Psychological Review
1961, Vol. 68, No. 5, 301–340

DECISION PROCESSES IN PERCEPTION[1]

JOHN A. SWETS
Massachusetts Institute of Technology
WILSON P. TANNER, JR., AND THEODORE G. BIRDSALL
University of Michigan

About 5 years ago, the theory of statistical decision was translated into a theory of signal detection.[2] Although the translation was motivated by problems in radar, the detection theory that resulted is a general theory for, like the decision theory, it specifies an ideal process. The generality of the theory suggested to us that it might also be relevant to the detection of signals by human observers. Beyond this, we were struck by several analogies between this description of ideal behavior and various aspects of the perceptual process. The detection theory seemed to provide a framework for a realistic description of the behavior of the human observer in a variety of perceptual tasks.

The particular feature of the theory that was of greatest interest to us was the promise that it held of solving an old problem in the field of psychophysics. This is the problem of controlling or specifying the criterion that the observer uses in making a perceptual judgment. The classical methods of psychophysics make effective provision for only a single free parameter, one that is associated with the sensitivity of the observer. They contain no analytical procedure for specifying independently the observer's criterion. These two aspects of performance are confounded, for example, in an experiment in which the dependent variable is the intensity of the stimulus that is required for a threshold response. The present theory provides a quantitative measure of the criterion. There is left, as a result, a relatively pure measure of sensitivity. The theory, therefore, promised to be of value to the student of personal and social processes in perception as well as to the student of sensory functions. A second feature of the theory that attracted us is that it is a normative theory. We believed that having a standard with which to compare the behavior of the human observer would aid in the description and in the interpretation of experimental results, and would be fruitful in suggesting new experiments.

This paper begins with a brief review of the theory of statistical decision and then presents a description of the elements of the theory of signal detection appropriate to human observers.

[1] This paper is based upon Technical Report No. 40, issued by the Electronic Defense Group of the University of Michigan in 1955. The research was conducted in the Vision Research Laboratory of the University of Michigan with support from the United States Army Signal Corps and the Naval Bureau of Ships. Our thanks are due H. R. Blackwell and W. M. Kincaid for their assistance in the research, and D. H. Howes for suggestions concerning the presentation of this material. This paper was prepared in the Research Laboratory of Electronics, Massachusetts Institute of Technology, with support from the Signal Corps, Air Force (Operational Applications Laboratory and Office of Scientific Research), and Office of Naval Research. This is Technical Report No. ESD-TR-61-20.

[2] For a formal treatment of statistical decision theory, see Wald (1950) ; for a brief and highly readable survey of the essentials, see Bross (1953). Parallel accounts of the detection theory may be found in Peterson, Birdsall, and Fox (1954) and in Van Meter and Middleton (1954).

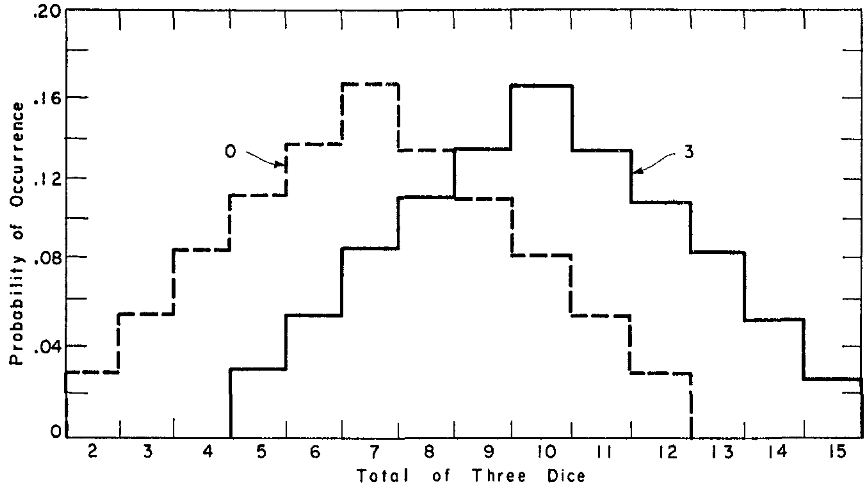

FIG. 1. The probability distributions for the dice game.

Following this, the results of some experimental tests of the applicability of the theory to the detection of visual signals are described.

The theory and some illustrative results of one experimental test of it were briefly described in an earlier paper (Tanner & Swets, 1954). The present paper contains a more nearly adequate description of the theory, a more complete account of the first experiment, and the results of four other experiments. It brings together all of the data collected to date in vision experiments that bear directly on the value of the theory.[3]

THE THEORY

Statistical Decision Theory

Consider the following game of chance. Three dice are thrown. Two of the dice are ordinary dice. The third die is unusual in that on each of three of its sides it has three spots, whereas on its remaining three sides it has no spots at all. You, as the player of the game, do not observe the throws of the dice. You are simply informed, after each throw, of the total number of spots showing on the three dice. You are then asked to state whether the third die, the unusual one, showed a 3 or a 0. If you are correct—that is, if you assert a 3 showed when it did in fact, or if you assert a 0 showed when it did in fact—you win a dollar. If you are incorrect—that is, if you make either of the two possible types of errors—you lose a dollar.

How do you play the game? Certainly you will want a few minutes to make some computations before you begin. You will want to know the probability of occurrence of each of the possible totals 2 through 12 in the event that the third die shows a 0, and you will want to know the probability of occurrence of each of the possible totals 5 through 15 in the event that the third die shows a 3. Let us ignore the exact values of these probabilities, and grant that the two probability distributions in question will look much like those sketched in Figure 1.

Realizing that you will play the game many times, you will want to establish

[3] Reports of several applications of the theory in audition experiments are available in the literature; for a list of references, see Tanner and Birdsall (1958).

a policy which defines the circumstances under which you will make each of the two decisions. We can think of this as a *criterion* or a cutoff point along the axis representing the total number of spots showing on the three dice. That is, you will want to choose a number on this axis such that whenever it is equaled or exceeded you will state that a 3 showed on the third die, and such that whenever the total number of spots showing is less than this number, you will state that a 0 showed on the third die. For the game as described, with the a priori probabilities of a 3 and a 0 equal, and with equal values and costs associated with the four possible decision outcomes, it is intuitively clear that the optimal cutoff point is that point where the two curves cross. You will maximize your winnings if you choose this point as the cutoff point and adhere to it.

Now, what if the game is changed? What, for example, if the third die has three spots on five of its sides, and a 0 on only one? Certainly you will now be more willing to state, following each throw, that the third die showed a 3. You will not, however, simply state more often that a 3 occurred without regard to the total showing on the three dice. Rather, you will lower your cutoff point: you will accept a smaller total than before as representing a throw in which the third die showed a 3. Conversely, if the third die has three spots on only one of its sides and 0's on five sides, you will do well to raise your cutoff point—to require a higher total than before for stating that a 3 occurred.

Similarly, your behavior will change if the values and costs associated with the various decision outcomes are changed. If it costs you 5 dollars every time you state that a 3 showed when in fact it did not, and if you win 5 dollars every time you state that a 0 showed when in fact it did (the other value and the other cost in the game remaining at one dollar), you will raise your cutoff to a point somewhere above the point where the two distributions cross. Or if, instead, the premium is placed on being correct when a 3 occurred, rather than when a 0 occurred as in the immediately preceding example, you will assume a cutoff somewhere below the point where the two distributions cross.

Again, your behavior will change if the amount of overlap of the two distributions is changed. You will assume a different cutoff than you did in the game as first described if the three sides of the third die showing spots now show four spots rather than three.

This game is simply an example of the type of situation for which the theory of statistical decision was developed. It is intended only to recall the frame of reference of this theory. Statistical decision theory—or the special case of it which is relevant here, the theory of testing statistical hypotheses—specifies the optimal behavior in a situation where one must choose between two alternative statistical hypotheses on the basis of an observed event. In particular, it specifies the optimal cutoff, along the continuum on which the observed events are arranged, as a function of (*a*) the a priori probabilities of the two hypotheses, (*b*) the values and costs associated with the various decision outcomes, and (*c*) the amount of overlap of the distributions that constitute the hypotheses.

According to the mathematical theory of signal detectability, the problem of detecting signals that are weak relative to the background of interference is like the one faced by the player of our dice game. In short, the detection problem is a problem in statistical decision; it requires testing statistical hy-

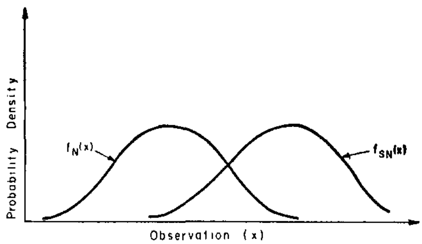

Fig. 2. The probability density functions of noise and signal plus noise.

potheses. In the theory of signal detectability, this analogy is developed in terms of an idealized observer. It is our thesis that this conception of the detection process may apply to the human observer as well. The next several pages present an analysis of the detection process that will make the bases for this reasoning apparent.[4]

Fundamental Detection Problem

In the fundamental detection problem, an observation is made of events occurring in a fixed interval of time, and a decision is made, based on this observation, whether the interval contained only the background interference or a signal as well. The interference, which is random, we shall refer to as *noise* and denote as *N;* the other alternative we shall term *signal plus noise, SN*. In the fundamental problem, only these two alternatives exist—noise is always present, whereas the signal may or may not be present during a specified observation interval. Actually, the observer, who has advance knowledge of the ensemble of signals to be presented, says either "yes, a signal was present" or "no, no signal was present" following each observation. In the experiments reported below, the signal consisted of a small spot of light flashed briefly in a known location on a uniformly illuminated background. It is important to note that the signal is always observed in a background of noise; some, as in the present case, may be introduced by the experimenter or by the external situation, but some is inherent in the sensory processes.

[4] It is to be expected that a theory recognized as having a potential application in psychophysics, although developed in another context, will be similar in many respects to previous conceptions in psychophysics. Although we shall not, in general, discuss explicitly these similarities, the strong relationship between many of the ideas presented in the following and Thurstone's earlier work on the scaling of judgments should be noted (see Thurstone, 1927a, 1927b). The present theory also has much in common with the recent work of Smith and Wilson (1953) and of Munson and Karlin (1956). Of course, for a new theory to arouse interest, it must also differ in some significant aspects from previous theories—these differences will become apparent as we proceed.

Representation of Sensory Information

We shall, in the following, use the term *observation* to refer to the sensory datum on which the decision is based. We assume that this observation may be represented as varying continuously along a single dimension. Although there is no need to be concrete, it may be helpful to think of the observation as some measure of neural activity, perhaps as the number of impulses arriving at a given point in the cortex within a given time. We assume further that any observation may arise, with specific probabilities, either from noise alone or from signal plus noise. We may portray these assumptions graphically, for a signal of a given amplitude, as in Figure 2. The observation is labeled x and plotted on the abscissa. The left-hand distribution, labeled $f_N(x)$, represents the probability density that x will result given the occurrence of noise alone. The right-hand distribution, $f_{SN}(x)$, is the probability density function of x given the occurrence of signal plus noise. (Probability density functions are used, rather than probability functions, since x is assumed to be continuous.) Since the observations will tend to be of greater magnitude when a signal is presented, the mean of the SN distribution will be greater than the mean of the N distribution. In general, the greater the amplitude of the signal, the greater will be the separation of these means.

Observation as a Value of Likelihood Ratio

It will be well to question at this point our assumption that the observation may be represented along a single axis. Can we, without serious violation, regard the observation as unidimensional, in spite of the fact that the response of the visual system probably has many dimensions? The answer to this question will involve some concepts that are basic to the theory.

One reasonable answer is that when the signal and interference are alike in character, only the magnitude of the total response of the receiving system is available as an indicator of signal existence. Consequently, no matter how complex the sensory information is in fact, the observations may be represented in theory as having a single dimension. Although this answer is quite acceptable when concerned only with the visual case, we prefer to advance a different answer, one that is applicable also to audition experiments, where, for example, the signal may be a segment of a sinusoid presented in a background of white noise.

So let us assume that the response of the sensory system does have several dimensions, and proceed to represent it as a point in an m-dimensional space. Call this point y. For every such point in this space there is some probability density that it resulted from noise alone, $f_N(y)$, and, similarly, some probability density that it was due to signal plus noise, $f_{SN}(y)$. Therefore, there exists a likelihood ratio for each point in the space, $\lambda(y) = f_{SN}(y)/f_N(y)$, expressing the likelihood that the point y arose from SN relative to the likelihood that it arose from N. Since any point in the space, *i.e.*, any sensory datum, may be thus represented as a real, nonzero number, these points may be considered to lie along a single axis. We may then, if we choose, identify the observation x with $\lambda(y)$; the decision axis becomes likelihood ratio.[5]

Having established that we may identify the observation x with $\lambda(y)$, let us note that we may equally well identify x with any monotonic transformation of $\lambda(y)$. It can be shown that we lose nothing by distorting the linear continuum as long as order is maintained. As a matter of fact we may gain if, in particular, we identify x with some transformation of $\lambda(y)$ that results in Gaussian density functions on x. We have assumed the existence of such a transformation in the representation of the density functions, $f_{SN}(x)$ and $f_N(x)$, in Figure 2. We shall see shortly that the assumption of normality simplifies the problem greatly. We shall also see that this assumption is subject to experimental test. A further assumption incorporated into the picture of Figure 2, one made quite tentatively, is that the two density functions are of equal variance. This is equivalent to the assumption that the SN function is a simple translation of the N function, or that adding a signal to the noise merely adds a constant to the N function. The re-

[5] Thus the assumption of a unidimensional decision axis is independent of the character of the signal and noise. Rather, it depends upon the fact that just two decision alternatives are considered. More generally, it can be shown that the number of dimensions required to represent the observation is $M-1$, where M is the number of decision alternatives considered by the observer.

sults of a test of this assumption are also described below.

To summarize the last few paragraphs, we have assumed that an observation may be characterized by a value of likelihood ratio, $\lambda(y)$, i.e., the likelihood that the response of the sensory system y arose from SN relative to the likelihood that it arose from N. This permits us to view the observations as lying along a single axis. We then assumed the existence of a particular transformation of $\lambda(y)$ such that on the resulting variable, x, the density functions are normal. We regard the observer as basing his decisions on the variable x.

Definition of the Criterion

If the representation depicted in Figure 2 is realistic, then the problem posed for an observer attempting to detect signals in noise is indeed similar to the one faced by the player of our dice game. On the basis of an observation, one that varies only in magnitude, he must decide between two alternative hypotheses. He must decide from which hypothesis the observation resulted; he must state that the observation is a member of the one distribution or the other. As did the player of the dice game, the observer must establish a policy which defines the circumstances under which the observation will be regarded as resulting from each of the two possible events. He establishes a criterion, a cutoff x_c on the continuum of observations, to which he can relate any given observation x_i. If he finds for the ith observation, x_i, that $x_i > x_c$, he says "yes"; if $x_i < x_c$, he says "no." Since the observer is assumed to be capable of locating a criterion at any point along the continuum of observations, it is of interest to examine the various factors that, according to the theory, will influence his choice of a particular criterion. To do so requires some additional notation.

