and $k = 2$, 3, 4, 5, and 8. Assuming equal biases, so that Eqs. 23 and 27 can be used, the detectability and choice models can be compared. The results are shown in Fig. 13, and they clearly favor the detectability model. To what extent this conclusion depends upon the assumption of equal biases is not

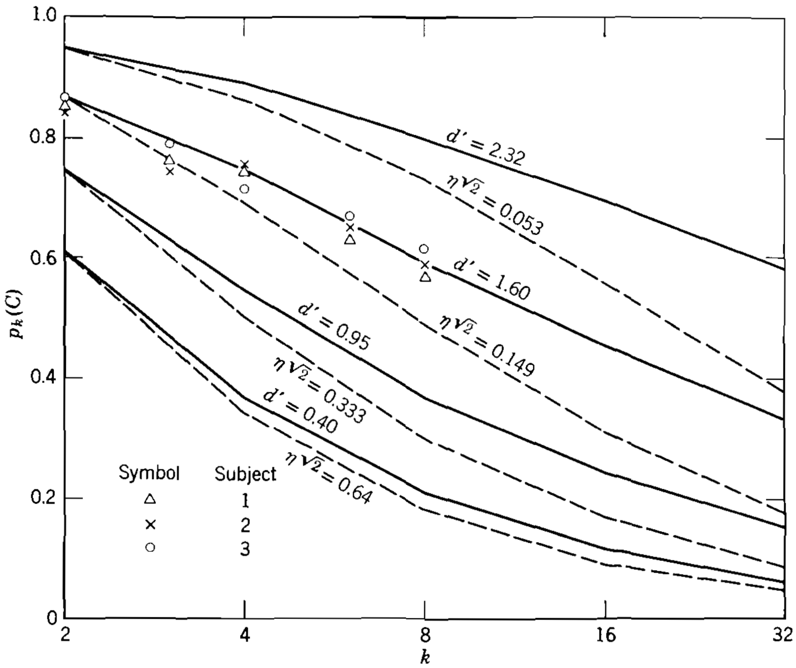

Fig. 13. Typical predictions of the signal detectability (solid curves) and choice (dotted curves) models for the unbiased k-alternative forced-choice design. The data points for these subjects are from Swets (1959).

clear. If, for example, the biases were U-shaped, so that the relative frequency of the first and last responses was in excess of $1/k$ and of the middle ones less than $1/k$, then it is quite possible that $p_3(C)$ would be artificially inflated and that $p_k(C)$, $k \geqslant 5$, would be artificially deflated. If that were the case, the data surely would not support the choice model and, depending upon the magnitude of the effect, might very well not support the signal detectability model. On the other hand, if the biases formed an inverted U, then $p_3(C)$ could easily be deflated and $p_k(C)$, $k \geqslant 5$, inflated. If the effect were large enough, this could cause us to accept the choice and reject the detection model. Swets does not indicate the nature of the biases in his data, but my best guess (based upon biases in recognition confusion matrices) is that they were of the first type, in which case the choice model is inadequate to account for these data. Until more detailed data are available, however, no very certain decision is possible.

Without Yes-No data on the same subjects under the same conditions, it is impossible to predict $p_k(C)$ uniquely using the threshold model. The best we can do is to calculate the extreme limits of Eq. 30 for several different values of $p_2(C)$. These are shown in Fig. 14 along with the data points again.

Swets, Shipley, McKey, and Green (1959) reported an acoustic study of the multiple-look four-alternative forced-choice design, using $m = 1$, 2, 3, 4, and 5. Assuming again that the biases are equal, η can be estimated from the probability of a correct response when $m = 1$, and then the other values are predicted from

$$p(r \mid \langle s_r, s_r, \ldots, s_r \rangle) = \frac{1}{1 + 3\eta^{\sqrt{2m}}},$$

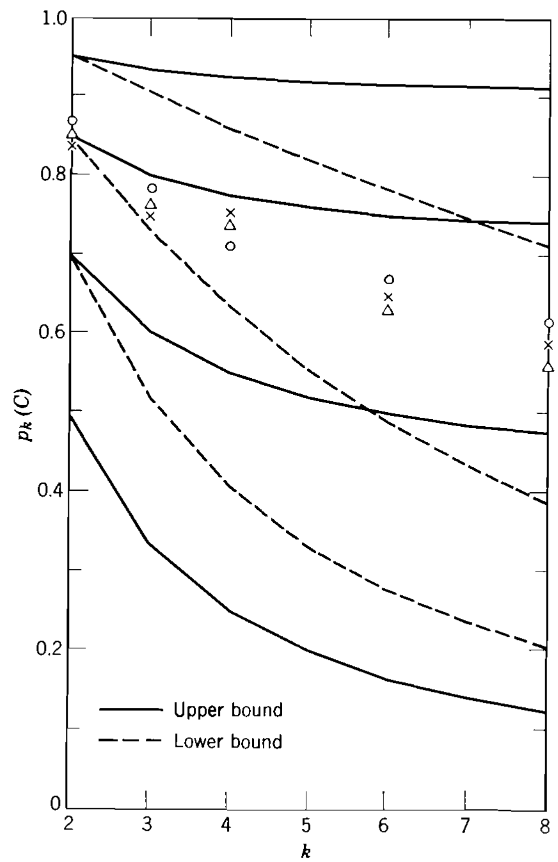

Fig. 14. Typical upper- and lower-bound predictions of the threshold model for an unbiased k-alternative forced-choice design. The data points for these subjects are the same as those shown in Fig. 13.

which follows from Eq. 28. The signal detectability analysis is similar, except that Eq. 24 is used. The results are shown in Table 6. The detectability model predicts a somewhat more rapid improvement in correct responses than is shown by the subjects; the choice model is closer to the behavior. Because each probability is estimated from 600 observations and because the theoretical probabilities are so near 1, some of the deviations of the detectability model are significant if not spectacular.

Table 6 Multiple-Look Four-Alternative Forced-Choice Data Reported by Swets, Shipley, McKey, and Green (1959)

Subject	Number of Observations	Observed Data	Predicted from $m = 1$ data: Signal Detectability	Choice Model
1	1	0.81	—	—
	2	0.92	0.937	0.925
	3	0.97	0.979	0.965
	4	0.99	>0.99	0.982
	5	0.99	>0.99	0.990
2	1	0.80	—	—
	2	0.89	0.930	0.918
	3	0.95	0.974	0.961
	4	0.96	>0.99	0.976
	5	0.98	>0.99	0.989
3	1	0.82	—	—
	2	0.95	0.942	0.931
	3	0.97	0.981	0.968
	4	0.98	>0.99	0.984
	5	0.99	>0.99	0.991

The noise was 35 db re 0.0002 d/cm^2 and the signal a 1000-cps tone at 12.5 db measured in terms of $10 \log_{10} E/N_0$.

In connection with the independence assumption that plays such a significant role in all three models, the following result is of considerable importance. Swets et al. repeated the last experiment using a tape recording of the stimulus plus noise or of the noise alone for the several presentations. Instead of gradually improving, as they do with uncorrelated noise (Table 6) and as is predicted by the theories, the subjects exhibited little or no improvement beyond two presentations. This suggests that it is perfectly possible experimentally to render the independence assumption incorrect to a degree that is quite noticeable.

In evaluating these studies, it should be kept in mind that none of them was designed to test among the three models but rather to decide about the adequacy of the signal detectability one. So far as these data are concerned, there is nothing in my opinion that clearly favors one model over another. There is some suggestion in the k-alternative forced-choice data that the choice model is inferior to the detectability one, but the reverse is true for the multiple-look four-alternative data. Because of the threshold model's larger number of parameters, none of these experiments adequately taxes it.

5. THEORIES OF THE BIAS PARAMETERS

To a traditional psychophysicist, what we have been doing so far in this chapter must seem strange, if not totally irrelevant to his interests. He wants to know the laws relating responses to well-controlled, specifiable stimuli, and yet nothing at all has been said about them. The reason is that many contemporary psychophysicists do not believe that this problem is nearly so straight-forward as it seems. The current view is that it should be divided into three distinct parts. The first is a theory of responses that relates responses to responses, not to stimuli. Such theories—they are what we have discussed so far—contain estimable parameters, such as η, d', or $q(s)$ and $q(n)$ and b, c, or t and u, which are thought to depend upon and to summarize the relevant decision-making effects of the stimulating and reward conditions of the experiment. Because such parameters can be estimated from the response data, there is actually no need to measure the physical properties of the stimuli or the characteristics of the outcome structure of the experiment; they need only be under control and reproducible at will. This sort of theory, as we have seen, uses the data from one experiment to predict the results of others having different designs but involving the same stimuli, background, and residual environment.

Once such a theory is developed and has received enough confirmation so that one feels that it may be approximately correct, one can begin to look into the other two problems: first, relations between the stimulus parameters of the theory and measurable properties of the stimuli, and second, relations between bias parameters and other aspects of the experimental conditions. There is precious little point, however, in trying to establish such relations until the response-response theory has been rather carefully tested.

If we are correct in supposing that parameters of the one class measure the subject's sensitivity to the stimuli and those of the other measure response biases that are under his control, then we must anticipate separate theories relating each to certain aspects of the experimental situation. This section presents two quite different theories for the bias parameters. The next discusses theories of the sensitivity parameters.

5.1 Expected Value

As we have seen in Sec. 3 on isosensitivity curves, experimental manipulations of either the presentation probability or of the payoff matrix appreciably affect the response probabilities, even when the stimulating conditions are fixed. This suggests that a theory of the bias parameters must involve at least these two experimental factors. Broadly speaking, mathematical psychologists have come up with two ideas about this dependence. The one that we look into in this subsection stems mainly from the economic and statistical literature. It says that subjects choose the parameters to optimize something. The other, which is discussed in the next subsection, says that subjects continually adjust the parameters in an "adaptive" fashion—they learn.

Suppose that the presentation and payoff structure in the Yes-No design is

Presentation Probability	Stimulus Presentation	Response Y	Response N
P	s	o_{11}	o_{12}
$1 - P$	n	o_{21}	o_{22}

where the o_{ij} are sums of money. One reasonably sensible criterion that a subject might use is to select that bias parameter that maximizes his total expected money return during the course of the experiment. Because the trials are assumed to be independent and because the response probabilities are assumed to be constant, this is the same as selecting it to maximize the expected value of a single trial. This assumption is criticized later.

The expected outcome, $E(o)$, is simply the money value of each of the four possible presentation-response conditions weighted by their respective probabilities of occurring:

$$\begin{aligned} E(o) &= Pp(Y \mid s)o_{11} + Pp(N \mid s)o_{12} + (1 - P)\, p(Y \mid n)o_{21} + (1 - P)\, p(N \mid n)o_{22} \\ &= [p(Y \mid s) - \beta p(Y \mid n)]\, P(o_{11} - o_{12}) + Po_{12} + (1 - P)o_{22}, \qquad (31) \end{aligned}$$

where

$$\beta = \left(\frac{1 - P}{P}\right)\left(\frac{o_{22} - o_{21}}{o_{11} - o_{12}}\right). \qquad (32)$$

If the response probabilities depend upon a single bias parameter z, then to find that value of z that maximizes $E(o)$ we set the derivative of $E(o)$ with respect to z equal to 0 and solve for z:

$$\frac{dE(o)}{dz} = 0 = \frac{dp(Y \mid s)}{dz} - \beta \frac{dp(Y \mid n)}{dz},$$

so

$$\beta = \frac{dp(Y \mid s)/dz}{dp(Y \mid n)/dz}. \tag{33}$$

For the signal detectability model, we simply calculate the derivative of Eqs. 8 with respect to c and find that

$$\beta = \frac{p(c \mid s)}{p(c \mid n)}. \tag{34}$$

Thus, given the payoffs and presentation probabilities, we can calculate β and from this determine c via Eq. 34, provided that we know the forms of $p(\cdot \mid s)$ and $p(\cdot \mid n)$. An exactly parallel development holds for the two-alternative forced-choice design, except that β equals the ratio of the difference density for $\langle s, n\rangle$ to the difference density for $\langle n, s\rangle$.

Equation 33 says that the slope of the isosensitivity curve should equal the optimum β defined in Eq. 32, so that one comparison we can make is between these two quantities, using, say, the theoretical signal detectability curve to estimate the slope. Green (1960) has done this, and the results are shown in Fig. 15. It is clear that the data depart considerably from the

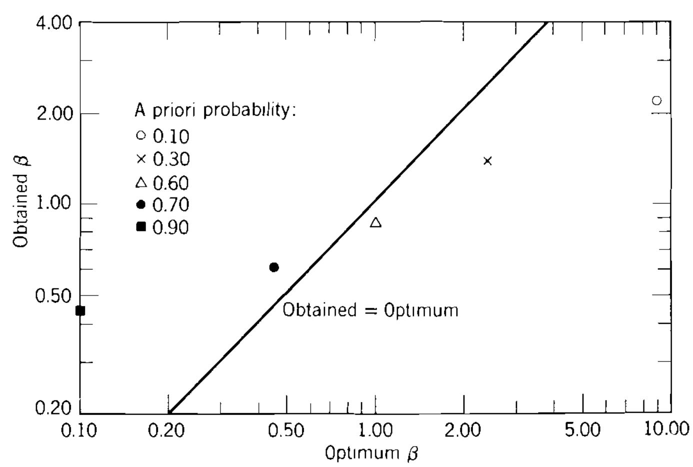

Fig. 15. Obtained versus optimum values of β assuming a maximization of expected value and the detectability Yes-No model. Adapted with permission from Green (1960, p. 1195).

theory. Green argues that the insensitivity of the expected value to changes in the response probabilities probably accounts for the poor predictions. For example, with $P = 0.5$, a symmetric payoff matrix, and $d' = 1$, the expected payoff is within 90 per cent of the maximum as long as $p(Y \mid n)$ is between 0.15 and 0.50. Nevertheless, the departure from the optimum β curve is systematic and needs to be explained.

