Psychological Bulletin
1990, Vol. 107, No. 3, 401–413

0033-2909/90/$00.75

Response Bias: Characteristics of Detection Theory, Threshold Theory, and "Nonparametric" Indexes

Neil A. Macmillan
Brooklyn College of the City University of New York

C. Douglas Creelman
University of Toronto
Toronto, Ontario, Canada

Models of discrimination based on statistical decision theory distinguish sensitivity (the ability of an observer to reflect a stimulus–response correspondence defined by the experimenter) from response bias (the tendency to favor 1 response over others). Measures of response bias have received less attention than those of sensitivity. Bias measures are classified here according to 2 characteristics. First, the distributions assumed or implied to underlie the observer's decision may be normal, logistic, or rectangular. Second, the bias index may measure criterion location, criterion location relative to sensitivity, or likelihood ratio. Both parametric and "nonparametric" indexes are classified in this manner. The various bias statistics are compared on pragmatic and theoretical grounds, and it is concluded that criterion location measures have many advantages in empirical work.

In a simple but important type of discrimination experiment, a stimulus from one of two classes (S_1 or S_2) is presented on each trial, and an observer makes one of two corresponding responses ("no" or "yes"). Examples of possible stimulus classes are noise and signal-plus-noise in a detection task; new and old words in a recognition memory experiment; and normal x-rays and those displaying abnormalities in medical diagnosis.

The data from such an experiment can be summarized by two proportions. The hit rate H is the proportion of "yes" responses on S_2 trials, and the false-alarm rate F is the proportion of "yes" responses on S_1 trials. It is often useful, however, to summarize the data by using two other statistics, a measure of *sensitivity* and a measure of *response bias*. Sensitivity is the observer's ability to respond in accordance with the stimulus presented, "yes" to S_2 and "no" to S_1. Response bias is the observer's tendency to use one of the two responses.

In this article, we compare alternative measures of response bias. To begin, we propose some desirable attributes of a bias measure. We then present specific measures that derive from one of three theoretical approaches: detection theory, threshold theory, and "nonparametric" analysis. In all cases, we consider models that assume that trials are independent, making no effort to deal with sequential effects. We describe the relations among the bias measures and their relations to indexes of sensitivity. In a final section, we consider several possible rationales for deciding among bias measures and apply them to the candidates nominated earlier.

We are grateful to Howard Kaplan for many useful discussions of these issues. We also thank Doris Aaronson, reviewers A. E. Dusoir and J. G. Snodgrass, and an anonymous reviewer for helpful comments on previous drafts. The work was supported by a PSC-CUNY Research Award to Neil A. Macmillan.

Correspondence concerning this article should be addressed to Neil A. Macmillan, Department of Psychology, Brooklyn College of the City University of New York, Brooklyn, New York 11210.

Dusoir (1975) discussed some of these measures, distinguishing those based on detection theory from nonparametric indexes. We extend his theoretical work by placing the various indexes in a more systematic psychophysical framework, showing that each measure of bias can be interpreted in terms of an assumed underlying decision space.

Our attitude toward the problem of selecting an appropriate measure of response bias closely resembles the stance of Swets (1986b) in evaluating sensitivity indexes. Like Swets, we argue that all measures have psychophysical consequences and make distributional assumptions; none is nonparametric. Swets's recommendations about sensitivity were ultimately empirical (see Swets, 1986a). Because of some important differences between sensitivity and bias, our conclusions depend more on practical and theoretical considerations.

General Characteristics of Sensitivity and Bias Measures

Treatments of bias and of sensitivity are parallel in some ways, divergent in others. Although our interest is in bias, it is useful to begin with some brief comments about sensitivity.

Sensitivity

Consider a modest requirement for sensitivity indexes: If two hit/false-alarm pairs have the same value of F, then the pair with the higher hit rate has higher sensitivity; if two pairs have the same value of H, then the pair with the lower false-alarm rate has higher sensitivity. When this condition holds, the measure of sensitivity is monotonically increasing with H and monotonically decreasing with F. All proposed sensitivity measures that we are aware of have this property. A more stringent condition requires that H and F be treated symmetrically: Sensitivity is the difference between H and F, each transformed by a monotonic function u, the total possibly transformed by another monotonic function v:

$$\text{sensitivity} = v[u(H) - u(F)]. \quad (1)$$

The interesting characteristics of a sensitivity measure can be summarized by its receiver operating characteristic (ROC) curve, the function relating H to F when sensitivity is constant but bias varies. Sensitivity indexes that monotonically increase in H and decrease in F have monotonically increasing ROCs. Those that take the form of Equation 1, as most do, are symmetric around the minor diagonal of ROC space (that is, the unit square).