In the language of statistical decision theory the observer chooses a subset of all of the observations, namely the Critical Region A, such that an observation in this subset leads him to accept the Hypothesis SN, to say that a signal was present. All other observations are in the complementary Subset B; these lead to rejection of the Hypothesis SN, or, equivalently, since the two hypotheses are mutually exclusive and exhaustive, to the acceptance of the Hypothesis N. The Critical Region A, with reference to Figure 2, consists of the values of x to the right of some criterion value x_c.

As in the case of the dice game, a decision will have one of four outcomes: the observer may say "yes" or "no" and may in either case be *correct* or *incorrect*. The decision outcome, in other words, may be a *hit* ($SN \cdot A$, the joint occurrence of the Hypothesis SN and an observation in the Region A), a *miss* ($SN \cdot B$), a *correct rejection* ($N \cdot B$), or a *false alarm* ($N \cdot A$). If the a priori probability of signal occurrence and the parameters of the distributions of Figure 2 are fixed, the choice of a criterion value x_c completely determines the probability of each of these outcomes.

Clearly, the four probabilities are interdependent. For example, an increase in the probability of a hit, $p(SN \cdot A)$, can be achieved only by accepting an increase in the probability of a false alarm, $p(N \cdot A)$, and decreases in the other probabilities, $p(SN \cdot B)$ and $p(N \cdot B)$. Thus a given criterion yields a particular balance among the probabilities of the four possible outcomes; conversely, the balance desired by an observer in any instance will determine the optimal location of his criterion. Now the observer may desire the balance that maximizes the expected value of a decision in a situation where the four possible outcomes of a decision have individual values, as did the player of the dice game. In this case, the location of the best criterion is determined by the same parameters that determined it in the dice game. The observer, however, may desire a balance that maximizes some other quantity—i.e., a balance that is optimum according to some other definition of optimum—in which case a different criterion will be appropriate. He may, for example, want to maximize $p(SN \cdot A)$ while satisfying a restriction on $p(N \cdot A)$, as we typically do when as experimenters we assume an .05 or .01 level of confidence. Alternatively, he may want to maximize the number of correct decisions. Again, he may prefer a criterion that will maximize the reduction in uncertainty in the Shannon (1948) sense.

In statistical decision theory, and in the theory of signal detectability, the optimal criterion under each of these definitions of optimum is specified in terms of the likelihood ratio. That is to say, it can be shown that, if we define the observation in terms of the likelihood ratio, $\lambda(x) = f_{SN}(x)/f_N(x)$, then the optimal criterion can always be specified by some value β of $\lambda(x)$. In other words, the Critical Region A that corresponds to the criterion contains all observations with likelihood ratio greater than or equal to β, and none of those with likelihood ratio less than β.

We shall illustrate this manner of specifying the optimal criterion for just one of the definitions of optimum proposed above, namely, the maximization of the total expected value of a decision in a situation where the four possible outcomes of a decision have individual values associated with them. This is the definition of optimum that we assumed in the dice game. For this purpose we shall need the concept of *conditional probability* as opposed to the *probability of joint occurrence* introduced above. It should be stated that conditional probabilities will have a place in our discussion beyond their use in this illustration; the ones we shall introduce are, as a matter of fact, the fundamental quantities in evaluating the observer's performance.

There are two conditional probabilities of principal interest. These are the conditional probabilities of the observer saying "yes": $p_{SN}(A)$, the probability of a Yes decision *conditional upon,* or *given,* the occurrence of a signal, and $p_N(A)$, the probability of a Yes decision given the occurrence of noise alone. These two are sufficient, for the other two are simply their complements: $p_{SN}(B) = 1 - p_{SN}(A)$ and $p_N(B) = 1 - p_N(A)$. The conditional and joint probabilities are related as follows:

$$p_{SN}(A) = \frac{p(SN \cdot A)}{p(SN)}$$
$$p_N(A) = \frac{p(N \cdot A)}{p(N)} \qquad [1]$$

where: $p(SN)$ is the a priori probability of signal occurrence and $p(N) = 1 - p(SN)$ is the a priori probability of occurrence of noise alone.

Equation 1 makes apparent the convenience of using conditional rather than joint probabilities—conditional probabilities are independent of the a priori probability of occurrence of the signal and of noise alone. With reference to Figure 2, we may define $p_{SN}(A)$, or the conditional probability of a hit, as the integral of $f_{SN}(x)$ over the Critical Region A, and $p_N(A)$, the conditional probability of a false alarm, as the integral of $f_N(x)$ over A. That is, $p_N(A)$ and $p_{SN}(A)$ represent, respectively, the areas under the two curves of Figure 2 to the right of some criterion value of x.

To pursue our illustration of how an optimal criterion may be specified by a critical value of likelihood ratio β, let us note that the expected value of a decision (denoted EV) is defined in statistical decision theory as the sum, over the potential outcomes of a decision, of the products of probability of outcome and the desirability of outcome. Thus, using the notation V for *positive* individual values and K for costs or *negative* individual values, we have the following equation:

$$EV = V_{SN \cdot A}p(SN \cdot A) + V_{N \cdot B}p(N \cdot B) - K_{SN \cdot B}p(SN \cdot B) - K_{N \cdot A}p(N \cdot A) \quad [2]$$

Now if a priori and conditional probabilities are substituted for the joint probabilities in Equation 2 following Equation 1, for example, $p(SN)p_{SN}(A)$ for $p(SN \cdot A)$, then collecting terms yields the result that maximizing EV is equivalent to maximizing:

$$p_{SN}(A) - \beta p_N(A) \quad [3]$$

where

$$\beta = \frac{p(N)}{p(SN)} \cdot \frac{(V_{N \cdot B} + K_{N \cdot A})}{(V_{SN \cdot A} + K_{SN \cdot B})} \quad [4]$$

It can be shown that this value of β is equal to the value of likelihood ratio, $\lambda(x)$, that corresponds to the optimal criterion. From Equation 3 it may be seen that the value β simply weights the hits and false alarms, and from Equation 4 we see that β is determined by the a priori probabilities of occurrence of signal and of noise alone and by the values associated with the individual decision outcomes. It should be noted that Equation 3 applies to all definitions of optimum. Equation 4 shows the determinants of β in only the special case of the expected-value definition of optimum.

Return for a moment to Figure 2, keeping in mind the result that β is a critical value of $\lambda(x) = f_{SN}(x)/f_N(x)$. It should be clear that the optimal cutoff x_c along the x axis is at the point on this axis where the ratio of the ordinate value of $f_{SN}(x)$ to the ordinate value of $f_N(x)$ is a certain number, namely β. In the symmetrical case, where the two a priori probabilities are equal and the four individual values are equal, $\beta = 1$ and the optimal value of x_c is the point where $f_{SN}(x) = f_N(x)$, where the two curves cross. If the four values are equal but $p(SN) = 5/6$ and $p(N) = 1/6$, another case described in connection with the dice game, then $\beta = 1/5$ and the optimal value of x_c is shifted a certain distance to the left. This shift may be seen intuitively to be in the proper direction —a higher value of $p(SN)$ should lead to a greater willingness to accept the Hypothesis SN, i.e., a more lenient cutoff. To consider one more example from the dice game, if $p(SN) = p(N) = 0.5$, if $V_{N \cdot B}$ and $K_{N \cdot A}$ are set at 5 dollars and $V_{SN \cdot A}$ and $K_{SN \cdot B}$ are equal to 1 dollar, then $\beta = 5$ and the optimal value of x_c shifts a certain distance to the right. Again intuitively, if it is more important to be correct when the Hypothesis N is true, a high, or strict, criterion should be adopted.

In any case, β specifies the optimal weighting of hits relative to false alarms: x_c should always be located at the point on the x axis corresponding to β. As we pointed out in discussing the dice game, just where this value of x_c will be with reference to the x axis depends not only upon the a priori probabilities and the values but also upon the overlap of the two density functions, in short, upon the signal strength. We shall define a measure of signal strength within the next few pages. For now, it is important to note that for any detection goal to which the observer may subscribe, and for any set of parameters that may characterize a detection situation (such as a priori probabilities and values associated with decision outcomes), the optimal criterion may be specified in terms of a single number, β, a critical value of likelihood ratio.[6]

Receiver-Operating-Characteristic

Whatever criterion the observer actually uses, even if it is not one of the optimal criteria, can also be described by a single number, by some value of likelihood ratio. Let us proceed to a consideration of how the observer's performance may be evaluated with respect to the location of his criterion, and, at the same time we shall see how his performance may be evaluated with respect to his sensory capabilities.

As we have noted, the fundamental quantities in the evaluation of performance are $p_N(A)$ and $p_{SN}(A)$, these quantities representing, respectively, the areas under the two curves of Figure 2 to the right of some criterion value of x. If we set up a graph of $p_{SN}(A)$ versus $p_N(A)$ and trace on it the curve resulting as we move the decision criterion along the decision axis of Figure 2, we sketch one of the arcs shown in Figure 3. Ignore, for a moment, all but one of these arcs. If the decision criterion is set way at the left in Figure 2, we obtain a point in the upper right-hand corner of Figure 3: both $p_{SN}(A)$ and $p_N(A)$ are unity. If the criterion is set at the right end of the decision axis in Figure 2, the point at the other extreme of Figure 3, $p_{SN}(A) = p_N(A) = 0$, is obtained. In between these extremes lie the criterion values of more practical interest. It should be noted that the exact form of the curve shown in Figure 3 is not the only form which might result, but it is the form which will result if the observer chooses a criterion in terms of likelihood ratio, and the probability density functions are normal and of equal variance.

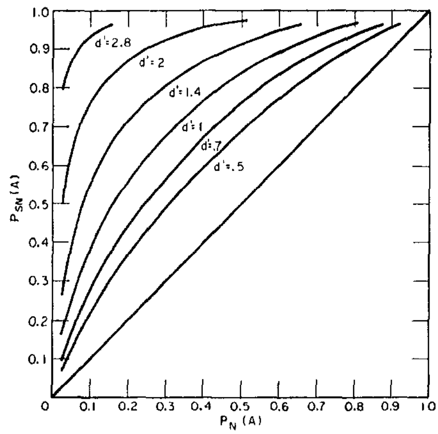

FIG. 3. The receiver-operating-characteristic curves. (These curves show $p_{SN}(A)$ vs. $p_N(A)$ with d' as the parameter. They are based on the assumptions that the probability density functions, $f_N(x)$ and $f_{SN}(x)$, are normal and of equal variance.)

This curve is a form of the *operating characteristic* as it is known in statistics; in the context of the detection problem it is usually referred to as the *receiver-operating-characteristic,* or ROC, curve. The optimal "operating level" may be seen from Equation 3 to

[6] We have reached a point in the discussion where we can justify the statement made earlier that the decision axis may be equally well regarded as likelihood ratio or as any monotonic transformation of likelihood ratio. Any distortion of the linear continuum of likelihood ratio, that maintains order, is equivalent to likelihood ratio in terms of determining a criterion. The decions made are the same whether the criterion is set at likelihood ratio equal to β or at the value that corresponds to β of some new variable. To illustrate, if a criterion leads to a Yes response whenever $\lambda(y) > 2$, if $x = [\lambda(y)]^2$ the decisions will be the same if the observer says "yes" whenever $x > 4$.

be at the point of the ROC curve where its slope is β. That is, the expression $p_{SN}(A) - \beta p_N(A)$ defines a utility line of slope β, and the point of tangency of this line to the ROC curve is the optimal operating level. Thus the theory specifies the appropriate hit probability and false alarm probability for any definition of optimum and any set of parameters characterizing the detection situation.

It is now apparent how the observer's choice of a criterion in a given experiment may be indexed. The proportions obtained in an experiment are used as estimates of the probabilities, $p_N(A)$ and $p_{SN}(A)$; thus, the observer's behavior yields a point on an ROC curve. The slope of the curve at this point corresponds to the value of likelihood ratio at which he has located his criterion. Thus we work backward from the ROC curve to infer the criterion that is employed by the observer.

There is, of course, a family of ROC curves, as shown in Figure 3, a given curve corresponding to a given separation between the means of the density functions $f_N(x)$ and $f_{SN}(x)$. The parameter of these curves has been called d', where d' is defined as the difference between the means of the two density functions expressed in terms of their standard deviation, i.e.:

$$d' = \frac{M_{f_{SN}(x)} - M_{f_N(x)}}{\sigma_{f_N(x)}} \qquad [5]$$

Since the separation between the means of the two density functions is a function of signal amplitude, d' is an index of the detectability of a given signal for a given observer.

Recalling our assumptions that the density functions $f_N(x)$ and $f_{SN}(x)$ are normal and of equal variance, we may see from Equation 5 that the quantity denoted d' is simply the familiar normal deviate, or x/σ measure. From the pair of values $p_N(A)$ and $p_{SN}(A)$ that are obtained experimentally, one may proceed to a published table of areas under the normal curve to determine a value of d'. A simpler computational procedure is achieved by plotting the points $[p_N(A), p_{SN}(A)]$ on graph paper having a probability scale and a normal deviate scale on both axes.

We see now that the four-fold table of the responses that are made to a particular stimulus may be treated as having two independent parameters—the experiment yields measures of two independent aspects of the observer's performance. The variable d' is a measure of the observer's sensory capabilities, or of the effective signal strength. This may be thought of as the object of interest in classical psychophysics. The criterion β that is employed by the observer, which determines the $p_N(A)$ and $p_{SN}(A)$ for some fixed d', reflects the effect of variables which have been variously called the set, attitude, or motives of the observer. It is the ability to distinguish between these two aspects of detection performance that comprises one of the main advantages of the theory proposed here. We have noted that these two aspects of behavior are confounded in an experiment in which the dependent variable is the intensity of the signal that is required for a threshold response.

Relationship of d′ to Signal Energy

We have seen that the optimal value of the criterion, β, can be computed. In certain instances, an optimal value of d', i.e., the sensitivity of the mathematically ideal device, can also be computed. If, for example, the exact wave form and starting time of the signal are determinable, as in the case of an auditory signal, then the optimal value of d' is equal to $\sqrt{2E/N_o}$, where E is the signal energy and N_o is the noise power in a one-cycle band (Peterson, Birdsall, & Fox, 1954). A specification of the optimal value of d' for visual signals has been developed very recently.[7] Although we shall not elaborate the point in this paper, it is worth noting that an empirical index of detectability may be compared with ideal detectability, just as observed and optimal indices of decision criteria may be compared. The ratio of the squares of the two detectability indices has been taken as a measure of the observer's sensory efficiency. This measure has demonstrated its usefulness in the study of several problems in audition (Tanner & Birdsall, 1958).