Note that the quantity $p(c \mid s)/p(c \mid n)$ is the likelihood ratio of the stimulus density to the noise density; it is usually denoted by $l(c)$. It is a measure of the relative likelihood that a given observation is due to s or to n, and the decision rule that maximizes expected value is to say Yes if the observation x is such that $l(x) > \beta$ and to say No if $l(x) < \beta$. Thus, if we had simply postulated a decision axis, we would have been forced by this bias theory to the position that it is the likelihood ratio axis. So, this biasing model is a possible defense for the original assumption that likelihood ratios are involved.

A variety of other decision rules has been explored (Birdsall, 1955; Peterson, Birdsall, & Fox, 1954), all of which lead to the same general structure: the likelihood ratio is compared with some function of P and the o_{ij}. I shall not go into these here.

The analysis for the choice theory is little different. Through Eq. 33 no response theory is assumed. At that point we calculate the derivatives of the Yes-No choice model probabilities (Eq. 12)

$$p(Y \mid s) = \frac{1}{1 + \eta b} \quad \text{and} \quad p(Y \mid n) = \frac{\eta}{\eta + b}$$

with respect to b, substitute in Eq. 33, and solve for b:

$$b = \frac{\sqrt{\beta} - \eta}{1 - \eta\sqrt{\beta}}. \tag{35}$$

The forced choice solution is the same, except that η is replaced by $\eta^{\sqrt{2}}$.

Similar, but more complex, calculations can be made for designs having three or more responses. Partial differentiation with respect to each of the bias parameters results in a set of simultaneous equations for these parameters, which may be very difficult to solve explicitly, but numerical solutions can always be obtained when needed.

In both models the optimum bias parameter depends continuously upon β, which in turn is a continuous function of P and the o_{ij}, except when $P = 0$ or when $o_{11} = o_{12}$. The same maximization analysis leads to quite different results when it is applied to the threshold model. Because the response probabilities depend linearly upon the bias parameter—t on the lower limb and u on the upper—the maximum of $E(o)$ occurs either at $t = 0$ or 1 or at $u = 0$ or 1, depending upon the particular values of P and the o's. In terms of the isosensitivity curves, this means that the response data must fall at either $\langle 0, 0\rangle$, $\langle q(n), q(s)\rangle$, or $\langle 1, 1\rangle$, which is clearly not what happens in Figs. 8 and 9. Thus it is certain that the threshold model together with the maximization of expected money criterion is wrong, but which of the two is at fault is not certain.

No very serious testing of the expected value model has been carried out within the detection context, but we know from preference and utility studies (*see* Edwards, 1954, 1961) that to have any hope of predicting behavior we must convert it into a subjective expected utility model in which subjective presentation probabilities replace P and utilities of outcomes replace the o_{ij}. Whether or not this change results in an adequate bias theory for the detectability and choice models, it leaves unaffected the unacceptable results for the threshold model. So we must consider whether there is an acceptable alternative for the threshold theory.

5.2 Asymptotic Learning

It is well known that when information feedback is used a period of pretraining must be included before the responses settle down to their "asymptotic" values. Presumably, the subject is gaining some information relevant to his responses and he is using it to alter his behavior. One possibility is that he is discovering empirically, as it were, what the presentation probability P and his own response probabilities $p(Y \mid s)$ and $p(Y \mid n)$ are so that he can calculate an optimum bias parameter from Eq. 33 or something analogous to it. Another possibility, which some feel is a bit more likely, is that he is engaged in a learning process during which he alters his biases one way or the other, depending upon the trial-by-trial outcomes. This suggests that we set up a stochastic learning process of the sort discussed in Chapter 9 of Vol. II, in which different operators are applied to the biases, depending upon the outcomes. Its asymptotic properties describe the subject when we, as psychophysicists, observe his behavior.

In deciding on a given trial how to modify the bias—from the theorists' viewpoint, in deciding what learning operator to use—three classes of events might be taken into account: the stimulus presented, which is revealed to the subject by the information feedback, his internal observation (if any) resulting from the presentation, and the response made. It seems clear that the operator applied should depend upon the presentation. One might also suspect that it should depend upon the response—in the choice model this is the only other possibility. If so, then the probability of applying an operator depends upon the product of the presentation probability, which is constant during the experimental run, and the response probability, which is not. Because the probability that a particular operator will be applied is changing over trials, the resulting stochastic process is exceedingly complicated. At present insufficient is known about its asymptotic properties for it to be of any use to us. This is, of course, a limitation in practice, not in principle.

So we confine our attention to models in which the subject decides how to change his bias on the basis of the presentation and the internal observation resulting from it. These models are called experimenter-controlled in learning theory. In many ways the internal observation seems a much more relevant event than the subject's response, for it is these observations that he must use in the future to decide what responses to make. By assumption, the conditional probability of an internal effect occurring is constant over trials, so the probability of applying a given learning operator is also constant, which eliminates the major difficulty mentioned above.

With the signal detectability model, however, a problem still remains, namely that there is a continuum of effects. Although Suppes (1959, 1960) has begun work on such learning models, insufficient is currently known about them to arrive at a theory of biasing. The choice model does not suffer from this difficulty because there are no internal observations, nor does the threshold model because there are only two observation states, D and $\bar{D}$.

The choice theory learning model has already been presented in Sec. 1.2 as an argument for assuming the choice theory. As far as the biases are concerned, we found that

$$b(r) = P[\iota(r)]\,\theta(r),$$

where ι is the identification function, P the presentation probability, and θ a learning rate parameter. Assuming that this learning model is correct, the major unsolved bias problem is how learning rates depend on the payoffs and whatever else they depend on.

A somewhat similar analysis can be given for the threshold model. Suppose, first, that the subject is operating upon the lower limb of the isosensitivity curve; that is, he is saying Yes only to a proportion t of the D observations and No, otherwise. He is adjusting t on the basis of his experiences. It is surely inappropriate for him to change it on those trials when a $\bar{D}$ observation occurs. (Such an observation may, of course, influence his decision to shift from the lower to the upper limb.) So suppose a D observation occurs. If it resulted from an s presentation, he should increase his tendency t to say Yes to D observations; whereas, if it resulted from an n observation, he should decrease t. With this in mind and assuming linear operators, we postulate that

$$t_{i+1} = \begin{cases} t_i + \theta(1 - t_i), & \text{if } s \text{ and } D \text{ occur on trial } i \\ t_i - \theta' t_i, & \text{if } n \text{ and } D \text{ occur on trial } i \\ t_i, & \text{if } \bar{D} \text{ occurs on trial } i, \end{cases} \tag{36}$$

where t_i is the bias on trial i. It follows that the expected value of t_{i+1} given t_i is

$$\begin{aligned} E(t_{i+1} \mid t_i) &= [t_i + \theta(1 - t_i)]\, Pq(s) + (1 - \theta')t_i(1 - P)\, q(n) \\ &\quad + t_i P[1 - q(s)] + t_i(1 - P)[1 - q(n)] \\ &= t_i\{(1 - \theta)\, Pq(s) + (1 - \theta')(1 - P)q(n) + P[1 - q(s)] \\ &\quad + (1 - P)[1 - q(n)]\} + \theta\, Pq(s). \end{aligned}$$

Because all of the probabilities on the right are trial-independent, we can take expectations over t_i:

$$E(t_{i+1}) = E(t_i)\{-Pq(s)\theta - (1 - P)\, q(n)\theta' + 1\} + \theta\, Pq(s). \tag{37}$$

If we assume that the asymptotic expectation of t_i, call it t_∞, exists, then by taking the limit of Eq. 37 as i goes to infinity we may solve for t_∞:

$$t_\infty = \frac{q(s)}{q(s) + q(n)b}, \tag{38}$$

where

$$b = \left(\frac{1 - P}{P}\right)\left(\frac{\theta'}{\theta}\right). \tag{39}$$

Note that, as in the choice model, the bias parameters are the product of the presentation probability and the corresponding learning rate parameters.

The parallel model for the upper limb assumes

$$u_{i+1} = \begin{cases} u_i + \theta(1 - u_i), & \text{if } s \text{ and } \bar{D} \text{ occur on trial } i \\ u_i - \theta' u_i, & \text{if } n \text{ and } \bar{D} \text{ occur on trial } i \\ u_i, & \text{if } D \text{ occurs on trial } i, \end{cases} \tag{40}$$

and it results in the asymptotic expectation of u_n

$$u_\infty = \frac{1 - q(s)}{1 - q(s) + [1 - q(n)]b}. \tag{41}$$

The quantity b is formally similar to β (Eq. 32) in that the presentation probability enters in the same way. Presumably, the learning-rate parameters depend in some fashion upon the payoffs, but no one has yet reported a theory for this dependence. Much research is needed to determine whether this sort of model is adequate and to understand the relation between learning rates and payoffs.

An interesting feature of these asymptotic results for the threshold model is that the response probabilities can approach the true detection probabilities only under very special conditions. If the subject is operating on the lower limb and b has a moderate value somewhere in the neighborhood of 1, then t_∞ approaches 1 only as $q(n)$ approaches 0. On the upper limb u_∞ approaches 0 only as $q(s)$ approaches 1. Thus, if, as in the data of Figs. 8 and 9, $q(n) > 0$ and $q(s) < 1$, the theory predicts that no data points lie in the immediate neighborhood of $\langle q(n), q(s)\rangle$, and none seems to. In other words, one effect of information feedback, according to this model, is to prevent the subject from revealing directly the true detection probabilities. It is not known what he does when there is no information feedback, but it certainly should not be assumed that $p(Y \mid s) = q(s)$ and $p(Y \mid n) = q(n)$ without careful investigation.

A second point of interest is that at asymptote the response probabilities are still fluctuating under the processes described by Eqs. 36 and 40. An expression can be derived for the variance of the response probability at asymptote which shows that the more rapid the learning, the larger the variance. Most experimenters feel that there is more than binomial variability in much psychophysical data, and learning may very well be one source. If so, considerable care must be exercised in applying the standard tests of significance that postulate constant underlying probabilities.

A similar learning model can be developed for the biases v and w of the two-alternative forced-choice design (see Luce, 1963). Suffice it to say that $v_\infty = w_\infty = 1/(1 + b)$. Note that, when $b = 1$, $v_\infty = w_\infty = \frac{1}{2}$, which implies $p(1 \mid \langle s, n\rangle) = p(2 \mid \langle n, s\rangle)$. For $P = \frac{1}{2}$, $b = 1$ if and only if $\theta = \theta'$. Thus the apparent tendency toward behavioral symmetry when $P = \frac{1}{2}$ and the payoff matrix is symmetric suggests that the learning rates corresponding to symmetric payoffs are approximately equal. In that case $b = (1 - P)/P$.

Assuming this, we may use Eqs. 38 and 41 to predict the data shown earlier in Fig. 8. The predicted values, which correspond to the points by pairs as one sweeps around the isosensitivity curve, are shown as crosses in Fig. 8 (p. 134).

6. THEORIES OF THE STIMULUS PARAMETERS

Relatively little is yet known about the way in which the stimulus parameters of the several theories depend upon physical measures of the stimuli. The research is scattered and incomplete. Tanner and his colleagues have worked intensively on this problem for detectability theory during the last three or four years, and considerable data have been collected, but in my view no adequate theory has yet evolved. Some indication of their direction is given in Sec. 6.1. No work at all has yet been done in connection with the choice theory. The quantal studies of Békésy (1930) and Stevens, Morgan, and Volkmann (1941) can, and I think should, be viewed as a stimulus-parameter theory for the threshold model when applied to situations in which the background and stimulus differ on only one physical dimension, such as energy or frequency. Although I shall avoid the details, it is not difficult to modify their model to give a threshold theory for the detection of a stimulus in noise. The neural quantum theory is examined in Secs. 6.2 and 6.3.

6.1 Ideal Observers

Peterson, Birdsall, and Fox (1954), in their presentation of detectability theory as a physical—not a psychophysical—theory, treated the question of the optimum detectability possible for physical signals in noise when one has different amounts of information about the signals and noise. They were not concerned with what people do but with what an optimum detection device can possibly do. The best known of the several sets of assumptions they looked into is the case of the so-called "signal-known-exactly." The signal and noise are both assumed to be limited to some band of frequencies, and the noise is assumed to have equal power at all of its frequencies and to have normally distributed amplitudes (so-called white Gaussian noise). The ideal detector knows everything there is to know about the signal: frequencies, phase relationships, amplitudes, time of onset, etc. Under these assumptions, they showed that the logarithm of the likelihood ratio is normally distributed with variance $2E/N_0$, where E denotes the stimulus energy and N_0 the noise power per unit bandwidth. When noise alone is presented, the mean is $-E/N_0$, and when stimulus plus noise is presented it is E/N_0. Thus we see that $d' = (2E/N_0)^{1/2}$. It was primarily this result that suggested the normality and equal variance assumptions typical of detectability theory. I shall not attempt to reproduce the argument leading to it. Peterson et al. explored a variety of other cases in which different assumptions are made about the information available to the detector.