Response Bias

Parallel to our requirement for sensitivity statistics is a proposal about bias measures: If two hit/false-alarm pairs have the same value of F, then the pair with the higher hit rate displays the greater bias toward saying "yes"; if two pairs have the same value of H, then the pair with the higher false-alarm rate shows greater bias toward "yes". When this constraint is satisfied, the measure of response bias is monotonic with H and F in the same direction (both increasing or both decreasing). This *monotonicity condition* implies that when sensitivity declines, an observer who is maintaining the "same bias" should show both a decreased hit rate and an increased false-alarm rate. Although this property seems as modest as the corresponding one for sensitivity, many indexes satisfy it for only some values of H and F. The false-alarm rate itself, for example, fails to satisfy the monotonicity condition, which requires that response rates to both S_1 and S_2 contribute to an assessment of bias.

A more restrictive condition, similar to Equation 1, is

$$\text{bias} = v[u(H) + u(F)], \quad (2)$$

where u and v are monotonic functions. We consider some indexes that satisfy Equation 2, some that satisfy the monotonicity condition but not Equation 2, and some that violate even that requirement in some parts of their range.

The *isobias curve* of a bias statistic, analogous to the ROC of a sensitivity measure, is the locus of (H, F) points for which bias is constant but sensitivity varies. The monotonicity condition implies monotonically decreasing isobias curves.

All significant information about a bias measure is captured by its isobias curve. If a measure m_1 is related to another measure m_2 by any one-to-one transformation, both indexes have the same isobias curves, for hit/false-alarm pairs that have a constant value of m_1 also have a constant (in general, different) value of m_2. We call measures having the same isobias curve *equivalent;* several instances of equivalent indexes are discussed later. For comparing point estimates of bias across conditions, choices among equivalent measures are arbitrary; for other purposes, secondary considerations (such as the availability of statistical analysis) may render one particular transformation of an index more useful than others.

Detection Theory Measures of Bias

Using the machinery of statistical decision theory, Equations 1 and 2 can be translated into statements about the observer's decision space. According to detection theory, both S_1 and S_2 lead to underlying probability densities that are continuous. Two classes of detection-theoretic models have been especially important: those in which the distributions are Gaussian

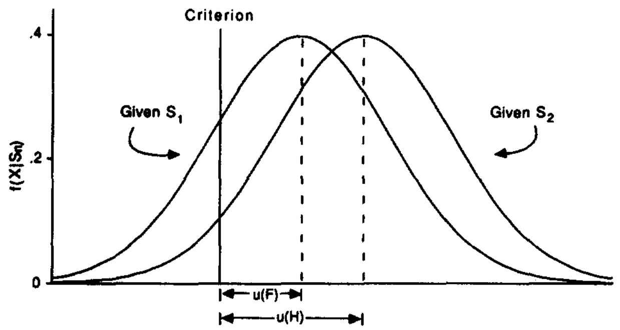


Figure 1. Decision space for the yes–no experiment, showing likelihood distributions for stimulus classes S_1 and S_2 according to detection theory; the observer responds "yes" (classifies the stimulus as S_2) for events above the criterion.

(Green & Swets, 1966) and those in which they are logistic (as in Choice Theory; Luce, 1959, 1963a). Historically, the term *Signal Detection Theory* was used by the developers of Gaussian models; we sometimes use the abbreviation SDT for such models.

If S_1 and S_2 give rise to densities that form a *shift family,* that is, differ only in mean, then the transformation u in Equations 1 and 2 is the inverse of the shared distribution function. Figure 1 shows the decision space for a simple discrimination experiment, according to such a model. The observer is assumed to respond "yes" for observations above a fixed criterion and "no" for those below that criterion. The function u converts hit and false-alarm proportions into distances from the criterion to the means of the S_2 and S_1 distributions, and the difference between these transformed proportions equals the difference between the means of S_2 and S_1. (The function v in Equation 1 is the identity function in this case.)

Three possible measures of response bias can be described within the scheme of Figure 1: (a) the location of the criterion, (b) the location of the criterion divided by sensitivity, and (c) the likelihood ratio, the height of the S_2 distribution at the criterion divided by the height of the S_1 distribution. The three indexes yield very different isobias curves whose exact shape depends on the assumed form of the underlying distributions.

Underlying Gaussian Distributions

Consider first the best-known SDT model, in which the underlying normal densities have equal variance. The mean difference d' is found from

$$d' = z(H) - z(F), \quad (3)$$

so that z, the inverse of the normal distribution function, is the function u of Equation 1. In Figure 2, the origin is placed at the midpoint between the two densities, so that the distribution means are at $\pm d'/2$.

Criterion location is measured relative to the origin, the crossover point of the two distributions. Although this is to some degree an arbitrary choice, two other possible reference points, the means of S_1 and S_2, can be ruled out. If measured relative to the mean of S_1, criterion location depends only on

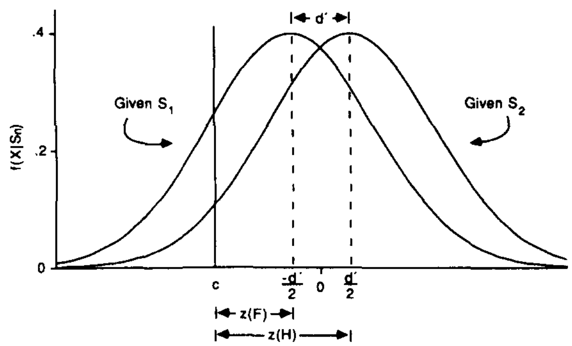


Figure 2. Decision space under Gaussian assumptions. (The decision axis is measured in units of the common standard deviation of the two distributions [z scores]. The point of zero bias is arbitrarily set to zero.)