Use of Ideal Descriptions as Models

It might be worthwhile to describe at this point some of the reasons for the emphasis placed here on optimal measures, and, indeed, the reasons for the general enterprise of considering a theory of ideal behavior as a model for studies of real behavior.[8] In view of the deviations from any ideal which are bound to characterize real organisms, it might appear at first glance that any deductions based on ideal premises could have no more than academic interest. We do not think this is the case. In any study, it is desirable to specify rigorously the factors pertinent to the study. Ideal conditions generally involve few variables and permit these to be described in simple terms. Having identified the performance to be expected under ideal conditions, it is possible to extend the model to include the additional variables associated with real organisms. The ideal performance, in other words, constitutes a convenient base from which to explore the complex operation of a real organism.

In certain cases, as in the problem at hand, values characteristic of ideal conditions may actually approximate very closely those characteristics of the organism under study. The problem then becomes one of changing the ideal model in some particular so that it is slightly less than ideal. This is usually accomplished by depriving the ideal device of some particular function. This method of attack has been found to generate useful hypotheses for further studies. Thus, whereas it is not expected that the human observer and the ideal detection device will behave identically, the emphasis in early studies is on similarities. If the differences are small, one may rule out entire classes of alternative models, and regard the model in question as a useful tool in further studies. Proceeding on this assumption, one may then in later studies emphasize the differences, the form and extent of the differences suggesting how the ideal model may be modified in the direction of reality.

Alternative Conceptions of the Detection Process

The earliest studies that were undertaken to test the applicability of the decision model to human observers were quite naturally oriented toward determining its value relative to existing psychophysical theory. As a result, some of the data presented below are meaningful only with respect to differences in the predictions based upon different theories. We shall, therefore, briefly consider alternative theories of the detection process.

Although it is difficult to specify with precision the alternative theories of detection, it is clear that they generally involve the concept of the *threshold* in an important way. The development of the threshold concept is fairly obscure. It is differently conceived by

[7] W. P. Tanner, Jr. & R. C. Jones, personal communication, November 1959.

[8] The discussion immediately following is, in part, a paraphrase of one in Horton (1957).

different people, and few popular usages of the concept benefit from explicit statement. One respect, however, in which the meaning of the threshold concept is entirely clear is its assertion of a lower limit on sensitivity. As we have just seen, the decision model does not include such a boundary. The decision model specifies no lower bound on the location of the criterion along the continuous axis of sensory inputs. Further, it implies that any displacement of the mean of $f_{SN}(x)$ from the mean of $f_N(x)$, no matter how small, will result in a greater value of $p_{SN}(A)$ than $p_N(A)$, irrespective of the location of the criterion.

To permit experimental comparison of decision theory and threshold theory, we shall consider a special version of threshold theory (Blackwell, 1953). Although it is a special version, we believe it retains the essence of the threshold concept. In this version, the threshold is described in the same terms that are used in the description of decision theory. It is regarded as a cutoff on the continuum of observations (see Figure 2) with a fixed location, with values of x above the cutoff always evoking a positive response, and with discrimination impossible among values of x below the cutoff. This description of a threshold in terms of a fixed cutoff and a stimulus effect that varies randomly, it will be noted, is entirely equivalent to the more common description in terms of a randomly varying cutoff and a fixed stimulus effect. There are several reasons for assuming that the hypothetical threshold cutoff is located quite high relative to the density function $f_N(x)$, say at approximately $+3\sigma$ from the mean of $f_N(x)$. We shall compare our data with the predictions of such a "high threshold" theory, and shall indicate their relationship to predictions from a theory assuming a lower threshold. We shall, in particular, ask how low a threshold cutoff would have to be to be consistent with the reported data. It may be noted that if a high threshold exists, the observer will be incapable of ordering values of x likely to result from noise alone, and hence will be incapable of varying his criterion over a significant range.

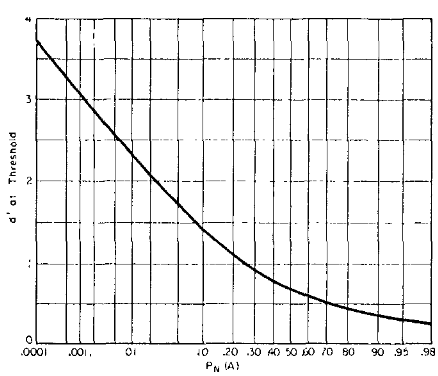

FIG. 4. The relationship between d' and $p_N(A)$ at threshold.

If a threshold exists that is rarely exceeded by noise alone, this fact will be immediately apparent from the ROC curves (see Figure 3) that are obtained experimentally. It can be shown that the ROC curves in this case are straight lines from points on the left-hand vertical axis—$p_{SN}(A)$—to the upper right-hand corner of the plot. These straight line curves represent the implication of a high threshold theory that an increase in $p_N(A)$ must be effected by responding "yes" to a random selection of observations that fail to reach the threshold, rather than by a judicious selection of observations, i.e., a lower criterion level. If we follow the usual procedure of regarding the stimulus threshold as the signal intensity yielding a value of $p_{SN}(A) = 0.5$ for $p_N(A) = 0.0$, then an appreciation of the relationship between d' and $p_N(A)$ at threshold may be gained by visualizing a straight line in Figure 3 from this point to the upper right-hand corner. If we note which of the ROC curves drawn in Figure 3 are intersected by the visualized line, we see that the threshold decreases with increasing $p_N(A)$. For example, a response procedure resulting in a $p_N(A) = 0.02$ requires a signal of $d' = 2.0$ to reach the threshold, whereas a response procedure yielding a $p_N(A) = 0.98$ requires a signal of $d' < 0.5$ to reach the threshold. A graph showing what threshold would be calculated as a function of $p_N(A)$ is plotted in Figure 4. The calculated threshold is a strictly monotonic function of $p_N(A)$ ranging from infinity to zero.

The fundamental difference between the threshold theory we are considering and decision theory lies in their treatment of false alarm responses. According to the threshold theory, these responses represent guesses determined by nonsensory factors; i.e, $p_N(A)$ is independent of the cutoff which is assumed to have a fixed location. Decision theory assumes, on the other hand, that $p_N(A)$ varies with the temporary position of a cutoff under the observer's control; that false alarm responses arise for valid sensory reasons, and that therefore a simple correction will not eliminate their effect on $p_{SN}(A)$. A similar implication of Figure 4 that should be noted is that reliable estimates of $p_{SN}(A)$ or of the stimulus threshold are not guaranteed by simply training the observer to maintain a low, constant value of $p_N(A)$. Since extreme probabilities cannot be estimated with reliability, the criterion may vary from session to session with the variation having no direct reflection in the data. Certainly, false alarm rates of 0.01, 0.001, and 0.0001, are not discriminable in an experimentally feasible number of observations; the differences in the calculated values of the threshold associated with these different values of $p_N(A)$ may be seen from Figure 4 to be sizeable. The experiments reported in the following were designed, in large measure, to clarify the relationship that exists between $p_N(A)$ and $p_{SN}(A)$, to show whether or not the observer is capable of controlling the location of his criterion for a Yes response.

SOME EXPERIMENTS

Five experiments are reported in the following. They are the first experiments that were undertaken to test the applicability of decision theory to psychophysical tasks, and it must be emphasized that they were intended to explore only the general relationships specified in the theory. We shall refer also to more recent experiments conducted within the framework of decision theory. The later experiments, although not focused as directly on testing the validity of the theory, support the principal thesis of this paper.

The experiments reported here are devoted to answering the two principal questions suggested by a consideration of decision theory. The first of these may be stated in this way: is sensory information (or the decision axis) continuous, i.e., is the observer capable of discriminating among observations likely to result from noise alone? The alternative we consider is that there exists a threshold cut, on the decision axis, that is unlikely to be exceeded by observations resulting from noise, and below which discrimination among observations is impossible. The second question has two parts: is the observer capable of using different criteria, and, if so, does he change his criterion appropriately when the variables that we expect will determine his criterion (probabilities, values, and costs) are changed?

Three of the five experiments to be described pose for the observer what we have called the fundamental detection problem, the problem that occupied our attention throughout the theoretical discussion. Of these, two test the observer's ability to use the criterion that maximizes the expected value of a decision. The a priori probability of a signal occurrence and the individual values associated with the four possible decision outcomes are varied systematically, in order to determine the range over which the observer can vary his criterion and the form of the resultant ROC curve. A third experiment tests the observer's ability to maximize the proportion of hits while satisfying a restriction on the proportion of false alarms. This experiment is largely concerned with the degree of precision with which the observer can locate a criterion.

The remaining two experiments differ in that the tasks they present to the observer do not require him to establish a criterion, that is, they do not require a Yes or No response. They test certain implications of decision theory that we have not yet treated explicitly, but they will be seen to follow very directly from the theory and to contribute significantly to an evaluation of it. In one of these the observer is asked to report after each observation interval his subjective probability that the signal existed during the interval. This response is a familiar one; it is essentially a rating or a judgment of confidence. The report of "a posteriori probability of signal existence," as it is termed in detection theory, may be regarded as reflecting the likelihood ratio of the observation. This case is of interest since an estimate of likelihood ratio preserves more of the information contained in the observation than does a report merely that the likelihood ratio fell above or below a critical value. We shall see that it is also possible to construct the ROC curve from this type of response.

The other experiment not requiring a criterion employs what has been termed the temporal forced-choice method of response. On each trial a signal is presented in exactly one of n temporal intervals, and the observer states in which interval he believes the signal occurred. The optimal procedure for the observer to follow in this case, if he is to maximize the probability of a correct response, is to make an observation x in each interval and to choose the interval having the greatest value of x associated with it. Since decision theory specifies how the proportion of correct responses obtained with the forced-choice method is related to the detectability index d', the internal consistency of the theory may be evaluated. That is to say, if the observer follows the optimal procedure, then the estimate of the detectability of a signal of a given strength that is based on forced-choice data will be comparable to that based on yes-no data. The forced-choice method may also be used to make a strong test of a fundamental assumption of decision theory, namely, that sensory information is continuous, or that sensory information does not exhibit a threshold cutoff. For an experiment requiring the observer to rank the n intervals according to their likelihood of containing the signal, the continuity and threshold assumptions lead to very different predictions concerning the probability that an interval ranked other than first will be the correct interval.

All of the experiments reported in the following employed a circular signal with a diameter of 30 minutes of visual angle and a duration of $\frac{1}{100}$ of a second. The signal was presented on a large uniformly illuminated background having a luminance of 10 footlamberts. Details of the apparatus have been presented elsewhere (Blackwell, Pritchard, & Ohmart, 1954).

Maximizing the Expected Value of a Decision—An Experimental Analysis

A direct test of the decision model is achieved in an experiment in which the a priori probability of signal occurrence or the values of the decision outcomes, or both, are varied from one group of observations to another—in short, in which β (Equations 3 and 4) assumes different values. The observer, in order to maximize his expected value, or his payoff, must vary his willingness to make a Yes response, in accordance with the change in β. Variations in this respect will be indicated by the proportion of false alarms, $p_N(A)$. The point of interest is how $p_{SN}(A)$, the proportion of hits, varies with changes in $p_N(A)$, i.e., in the form of the observer's ROC curve. If the experimental values of $p_N(A)$ reflect the location of the observer's criterion, if the observer responds on the basis of the likelihood ratio of the observation, and if the density functions (Figure 2) are normal and of equal variance, the ROC curve of Figure 3 will result. If, on the other hand, the location of the criterion is fixed in such a position that it is rarely exceeded by noise alone, then the resulting ROC curve will be a straight line, as we have indicated above. We shall examine some empirical ROC curves with this distinction in mind.

This experiment can be made to yield another and, in one sense, a stronger test of these two hypotheses, by employing several values of signal strength within a single group of observations, i.e., while a given set of probabilities and values are in effect. For in this case stimulus thresholds can be calculated, and correlational techniques can be used to determine whether the calculated threshold is dependent upon $p_N(A)$ as predicted by decision theory, or independent of $p_N(A)$ as predicted by what we have termed the high threshold theory. We will grant that presenting more than one value of signal strength, within a single group of observations to which fixed probabilities and values apply, is not, conceptually, the simplest experiment that could have been performed to test our hypotheses. Nevertheless, a little reflection will show that this experimental procedure is entirely legitimate from any of our present points of view. We simply associate several values of $p_{SN}(A)$ with a given value of $p_N(A)$, and thereby obtain at once a point on each of several ROC curves and an estimate of the stimulus threshold that is associated with that value of $p_N(A)$.

First Expected-Value Experiment. The first of the two expected-value experiments that were performed employed four values of signal strength.

Three observers, after considerable practice, served in 16 2-hour sessions. In each session, signals at four levels of intensity (0.44, 0.69, 0.92, and 1.20 foot-lamberts) were presented along with a "blank" or "no-signal" presentation. The order of presentation was random within a restriction placed upon the total number of occurrences of each signal intensity and the blank in a given session. Each of the signal intensities occurred equally often within a session. The proportion of trials on which a signal (of any intensity) was presented, $p(SN)$, was either 0.80 or 0.40 in the various sessions. In all, there were 300 presentations in each session—six blocks of 50 presentations, separated by rest periods. Thus each estimate of $p_N(A)$ is based on either 60 or 180 observations, and each estimate of $p_{SN}(A)$ is based on 30 or 60 observations, depending upon $p(SN)$.