Given that results of this sort can be found, a possible approach to the question of a stimulus-parameter theory is suggested. We suppose that the person is in fact an optimum detection device operating on certain of the information that he has available—*an ideal observer.*

Thus, if the human observer were to perform as an ideal observer the following would be necessary: (1) he would have no source of internal noise. That is, the input signal would have to be transformed to a different type of energy by the end organ and transmitted by the nervous system, all with perfect fidelity. (2) He would have perfect memory for the signal parameters and the noise parameters. At any time t within the observation interval he must know the exact amplitude of the signal waveform. (3) He would be capable of calculating likelihood ratio or some monotonic transformation of likelihood ratio.

These are some of the requirements which must be met by the human observer if he is to perform as well as the ideal observer. Clearly, the human observer does not meet these specifications. However, it is possible to determine experimentally the manner and degree to which the human observer fails to meet these requirements and thus obtain a better understanding of the human observer. (Tanner, Birdsall, & Clarke, 1960, pp. 19–20).

As well as I can make out, Tanner proposes to search for assumptions about the information that is available to the ideal observer until he finds a set for which the optimum behavior predicted is that of the human being. Although much of this work is not yet published, or published only in summary form (Tanner, 1960, 1961), it seems that two lines are being developed: modifications of the experiments to fit the model and modifications of the model to fit the experiments. For example, the signal-known-exactly model postulates that the subject knows, among other things, the frequency and time of onset of the stimulus. The poorer performance of the subject may reflect a failure of these assumptions, and so by various experimental devices information about frequency and onset are presented to the subject to see whether the availability of this information improves his performance. The comparison measure used is the square of the ratio of the observed d' to that of the ideal observer detecting a signal that is completely known. This ratio is called the *efficiency* of the observer. Unfortunately, considerable good data are being reported only in terms of efficiencies and d''s, which, it is conceivable, may one day be of little more than historical interest.

The other approach is to change the information assumptions about the ideal observer to see whether an ideal observer with a less perfect memory more nearly approaches human behavior. Examples of this approach can be found in Green (1960), where the whole notion of the ideal observer is carefully described, and in Tanner (1961), who summarizes some of the memory aspects now believed to be relevant.

As the signal detectability theorists recognize, this program may eventually run afoul of difficulties that cannot easily be overcome within the framework of the ideal observer. In addition to introducing restrictions upon the physical information that is usable, the subject may very well add his own noise and other distortions to the presented information, in which case it may be quite impossible to find an ideal observer that performs as he does—assuming that ideal observers continue to be defined to have properties such as those listed in the foregoing quotation. A somewhat cruder approach that nonetheless may merit attention involves parametric studies in which one or at most two physical variables are manipulated and the model parameters are calculated from the data to see what, if any, simple relations appear to exist.

6.2 The Neural Quantum Model

Suppose that we have a simple background, such as a tone, and that a stimulus involves a short duration change of the background on one dimension, such as energy. The Békésy-Stevens neural quantum model attempts to relate the change in the number of neural quanta excited by the stimulus to a physical measure of the increment (or decrement) introduced in the background. The model supposes that this physical dimension can be partitioned at any instant into nonoverlapping intervals that correspond to the neural quanta. Thus two different levels of stimulation lying within one interval excite the same number of quanta, whereas two in different intervals excite different numbers of quanta. We may think of the subject as imposing a quantal grid over the physical dimension.

Over time, the quantal grid is assumed to fluctuate slowly as the result of changes internal to the subject, and so the number of quanta excited by a constant stimulus also fluctuates, sometimes increasing, at other times decreasing. Although it is generally felt that it is the grid that shifts, it is more convenient mathematically to view the grid as fixed and to suppose that the physical measure corresponding to the stimulus does the fluctuating. The two ways of viewing the matter are completely equivalent as long as the grid is equally spaced, as we shall assume. In terms of a fixed grid, a given background will have some distribution, such as that shown in Fig. 16.

Suppose that just prior to presenting stimulus s, which, it will be recalled, is simply an increment (or decrement) in the background, the effect of the background is $\mathbf{X}$. This effect is a random variable distributed in some manner, as shown in Fig. 16. The addition of s is assumed to change the effect from $\mathbf{X}$ to $\mathbf{X} + \Delta(s)$, where $\Delta(s)$ depends only upon s. Thus, whenever s is presented, the same increment is always added. This is an important point. We are saying that the effect of the background just before stimulation and the effect of the background plus the stimulus are perfectly correlated. This assumption is quite different from the independence assumptions we have repeatedly made when discussing the detection of stimuli in noise. The correlation assumption is interlocked

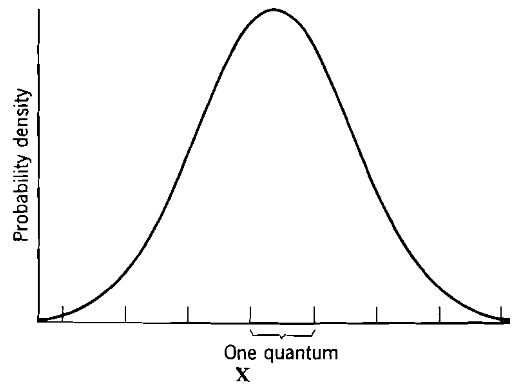

Fig. 16. A typical density of assumed stimulus effects in the quantal model.

with our earlier supposition that the grid fluctuates slowly—in order that a perfect correlation may exist, the stimulus presentation must be of sufficiently short duration so that little or no change in the grid location will take place during the presentation. In practice, a duration of the order of 100 ms has been deemed sufficiently short.

It will be recalled (Sec. 1.3) that we set up the decision rule that a change in stimulation is noted when it equals or exceeds some number k of quanta. Thus, if the physical increment corresponding to s, $\Delta(s)$, is less than the physical increment corresponding to $k - 1$ quanta, it fails to produce a detection observation. If, however, it equals or exceeds that corresponding to k quanta, then it will always be detected. And when it is between that corresponding to $k - 1$ and k quantal intervals, a detection observation may or may not occur. To be specific, suppose that $\Delta(s)$ corresponds to $\frac{1}{3}$ of a quantum interval more than $k - 1$ quanta; then, if the random variable $\mathbf{X}$ overflows an integral number of intervals by less than $\frac{2}{3}$, the stimulus cannot excite the necessary k quanta. However, if it overflows $\frac{2}{3}$ or more, then s excites the required k quanta. So the probability that a presentation of s produces a detection observation depends upon the probability that the background *residue*, as it is called, is greater than $\frac{2}{3}$.

It follows, then, that the probability of a detection observation occurring depends upon the distribution of residues. To talk about this distribution without specifying just how many quanta are excited by the background, as we have been doing, makes sense only if the physical measure we are using has the property that all quantal intervals are of the same size. It is not obvious that the usual physical measures have this property, but under very general conditions it is possible to find a continuous monotonic transformation that has. We assume that this is the measure we are using.

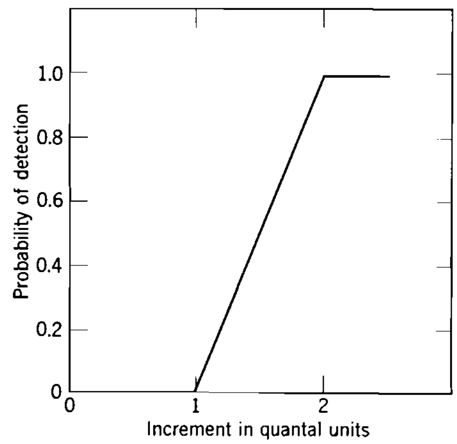

Fig. 17. Predicted true probability of detection versus a measure of the stimulus increment in quantal units (i.e., the true psychometric function) assuming the quantal model, a uniform distribution of residues, and a two-quantum criterion.

Now, if the distribution of residues is uniform in that measure, as has generally been assumed in discussions of neural quantum theory, then the probability of detection is easily seen to be rectilinear, as in Fig. 17.

Stevens, Morgan, and Volkmann (1941) attempted to argue verbally, and Corso (1956) alleged that it follows from Bayes's theorem, that the distribution of residues is uniform independent of the distribution of **X**. This is false. For example, suppose that **X** is distributed according to

$$\Pr(\mathbf{X} = x) = \frac{\lambda}{2} e^{-\lambda|x|}$$

and that the quantal boundaries are located at the points iq, where $i = 0, \pm 1, \pm 2, \ldots$, and $q > 0$ is the size of one quantum. If **R** denotes the residue random variable, then its distribution for $0 \leqslant r \leqslant q$ is given by

$$\begin{aligned}
\Pr(\mathbf{R} = r) &= \sum_{i=-\infty}^{\infty} \Pr(\mathbf{X} = iq + r) \\
&= \sum_{i=1}^{\infty} [\Pr(\mathbf{X} = iq + r) + \Pr(\mathbf{X} = -iq + r)] + \Pr(\mathbf{X} = r) \\
&= \frac{\lambda}{2}(e^{-\lambda r} + e^{\lambda r}) \sum_{i=1}^{\infty} (e^{-\lambda q})^i + \frac{\lambda}{2} e^{-\lambda r} \\
&= \frac{\lambda}{2}\left(\frac{e^{-\lambda r} + e^{\lambda r} e^{-\lambda q}}{1 - e^{-\lambda q}}\right),
\end{aligned}$$

and elsewhere it is zero. It is simple to show that the nonzero portion of this function has a minimum at $r = q/2$; hence, to get an idea of the departure from uniformity, we look at the ratio

$$\frac{\Pr(\mathbf{R} = 0)}{\Pr(\mathbf{R} = q/2)} = \frac{1 + e^{-\lambda q}}{2(e^{-\lambda q})^{1/2}}.$$

Thus, if $e^{-\lambda q} = \frac{1}{4}$, the ratio is 5/4. It is clear that as the distribution of $\mathbf{X}$ becomes flat relative to the quantum size, that is, as λq approaches 0, the more nearly the distribution of the residues approaches the uniform.

In general, however, this distribution is not uniform, and, to the extent that it deviates from uniformity, the transition from 0 to 1 in Fig. 17 must deviate from linearity. This point is important because much of the controversy in the literature over the quantal hypothesis has centered on the prediction of linearity and whether or not appropriate statistical tests have been performed to decide between it and an ogive. In my view, the theory as presently stated does not really make this prediction. I have already given one reason, another will be given now, and a third is presented in the next section.

Suppose that x denotes the physical value of a stimulus in a measure for which the quantal increments are equal and suppose that the distribution of residues is uniform in that measure. Thus a plot of the detection probability versus stimulus increments in this measure is rectilinear. Because we do not know what measure this is, we use instead some "natural" physical measure for which that value corresponding to x is y, the functional relation being $f(x) = y$. In general, we must expect f to be nonlinear, and so the straight line plot is somewhat warped when the natural physical measure is used; however, because the estimated size of the neural quantum is small, this effect is hardly noticeable for moderately nonlinear functions such as the logarithm.

But if not the straight line prediction, what then is there to test? The only other prediction as far as I can see is that the $p = 1$ and $p = 0$ intercepts stand in the ratio $k:(k - 1)$, where k is an integer. This result is independent of the distribution of the residues, but, of course, it is not independent of the independent variable that we use. Fortunately, if the transformation f is moderate in the sense that its derivative is nearly constant over a k quantum interval, then it does not matter much whether we use x or y. To show this, we use the well-known mean-value theorem, namely that if f is differentiable then there exists an x^* such that

$$y = f(x_b) + f'(x^*)(x - x_b),$$

where x_b is the background level and $x_b \leqslant x^* \leqslant x$. Thus, if x_0 denotes the

$p = 0$ intercept and x_1, the $p = 1$ intercept, we have

$$\frac{y_1 - y_b}{y_0 - y_b} = \frac{f(x_1) - f(x_b)}{f(x_0) - f(x_b)}$$
$$= \frac{f'(x_1^*)(x_1 - x_b)}{f'(x_0^*)(x_0 - x_b)}$$
$$= \left[\frac{f'(x_1^*)}{f'(x_0^*)}\right]\left(\frac{k}{k-1}\right).$$

Hence, if $f'(x_1^*)/f'(x_0^*)$ is approximately 1, which we expect because the quantal increments are thought to be small, the integral relation between the two intercepts is little changed.

6.3 The Neural Quantum Experiment

The neural quantum hypothesis and the experimental studies undertaken to test it have generated considerable controversy, much of which is described by Corso (1956). As has been indicated, a good deal of it has centered on the linearity hypothesis, but this is not really an essential feature of the theory. Much of the rest centers on the design of the so-called quantal experiment.

The proponents of the theory have emphasized how easy it is not to confirm the theory, and anyone who has tried is only too aware of the difficulties. Anything in the experimental design that makes the contribution of the stimulus, $\Delta(s)$, a random variable, so that $\mathbf{X}$ and $\mathbf{X} + \Delta(s)$ are not perfectly correlated, generates ogival detection functions that have no simple integral relation between intercepts. (It is easy to see this. The model is substantially the same as the detectability one with two independent random variables, the difference being that a response occurs only if the observations differ by some fixed amount corresponding to k quanta.) Apparently, any sort of distraction is likely to uncorrelate the presentations, and thus quantal results often are not found. This sort of vague notion of an acceptable experiment unfortunately makes the theory nearly immune to rejection. Any failure of the data to confirm it is likely to be taken as prima facie evidence that something was wrong with the experiment.