In the first four sessions, no values were associated with the various decision outcomes. For the first and fourth sessions the observers were informed that $p(SN) = 0.80$ and, for the second and third sessions, that $p(SN) = 0.40$. The average value of $p_N(A)$ obtained in the sessions with $p(SN) = 0.80$ was 0.43, and, in the sessions with $p(SN) = 0.40$, it was 0.15—indicating that the observer's willingness to make a Yes response

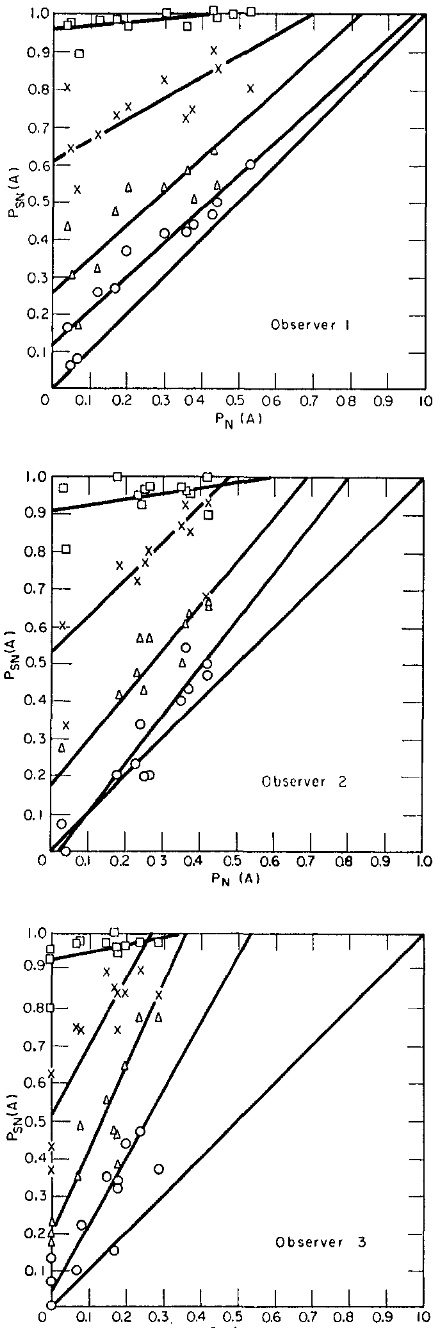

Fig. 5. Empirical receiver-operating-characteristic curves obtained from three observers in the first expected-value experiment.

is significantly affected by changes in $p(SN)$ alone. In the remaining 12 sessions, these two values of $p(SN)$ were used in conjunction with a variety of values placed on the decision outcomes. In the fifth session, for example, the observers were told that $p(SN) = 0.80$ and were, in addition, given the following payoff matrix:

	No	Yes
Signal	-1 $K_{SN \cdot B}$	$+1$ $V_{SN \cdot A}$
No Signal	$+2$ $V_{N \cdot B}$	-2 $K_{N \cdot A}$

A variety of simple matrices was used. These included, reading from left to right across the top and then the bottom row: $(-1, +1, +3, -3)$ and $(-1, +1, +4, -4)$ with $p(SN) = 0.80$, and $(-1, +1, +2, -2)$, $(-1, +1, +1, -1)$, $(-2, +2, +1, -1)$, and $(-3, +3, +1, -1)$ with $p(SN) = 0.40$. By reference to Equation 4, it may be seen that these matrices and values of $p(SN)$ define values of β ranging from 0.25 to 3.00. The observers were actually paid in accordance with these payoff matrices, in addition to their regular wage. The values were equated with fractions of cents, these fractions being adjusted so that the expected earnings per session remained relatively constant, at approximately one dollar.

The obtained values of $p_N(A)$ varied in accordance with changes in the values of the decision outcomes as well as with changes in the a priori probability of signal occurrence. Just how closely the obtained values of $p_N(A)$ approached those specified as optimal by the theory, we shall discuss shortly. For now, we may note that the range of values of $p_N(A)$ obtained from the three observers is shown in Figure 5. The parts of this figure also show four values of $p_{SN}(A)$ corresponding to each value of $p_N(A)$; the four values of $p_{SN}(A)$, one for each signal strength, are indicated by different symbols. We have, then, in the parts of Figure 5, four ROC curves.

Although entire ROC curves are not precisely defined by the data of the first experiment, these data will contribute to our purpose of distinguishing between the predictions of decision theory and the predictions of a high threshold theory. It is clear, for example, that the straight lines fitted to the data do not intersect the upper right-hand corner of the graph, as required by the concept of a high threshold.

We have mentioned that another analysis of the data is of interest in distinguishing the two theories we are considering. As we have indicated earlier in this paper, and developed in more detail elsewhere (Tanner & Swets, 1954), the concept of a high threshold leads to the prediction that the stimulus threshold is independent of $p_N(A)$, whereas decision theory predicts a negative correlation between the stimulus threshold and $p_N(A)$. Within the framework of the high threshold model that we have described, the stimulus threshold is defined as the stimulus intensity that yields a $p_{SN}(A) = 0.50$ for $p_N(A) = 0.0$. This stimulus intensity may be determined by interpolation from psychometric functions—$p_{SN}(A)$ vs. signal intensity—that are normalized so that $p_N(A) = 0.0$. The normalization is effected by the equation:

$$p_{SN}(A)_{\text{corrected}} = \frac{p_{SN}(A) - p_N(A)}{1 - p_N(A)} \qquad [6]$$

commonly known as the "correction for chance success." The intent of the correction is to remove what has been regarded as the spurious element of $p_{SN}(A)$ that is contributed by an observer's tendency to make a Yes response in the absence of any sensory indication of a signal, i.e., to make a Yes response following an observation that fails to reach the threshold level. It can be shown that the validity of this correction procedure is implied by the assumption of what we have termed a high threshold. The decision model, as we have indicated, differs in that it regards sensory information as thoroughly probabilistic, without a fixed cutoff—it asserts that the presence and absence of some sensory indication of a signal are not separable categories. According to the decision model, the observer does not achieve more Yes responses by responding positively to a random selection of observations that fall short of the fixed criterion level, but by lowering his criterion. In this case, the chance correction is inappropriate; the stimulus threshold will not remain invariant with changes in $p_N(A)$.

The relationship of the stimulus threshold to $p_N(A)$ in this first experiment is illustrated by Figures 6 and 7. The portion of data comprising each of the curves in these figures was selected to be relatively homogeneous with respect to $p_N(A)$. The curves are average curves for the three observers.

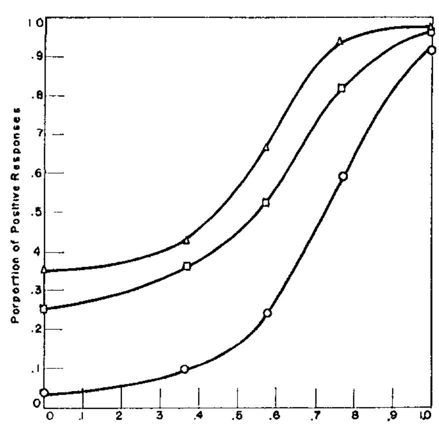

FIG. 6. The relationship between the stimulus threshold and $p_N(A)$ with the proportion of positive responses to four positive values of signal intensity, $p_{SN}(A)$, and to the blank or zero-intensity presentation, $p_N(A)$, at three values of $p_N(A)$.

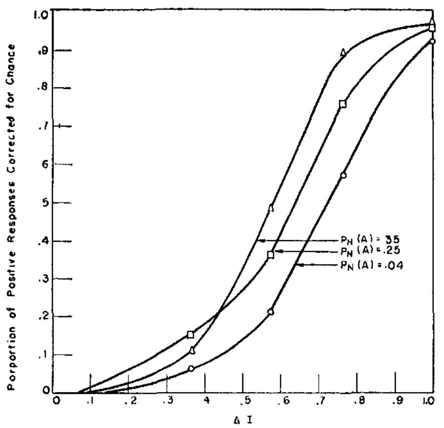

FIG. 7. The relationship between the stimulus threshold and $p_N(A)$ with the three curves corrected for chance success, by Equation 6.

Figure 6 shows $p_N(A)$ and $p_{SN}(A)$ as a function of the signal intensity, ΔI. The intercepts of the three curves may be seen to indicate values of $p_N(A)$ of 0.35, 0.25, and 0.04, respectively. Figure 7 shows the *corrected* value of $p_{SN}(A)$ plotted against signal intensity. It may be seen in Figure 7 that the stimulus threshold—the value of ΔI corresponding to a corrected $p_{SN}(A)$ of 0.50—is dependent upon $p_N(A)$ in the direction predicted by decision theory.[9]

Figures 6 and 7 portray the relationship in question in a form to which many of us are accustomed; they are presented here only for illustrative purposes. We can, of course, achieve a stronger test by computing the coefficients of correlation between $p_N(A)$ and the calculated threshold. We have made this computation, and have in the process avoided the averaging of data obtained from different observers and different experimental sessions. The product-moment coefficients for the three observers are $-.37$ ($p = 0.245$), $-.60$ ($p = 0.039$), and $-.81$ ($p = 0.001$), respectively. For the three observers combined, $p = 0.0008$. The implication of these correlations is the same as that of the straight lines fitted to the data of Figure 5, namely, that a dependence exists between the conditional probability that an observation arising from *SN* will exceed the criterion and the conditional probability that an observation arising from *N* will exceed the criterion. Stated otherwise, the correlations indicate that the observer's decision function is likelihood ratio or some monotonic function of it and that he is capable of adopting different criteria.

[9] ΔI is plotted in Figures 6 and 7 in terms of the transmission values of the filters that were placed selectively in the signal beam to yield different signal intensities. These values (0.365, 0.575, 0.765, 1.000) are converted to the signal values in terms of foot-lamberts that we have presented above, by multiplying them by 1.20, the value of the signal in foot-lamberts without selective filtering.

Second Expected-Value Experiment. A second expected-value experiment was conducted to obtain a more precise definition of the ROC curve than that provided by the experiment just described. In the second experiment greater definition was achieved by increasing the number of observations on which the estimates of $p_{SN}(A)$ and $p_N(A)$ were based, and by increasing the range of values of $p_N(A)$.

In this experiment only one signal intensity (0.78 foot-lamberts) was employed. Each of 13 experimental sessions included 200 presentations of the signal, and 200 presentations of noise alone. Thus, $p(SN)$ remained constant at 0.50 throughout this experiment. Changes in the optimal criterion β, and thus in the obtained values of $p_N(A)$, were effected entirely by changes in the values associated with the decision outcomes. These values were manipulated to yield β's (Equation 4) varying from 0.16 to 8.00. A different set of observers served in this experiment.

The results are portrayed in Figure 8. It may be seen that the experimentally determined points are fitted quite well by the type of ROC curve that is predicted by decision theory. It is equally apparent, excepting Observer 1, that the points do not lie along a straight line intersecting the point $p_N(A) = p_{SN}(A) = 1.00$, as predicted by the high threshold model.

One other feature of these figures is worthy of note. It will be recalled that in our presentation of decision theory we tentatively assumed that the density functions of noise and of signal plus noise, $f_N(x)$ and $f_{SN}(x)$, are of equal variance. Although we did not, in order to preserve the continuity of the discussion, we might have acknowledged at that point that the assumption of equal variance is not necessarily the best one. In particular, one might rather expect the variance of $f_{SN}(x)$ to be proportional to its mean. At any rate, the assumption made about variances represents a degree of freedom of the theory that we have not emphasized previously. We have, however, used this degree of freedom in the construction of the theoretical ROC curves of Figure 8. Notice that these curves are not symmetrical about the diagonal, as are the curves of Figure 3 that are predicated on equal variance. The

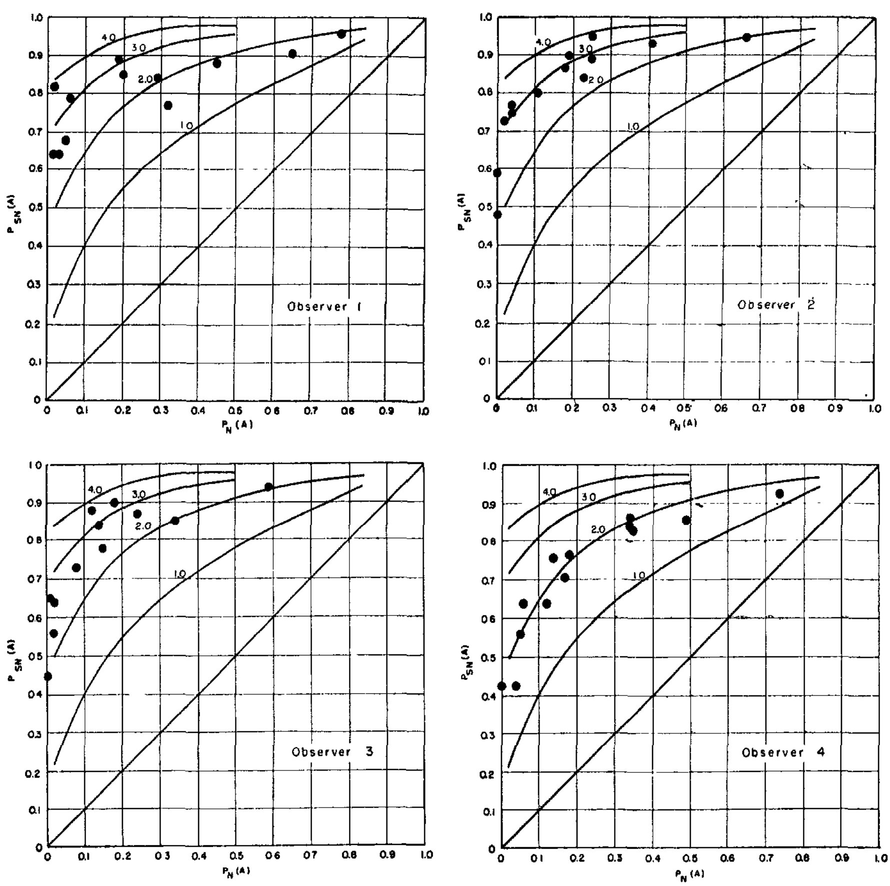

FIG. 8. Empirical receiver-operating-characteristic curves for four observers in the second expected-value experiment.

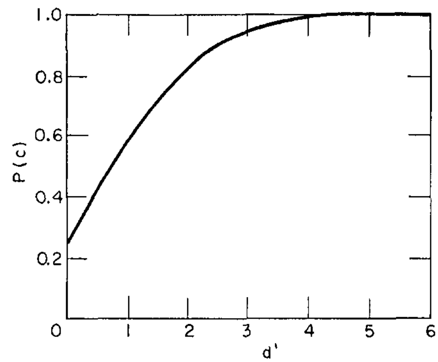

Fig. 9. The probability of a correct choice in a four-alternative forced-choice experiment as a function of d'.

curves of Figure 8 are based on the assumption that the ratio of the increment of the mean of $f_{SN}(x)$ to the increment of its standard deviation is equal to 4, $\Delta M/\Delta\sigma = 4$. A close look at these figures suggests that ROC curves calculated from a still greater ratio would provide a still better fit. Since other data presented in the following bear directly on this question of a dependence between variance and signal strength, we shall postpone further discussion of it. We shall also consider later whether the exact form of the empirical ROC curves supports the assumption of normality of the density functions $f_N(x)$ and $f_{SN}(x)$. For now, the main point is that decision theory predicts the curvilinear form of the ROC curves that are yielded by the observers.

Forced-Choice Experiments

We have indicated above that an extension of the decision model may be made to predict performance in a forced-choice test. On each trial of a typical forced-choice test, the signal is presented in one of n temporal intervals, and the observer selects the interval he believes to have contained the signal. It will be intuitively clear that, to behave optimally, in the sense of maximizing the probability of a correct response, the observer must make an observation x in each interval, and choose the interval having the greatest value of x associated with it. Equivalently, he may rank the intervals according to their values of likelihood ratio and choose that interval yielding the greatest value of likelihood ratio.