The feature of the "standard" quantal design that has received most criticism is the fact that the subject knows the presentation schedule. In order that the subject get properly "set," he is permitted to listen to repetitions of the same stimulus increment until he says he is ready, and then a run of identical increments is presented. The subject responds to each of these. Thus he knows in advance that Yes is the correct answer on every trial. The possibility for biasing seems great.

One school has argued as follows. Suppose that the true detection function is a smooth ogive. For a stimulus with a high detection probability, say 0.9, there is a tendency for the number of Yes responses to be inflated artificially because of two factors. One is that the subject knows that Yes is the correct response, and the other is that he is assumed to have a tendency to perseverate his responses, most of which have been Yes. This means, then, that the data function must be above the true function, and it intercepts the $p = 1$ line in much the same way as a linear function does. For a stimulus with a low true-detection probability, say 0.1, the argument is less clear because the perseveration tendency decreases the number of Yes's, whereas his knowledge of the presentation schedule tends to increase the number. So, according to this argument, we may expect the upper intercept to confirm the neural quantum model, but the lower one should vary from subject to subject and, on the whole, be more rounded. Although it is difficult to prove formally, inspection of the published as well as of considerable unpublished data suggests that just the opposite is true: the lower intercept seems more stable and more in line with the quantal model than the upper one.

Assuming that the rectilinear quantal model correctly describes the dependence of the true detection probability upon the stimulus magnitude, the learning model of Sec. 5.2 suggests that the observed responses should distort this function, especially at the upper intercept when there are no or only a few catch trials (Luce, 1963). Specifically, let us assume that a detection observation occurs when and only when a two-quanta change occurs, as suggested by the data. In addition, however, let us suppose that a conservative lower-limb bias is used by the subject when the detection observation is based upon a change of only two neural quanta, whereas with three or more he uses an upper-limb bias. Thus, for any stimulus of magnitude less than two quantal units, a lower-limb bias is in force, and so, by Eqs. 15 and 38,

$$P(Y \mid s) = t_\infty q(s)$$
$$= \frac{q(s)^2}{q(s) + q(n)b}.$$

For stimuli two quantal units or larger, $q(s) = 1$. For such stimuli, the foregoing equation yields $p(Y \mid s) = 1/[1 + q(n)b]$ for the lower limb and, by Eq. 16, $p(Y \mid s) = 1$ for the upper limb. The probability that a lower-limb bias is used decreases from 1 to 0 linearly as the stimulus magnitude increases from two to three quantal units. In summary, then, if s denotes the stimulus magnitude in quantal units, the response function is

$$p(Y \mid s) = \begin{cases} 0, & \text{if } 0 \leqslant s < 1 \\ \dfrac{(s-1)^2}{s - 1 + q(n)b}, & \text{if } 1 \leqslant s < 2 \\ \dfrac{1 + (s-2)\,q(n)b}{1 + q(n)b}, & \text{if } 2 \leqslant s < 3 \\ 1, & \text{if } 3 \leqslant s. \end{cases}$$

Plots of this function for three values of $q(n)b$ are shown in Fig. 18. We

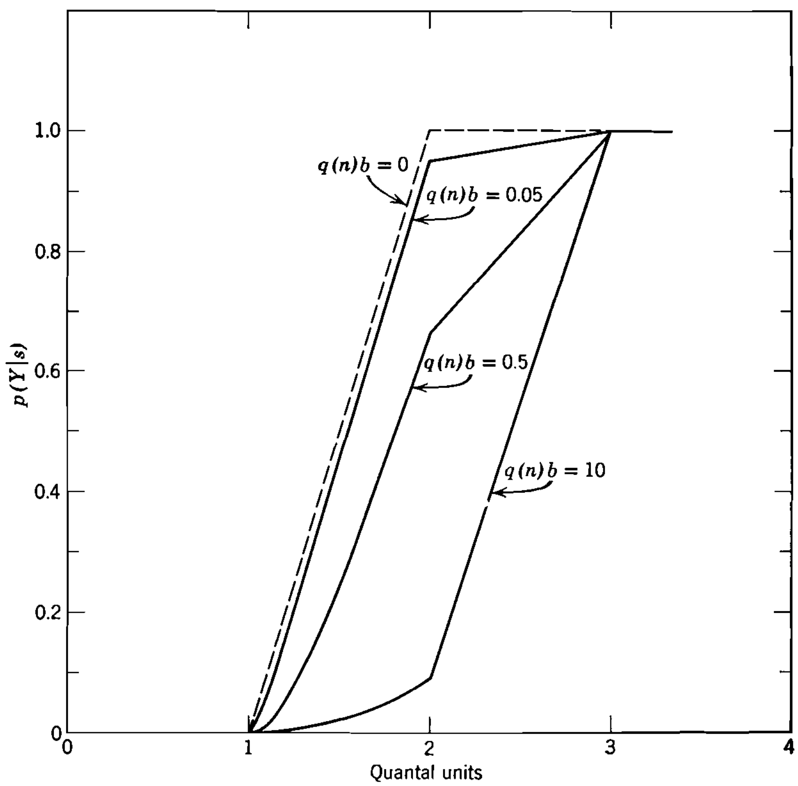

Fig. 18. Predicted observed probability of detection versus a measure of the stimulus increment in quantal units assuming a true underlying rectilinear function and the response biasing model described in the text. The parameter $q(n)$ is the true false alarm rate and b is a quantity that depends on the frequency of "catch" trials and the learning rates of the subject.

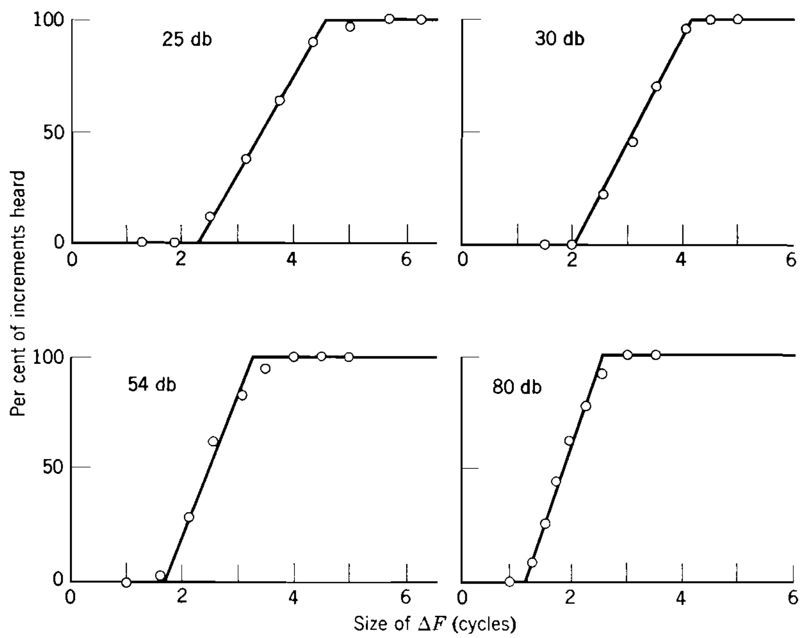

Fig. 19. Quantal data for the detection of frequency increments by one subject at four levels of sound intensity. Each data point is based on 100 observations. The theoretical curves were drawn subject to the condition that the intercepts stand in the relation of 2 to 1. Adapted with permission from Stevens, Morgan, and Volkmann (1941, p. 327).

see that for small values of $q(n)b$, which, for example, corresponds to a small proportion of catch trials, the only effect is a slight distortion of the true quantal function near the upper intercept. As $q(n)b$ becomes larger, we obtain a function that is approximately a straight line with 3:1 intercepts, and as $q(n)b$ becomes still larger the function approaches a 3:2 line.

In spite of all the arguments why the observed functions should not be rectilinear, the surprising thing is how linear they are. In Fig. 19 are data for one subject detecting frequency increments at different levels of intensity. The theoretical lines have 2:1 intercepts. Similar data for the detection of intensity differences of a pure tone for two subjects are shown in Fig. 20. Again 2:1 lines are shown.

In my opinion, the main challenge of these results for those who do not believe that thresholds exist is to explain, using a continuous theory, why the apparent intercepts should exhibit a 2:1 ratio. This has yet to be done.

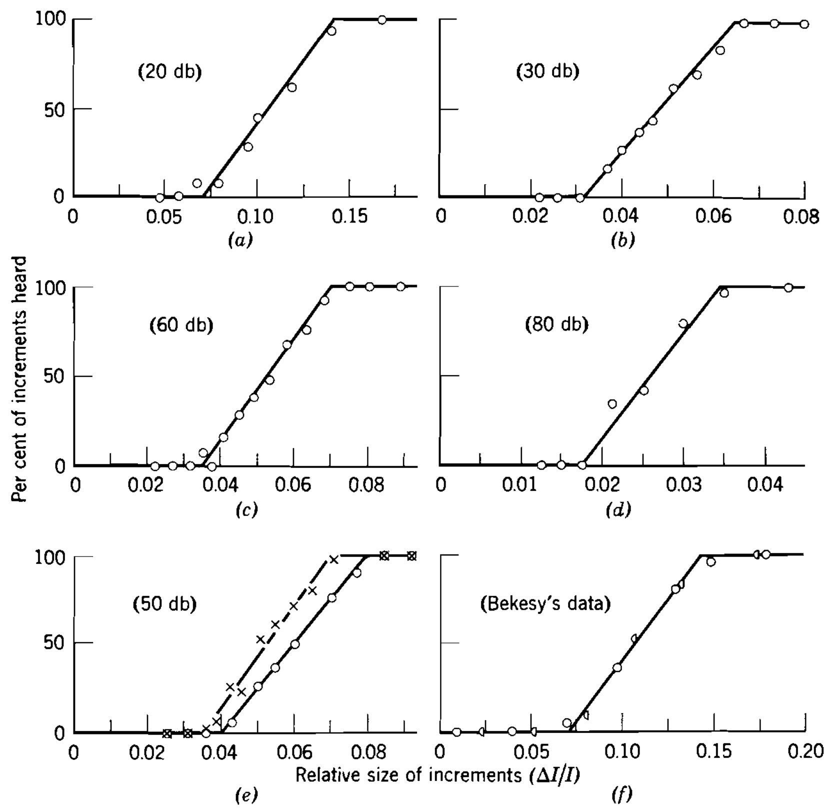

Fig. 20. Quantal data for the detection of intensity increments of a 1000-cps tone by one subject at five levels of sound intensity and in plot *f* Bekesy's intensity data for both increments and decrements. Each data point is based on 50 to 100 observations. The theoretical curves were drawn subject to the condition that the intercepts stand in the relation of 2 to 1. Adapted with permission from Stevens, Morgan, and Volkmann (1941, p. 323).

7. PURE RECOGNITION

7.1 Introduction

As mentioned at the beginning of this chapter, a recognition experiment is a complete identification design in which the presentation set has at least two stimuli in addition to the null one. We make the further distinction that it is a *pure recognition experiment* when $\emptyset \notin S$, and that it is a *simultaneous detection and recognition experiment* when $\emptyset \in S$.

Given a set of stimuli, it is the experimenter's decision whether to perform a pure recognition or a simultaneous detection and recognition study. To be sure, he would be considered foolish to use a simultaneous detection and recognition design with perfectly detectable stimuli, but it has not been uncommon to use pure recognition designs when the stimuli are difficult to detect. In such experiments there are bound to be trials when the subject would prefer, if permitted, to say that no stimulus was presented. That not being allowed, what happens? Broadly speaking, there are two possibilities. Either the subject has some information about the stimulus which he then uses when forced to recognize, in which event he is more often correct than not, even though he does not believe a stimulus was presented; or he does not have any information at all, and so he can only choose arbitrarily, possibly with a bias, among the recognition responses. If the first possibility is correct, then the pure recognition experiment may well tell us directly about his ability to recognize the stimuli, but if the second is correct, these simple experiments cannot produce simple data. The response frequencies are a compound of his ability to recognize when he has detected and of his arbitrary assignment of undetected presentations to the recognition responses.

The necessary data to decide this point are included in Shipley's (1961) thesis. On each trial stimulus s, stimulus s', or noise n was presented, and the subjects were required both to detect, Y or N, and, no matter what the detection response, to recognize s or s'. The stimuli were tones differing both in frequency and intensity. (Because it is conceivable that there is an interaction between the detection and forced-recognition responses, the same experiment was run except that the subjects were not required to recognize when they failed to detect. No interaction was found.) We look to see whether the recognition of "No stimulus" responses depends upon the stimulus presentation. The conditional percentages of s responses when no detection is reported and, for comparison, when detection is reported are shown in Table 7 for each of the three presentations. Also shown are the comparable data for recognition in the two-alternative forced-choice design. Apparently these three subjects were unable differentially to recognize the stimuli when they failed to detect them. There are strong biases, but there is no correlation with the stimulus presented (except possibly a slight negative one for subject 3 in the Yes-No design).

It is important before proceeding further to assess the significance of these data. Above all, they suggest that a recognition experiment involving barely detectable stimuli should not employ a forced-choice design. Two such studies performed in different laboratories with slightly different levels

Table 7 Percentage of Detected and of Undetected Responses Recognized as Stimulus *s* (Shipley, 1961)

Stimulus Presentation	Subject 1		Subject 2		Subject 3	
	Detected	Undetected	Detected	Undetected	Detected	Undetected
s	90.3	77.5	88.7	73.8	87.8	31.2
s′	7.6	75.7	19.7	74.7	11.7	41.5
n	40.8	72.2	56.8	78.6	47.9	31.5

Yes-No Design

Stimulus Presentation	Subject 1		Subject 2		Subject 3	
	Correct Detection	Incorrect Detection	Correct Detection	Incorrect Detection	Correct Detection	Incorrect Detection
s	87.6	43.4	87.2	63.3	87.3	42.7
s′	14.3	46.6	24.6	62.8	12.7	40.9

Forced-Choice Design

of detectability could easily lead to apparent differences in recognition, even when none existed. The relative stability, both experimentally and theoretically, of the forced-choice as against the three-response category scheme in discrimination studies (see Chapter 4, Sec. 5.2) must not be interpreted as a blanket recommendation for forced-choice designs in other types of experiments.

These data raise again the question of a threshold, for, when symmetric payoffs are used, the "no stimulus" response contains no residual information about the identity of the presented stimulus. This is certainly consistent with the notion of a detection threshold. It does not, however, prove that one exists. Another interpretation is that all information about the presentation is lost once the subject decides that no stimulus was presented. It should be possible to decide between these two hypotheses by running a simultaneous detection and recognition experiment with various asymmetric payoff matrices. If there is a threshold, the recognition of detected stimuli will be degraded in a predictable fashion as the detection frequency is increased by changing the payoffs; whereas, if we are witnessing a decision phenomenon, the recognition of detected stimuli will be independent of the frequency of detection responses. This study has not been performed.

Having made clear that the two types of recognition experiments must be treated separately, the remainder of this section is devoted to pure recognition studies.

7.2 Information Theory Analysis

Aside from traditional statistics of contingency tables, the main mathematical tool that has come to be used to study pure recognition with more than two or three stimulus presentations is information theory. It is impossible to devote the space needed for a complete review of Shannon's theory (Shannon & Weaver, 1949) or even of its varied uses in psychophysics, but fortunately several suitable summaries with extensive bibliographies already exist (Attneave, 1959; Luce, 1960; Miller, 1953, 1956).

Except for the words used, the description of a communication system assumed by the information theorists is identical to our complete identification design. They interpret S as the set of elementary signals that can be transmitted by the system, R as the set of signals that can be received, and ι as a given one-to-one correspondence between them. Thus S might be the ordinary alphabet, R the sequences of dots and dashes used in the Morse code, and ι the code relating them. A probability distribution p over $S \times R$ is assumed to exist—$p(s, r)$ is interpreted as the joint probability that signal s is transmitted and r is received. If we define

$$P(s) = \sum_{r \in R} p(s, r),$$

$$p(r) = \sum_{s \in S} p(s, r),$$

and

$$p(r \mid s) = \frac{p(s, r)}{P(s)},$$

then $P(s)$ is the probability that signal s is transmitted, $p(r)$ the unconditional probability that r is received, and $p(r \mid s)$ the conditional probability that r is received when s is transmitted. In a complete identification experiment, $P(s)$ is the probability that s is presented, $p(r)$, the unconditional probability of response r, and $p(r \mid s)$, the conditional probability of response r given stimulus s. In the communication terminology the matrix of conditional probabilities $p(r \mid s)$ is called a *noise matrix*, for by definition that which prevents communication from being perfect is *noise*; in a complete identification experiment it is called a confusion matrix.

Information theorists undertook to state by means of a single summary number the average information-transmitting characteristics of such a system. It was to be a measure that would satisfy certain a priori criteria and permit one to capture in precise theorems certain known empirical results concerning channel capacity, information transmission, and error correction. The major a priori requirement imposed by Shannon was this. Suppose that several signals are, in the statistical sense, independently selected and transmitted; then the average amount of information created by their joint selection shall be the sum of the average amounts of information created by their separate selections, that is, average information is postulated to be additive when the selections are independent. He showed that this coupled with other much weaker conditions implies that the measure must be of the form

$$H(S) = -\sum_{s \in S} P(s) \log P(s). \tag{42}$$

Usually, the base of the logarithm is chosen to be 2, thereby setting the unit of measure. Following a suggestion by J. Tukey, this unit is called a *bit*. A choice between two equally likely alternatives creates one bit; among 4, two bits; among 8, three bits; etc.

Two features of this measure should be noted. First, it is nonnegative and has the value 0 when and only when one of the probabilities is 1 (and so all the rest are 0). That is to say, no information is generated by the selection of an alternative that is certain to be selected; this agrees, for example, with the view that little or no information is transmitted by the conventional replies to conventional greetings. Second, the measure has its maximum value when all of the probabilities are equal; if there are k alternatives, the maximum is $\log_2 k$.

In like manner, we have as the information measure of the responses

$$H(R) = -\sum_{r \in R} p(r) \log_2 p(r), \tag{43}$$

and, as the conditional measure of the response given the stimulus,

$$H(R \mid S) = \sum_{s \in S} \sum_{r \in R} p(s, r) \log_2 p(r \mid s). \tag{44}$$

The quantity

$$T(S; R) = H(R) - H(R \mid S) \tag{45}$$

is called the *information transmitted* from the stimulus to the response. It not only plays a significant role in information theory itself, but it has proved to be a useful measure in psychology. It is not difficult to show that

$0 \leqslant T(S; R) \leqslant H(S)$; that T is 0 when and only when the responses are statistically independent of the stimuli, that is, when $p(r \mid s) = p(r)$; and that it is equal to $H(S)$, its maximum, when there is no confusion, that is, when $p[r \mid \iota(r)] = 1$ for all $r \in R$. Thus T is an inverse measure of how much degrading, on the average, is introduced by the subject.

The principal empirical results stemming from information analyses of recognition experiments are described in an excellent survey article by Miller (1956), so we need only summarize them very briefly here. Suppose that we select k stimuli on some unidimensional continuum, such as sound energy, so that they cover most of the sensible range and are more or less equally spaced; then $T(S; R)$ is approximately equal to $H(S)$ (which equals $\log_2 k$ when the stimuli are equally likely) in the range from 0 to roughly 2 bits. For $H(S)$ greater than 2 bits, the rate of increase of the transmitted information diminishes sharply, reaching a peak between 1.6 and 3.9 bits, depending upon the continuum. Increasing $H(S)$ beyond that point may in fact cause $T(S; R)$ to decrease. Moreover, for pitch at least, Pollack (1952) has shown that the range of frequencies can be varied by a factor of at least 20 with relatively little effect on the maximum amount of information transmitted by a fixed number of equally spaced stimuli.

If the stimuli are multidimensional, the maximum value of T can be increased considerably from what it is for any one of their dimensions, but the maximum is always less than the sum of the values of the separate component dimensions. Nonetheless, more seems to be gained by adding another dimension than by refining the categories per dimension. (See Beebe-Center, Rogers, & O'Connel, 1955; Halsey & Chapanis, 1954; Klemmer & Frick, 1953; Pollack, 1953; Pollack & Ficks, 1954).

The regularity and generality of these results is impressive, and much has been made of the relatively small values of transmitted information. On the one hand, it is well to know such facts when designing certain types of systems in which men must interact with information generating or receiving machines, and, on the other hand, they stand as summary statements in need of detailed scientific explanation and, thereby, refined restatement. Some writers have, I believe, taken the view that the behavioral regularities expressed in terms of the information measures themselves constitute a theory, but I am inclined to class them simply as empirical generalizations requiring theoretical analysis. It is often not easy to know when a particular relation stemming from experiments should be considered a generalization in need of explanation and when it should be introduced as an unanalyzed assumption of a theoretical system, but two features of these relations lead me to class them as generalizations. First, they are not really simple, certainly not in the sense that linearity, additivity, and independence assumptions are simple.

Second, they are statements about averages—not just averages of data, which are often used as estimates of probabilities, but averages over distinct classes of responses—and so the observed regularities are bound to mask much of the possibly interesting fine detail of the behavior.

McGill (1954, 1955a,b) and Garner and McGill (1956) extended the decomposition of transmitted information (Eq. 45) to more complex stimulus presentations. McGill's idea was, roughly, to ascertain how much each of the various possible determiners of the response, such as previous responses and stimulus presentations, contribute to the total information transmitted. He devised an additive decomposition in terms of the contributions of each variable and the various possible interactions among them. This constitutes an information theoretic, and hence non-parametric, analogue of the analysis of variance, and as such it is a useful device in the study of sequential dependencies among the responses and in discovering which events are determiners of the responses. For a summary of the ideas and applications, see Luce (1960).

7.3 A Choice Analysis of the Results of Information Theory

Little has been published attempting to account for the information theory findings in terms of the three response theories we have been considering. To show that work of this sort is possible, a choice theory analysis is given here. Also, see Luce (1959).

Consider stimuli that differ on only one physical dimension, such as energy. In terms of the distance measure (Eq. 6) introduced in Sec. 1.2, it is plausible that distance is simply additive for such stimuli. In terms of the η-scale, this amounts to assuming that

$$\eta(s, s'') = \eta(s, s')\,\eta(s', s'') \tag{46}$$

when s, s' and s'' differ on one physical dimension and $s < s' < s''$. Let us suppose further that we choose k stimuli that are ordered $s_1 < s_2 < \ldots < s_k$ and spaced on that dimension so that successive pairs are equally recognizable in the sense that

$$\eta(s_i, s_{i+1}) = \eta, \qquad \text{for } i = 1, 2, \ldots, k-1. \tag{47}$$

Performing a simple induction on Eqs. 46 and 47, it is easy to see that

$$\eta(s_i, s_j) = \eta^{|i-j|}.$$

Thus the confusion matrix of scale values is of the form

$$
\text{Stimulus Presentation}\quad
\begin{array}{c}
 \\ s_1 \\ s_2 \\ s_3 \\ \cdot \\ \cdot \\ \cdot \\ s_k
\end{array}
\overset{\text{Response}}{\begin{array}{ccccc}
1 & 2 & 3 & \dots & k
\end{array}}
\begin{bmatrix}
1 & \eta & \eta^2 & \dots & \eta^{k-1} \\
\eta & 1 & \eta & \dots & \eta^{k-2} \\
\eta^2 & \eta & 1 & \dots & \eta^{k-3} \\
 & & & & \\
\dots & \dots & \dots & \dots & \dots \\
 & & & & \\
\eta^{k-1} & \eta^{k-2} & \eta^{k-3} & \dots & 1
\end{bmatrix}, \tag{48}
$$

where the bias parameters are omitted.

The first thing to note is that the model has the often observed U-shape when the probabilities $p(r \mid s_r)$ are plotted against r. In Table 8 the predicted probabilities for the end and middle stimuli are presented for several small k's and for several plausible values of η. The dip is evident. Of course, the bias parameters affect the exact form of this U-shaped function.

Table 8 Comparison of the Theoretical Probability of Correct Identification for End and Middle Stimuli

k		η = 0.50	0.25	0.10
2	$p(1 \mid s_1)$	0.667	0.800	0.909
3	$p(1 \mid s_1)$	0.571	0.762	0.901
	$p(2 \mid s_2)$	0.500	0.667	0.833
5	$p(1 \mid s_1)$	0.547	0.751	0.900
	$p(3 \mid s_3)$	0.400	0.615	0.820
7	$p(1 \mid s_1)$	0.546	0.750	0.900
	$p(4 \mid s_4)$	0.381	0.604	0.820

Next we look into the question of transmitted information. By what we have assumed for our stimuli, we know that the matrix of scale values for the $k = 2$ recognition design is

$$\begin{array}{c} \\ s_i \\ s_{i+1} \end{array}\begin{array}{c} \begin{array}{cc} s_i & s_{i+1} \end{array} \\ \begin{bmatrix} 1 & \eta \\ \eta & 1 \end{bmatrix} \end{array},$$

again omitting biases. Assuming the independence condition, Assumption 4 on p. 114, the parallel forced-choice design has the matrix of scale values

$$\begin{array}{c} \\ \langle s_i, s_{i+1} \rangle \\ \langle s_{i+1}, s_i \rangle \end{array}\begin{array}{c} \begin{array}{cc} 1 & 2 \end{array} \\ \begin{bmatrix} 1 & \eta^{\sqrt{2}} \\ \eta^{\sqrt{2}} & 1 \end{bmatrix} \end{array}.$$

In the light of our discussion of the quantal model, it is not clear whether the independence assumption is justified, but in order to continue the discussion we accept it. In discrimination work (see Chapter 4, Sec. 1.2) two stimuli are said to be one jnd (one just noticeable difference) apart if $p(1 \mid \langle s_i, s_{i+1} \rangle) = \frac{3}{4}$, in which case the forced-choice model yields

$$\eta = (\tfrac{1}{3})^{1/\sqrt{2}}.$$

For stimuli that are m-jnds apart in the sense that $m - 1$ stimuli can be found between them such that successive ones are one jnd apart, Eq. 46 implies

$$\eta = (\tfrac{1}{3})^{m/\sqrt{2}}. \tag{49}$$

For our calculations, let us consider k stimuli so spaced that successive ones are m jnds apart; thus the total range of stimuli is $(k - 1)m$ jnds. Specifically, let us fix the range at $2^6 = 64$ jnds and let $k = 5, 9, 17$, and 33 stimuli, which means that successive ones are separated by $m = 16, 8, 4$, and 2 jnds, respectively. For each k the confusion matrix of probabilities can be determined from the scale values given in Eq. 48, using Eq. 49 to determine η. Assuming that the stimuli are equally likely, the information transmitted is calculated using the formulas in Sec. 7.2. The results are shown in the last column of Table 9. Up to something just over three bits presented, the information transmitted is nearly equal to the stimulus information. Increasing the stimulus information further, the transmitted information increases less rapidly, reaching a maximum of about 3.6 bits. Not only does this correspond qualitatively to the data, but it is in about the right range of values. The data, however, appear to have arisen from a somewhat broader range of stimulus values and to have resulted in

Table 9 Choice Theory Predictions of Information Transmitted versus Information Presented

Number of Equally Likely Alternatives	Bits Presented	Bits Transmitted for Stimulus Range in jnds	
		16	64
3	1.58	1.56	1.58
5	2.32	1.90	2.32
9	3.17	1.53	3.12
17	4.09		3.59
33	5.04		3.45

somewhat smaller maxima, which suggests that we are using too small a value for η. It is obvious that we could select a value of η that would yield quantitatively the same summary results as the data; it is a much more subtle question whether this model can reproduce the whole confusion matrix in detail, and that has not been thoroughly investigated.