If the observer behaves optimally, then the probability that a correct answer will result, $p(c)$, for a given value of d', is expressed by:

$$p(c) = \int_{-\infty}^{+\infty} [f(x)]^{n-1} g(x) dx \quad [7]$$

where: $f(x)$ is the area of the noise function to the left of x, $g(x)$ is the ordinate of the signal-plus-noise function, and n is the number of intervals used in the test.

This is simply the probability that one drawing from the distribution due to signal plus noise is greater than the greatest of n-1 drawings from the distribution due to noise alone.

It is intuitively clear that if the signal produces a large shift in the noise function, i.e., if d' is large, then the probability that the greatest value of x will be obtained in the interval that contains the signal is also large, and conversely—indeed (for a fixed number of intervals) $p(c)$ is a monotonic function of d'. Equation 7 can be seen to be a function of d' by noting that, under the assumption of equal variance, the signal-plus-noise function is simply the noise function shifted by d', i.e., $g(x) = f(x - d')$. Thus d' may be defined in a forced-choice experiment by determining a value of $p(c)$ for some signal intensity and then using Equation 7 to determine d'. A plot of $p(c)$ versus d', for the case of four intervals, and under the assumption of equal variance, is shown in Figure 9.

Estimates of Signal Detectability Obtained from Different Procedures. According to detection theory, the estimates of d' for a signal and background of given intensities should be the same irrespective of the psychophysical procedure used to collect the data. Thus we may check the internal consistency of the theory by comparing estimates of d' based on yes-no and on forced-choice data. The results of such a comparison have been reported in another paper (Tanner & Swets, 1954). It was shown there that estimates of d' based on the data of the first expected-value experiment that we have presented above, and on forced-choice tests conducted in conjunction with it, are highly consistent with each other. Comparable estimates of d' have also been obtained in auditory experiments—from yes-no and forced-choice procedures, and from forced-choice procedures with from two to eight alternatives (Swets, 1959). Hence, decision theory provides a unification of the data obtained with different procedures; it enables one to predict the performance in one situation from data collected in another.

It is a commonplace that calculated values of the stimulus threshold are not independent of the psychophysical procedure that is employed (Osgood, 1953). Of particular relevance to our present concern is the finding that thresholds obtained with the forced-choice procedure are lower than those obtained with the yes-no procedure (Blackwell, 1953). This finding is accounted for, in terms of decision theory, by the fact that the calculated threshold varies monotonically with the false alarm rate (see Figure 4)—with high thresholds corresponding to low false alarm rates such as were obtained in these experiments. The dependence of the stimulus threshold upon the false alarm rate, however the threshold is calculated, precludes the existence of a simple relationship between thresholds obtained with the yes-no procedure and those obtained with other response procedures. It is also the case that the normalization of the psychometric function provided by the correction for chance, or the normalization achieved by defining the threshold as the stimulus intensity yielding a proportion of correct responses halfway between chance performance and perfect performance, does not serve to relate forced-choice thresholds obtained with different numbers of alternatives.

Theoretical and Experimental Analysis of Second Choices. As we have indicated, a variation of the forced-choice procedure—in which the observer indicates his second choice as well as his first—provides a powerful test of a basic difference between the decision model and the high threshold model. If the observer is capable of discriminating among values of the observations x that fail to reach what we have termed the threshold, i.e., a criterion fixed at approximately $+3\sigma$ from the mean of the noise function, then the proportion of second choices that are correct will be considerably higher than if he is not.[10]

According to the high threshold model, only very infrequently will more than one of the n observations of a forced-choice trial exceed the threshold. Since the observations which do not exceed the threshold are assumed by the model to be indiscriminable, the second choice will be made among the $n - 1$ alternatives on a chance basis.

[10] This experiment was suggested to us by R. Z. Norman, formerly a member of the Electronic Defense Group, now at Princeton University. The general rationale of this experiment, and the results of its application to the perception of words exposed for short durations, have been presented by Bricker and Chapanis (1953) and by Howes (1954).

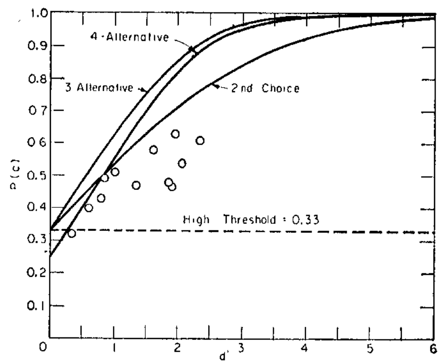

FIG. 10. The results of the second-choice experiment. (The proportions of correct second choices are plotted against d'. The curve labeled "2nd Choice" represents the prediction of decision theory, assuming the density functions to be normal and of equal variance. The prediction of the high threshold theory is shown by the dotted line.)

Thus, for a four-alternative experiment as described in the following, the high threshold model predicts that, when the first choice is incorrect, the probability that the second choice will be correct is 0.33. This predicted value, it may be noted, is independent of the signal strength.

Decision theory, on the other hand, implies that the observer is capable of ordering the four alternatives according to their likelihood of containing the signal. If this is the case, the proportion of correct second choices will be greater than .33. Should one of the samples of the noise function be the greatest of the four, leading to an incorrect first choice, the probability that the observation from the signal-plus-noise distribution will be the second greatest is larger than the probabilities that either of the observations of the noise distribution will be the second greatest. Again, it is intuitively clear that this probability is a function of d', or of signal strength—i.e., the probability that the observation of the signal-plus-noise value will be greater than two of the observations of noise increases with increases in d'. Specifically, the probability of a correct second choice in a four-alternative, forced-choice test, for a given value of d', is given by the expression:

$$\frac{3\int_{-\infty}^{+\infty}[f(x)]^2[1-f(x)]g(x)dx}{1-\int_{-\infty}^{+\infty}[f(x)]^3g(x)dx} \quad [8]$$

where the symbols have the same meaning as in Equation 7. This relationship is plotted in Figure 10 under the assumptions that the density functions of noise and signal plus noise are Gaussian and of equal variance. (The function predicted by decision theory for the proportion of correct first choices in a three-alternative situation is included in Figure 10 to show that this function is not the same as the predicted function of the probability of of a correct second choice, given an incorrect first choice, for the four-alternative situation).

To distinguish between the two predictions, data were collected from four observers; two of them had served previously in the second expected-value experiment, whereas the other two had received only routine force-choice training. Each of the observers served in three experimental sessions. Each session included 150 trials in which both a first and second choice were required.

The resulting 12 proportions of correct second choices are plotted against d' in Figure 10. The values of d' were determined by using the proportions of correct first choices as estimates of the probability of a correct choice, $p(c)$, and reading the corresponding values of d' from the middle curve of Figure 10, which is the same curve shown in Figure 9. Although just one value of signal intensity was used (0.78 foot-lamberts as in the second expected-value experiment), the values of d' differed sufficiently from one observer to another to provide an indication of the agreement of the data with the two predicted functions. Additional variation in the estimates of d' resulted from the fact that, for two observers, a constant distance from the signal was not maintained in all three of the experimental sessions.

A systematic deviation of the data from a proportion of 0.33 clearly exists. Considering the data of the four observers combined, the proportion of correct second choices is 0.46. Further, a correlation between the proportion of correct second choices and d' is evident.

Two control conditions aid in interpreting these data. The first of these allowed for the possibility that requiring the observer to make a second choice might depress his first-choice performance. During the experiment, blocks of 50 trials in which only a first choice was required were alternated with blocks of 50 trials in which both a first and a second choice were required. Pooling the data from the four observers, the proportions of correct first choices for the two conditions are 0.650 and 0.651, a difference that is obviously not significant. A preliminary experiment in which data were obtained from a single observer for five values of signal intensity also serves as a control. In that experiment, 150 observations were made at each value of signal intensity. The relative frequencies of correct second choices for the lowest four values of signal intensity were, in increasing order of signal intensity: 26/117 (0.22), 33/95 (0.35), 30/75 (0.40), and 20/30 (0.67). For the highest value of signal intensity, none of five second choices was correct. In this experiment, then, the proportion of correct second choices is seen to be correlated with a physical measure of signal intensity as well as with the theoretical measure d'—this eliminates the possibility that the correlation found with a constant value of signal intensity, involving d' as one of the variables (Figure 10), is an artifact of theoretical manipulation.

It may be seen from Figure 10 that the second-choice data also deviate systematically from the predicted function derived from decision theory. This discrepancy, as will be seen, results from the inadequacy of the assumption—of equal variance of the noise and signal-plus-noise density functions—upon which the predicted functions in Figure 10 are based. It was pointed out above that the data obtained in the second expected-value experiment (see Figure 8 and accompanying text) indicate that a better assumption would be that the ratio of the increment in the mean of the signal-plus-noise function to the increment in its standard deviation is equal to 4. Figure 11 shows the second-choice data and the predicted four-alternative and second-choice curves derived from the theory under this assumption that $\Delta M/\Delta\sigma = 4$. In view of the variance associated with each of the points (each first-choice d' was estimated on the basis

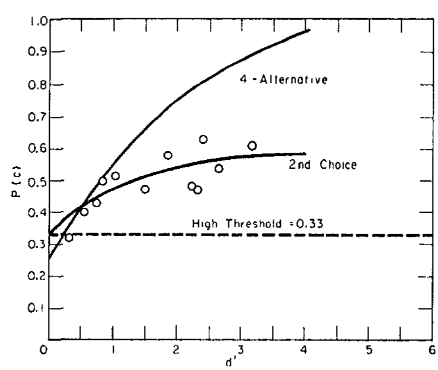

FIG. 11. The results of the second-choice experiment calculated under another assumption. (The predictions from decision theory for first and second choices are plotted under the assumption that $\Delta M/\Delta\sigma = 4$.

of 300 observations and each second-choice proportion on less than 100 observations), the agreement of the data and the predicted function shown in Figure 11 is quite good.

The conclusion to be drawn from these results of the second-choice experiment, though perhaps more obvious here, is the same as that drawn from the yes-no, or expected-value, experiments: the sensory information, or the decision axis, is continuous over a greater range than allowed for by the high threshold model. If a threshold cutoff, below which there is no discrimination among observations, exists at all, it is located in such a position that it is exceeded by much of the noise distribution.

Note on the Variance Assumption

Before considering the two remaining experiments, we should pause briefly to take up the problem of the relative sizes of the variances of the noise and signal-plus-noise distributions. We have seen, as indicated in the theoretical discussion, that an assumption concerning these variances may be tested by experiment. We have found that two sets of data, from yes-no and forced-choice experiments, support the assumption that the variance of the signal-plus-noise distribution increases with its mean. In particular, the assumption that $\Delta M/\Delta\sigma = 4$ is seen to fit those data reasonably well, and noticeably better than the assumption of equal variance. We should like to point out three aspects of this topic in the following paragraphs: first, the assumption of $\Delta M/\Delta\sigma = 4$ is probably not generally applicable; second, that we have good reason to suspect in advance of experimentation, in the visual case, that the variance of the signal-plus-noise distribution is greater than that of the noise distribution; and, third, that the very assumption of unequal variances requires that we qualify a statement made earlier in this paper.

It will be apparent that if the variance of these sampling distributions is a function of sample size, then their variances will differ as a function of the duration and the area of the signal. The assumption of $\Delta M/\Delta\sigma = 4$ will probably not fit the results of experiments with different physical parameters. Further, as we have indicated, we have not explored the extent of agreement between other specific assumptions and our present data. It appears likely that more precise data will be required to determine the relative adequacy of different assumptions about the increase in variance with signal strength.

Peterson, Birdsall, and Fox (1954), after developing the general theory of signal detectability, spelled out the specific forms it takes in a variety of different detection problems. By way of illustration, we may mention the problems in which *the signal is known exactly, the signal is known exactly except for phase,* and *the signal is a sample of white Gaussian noise.* A principal difference among these problems lies in the shape of the expected ROC curve. For our present purposes, we may regard these problems as differing in the degree of variance contributed by the signal itself. For the first case mentioned, the signal contributes no variance—the signal-plus-noise distribution is simply a translation of the noise distribution, the two have equal variances. In the other two cases, the signal itself has a variability which increases with its strength.

Clearly, if we are to select one of the specific models incorporated within the theory of signal detectability to apply to a visual detection problem, we would not select the one that assumes that the signal is known exactly, for the visual signal does not contain phase information. Thus, the second model is more likely to be applicable than the first. Actually, the third model, which assumes that the signal is a sample of noise, is the best representation of a visual signal. The fundamental point here is that either of the last two models leads to predicted results quite similar to those that are predicted under the assumption that $\Delta M/\Delta\sigma = 4$. Further discussion of this point would lead us too far off the path; we would like simply to note here that a specific form of the theory of signal detectability, which on a priori grounds is most likely to be applicable to vision experiments, predicts results very similar to those obtained. It is interesting to note in this connection that the results of auditory experiments using pure tones as signals are in close agreement with the signal-known-exactly model, with the assumption of equal variance.

The discerning reader will have noted that the assumption of a variance of the signal-plus-noise distribution that increases with its mean is inconsistent with a statement made in the theoretical discussion. In particular, the assumption of a greater variance of $f_{SN}(x)$ than of $f_N(x)$ conflicts with the statement that the decision axis x may be regarded as a likelihood-ratio axis. It was stated above (see the discussion following Figure 2) that a multidimensional response of the sensory system, i.e., one that might be represented by a point y in a multidimensional space, could be mapped into a line by considering the likelihood that y arose from SN relative to the likelihood that y arose from N, or $\lambda(y) = f_{SN}(y)/f_N(y)$. We then stated that we could identify the observation variable x with some monotonic transformation of $\lambda(y)$. If, now, the variance of $f_{SN}(x)$ is greater than the variance of $f_N(x)$, then as x decreases from a high value, $\lambda(x)$ will decrease—but, at some point below the mean of the function $f_N(x)$, $\lambda(x)$ will begin to increase again, and will, as a matter of fact, become greater than unity. Thus, if we choose to maintain the assumption of a greater variance of $f_{SN}(x)$, then the variable x cannot be regarded, throughout its range, as a likelihood ratio. Given that we do want to maintain the assumption of increasing variance of $f_{SN}(x)$, for the time being at least, we may take any of several possible steps to correct the difficulty. We can, for example, assume that there exists a low threshold, near the mean of $f_N(x)$, such that values of x less than this threshold are not ordered by the observer, and hence the fact that x cannot be considered as a likelihood ratio below this point is of no consequence. Another alternative is to assume outright that the variable x is unidimensional, without recourse to the likelihood-ratio argument to make the assumption reasonable. Which particular solution we shall adopt will depend upon further experimentation.