A second major information result is that the increase in information transmitted is relatively slight as the range is increased (Pollack, 1952). Two ranges, differing by a factor of four, are shown in columns 3 and 4 of Table 9. A considerable difference in the maximum exists, strongly suggesting that the present model is inadequate to explain these results.

A third information theory result concerns the less than additive increase of the transmitted information as the number of dimensions per stimulus is increased. We consider the simplest possible case of two perfectly detectable stimuli differing on two equally recognizable dimensions. Suppose s and s' are the stimulus values on one dimension and t and t' on the other and that their confusion matrices of scale values are

$$\begin{array}{lc} & \text{Response} \\ & \begin{array}{cc} s & s' \end{array} \\ \begin{array}{r} \text{Stimulus} \quad s \\ \text{Presentation} \quad s' \end{array} & \begin{bmatrix} 1 & \eta \\ \eta & 1 \end{bmatrix} \end{array} \qquad \begin{array}{lc} & \text{Response} \\ & \begin{array}{cc} t & t' \end{array} \\ \begin{array}{r} t \\ t' \end{array} & \begin{bmatrix} 1 & \eta \\ \eta & 1 \end{bmatrix} \end{array}.$$

For the choice model with Assumption 4, the scale values for the composite stimuli are

$$\begin{array}{rc} & \begin{array}{cc} \langle s, t\rangle & \langle s', t'\rangle \end{array} \\ \begin{array}{r} \langle s, t\rangle \\ \langle s', t'\rangle \end{array} & \begin{bmatrix} 1 & \eta^{\sqrt{2}} \\ \eta^{\sqrt{2}} & 1 \end{bmatrix} \end{array}.$$

The ratio of the information transmitted in the composite case to that in the unidimensional case, assuming equally likely stimulus presentations, is shown in Table 10 for three values of η. The ratio is always less than 2, and it approaches 1 as the stimuli become more recognizable.

Table 10 Predicted Ratio of Information Transmitted for Two-Alternative Two-Dimensional Case to Two-Alternative One-Dimensional Case

η	0.50	0.25	0.10
ratio	1.90	1.66	1.38

Although there are some tentative indications that the recognition choice model may account for the information theory results, the only really satisfactory test is whether it accounts for the whole confusion matrix. Assuming that the stimuli are physically ordered and that i is the correct response for stimulus presentation s_i, we have the matrix of scale values

$$\begin{array}{c} \\ s_1 \\ s_2 \\ \cdot \\ \cdot \\ \cdot \\ s_k \end{array} \begin{array}{c} \begin{array}{cccc} 1 & 2 & \ldots & k \end{array} \\ \begin{bmatrix} \eta(1,1)b(1) & \eta(1,2)b(2) & \ldots & \eta(1,k)b(k) \\ \eta(2,1)b(1) & \eta(2,2)b(2) & \ldots & \eta(2,k)b(k) \\ & & & \\ \cdots & \cdots & \cdots & \cdots \\ & & & \\ \eta(k,1)b(1) & \eta(k,2)b(2) & \ldots & \eta(k,k)b(k) \end{bmatrix} \end{array}.$$

Assuming that Eq. 46 holds, then for $i < j$,

$$\eta(i, j) = \eta(i, i + 1)\eta(i + 1, i + 2) \ldots \eta(j - 1, j),$$

and, recalling that $\eta(i, j) = \eta(j, i)$ and that $\eta(i, i) = 1$, it follows that there are only $k - 1$ independent stimulus parameters, the $\eta(i, i + 1)$, and $k - 1$ bias parameters, the $b(i)$, to be estimated.

In practice, we can only be certain that the conditional probabilities on and near the main diagonal are appreciably larger than zero, and so any estimation scheme had better rely heavily upon these entries. One possibility—one that has no known statistical properties but that uses entries only from the main diagonal and adjacent cells—is

$$\eta(i, i+1)^2 = \left[\frac{\eta(i, i+1)b(i+1)}{b(i)}\right]\left[\frac{\eta(i, i+1)b(i)}{b(i+1)}\right]$$

$$= \frac{p(i+1 \mid i)}{p(i \mid i)} \frac{p(i \mid i+1)}{p(i+1 \mid i+1)}$$

and

$$\left[\frac{b(i+1)}{b(i)}\right]^2 = \left[\frac{\eta(i, i+1)b(i+1)}{b(i)}\right]\left[\frac{b(i+1)}{\eta(i, i+1)b(i)}\right]$$

$$= \frac{p(i+1 \mid i)}{p(i \mid i)} \frac{p(i+1 \mid i+1)}{p(i \mid i+1)}.$$

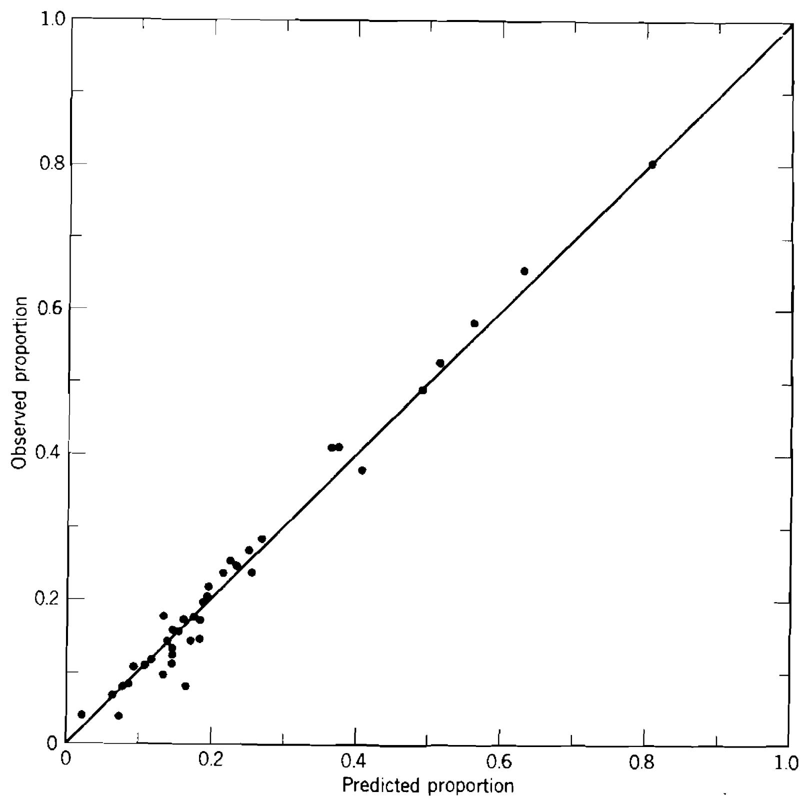

Fig. 21. Observed proportion versus proportions predicted by the choice model for McGuire's (Shepard, 1958b) size recognition data. The estimation scheme and the collapsing of the data are described in the text.

We apply this estimation scheme to McGuire's data on size recognition (reported in Shepard, 1958b), although they are not ideal because ten subjects are averaged together. Each subject responded to 80 presentations of each of nine circular areas. A χ^2 comparison of the predicted and observed proportions yields a value of 96.3 which with $8 \times 9 - 16 = 56$ degrees of freedom is highly significant. This is none too surprising because the estimation scheme completely ignores the small entries in the table, which, of course, contribute heavily to χ^2. That the estimates are probably nonoptimal is indicated, for example, by the fact that about half the total contribution to χ^2 comes from the first column. In the light of the failings of the estimation procedure, a more reasonable test of the model is to lump together all entries to the left of cell $i - 1$ in row i and all of those to the right of the $i + 1$ entry. This reduces χ^2 to 15.6 and the degrees of freedom to 14, yielding $0.2 < p < 0.5$. The observed versus predicted proportions for this collapsing are shown in Fig. 21. These results suggest that a better estimation method might very well result in a nonsignificant over-all χ^2.

8. SIMULTANEOUS DETECTION AND RECOGNITION

Even restricting our attention to the simplest simultaneous detection and recognition designs, namely $S = \mathscr{S} = \{s, s', \emptyset\}$, we find that relatively little work has been reported. Because the ideas are adaptations of those we have already discussed, it will suffice simply to outline them.

8.1 Signal Detectability Analysis

Following the general structure of the signal detectability model, there is a decision axis relating stimulus s to stimulus s', another relating s to noise, and a third relating s' to noise. Tanner (1956) assumed that they can be represented in the plane, as in Fig. 22. The three intersections are supposed to occur at the means of the distributions projected on the several axes. The noise and each of the stimulus-plus-noise distributions are assumed to be independent and normal, all with equal variance. The two detection axes are separated by some angle θ, not necessarily 90°.

Tanner (1956) used this structure to analyze the pure recognition experiment $\mathscr{S} = \{s, s'\}$. I did not present this in the last section because no testable conclusions seem to derive from it. Swets and Birdsall (1956) discussed the simultaneous detection and recognition experiment, proposing

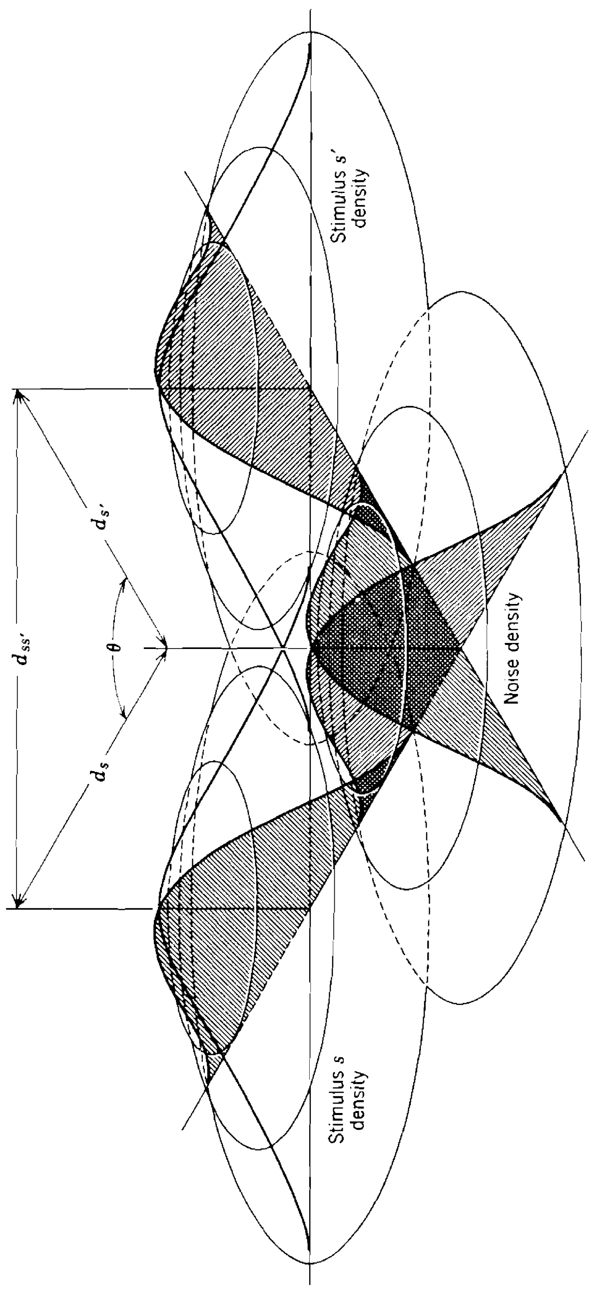

Fig. 22. Representation of stimulus effects in the signal detectability model for simultaneous detection and recognition of two stimuli s and s'. The two decision axes of the simple Yes-No design are assumed to be separated by an angle θ, not in general 90°.

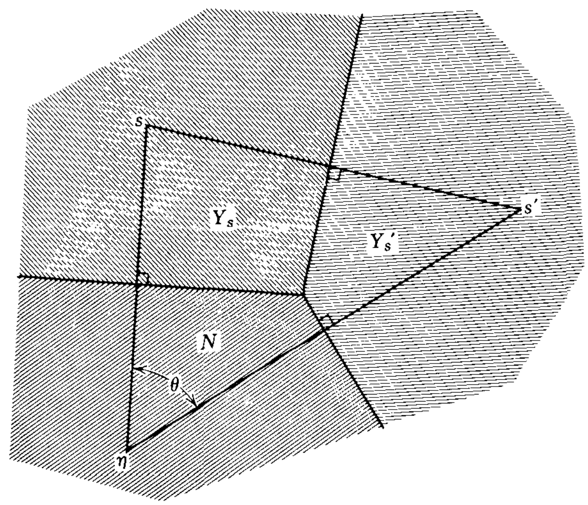

Fig. 23. A decision rule proposed by Swets and Birdsall (1956) for the two-stimulus simultaneous detection and recognition experiment.

the decision rule shown in Fig. 23. The three straight lines meet at a point in the triangle formed by connecting the means of the three distributions; each line is orthogonal to a side of the triangle.