Analysis of the Rating Scale

We have concluded from the experiments described above that the observer's decision axis is continuous over a large range, i.e., that he can order observations likely to result from noise alone. We might expect then, in the language of decision theory, that he will be able to report the a posteriori probability of signal existence, i.e., that he will be able to state, following an observation interval, the probability that a signal existed during the interval. In more familiar terms, we are expecting that the observer will be capable of reporting a subjective probability, or of employing a rating scale. Experimental verification of this hypothesis is required, of course, for a reasonable doubt remains whether the observer will be able to maintain the multiple criteria essential to the use of a rating scale. If, for example, six categories of a posteriori probability are used, or a six-point rating scale, the observer must establish five criteria instead of just one as in the yes-no procedure—this may be considerably more difficult.

The ability to make a probability or rating response is of interest, in part, because such a response is highly efficient—in principle, a probability response retains all of the information contained in the observation. In contrast, breaking up the observation continuum into Yes and No sections is a process that loses information. From a procedure forcing a binary response, one learns from the observer only that the observation fell above or below a critical value, and not how far above or below. In some practical detection problems, the finer-grain information gained from a probability response can be utilized to advantage: the observer may record a posteriori probability so that Yes and No decisions concerning the action to be taken can be made at a later time, or by someone else who may be more responsible or who may possess more information about the values and costs of the decision outcomes.

More to the point in terms of our present interests, an experimental test of the ability to make a rating response contributes to the evaluation of decision theory, and also to distinguishing between the adequacy of decision theory and the high threshold theory. Since the data obtained with a rating procedure may be used to construct ROC curves, this experiment attacks the same problem as those described above, i.e., whether the observer can discriminate among observations likely to result from noise alone. It is also the case, as pointed out by Egan, Schulman, and Greenberg (1959), that the rating procedure generates ROC curves, of a given reliability, with a considerable economy of time compared to the yes-no procedure. Therefore it is of interest, with respect to future applications of decision theory, to determine whether the observer can perform as well, as indexed by d', with the rating procedure as with the yes-no procedure.

The observer's task in this experiment was to place each observation in one of six categories of a posteriori probability. Four categories of equal size (0.2) were used in the range between 0.2 and 1.0; the other two categories were 0.0–0.04 and 0.05–0.19. The boundaries of the categories were chosen in conference with the observers; they believed that they would be able to operate reasonably within this particular scheme. Actually, the specific sizes of the categories used are not important for most purposes; we can as well think of a six-point rating scale and assume only the property of order.

The four observers in this experiment were those who served in the second expected-value experiment. Further, the same signal intensity (0.78 foot-lamberts) and the same a priori probabilities—$p(SN) = p(N) = 0.50$—that were employed in that experiment were employed in this one. The observers made a total of 1,200 observations in three experimental sessions.

Results. The raw data for each observer consist of the number of observations of signal plus noise and the number of observations of noise alone that were placed in each of the six categories of a posteriori probability. Before proceeding with more complex analyses, we shall first make a rough determination of the validity of the observers' use of the categories, i.e., of whether we are, in fact, dealing with a scale. This may be achieved by computing the proportion of the total number of observations placed in each category that were actually observations of a signal. If the categories were used properly, this proportion will increase with increases in the probabilities that define the categories.

The results of this analysis are shown in Figure 12. Five curves are plotted there, one for each of the four observers and one showing the average result. We may note, as an aside, that Observer 4 is considerably more cautious than the others. A look at the raw data reveals that he used the lowest category twice to four times as often as the other observers; as a matter of fact, he placed 60% of his observations in that category. We may look for this difference to reappear in other analyses of the data of this experiment. The major point here, however, is that three of the four individual curves are monotonic increasing, whereas the fourth shows only one reversal. This result indicates the feasibility of using a scaling procedure —it indicates that requiring an observer to maintain five criteria simultaneously in a detection problem is not unreasonable. The result is consistent with an ability to order completely the observations, those arising from noise alone as well as those arising from signal plus noise.

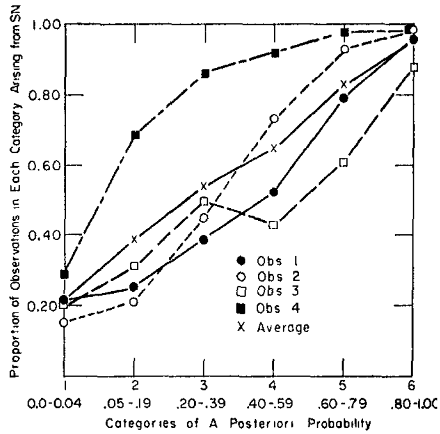

FIG. 12. The results of the rating experiment.

ROC Curves Obtained from the Rating Data. ROC curves can be generated from data obtained with the rating procedure since these data can be compressed to those of the binary-decision procedure with any of several criterion levels. That is to say, we can calculate the pair of values, $p_N(A)$ and $p_{SN}(A)$, ignoring all but one of the (five) criteria, or category boundaries, employed by the observer. We successively calculate five pairs of these values, each time singling out a different criterion, and thus trace out an ROC curve. In particular, we first compute the conditional probabilities that observations arising from noise alone and from signal plus noise will be placed in the top category; then these probabilities are computed with respect to the top two categories, and so forth. We assume, in these calculations, that observations placed in a particular category would fall above the criteria that define a lower category.

The ROC curves so obtained are shown in the upper left-hand portions of each part of Figure 13. (Ignore, for now, the other curves in Figure 13.) We may note that the data are well described by the type of ROC curve predicted from decision theory. As is the case with the empirical ROC data from yes-no experiments, they cannot be fitted well by a straight line intersecting the point $p_{SN}(A) = p_N(A) = 1.0$, the prediction made from the high threshold theory. This result indicates that the observers can discriminate among observations likely to result from noise alone, and are capable of maintaining the multiple criteria required for the rating response.

Comparison of ROC Curves Obtained from Ratings and Binary Decisions. It is intuitively clear that an estimate of d' of given reliability can be achieved with fewer observations by the rating procedure than by the yes-no procedure. This proposition is sup-

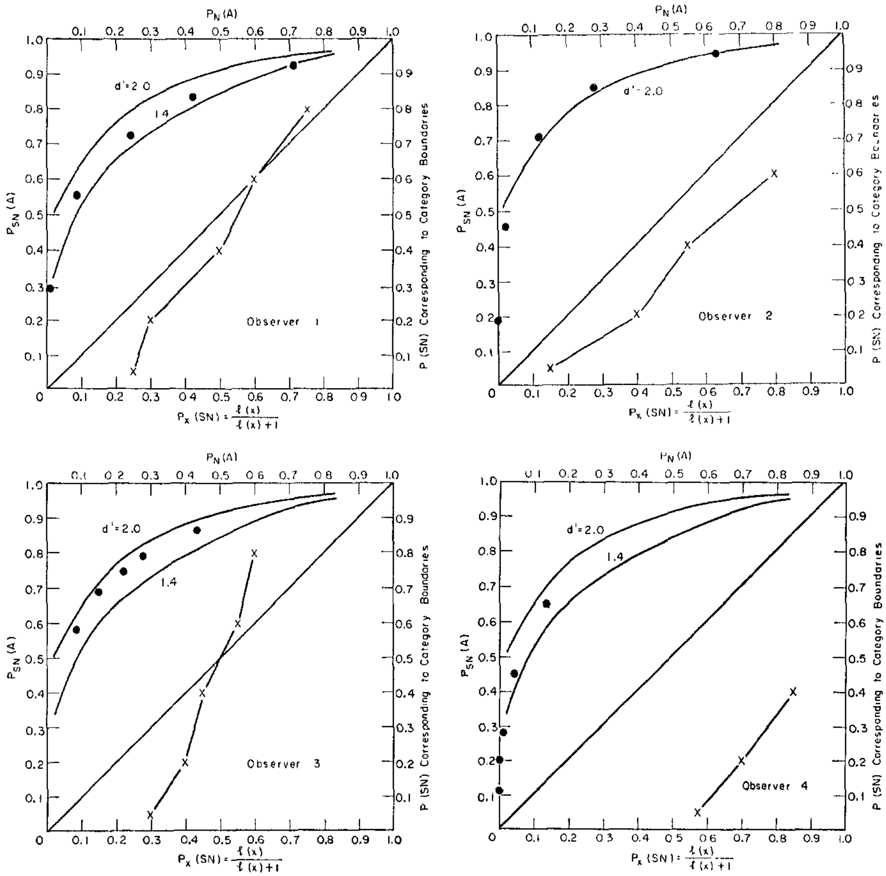

Fig. 13. Empirical receiver-operating-characteristic curves for four observers in the rating experiment—two alternative presentations.

ported by a comparison of the yes-no data shown in Figure 8 with the rating data shown in Figure 13—the rating data, which show considerably less variation, are based on 1,200 observations whereas the yes-no data are based on approximately 5,000 observations.

The economy provided by the rating procedure makes it desirable to determine whether the two procedures are equivalent means of generating the ROC curve. Unfortunately, to answer this question immediately, there are some clear differences between the ROC curves we have obtained with the two procedures. These differences are best illustrated by plotting the data on normal coordinates, i.e., on probability scales transformed so that the normal deviates are linearly spaced. These scales are convenient since on them the ROC curve specified by decision theory becomes a straight line. Further, the slope of this line represents the relative variances of the density functions, $f_N(x)$ and $f_{SN}(x)$, that underlie the ROC curve. In particular, it can be shown that the reciprocal of the slope (with respect to the normal deviate scales) is equal to the ratio $\sigma SN/\sigma N$.

The empirical ROC curves obtained

with the rating and yes-no procedures are shown on normal coordinates in Figure 14. It is immediately evident from this figure that a lower detectability resulted from the rating procedure for all four observers. We may see from the alternative presentations of these data in Figures 8 and 13 that the values of d' range from 2.0 to 3.0 for the yes-no data and from 1.5 to 2.0 for the rating data.[11] It is further apparent in Figure 14 that, consistent with the difference in d', the rating curve has a greater slope than the yes-no curve. This difference is small—the greater variance of $f_{SN}(x)$ under the yes-no procedure did not show clearly in the plots on linear probability axes—but it is regular. We may also note again, as this way of plotting the data makes very clear, that the rating data show considerably less scatter than the yes-no data.

The values and costs associated with the decision outcomes in this situation make us hesitant, on the basis of the data we obtained, to reject the hypothesis that the rating and yes-no procedures are equivalent means of generating ROC curves. It is possible, of course, that some undetected difference existed between the experimental conditions in the two experiments; one was conducted after the other was completed. Such a difference might easily account for the relatively small discrepancies observed. Again, it has recently been shown in an auditory experiment that the two procedures result in essentially the same ROC curve, both with respect to d' and to slope (Egan, Schulman, & Greenberg, 1959). Still, we cannot discount the present results on the basis of the auditory experiment, for we have noted several differences between visual and auditory data that are likely to be real—one perhaps relevant to this issue is that the ROC curves obtained with pure tones have slopes that are uniformly near one. We should perhaps be content, at this point, with the admittedly weak conclusion that no data exist to support the hypothesis that the two procedures are equivalent in the case of visual stimuli.[12]

Test of the Normality of the Density Functions. At this juncture, it is convenient to turn briefly, but explicitly, to a topic first considered in the theoretical discussion. It was stated there that we would assume the density functions on the observer's decision axis to be Gaussian in form, but that the assumption was subject to experimental test. A test of this assumption is provided by plotting the empirical ROC curves on normal coordinates. Having now introduced plots of the data in this form in Figure 14, we may use them for this purpose. If the observer's density functions are normal, then the empirical points of an ROC curve plotted on normal coordinates will be fitted best by a straight line. Clearly, a straight line provides an adequate description of the data in these figures. Thus the assumption of normality, an important one for the sake of simplicity of analysis, is supported by the data.

Approach to Optimal Behavior

In the presentation of experimental results thus far, we have concentrated on the continuity of the observer's decision axis, and on his ability to adopt

[11] Values of d' can, of course, be computed from the normal deviate scales of the plots in Figure 14. A problem arises, however, if the slope of the line fitted to the data is not unity. A solution to this problem is proposed in Clarke, Birdsall, and Tanner (1959).

[12] As this article goes to press we can report that in a repetition of this experiment with visual stimuli (unpublished) no reliable or regular differences were found between ROC curves obtained from ratings and binary decisions.

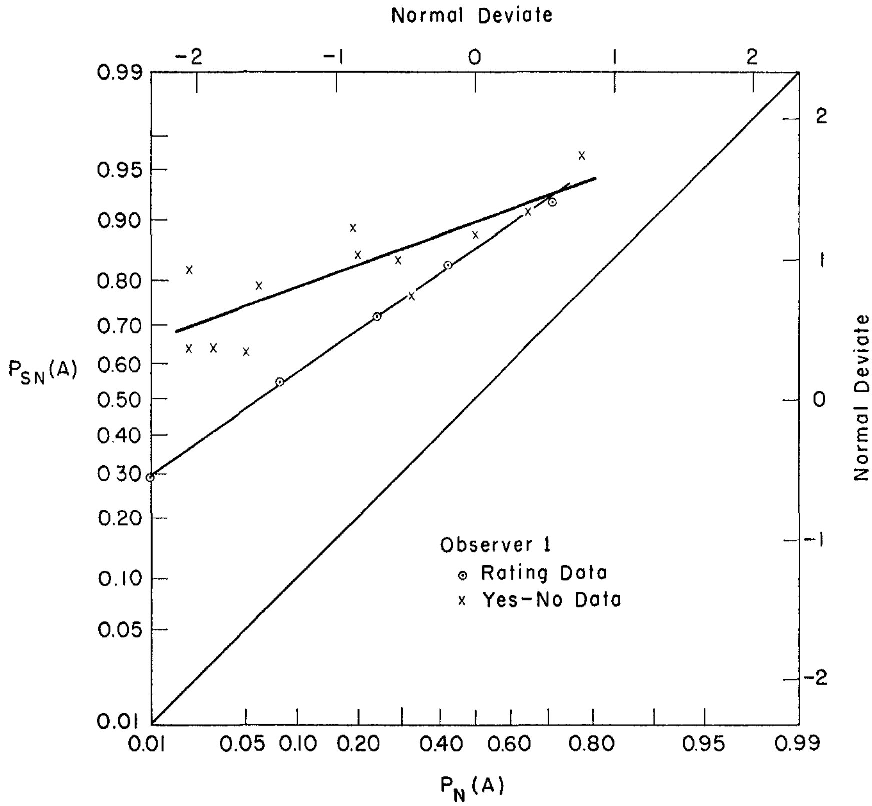

FIG. 14. Comparison of the receiver-operating-characteristic curves obtained from ratings and binary decisions.

various criteria along this axis. A remaining question is how closely the criteria he adopts correspond to those specified by decision theory as the optimal criteria. To answer this question we shall consider some further analyses of experimental results already described, and the results of an additional experiment.

It should be recalled that decision theory specifies as the optimal decision function either likelihood ratio, $\lambda(x)$, or some monotonic function of likelihood ratio, call it $\lambda(x)'$. That is to say, any transformation of the decision axis is acceptable as long as order is maintained. If the decision function is $\lambda(x)$, then the optimal criterion is the value of $\lambda(x)$ equal to β (Equation 3). If the decision function is $\lambda(x)'$, then the optimal criterion is the value of this function that corresponds to β, call it β'. The monotonic relationship means that $\lambda(x)' > \beta' \leftrightarrow \lambda(x) > \beta$. Thus to establish the applicability of decision theory, it is sufficient to demonstrate that the observer's criteria are monotonically related to β. If sampling error is taken into account, it is sufficient to demonstrate a significant correlation between the observer's criteria and β. It is of interest, however, to determine just how closely the observer's criteria do approach the optimal criteria as specified by β. In examining this question we shall make

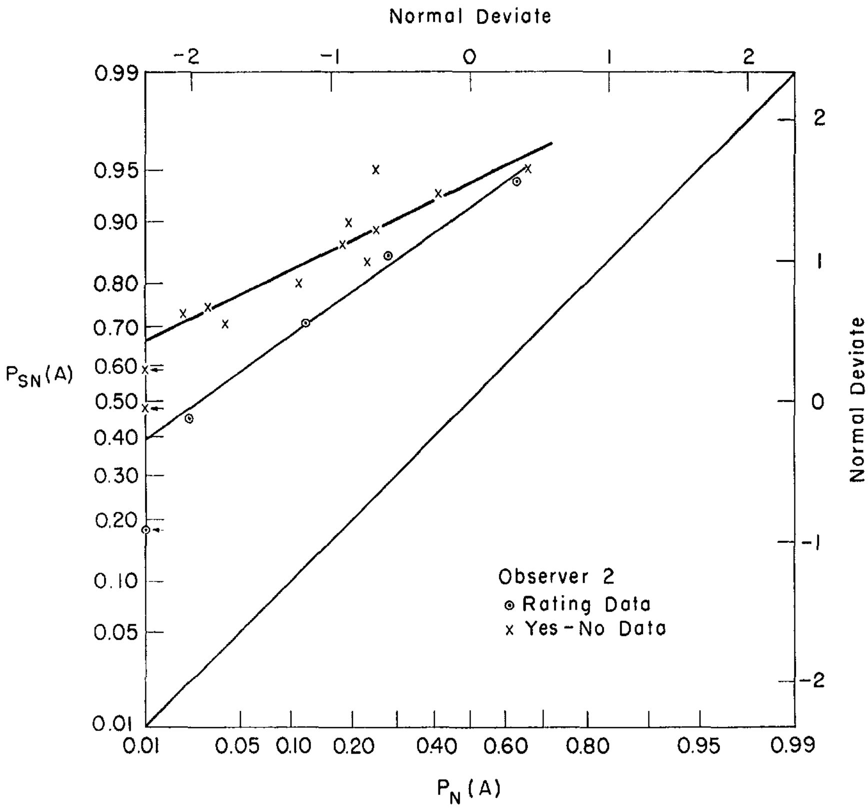

FIG. 14—*Continued*

use of the fact that, in order to index the observer's criterion, it is not strictly necessary to compute a value of likelihood ratio from the proportions of hits and false alarms; it is more convenient, and for purposes of interpretation, more direct, to take simply the proportion of false alarms as the index.

Criteria Employed in the Expected-Value Experiments. In the first expected-value experiment, the observers were told only the a priori probabilities of signal and noise and the values of the various decision outcomes that were in effect during each experimental session. They were not told that any combination of these factors can be expressed by a single number (β) which, in conjunction with a value of d', specifies the optimal criterion or the optimal false alarm rate. The rank-order correlations between β and the obtained proportions of false alarms that were computed from the data of this first study were .70, .46, and .71 for the three observers, respectively. A correlation of .68 is significant at the .01 level of confidence. This result indicates that the observer did not merely vary his criterion from one session to another, but that his criterion varied appropriately with changes in β.

In the second expected-value experiment, the observers were told the optimal proportion of false alarms for each session as well as the a priori probabilities and decision values. This

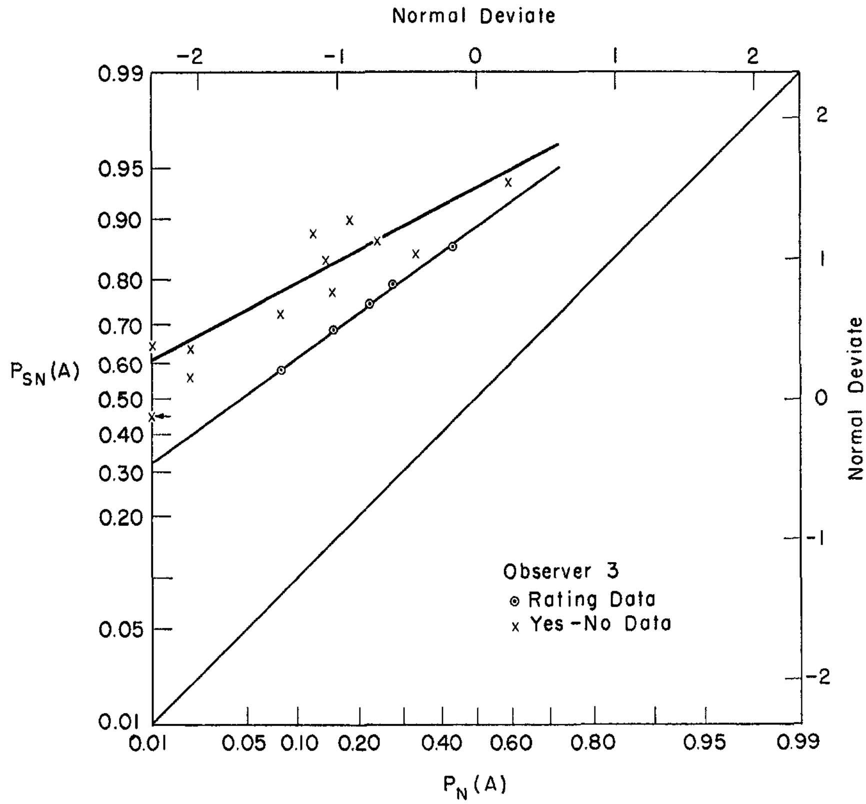

Fig. 14—*Continued*

information was available to the experimenter since values of d' had previously been determined by the forced-choice procedure during a training period. Thus, in the second study, we were asking how closely the observer would approach the optimal false alarm rate given knowledge of it. The rank-order correlations between the false alarm rates announced as optimal and the false alarm rates yielded by the four observers were .94, .97, .86, and .98. Again, a coefficient of .68 is significant at the .01 level of confidence. Data obtained later in an auditory experiment showed coefficients of this magnitude—as a matter of fact, the rank-order cofficient based on five pairs of measures for each of two observers in the auditory experiment was 1.0—when the observers were *not* informed of the optimal false alarm rate (Tanner, Swets, & Green, 1956).

Satisfying a Restriction on the Proportion of False Alarms. A more direct attack on the question of the observer's ability to reproduce a given false alarm rate is provided by an experimental procedure not previously described in detail, one involving a different definition of optimal behavior. Under this definition of optimal behavior, no values and costs are assigned the various decision outcomes; instead, a restriction is placed on the proportion of false alarms permitted. The optimal behavior is to maximize the proportion of hits while satisfying the

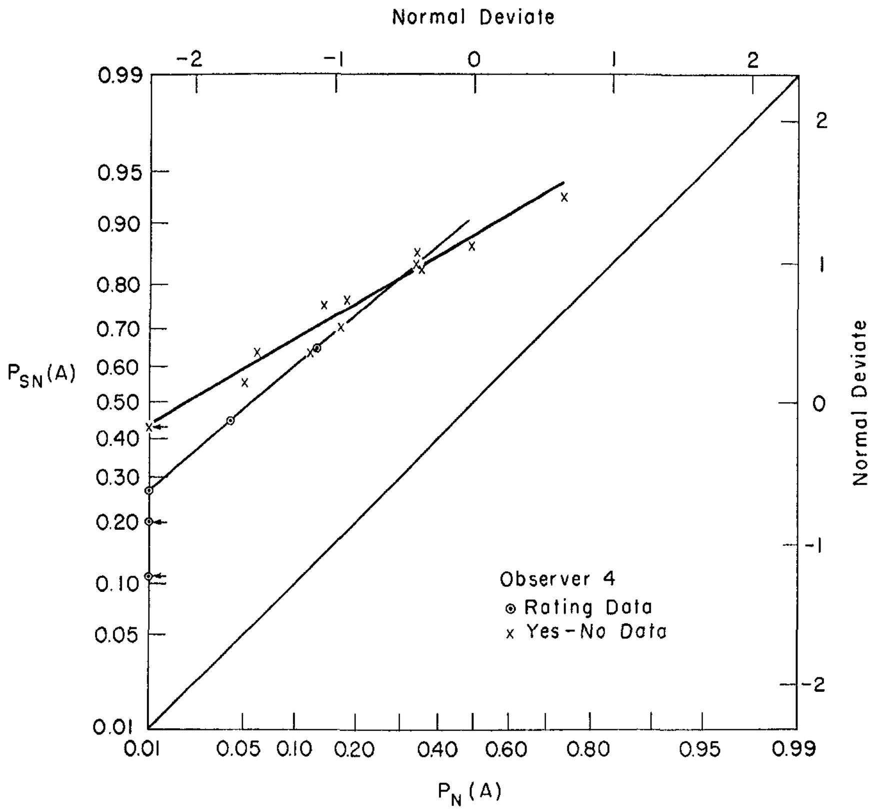

FIG. 14—*Continued*

restriction on false alarms. This, it will be recognized, is the procedure most popular among experimenters for testing statistical hypotheses.

An experiment using this procedure was conducted with a different set of four observers. The a priori probability of signal occurrence was 0.72 throughout the experiment. There were, then, 14 presentations of noise alone in a block of 50 presentations. There were four different experimental conditions, each extending over 18 blocks of 50 presentations. In each of these conditions, the observers were instructed to adopt a criterion that would result in Yes responses to approximately n or $n + 1$ of the 14 presentations of noise alone in a block of 50 presentations. For the four conditions of the experiment, n was equal to 0, 3, 6, and 9, respectively. Thus the acceptable range for the proportion of false alarms was .0–.07, .21–.28, .43–.50, or .64–.71. The primary data consist of four values of false alarm rate for each observer; each value is based on 252 presentations of noise alone.

The data are shown in Figure 15. The false alarm rates obtained are plotted against the restricted ranges of false alarm rate. The four observers are represented by different symbols; the vertical bars designate the acceptable range. It may be seen that the largest deviation from the range stipulated is .04. This result suggests that

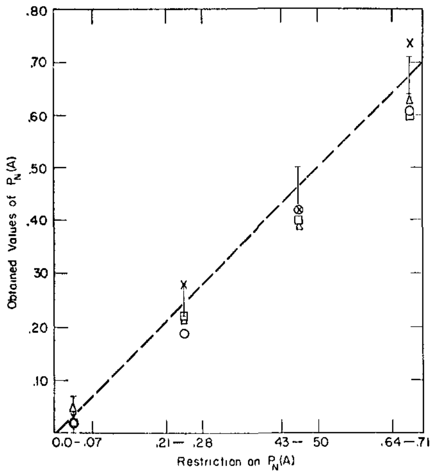

FIG. 15. The reproduction of a given false alarm rate.

the observer is able to adjust his criterion with considerable precision.

Two other pieces of information are needed, however, to interpret the data shown in Figure 15. For, of course, if the observer were given information about the correctness of his response after each response, these data could be obtained even if the observer were unable to vary his criterion. The observer could then approximate any false alarm rate by saying "yes" until the desired number of false alarms was achieved, and then by saying "no" on the remaining presentations. That procedure would entail a severe depression of d'. Actually, the observers were given information about correctness only after each block of 50 presentations, and the values of d' were not depressed. Thus the false alarm rates that were obtained may legitimately be regarded as reflecting the observer's criteria.

Criteria Employed in the Rating Scale Experiment. We may also investigate how closely the multiple criteria adopted by the observers in the rating scale experiment approach the optimal criteria. Stated otherwise, we may examine the relationship that existed between the subjective and objective probabilities of signal occurrence in that experiment. It may be noted in advance that an alternative presentation of the results, in Figure 12, gives an indication of the extent of agreement we may expect.

As stated earlier, the a posteriori probability of signal existence is a monotonic function of likelihood ratio. In particular, the optimal relationship between the two is:

$$p_x(SN) = \frac{\lambda(x)p(SN)}{\lambda(x)p(SN) + p(N)} \quad [9]$$

where: $p_x(SN)$ denotes the probability that the signal existed given the observation x (i.e., the a posteriori probability), $\lambda(x)$ is the likelihood ratio, and $p(SN)$ and $p(N)$ are the a priori probabilities (Peterson et al., 1954).

For our experiment, with $p(SN) = p(N) = 0.50$, this equation reduces to:

$$p_x(SN) = \frac{\lambda(x)}{\lambda(x) + 1} \quad [10]$$

As described above, a point on the ROC curve can be obtained for each of the boundaries of the six categories employed by the observer, i.e., for the five criteria he employed. Since, as we have also pointed out, the criterion value of $\lambda(x)$ corresponds to the slope of the ROC curve at the point in question, this criterion value of $\lambda(x)$ can be determined. Thus $p_x(SN) = \lambda(x)/\lambda(x) + 1$ can be computed for each of the criteria employed by the observer. Assuming now that the observer's decision function is likelihood ratio, then if he is behaving according to the optimal relationship between $p_x(SN)$ and $\lambda(x)$, the values of $\lambda(x)/\lambda(x) + 1$ computed from his data will correspond directly to probability values that were announced as defining the categories. In short, we know the values of $p_x(SN)$ that were announced as marking off the categories; by pursuing a route through the empirical ROC curve and $\lambda(x)$ we can calculate the values of $p_x(SN)$ that bound the categories the observer actually used—therefore we can assess how well the two sets of criterion values of $p_x(SN)$, the objective and subjective probabilities, agree.