8.2 Choice Analysis

The choice model analysis follows immediately from Eq. 5 and Assumptions 1 to 3:

$$\begin{array}{c} \\ s \\ s' \\ n \end{array} \begin{array}{c} \begin{array}{ccc} Y_s & Y_{s'} & N \end{array} \\ \left[\begin{array}{ccc} 1 & \lambda b & \eta c \\ \lambda & b & \eta' c \\ \eta & \eta' b & c \end{array}\right] \end{array}, \tag{50}$$

where $\eta = \eta(s, n)$, $\eta' = \eta(s', n)$, and $\lambda = \eta(s, s')$.

The adequacy of the choice model for this simultaneous identification design can be tested, using Shipley's (1961) data. The parameters η and η' are estimated as the mean of those obtained from the simple Yes-No and two-alternative forced-choice experiments (Table 2). The remaining

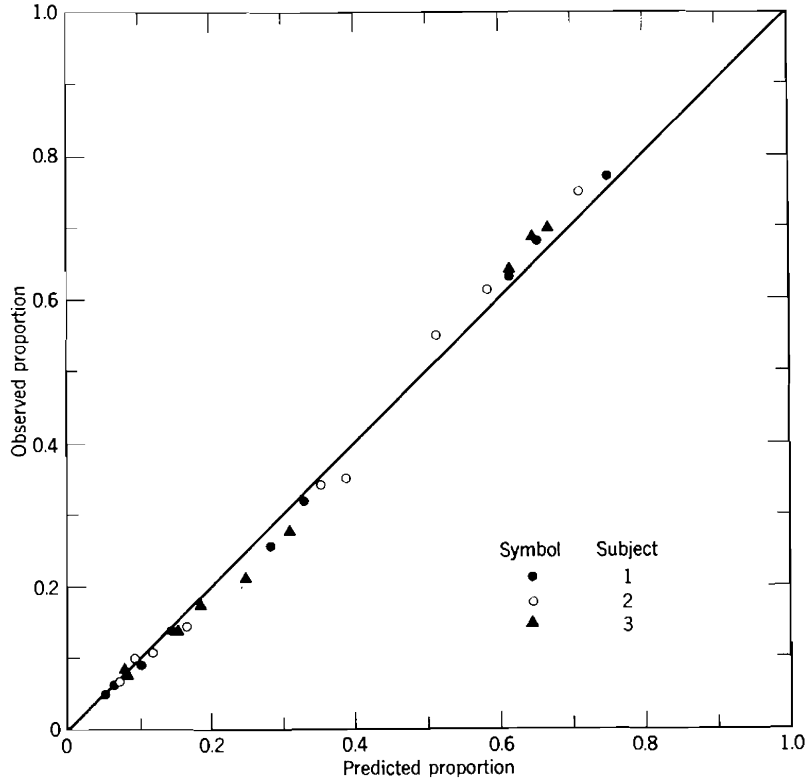

Fig. 24. Observed proportions versus those predicted by the choice model for a two-stimulus simultaneous detection and recognition design. The data are from Shipley (1961). The parameter values used were

Subject	η	η'	λ	b	c
1	0.240	0.325	0.089	1.068	1.785
2	0.372	0.358	0.139	0.767	1.610
3	0.298	0.389	0.124	0.958	1.241

three parameters are estimated from the data in question by

$$\lambda = \left[\frac{p(Y_{s'} \mid s)}{p(Y_s \mid s)} \frac{p(Y_s \mid s')}{p(Y_{s'} \mid s')}\right]^{1/2}$$

$$b = \left[\frac{p(N \mid s)}{p(Y_s \mid s)} \frac{p(Y_{s'} \mid s)}{p(Y_s \mid s)} \frac{p(Y_{s'} \mid s')}{p(Y_s \mid s')} \frac{p(Y_{s'} \mid s')}{p(N \mid s')} \frac{p(Y_{s'} \mid n)}{p(Y_s \mid n)}\right]^{1/4}$$

$$c = \left[\frac{p(N \mid s)}{p(Y_s \mid s)} \frac{p(Y_{s'} \mid s)}{p(Y_s \mid s)} \frac{p(N \mid s')}{p(Y_s \mid s')} \frac{p(N \mid n)}{p(Y_s \mid n)} \frac{p(N \mid n)}{p(Y_{s'} \mid n)}\right]^{1/4},$$

which follow immediately from Eq. 50. This leaves three degrees of freedom. The observed versus predicted proportions are shown in Fig. 24. Although the prediction is really quite good, the errors appear to be not entirely random: the predicted values are consistently less than the observed for the larger probabilities and consistently more for the middle values. Whether better estimates would eliminate this deviation is not known.

9. DETECTION OF AN UNKNOWN STIMULUS

A simple Yes-No detection experiment in which the stimulus presentation can be one of two or more different stimuli is said to involve the detection of an unknown stimulus. It is, of course, a partial identification experiment—the first to be examined in this chapter. For simplicity we restrict our attention to the case in which the unknown stimulus is one of two possibilities, s or s'. The existing data (Creelman, 1959; Green, 1958; Swets, Shipley, McKey, & Green, 1959; Tanner, Swets, & Green, 1956; Veniar, 1958a,b,c) indicate that the detectability of an unknown stimulus is less than that of either stimulus singly and that the decrement is an increasing function of the recognizability of the stimuli. Why?

One theory, suitable for pure tones, assumes that a subject can listen at any instant to only a narrow band of frequencies and that, when the signal is unknown, he must continually shift from filter to filter (Tanner, Swets, & Green, 1956). Because he sometimes listens through the wrong filter, the probability of detecting an unknown stimulus must be less than that for a known one.

A second acoustic theory, proposed by Green (1958), assumes that the subject can listen through as many filters as needed, each being centered on a different frequency. The effect of increasing the number of frequencies is to increase the total amount of noise heard without affecting the detectability of the stimulus in a single band. This produces an over-all reduction in the detectability. The theory states precisely how the detectability reduces with the number of unknown frequencies, the band width of the filter being a parameter in the model.

Detailed statements of these theories and some relevant acoustic data are given in Swets, Shipley, McKey, & Green (1959); on the whole, the scanning model comes off better than the multiband model. Veniar (1958a,b,c) suggests that neither is completely adequate. Whichever may be better for pure tones, neither is readily generalizable to other classes of stimuli for which, presumably, the same decrease in detectability occurs. Possibly each stimulus modality requires its own explanation, but, in the absence of any compelling data or arguments, one hopes that the phenomenon is of a more general response character and that it requires fewer specific assumptions than those made in the filter theories.

Shipley (1960) suggested such an idea, one that is applicable to any response model. She supposed that the subject is covertly recognizing the stimuli as well as overtly detecting them, even though he makes no recognition response. If so, and if we assume the choice model, the matrix of scale values is simply Eq. 50 with the two detection responses combined.

To show that this predicts results qualitatively similar to those observed, consider the special case in which the stimuli are equally detectable and there is no recognition bias, that is, $\eta = \eta'$ and $b = 1$. Then we can combine the two stimulus presentations into one because the two rows are the same:

$$\text{Stimulus Presentation} \quad \begin{array}{c} \\ s \text{ or } s' \\ n \end{array} \overset{\text{Response}}{\begin{array}{cc} Y & N \\ \end{array}} \begin{bmatrix} 1 + \lambda & \eta c \\ 2\eta & c \end{bmatrix}.$$

This is equivalent, in the sense of generating the same probabilities, to a matrix of scale values of the form

$$\begin{array}{c} s \text{ or } s' \\ n \end{array} \overset{\begin{array}{cc} Y & N \end{array}}{\begin{bmatrix} 1 & \zeta c' \\ \zeta & c' \end{bmatrix}},$$

where

$$\zeta = \frac{\eta}{[(1 + \lambda)/2]^{1/2}}. \tag{51}$$

This is the form of the simple Yes-No matrix discussed in Sec. 2.2, and so ζ is an apparent detection parameter and

$$c' = \frac{c/2}{[(1 + \lambda)/2]^{1/2}}. \tag{52}$$

is an apparent bias parameter.

The probability of detecting an unknown stimulus is smaller than that of a known stimulus if and only if $\zeta > \eta$, and that in turn holds if and only if $\lambda < 1$, as it must be for distinct stimuli. Moreover, as the stimuli become more recognizable, that is, as λ gets smaller, ζ gets larger and so detectability becomes poorer, as has been observed.

A parallel development exists for the forced-choice design. Omitting the biases, the simultaneous detection-recognition matrix of scale values is seen to be

$$
\begin{array}{cc} & \text{Responses} \\ & \begin{array}{cccc} 1s & 1s' & 2s & 2s' \end{array} \\
\text{Stimulus Presentation} \quad \begin{array}{c} \langle s, n\rangle \\ \langle s', n\rangle \\ \langle n, s\rangle \\ \langle n, s'\rangle \end{array} &
\begin{bmatrix} 1 & \lambda & \eta^{\sqrt{2}} & \delta \\ \lambda & 1 & \delta & \eta'^{\sqrt{2}} \\ \eta^{\sqrt{2}} & \delta & 1 & \lambda \\ \delta & \eta'^{\sqrt{2}} & \lambda & 1 \end{bmatrix},
\end{array}
$$

where δ is given by

$$(\log \delta)^2 = (\log \eta)^2 + (\log \eta')^2$$

and the other symbols have their previous meanings. To see what happens qualitatively, again suppose $\eta = \eta'$ and that the biases all equal 1; then we can collapse on both rows and columns:

$$
\begin{array}{cc} & \begin{array}{cc} 1 & 2 \end{array} \\
\begin{array}{c} \langle s \text{ or } s', n\rangle \\ \langle n, s \text{ or } s'\rangle \end{array} &
\begin{bmatrix} 1+\lambda & 2\eta^{\sqrt{2}} \\ 2\eta^{\sqrt{2}} & 1+\lambda \end{bmatrix}.
\end{array}
$$

The standard form for the unbiased forced-choice matrix of scale values is

$$
\begin{array}{cc} & \begin{array}{cc} 1 & 2 \end{array} \\
\begin{array}{c} \langle s, n\rangle \\ \langle n, s\rangle \end{array} &
\begin{bmatrix} 1 & \xi^{\sqrt{2}} \\ \xi^{\sqrt{2}} & 1 \end{bmatrix},
\end{array}
$$

so the effective stimulus parameter is

$$\xi = \frac{\eta}{[(1+\lambda)/2]^{1/\sqrt{2}}}. \tag{53}$$

As for the Yes-No design, there is an increasing reduction in detectability as the stimuli are made more identifiable, that is, as λ is made smaller. We observe that the apparent loss in detectability is greater in the forced-choice than in the Yes-No design because

$$\zeta = \frac{\eta}{[(1+\lambda)/2]^{1/2}} < \frac{\eta}{[(1+\lambda)/2]^{1/\sqrt{2}}} = \xi.$$

To see whether Shipley's idea has any possibility of being correct, I turn again to her (1961) data. Both the simultaneous detection-recognition and the detection of an unknown stimulus conditions were run, so we can collapse the first data matrix on the recognition responses and compare it with the second. This is done in Table 11 for the Yes-No experiment and in Table 12 for the forced-choice experiment. Although there are some differences, which may very well be due to different response biases, they seem in sufficient accord to warrant more study of the idea.

Table 11 Per Cent Yes Responses in the Yes-No Design When the Signal Is Unknown

Stimulus Presentation	Subject 1		Subject 2		Subject 3	
	Observed	Calculated	Observed	Calculated	Observed	Calculated
s	77.5	74.4	69.2	67.6	73.3	78.9
s'	65.2	68.0	70.4	65.2	55.1	72.2
n	27.8	23.1	33.3	25.1	34.6	31.6

The calculated columns are obtained by collapsing the recognition responses in the corresponding detection and recognition experiment (Shipley, 1961). See Table 2 for a description of the experimental conditions.

Table 12 Per Cent Correct Responses in Forced-Choice Design When Stimulus Is Known and When It Is Unknown

Stimulus Presentation	Subject 1			Subject 2			Subject 3		
	Stimulus Known	Stimulus Unknown	Calculated	Stimulus Known	Stimulus Unknown	Calculated	Stimulus Known	Stimulus Unknown	Calculated
$\langle s, n\rangle$	89.5	81.8	83.9	79.8	70.9	66.7	83.6	78.1	72.2
$\langle s', n\rangle$	83.5	76.2	76.8	79.5	73.8	67.7	79.1	61.7	67.8
$\langle n, s\rangle$	88.0	83.0	83.1	79.6	70.7	81.9	86.6	78.3	84.2
$\langle n, s'\rangle$	83.8	78.2	77.8	81.2	78.1	81.0	83.2	80.5	83.1

The calculated columns are obtained by collapsing the recognition responses for the four response detection and recognition experiment (Shipley, 1961). See Table 2 for a description of the experimental conditions.

10. CONCLUSIONS

Although detection and recognition experiments have long been performed—not always under those names—interest in them in theoretical circles has increased considerably during the last ten or fifteen years, and a healthy interaction of theory and experiment has evolved. There is every reason to expect a continued rapid accretion in our knowledge of these basic processes during the 1960's. Perhaps a brief indication of some of the possible paths of work is a good way to summarize the progress we have made.