The lower right-hand portions of Figure 13 show the probability values that were announced as defining the categories, plotted against the probability values that characterize the criteria actually employed by the observers, i.e., against $p_x(SN) = \lambda(x)/\lambda(x) + 1$ as determined from the data. (Some points are missing since $\lambda(x)$ is indeterminate at very low values of $p_N(A)$.) It is apparent from these plots that Observers 1, 2, and 3 are operating with a decision function similar to likelihood ratio and approximately according to the optimal relationship between $p_x(SN)$ and $\lambda(x)$. The pattern exhibited by Observers 1 and 3, that of overestimating small deviations from a probability of 0.50, will be familiar to those acquainted with the literature on subjective probability. Observer 4, as we noted earlier, is quite different from the others. His tendency, also evidenced but to a far lesser extent by Observer 2, is to consistently underestimate the a posteriori probability, i.e., to set all of his criteria too high.

To summarize our discussion of how nearly the criteria adopted by the observers in these several experiments correspond to the optimal criteria, we may say that the observer, for want of a better term, behaves in an "optimal fashion." He is responsive to changes in both the a priori probability of signal occurrence and the values of the decision outcomes; the criteria he adopts are highly correlated with the optimal criteria. Subjective transformations of the real probability scale and of the "real" value scale do, of course, exist, and differ somewhat from one observer to another. Undoubtedly, values also play a role in those experiments in which no values are explicitly assigned by the experimenter. Nevertheless, we have seen that the observer can adopt successively as many as 10 different criteria, on the basis of different combinations of probabilities and values presented to him, that are almost perfectly ordered. He can maintain simultaneously at least five criteria that are a reasonable facsimile of the optimal criteria. If he is told the optimal false alarm rate, he can, provided it is not very large or very small, approximate it with a small error.

SUMMARY, CONCLUSIONS, AND REVIEW OF IMPLICATIONS

We imagine the process of signal detection to be a choice between two Gaussian variables. One, having a mean equal to zero, is associated with noise alone; the other, having a mean equal to d', is associated with signal plus noise. In the most common detection problem the observer decides, on the basis of an observation that is a sample of one of these populations, which of the two alternatives existed during the observation interval. The particular decision that is made depends upon whether or not the observation exceeds a criterion value; the criterion, in turn, depends upon the observer's detection goal and upon the information he has about relevant parameters of the detection situation. The accuracy of the decision that is made is a function of the variable d' which is monotonically related to the signal strength.

This description of the detection process is an almost direct translation of the theory of statistical decision. The main thrust of this conception, and the experiments that support it, is that more than sensory information is involved in detection. Conveniently, a large share of the nonsensory factors are integrated into a single variable, the criterion. There remains a measure of sensitivity (d') that is purer than any previously available, a measure largely unaffected by other than physical variables. This separation of the factors that influence the observer's attitudes from those that influence his sensitivity is the major contribution of the psychophysical application of statistical decision theory.[18]

We have indicated several times in the preceding that another conception of the detection process, one involving what we termed a "high threshold," is inconsistent with the data reported. It should be noted, however, that these data, to the extent analyzed in this paper, do not preclude the existence of a lower threshold. The analyses presented do not indicate explicitly how far down into the noise the observations are being ordered, i.e., how low a threshold must be relative to the noise distribution in order to be compatible with the data. As it happens, further analyses of the yes-no and forced-choice results show them to be consistent with a threshold slightly above the mean of the noise distribution. If, for example, we examine the empirical ROC curves of Figures 8 and 13, we see that at values of $p_N(A)$ greater than 0.16, the curves are adequately fit by a straight line through the upper right-hand corner. Thus these data are consistent with a threshold cutoff that is located one sigma above the mean of the noise distribution.

Of course, a determination of the level at which a threshold may possibly exist is neither critical nor useful. A threshold well within the noise distribution is not a workable concept. Such a concept, since it is inconsistent with the correction for chance, complicates rather than facilitates the mathematical treatment of the data. Moreover, a threshold that is low is, for practical purposes, not measurable. The forced-choice experiment is a case in point; the observer conveys less information than he is capable of conveying if only a first choice is required. That the second choice contains a significant amount of information has been demonstrated; auditory experiments have shown that the fourth choice conveys information (Tanner et al., 1956). Thus it is difficult to determine when enough information has been extracted to yield a valid estimate of a *low* threshold. In addition, the existence of such a threshold is of little consequence for the application of the decision model—for example, yes-no data resulting from a suprathreshold criterion depend upon the criterion but are completely independent of the threshold value.

One of the major reasons for our concern with the threshold concept is

[18] It is interesting to note that the present account is not the first to model psychophysical theory after developments in the theory of statistical decision—as a matter of fact, Fechner was influenced by Bernoulli's suggestion that expectations might be expressed in terms of satisfaction units. As Boring (1950, p. 285) relates the story, Bernoulli's interest in games of chance led him to formulate the concept of "mental fortune"; he believed changes in mental fortune to vary with the ratio of the change in physical fortune to the total fortune. This mathematical relationship between mental and physical terms was the sort of relationship that Fechner sought to establish with his psychophysics. It should also be observed that Fechner anticipated the decision model under discussion in a much more direct way. His concept of "negative sensations," largely dismissed by subsequent workers in the field, denies the existence of such a cut in the continuum of observations that the magnitudes of observations below the cut are indiscriminable.

that this concept supports several common psychophysical procedures that are invalidated by the results we have described. The correction for chance success has already been mentioned as a technique that stems from a high threshold theory, and one that is inconsistent with the data. This correction is frequently applied to data collected with the method of constant stimuli. It is used implicitly whenever the threshold is defined as the stimulus intensity that yields a probability of correct response halfway between chance and perfect performance. The method of adjustment and the standard method of serial exploration are also inappropriate, given the mechanism of detection described above. When the method of serial exploration is used with the signal always present, or with insufficient "catch trials" to estimate the probability of a false alarm, the raw data will not permit separating the variation in the observer's criterion from variation in his sensitivity. Changes in an observer's criterion from one session to another can be estimated only if it is assumed that his sensitivity has not changed, and conversely. The same applies to data collected with the method of adjustment.

To be sure, unrecognized variations in the criterion are not important in many psychophysical measurements for they may be expected to contribute relatively little variation to the computed value of the threshold. Fairly large changes in the criterion will affect the threshold value by less than 3 db. in the case of vision, and by no more than 6 db. in the case of audition. This degree of reliability is acceptable in clinical audiometry, for example, in which the method of limits is usually employed. Neither would it distort appreciably curves of the course of dark adaptation. In many experiments, however—in experiments concerned with substantive as well as with theoretical problems—a reliability of less than 1 db. is required, and in these cases a knowledge of the criterion used by the observer is essential.

To illustrate the problems in which the threshold concept and its associated procedures may have led to improper conclusions, we may single out one of current interest, that of "subliminal perception." In most of the studies of this phenomenon, the evidence for it consists of the finding that subjects who first report seeing no stimulus can then identify the stimulus with greater-than-chance accuracy when forced to make a choice.[14] We have mentioned above as a typical result in psychophysical work that the forced-choice procedure yields lower threshold values than does the yes-no procedure. We have also suggested that this result may be accounted for by the fact that with the yes-no procedure the calculated value of the threshold varies directly with the observer's criterion, and that a strict criterion is usually employed by the observers under this procedure. That a strict criterion is usually used with the yes-no procedure is not surprising in view of the fact that observers are often instructed to avoid making false alarm responses. It is also likely that the stigma associated with "hallucinating" promotes the use of a strict criterion in the absence of an explicit caution against false alarms. Thus it may be expected that on many occasions when an observer does not choose to report the existence of the stimulus, he nevertheless possesses some information about it. It may be,

[14] This procedure was used explicitly in the earlier studies of subliminal perception; several of these studies are reviewed by Miller (1942). With minor variations, this procedure also underlies many of the more recent studies—see, for example, Bricker and Chapanis (1953).

therefore, that subliminal perception exists only when a high criterion is incorrectly identified as a limen.[15]

Having presented a theory of detection behavior and some detection experiments, and having just discussed the relationship of this work to "psychophysics," it remains to articulate with the title and the introductory paragraph of this paper, to consider the relationship of the work to the study of "perception."

In principle, the general scheme we have outlined may apply to perception as well as to detection. It seems reasonable to suppose that perception is also a choice among Gaussian variables. Consistent with the existence of many alternatives in the case of perception, we may imagine many critical regions to exist in the observation space. This space will have more dimensions than are involved in detection—as we have previously indicated, one less dimension than the number of alternatives considered. We may presume, in perception as in detection, that the boundaries of the critical regions are defined in terms of likelihood ratio, and are determined by the a priori probabilities of the alternatives and the relative values of the decision outcomes.

It may also be contended that what we have been referring to as a detection process is itself a perceptual process. Certainly, if *perceptual processes* are to be distinguished from *sensory processes* on the grounds that the former must be accounted for in terms of events presumed to occur at higher centers whereas the latter can be accounted for in terms of events occurring within the receptor systems, then the processes with which we have been concerned qualify as perceptual processes. Since, in detecting signals, the observer's detection goal and the information he possesses about probabilities and values play a major role, we must assume either that signal detection is a perceptual process, or that the foregoing distinction between sensory and perceptual processes is of little value.

Thus the thesis of the present paper is, in one of its aspects, another stage in the history of the notion that the process of perceiving is not merely one of passively reflecting events in the environment, but one to which the perceiver himself makes a substantial contribution. Various writers have suggested that our perceptions are based upon unconscious inferences, that sensory events are interpreted in terms of unconscious assumptions about their probable significance, that our responses to stimuli reflect the influence of our needs and expectancies, that we utilize cues in selectively placing sensory events in categories of identity, and so forth. The present view differs from these in regarding the observer as relating his sense data to information he has previously acquired, and to his goals, in a manner specified by statistical decision theory. The approach from decision theory has the advantage that it specifies the perceiver's contribution to perception at other than the conversational level; it provides quantitative relationships between the nonsensory factors and both the independent and dependent variables.

We submit then that the present paper, although confined to detection experiments, is aptly named. We may view detection and perception as made of the same cloth. Of course, signal detection is a relatively simple perceptual process, but it is exactly its simplicity that makes the detection setting most appropriate to a preliminary examination of the value of statistical

[15] This analysis of the problem of subliminal perception has been elaborated by Goldiamond (1958).

decision theory for the study of perception. Because detection experiments permit precise control over the variables specified by the theory as pertinent to the perceptual process, they provide the rigor desirable in the initial tests of a theory. Once these tests are passed, the theory may be extended and applied to more complex problems. Recent studies within the framework of decision theory include the recognition of one of two signals (Tanner, 1956), combined detection and recognition (Swets & Birdsall, 1956), problems in which a single decision is based on a series of observations (Swets, Shipley, McKey, & Green, 1959), problems in which the observer decides sequentially whether to make another observation before making a final decision (Swets & Green, in press), and the recognition of speech (Decker & Pollack, 1958; Egan, 1957; Egan & Clarke, 1956; Egan, Clarke, & Carterette, 1956; Pollack & Decker, 1958).

REFERENCES

BLACKWELL, H. R. Psychophysical thresholds: Experimental studies of methods of measurement. *Bull. Eng. Res. Inst. U. Mich.*, 1953, No. 36.

BLACKWELL, H. R., PRITCHARD, B. S., & OHMART, T. G. Automatic apparatus for stimulus presentation and recording in visual threshold experiments. *J. Opt. Soc. Amer.*, 1954, **44**, 322–326.

BORING, E. G. *A history of experimental psychology.* (2nd. ed.) New York: Appleton-Century-Crofts, 1950.

BRICKER, P. D., & CHAPANIS, A. Do incorrectly perceived tachistoscopic stimuli convey some information? *Psychol. Rev.*, 1953, **60**, 181–188.

BROSS, I. D. J. *Design for decision.* New York: Macmillan, 1953.

CLARKE, F. R., BIRDSALL, T. G., & TANNER, W. P., JR. Two types of ROC curves and definitions of parameters. *J. Acoust. Soc. Amer.*, 1959, **31**, 629–630.

DECKER, L. R., & POLLACK, I. Confidence ratings and message reception for filtered speech. *J. Acoust. Soc. Amer.*, 1958, **30**, 432–434.

EGAN, J. P. Monitoring task in speech communication. *J. Acoust. Soc. Amer.*, 1957, **29**, 482–489.

EGAN, J. P., & CLARKE, F. R. Source and receiver behavior in the use of a criterion. *J. Acoust. Soc. Amer.*, 1956, **28**, 1267–1269.

EGAN, J. P., CLARKE, F. R., & CARTERETTE, E. C. On the transmission and confirmation of messages in noise. *J. Acoust. Soc. Amer.*, 1956, **28**, 536–550.

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.

GOLDIAMOND, I. Indicators of perception: I. Subliminal perception, subception, unconscious perception: An analysis in terms of psychophysical indicator methodology. *Psychol. Bull.*, 1958, **55**, 373–411.

HORTON, J. W. *Fundamentals of sonar.* Annapolis: United States Naval Institute, 1957.

HOWES, D. H. A statistical theory of the phenomenon of subception. *Psychol. Rev.*, 1954, **61**, 98–110.

MILLER, J. G. *Unconsciousness.* New York: Wiley, 1942.

MUNSON, W. A., & KARLIN, J. E. The measurement of the human channel transmission characteristics. *J. Acoust. Soc. Amer.*, 1956, **26**, 542–553.

OSGOOD, C. E. *Method and theory in experimental psychology.* New York: Oxford Univer. Press, 1953.

PETERSON, W. W., BIRDSALL, T. G., & FOX, W. C. The theory of signal detectability. *IRE Trans.*, 1954, **PGIT-4**, 171–212.

POLLACK, I., & DECKER, L. R. Confidence ratings, message reception, and the receiver operating characteristic. *J. Acoust. Soc. Amer.*, 1958, **30**, 286–292.

SHANNON, C. E. The mathematical theory of communication. *Bell Sys. tech. J.*, 1948, **27**, 379–423.

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).

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

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.*, 1956, **IT-2**, 138–165.

Swets, J. A., & Green, D. M. Sequential observations by human observers of signals in noise. In C. Cherry (Ed.), *Fourth symposium on information theory.* London: Butterworth, in press.

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

Tanner, W. P., Jr. A theory of recognition. *J. Acoust. Soc. Amer.*, 1956, **28**, 882–888.

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., & Swets, J. A. A decision-making theory of visual detection. *Psychol. Rev.*, 1954, **61**, 401–409.

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

Thurstone, L. L. A law of comparative judgment. *Psychol. Rev.*, 1927, **34**, 273–286. (a)

Thurstone, L. L. Psychophysical analysis. *Amer. J. Psychol.*, 1927, **38**, 368–389. (b)

Van Meter, D., & Middleton, D. Modern statistical approaches to reception in communication theory. *IRE Trans.*, 1954, **PGIT-4**, 119–145.

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

(Received December 16, 1959)