1. As I have been at pains to point out, there are now at least three different response theories designed to account for detection behavior. One task, therefore, is to decide among them or, if need be, to develop better theories for at least several of the more important modalities, including visual and acoustic intensity.

2. No matter what response model is ultimately judged best, questions of the dependence and independence of effects are bound to exist. For example, we must know when a forced-choice design can be treated as an independent combination of several Yes-No designs. At present, we make *ad hoc* assumptions that apply only to certain extreme cases: a white noise background is assumed to result in independent effects, whereas a pure tone background is assumed to result in a perfect correlation of effects. Just why we should make these assumptions and what we should assume in intermediate cases is unclear; hence we need a detailed characterization, stated in terms of the physical nature of the stimulation, of the dependencies that are introduced.

3. Again, no matter how we resolve the question of the best response theory, we shall need a theory to relate the stimulus parameters to the physical properties of the stimuli and one to relate the bias parameters to various other objective features of the experiments, especially the payoffs. Considerable research is currently under way to uncover the stimulus-parameter relations for both the signal detectability and threshold models. Rather less effort is being devoted to the biasing problem, partly because it strikes a number of workers as less interesting than the stimulus problem, which they feel deals with the fundamental mechanisms of hearing and vision. Without questioning the importance of theories about the stimuli, it should not be forgotten that theories about the bias parameters are likely to get at fundamental issues in learning and cognition and so, in my view, deserve as much careful attention.

4. The theoretical analysis of stimulus recognition is less developed than that of detection, despite all of the interesting work that has resulted from applications of the information-theoretic measures. No one has yet effectively accounted for the information-theory findings in terms of any of the response theories we have discussed. My attempts to apply the choice theory are incomplete and are not entirely satisfactory. No really serious attempts have been made using the other theories, mainly because

of the severe conceptual difficulties that seem to arise when there are more than two or at most three stimuli.

Substantively, recognition seems to be somewhat different from detection, even though both are studied experimentally by means of complete identification designs. This apparent difference should not be forgotten by theorists, for it may mean that quite different response theories are needed. It would not surprise me if detection were a discrete threshold phenomenon, whereas recognition might turn out to be a continuous process or, at least, well approximated by one.

References

Attneave, F. *Applications of information theory to psychology.* New York: Holt, 1959.

Beebe-Center, J. B., Rogers, M. S., & O'Connel, D. N. Transmission of information about sucrose and saline solutions through the sense of taste. *J. Psychol.*, 1955, **39**, 157–160.

Békésy, G. von. Über das Fechner'sche Gesetz und seine Bedeutung für die Theorie der akustischen Beobachtungsfehler und die Theorie des Hörens. *Ann. Phys.*, 1930, **7**, 329–359.

Birdsall, T. G. The theory of signal detectability. In H. Quastler (Ed.), *Information theory in psychology*, Glencoe, Ill.: The Free Press, 1955. Pp. 391–402.

Birdsall, T. G. *Detection of a signal specified exactly with a noisy stored reference signal.* T.R. 93, Electronic Defense Group, Univer. of Mich., Ann Arbor, 1959.

Blackwell, H. R. *Psychophysical thresholds: experimental studies of methods of measurement.* Engineering Research Bulletin 36, Ann Arbor: Univer. of Michigan, 1953.

Bradley, R. A. Incomplete block rank analysis: on the appropriateness of the model for the method of paired comparisons. *Biometrics,* 1954, **10**, 375–390.(a)

Bradley, R. A. Rank analysis of incomplete block designs. II. Additional tables for the method of paired comparisons. *Biometrika*, 1954, **41**, 502–537.(b)

Bradley, R. A. Rank analysis of incomplete block designs. III. Some large sample results on estimation and power for a method of paired comparisons. *Biometrika*, 1955, **42**, 450–470.

Bradley, R. A., & Terry, M. E. Rank analysis of incomplete block designs. I. The method of paired comparisons. *Biometrika*, 1952, **39**, 324–345.

Bush, R. R., Luce, R. D., & Rose, R. M. Learning models for psychophysics. In R. C. Atkinson (Ed.), *Studies in mathematical psychology*, Vol. I. Stanford: Stanford Univer. Press, in press, 1963.

Clarke, F. R. Constant-ratio rule for confusion matrices in speech communication. *J. acoust. Soc. Amer.*, 1957, **29**, 715–720.

Clarke, F. R., Birdsall, T. G., & Tanner, W. P., Jr. Two types of R.O.C. curves and definitions of parameters. *J. acoust. Soc. Amer.*, 1959, **31**, 629–630.

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

Creelman, C. D. *Detection of signals of uncertain frequency.* T.M. 71, Electronic Defense Group, Univer. of Mich., Ann Arbor, 1959.

Creelman, C. D. *Detection of complex signals as a function of signal band width and duration.* T.R. 99, Electronic Defense Group, Univer. of Mich., Ann Arbor, 1960.

Edwards, W. The theory of decision making. *Psych. Bull.*, 1954, **51,** 380–417.

Edwards, W. Behavioral decision theory. In *Annual Reviews of Psychology*, Vol. 12. Palo Alto, Calif.: Annual Reviews, Inc. 1961. Pp. 473–498.

Egan, J. P., Schulman, A. I., & Greenberg, G. Z. Operating characteristics determined by binary decisions and by ratings. *J. acoust. Soc. Amer.*, 1959, **31,** 768–773.

Elliott, P. B. *Tables of d'.* T.R. 97, Electronics Defense Group, Univer. of Mich., Ann Arbor, 1959.

Garner, W. R., & McGill, W. J. Relation between uncertainty, variance, and correlation analysis. *Psychometrika*, 1956, **21,** 219–228.

Green, D. M. Detection of multiple component signals in noise. *J. acoust. Soc. Amer.*, 1958, **30,** 904–911.

Green, D. M. Psychoacoustics and detection theory. *J. acoust. Soc. Amer.*, 1960, **32,** 1189–1203.

Green, D. M., Birdsall, T. G., & Tanner, W. P., Jr. Signal detection as a function of signal intensity and duration. *J. acoust. Soc. Amer.*, 1957, **29,** 523–531.

Halsey, R. M., & Chapanis, A. On the number of absolutely identifiable spectra hues. *J. opt. Soc. Amer.*, 1954, **41,** 1057–1058.

Klemmer, E. T., & Frick, F. C. Assimilation of information from dot and matrix patterns. *J. exp. Psychol.*, 1953, **45,** 15–19.

Licklider, J. C. R. Three auditory theories. In S. Koch (Ed.), *Psychology: a study of a science*, Vol. 1. New York: McGraw-Hill, 1959. Pp. 41–144.

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

Luce, R. D. The theory of selective information and some of its behavioral applications. In R. D. Luce (Ed.), *Developments in mathematical psychology.* Glencoe, Ill.: The Free Press, 1960.

Luce, R. D. A threshold theory for simple detection experiments. *Psychol. Rev.*, in press, 1963.

McGill, W. J. Multivariate information transmission. *Psychometrika*, 1954, **19,** 97–116.

McGill, W. J. Isomorphism in statistical analysis. In H. Quastler (Ed.), *Information theory in psychology.* Glencoe, Ill.: The Free Press, 1955. Pp. 56–62.(a)

McGill, W. J. The relation between uncertainty and variance. *Proc.* 1954 *Conf. Test Probl. Educ. Test. Serv.*, 1955, 37–42.(b)

Miller, G. A. What is information measurement? *Amer. Psychol.*, 1953, **8,** 3–11.

Miller, G. A. The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychol. Rev.*, 1956, **63,** 81–97.

Neisser, U. R. G. *A methodological study of the quantal hypothesis in auditory psychophysics.* Unpublished doctoral dissertation, Harvard University, 1955.

Peterson, W. W., Birdsall, T. G., & Fox, W. C. The theory of signal detectability. *IRE Trans. Professional Group on Information Theory*, 1954, **4,** 171–212.

Pollack, I. Information of elementary auditory displays. *J. acoust. Soc. Amer.*, 1952, **24,** 745–750.

Pollack, I. The information of elementary auditory displays. II. *J. acoust. Soc. Amer.*, 1953, **25,** 765–769.

Pollack, I., & Ficks, L. Information of elementary multidimensional auditory displays. *J. acoust. Soc. Amer.*, 1954, **26,** 155–158.

Restle, F. *Psychology of judgment and choice.* New York: Wiley, 1961.

Senders, Virginia L. Further analysis of response sequences in the setting of a psychophysical experiment. *Amer. J. Psychol.*, 1953, **66,** 215–228.

Senders, Virginia L., & Sowards, A. Analysis of response sequences in the setting of a psychophysical experiment. *Amer. J. Psychol.*, 1952, **65,** 358–374.

Shannon, C. E., & Weaver, W. *The mathematical theory of communication.* Urbana: Univer. of Illinois Press, 1949.

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

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

Shepard, R. N. Stimulus and response generalization: tests of a model relating generalization to distance in psychological space. *J. exp. Psychol.*, 1958, **55,** 509–523.(b)

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

Shipley, Elizabeth F. *Detection and recognition with uncertainty.* Unpublished doctoral dissertation, Univer. of Pennsylvania, 1961.

Smith, M., & Wilson, Edna A. A model of the auditory threshold and its application to the problem of the multiple observer. *Psychol. Monogr.*, 1953, **67** (9, Whole No. 359).

Speeth, S. D., & Mathews, M. V. Sequential effects in the signal detection situation. *J. acoust. Soc. Amer.*, 1961, **33,** 1046–1054.

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

Suppes, P. A linear model for a continuum of responses. In R. R. Bush, & W. K. Estes (Eds.), *Studies in mathematical learning theory.* Stanford: Stanford Univer. Press, 1959. Pp. 400–414.

Suppes, P. Stimulus-sampling theory for a continuum of responses. In K. J. Arrow, S. Karlin, & P. Suppes (Eds.), *Mathematical methods in the social sciences*, 1959. Stanford: Stanford Univer. Press, 1960. Pp. 348–365.

Swets, J. A. Indices of signal detectability obtained with various psychophysical procedures. *J. acoust. Soc. Amer.*, 1959, **31,** 511–513.

Swets, J. A. Detection theory and psychophysics: a review. *Psychometrika*, 1961, **26,** 49–63.(a)

Swets, J. A. Is there a sensory threshold? *Science,* 1961, **134,** 168–177.(b)

Swets, J. A., & Birdsall, T. G. The human use of information III. Decision making in signal detection and recognition situations involving multiple alternatives. *IRE Trans. Professional Group on Inform. Theory*, **2,** 1956, 138–165.

Swets, J. A., Shipley, Elizabeth F., McKey, Molly J., & Green, D. M. Multiple observations of signals in noise. *J. acoust. Soc. Amer.*, 1959, **31,** 514–521.

Swets, J. A., Tanner, W. P., Jr., & Birdsall, T. G. *The evidence for a decision-making theory of visual detection.* T.R. 40, Electronic Defense Group, Univer. of Mich., Ann Arbor, 1955.

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

Tanner, W. P., Jr. On the design of psychophysical experiments. In H. Quastler (Ed.), *Information theory in psychology*, Glencoe, Ill.: The Free Press, 1955. Pp. 403–414.

Tanner, W. P., Jr. Theory of recognition. *J. acoust. Soc. Amer.*, 1956, **28,** 882–888.

Tanner, W. P., Jr. Theory of signal detectability as an interpretive tool for psychophysical data. *J. acoust. Soc. Amer.*, 1960, **32,** 1140–1147.

Tanner, W. P., Jr. Physiological implications of psychophysical data. *Ann. N.Y. Acad. Sci.*, 1961, **89,** 752–765.

Tanner, W. P., Jr., & Birdsall, T. G. Definitions of d' and η as psychophysical measures. *J. acoust. Soc. Amer.*, 1958, **30,** 922–928.

Tanner, W. P., Jr., Birdsall, T. G., & Clarke, F. R. *The concept of the ideal observer in* *psychophysics*. T.R. 98, Electronic Defense Group, Univer. of Mich., Ann Arbor, 1960.

Tanner, W. P., Jr., & Norman, R. Z. The human use of information. II. Signal detection for the case of an unknown signal parameter. *IRE Trans. Professional Group on Information Theory*, 1954, **4**, 222–226.

Tanner, W. P., Jr., & Swets, J. A. A decision making theory of visual detection. *Psychol. Rev.*, 1954, **61**, 401–409.(a)

Tanner, W. P., Jr., & Swets, J. A. The human use of information. I. Signal detection for the case of the signal known exactly. *IRE Trans. Professional Group on Information Theory*. 1954, **4**, 213–221.(b)

Tanner, W. P., Jr., Swets, J. A., & Green, D. M. *Some general properties of the hearing mechanism*. T.R. 30, Electronic Defense Group, Univer. of Mich., Ann Arbor, 1956.

Van Meter, D., & Middleton, D. Modern statistical approaches to reception in communication theory. *IRE Trans. Professional Group on Information Theory*, 1954, **4**, 119–145.

Veniar, Florence A. Effect of auditory cue on discrimination of auditory stimuli. *J. acoust. Soc. Amer.*, 1958, **30**, 1079–1081.(a)

Veniar, Florence A. Signal detection as a function of frequency ensemble. I. *J. acoust. Soc. Amer.*, 1958, **30**, 1020–1024.(b)

Veniar, Florence A. Signal detection as a function of frequency ensemble. II. *J. acoust. Soc. Amer.*, 1958, **30**, 1075–1078.(c)

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