F; if measured relative to the mean of S_2, only on H. Such indexes fail to satisfy the monotonicity condition.

For the Gaussian SDT model, Figure 2 shows that the criterion location relative to the origin is the average of the transformed hit and false-alarm rates:[1]

$$c = -0.5[z(H) + z(F)]. \tag{4}$$

Ingham (1970) was apparently the first to propose c as a bias measure. Clearly c satisfies Equation 2 and therefore the monotonicity condition. Isobias curves implied by c are shown in Figure 3a.

To normalize c by sensitivity, divide by d':

$$c' = c/d' = -0.5[z(H) + z(F)]/[z(H) - z(F)]. \tag{5}$$

Isobias curves for c', shown in Figure 3b, fan out from the point $H = F = .5$. They satisfy the monotonicity condition in the upper left quadrant of ROC space but not elsewhere.

The likelihood ratio β_G (G for Gaussian) is the ratio of two normal ordinates:

$$\beta_G = \exp[-0.5z^2(H)]/\exp[-0.5z^2(F)]. \tag{6}$$

The relation between likelihood ratio and c turns out to be extremely simple, for

$$\begin{aligned}\log(\beta_G) &= -0.5z^2(H) + 0.5z^2(F)\\ &= -0.5[z(H) + z(F)][z(H) - z(F)]\\ &= cd'.\end{aligned} \tag{7}$$

Log likelihood ratio equals the criterion location multiplied by sensitivity. Notice that likelihood ratio can be computed by using the z transformation only; values of normal ordinates are not required.

Isobias curves for β_G fan out from the perfect-performance point, as shown in Figure 3c. Like those for c', they satisfy the monotonicity condition only in the upper left quadrant.

Underlying Logistic Distributions

In Choice Theory, the underlying densities are logistic with variance $\pi^2/3$. The mean difference $2\log(\alpha)$ is found from[2]

$$\log(\alpha) = 0.5\{\log[H/(1 - H)] - \log[F/(1 - F)]\}. \tag{8}$$

The transformation from p to $\log[p/(1 - p)]$ is the inverse of the logistic distribution function. In Figure 4, the origin is placed at the midpoint between the two densities, and the distribution means are at $\pm\log(\alpha)$ (see McNicol, 1972; Noreen, 1977). The criterion location $\log(b)$ is defined by

$$\log(b) = -0.5\{\log[H/(1 - H)] + \log[F/(1 - F)]\}, \tag{9}$$

so this bias measure satisfies Equation 2.

To normalize $\log(b)$ by sensitivity,

$$b' = \log(b)/2\log(\alpha). \tag{10}$$

The likelihood ratio β_L can be expressed in terms of the criterion and sensitivity parameters,

$$\beta_L = [(1 + \alpha b)/(\alpha + b)]^2, \tag{11}$$

or in terms of the hit and false-alarm rates (see Appendix):

$$\begin{aligned}\log(\beta_L) &= \log[H(1 - H)] - \log[F(1 - F)]\\ \beta_L &= H(1 - H)/[F(1 - F)].\end{aligned} \tag{12}$$

Equation 11 displays relations among sensitivity, criterion location, and likelihood ratio that are similar to those found in Equation 7. For a criterion that is a fixed distance above the zero-bias point, likelihood ratio increases with sensitivity; for a negative criterion, the relation is inverse.

Because the logistic and Gaussian distribution functions are quite similar in shape, the isobias curves implied by the three measures are very much like those of Figure 3.

Threshold Theory Measures

Threshold theories (Luce 1963a; Krantz, 1969) assume that the decision space is characterized by a small number of discrete states, rather than the continuous dimensions of SDT and Choice Theory. Different threshold models are distinguished in the ways these states are evoked by stimulus classes and in the ways they lead to responses. For each of the two models presented here, we develop a *state diagram* that spells out these connections and determines the model's ROC.

A single ROC is, however, consistent with more than one set of underlying distributions. Each threshold ROC can be interpreted as arising from a decision space in which the S_1 and S_2 densities are rectangular (Green & Swets, 1966, also adopted this convention). These representations are not mathematically equivalent to the corresponding state diagrams, but because both generate the same ROC, both are legitimate representations of the model. The rectangular version has the important advantage of enabling comparison with the detection theory models already discussed.

We consider only two threshold theories. Both are characterized by "high" thresholds; that is, both assume that some internal state(s) arise only from one stimulus, whereas others may be activated by either. Although low-threshold theory (Luce,

[1] The minus sign is necessary to ensure that positive values of c correspond to locations above the equal-bias point.

[2] We use α to denote the inverse of Luce's statistic η, following the usage of McNicol (1972).

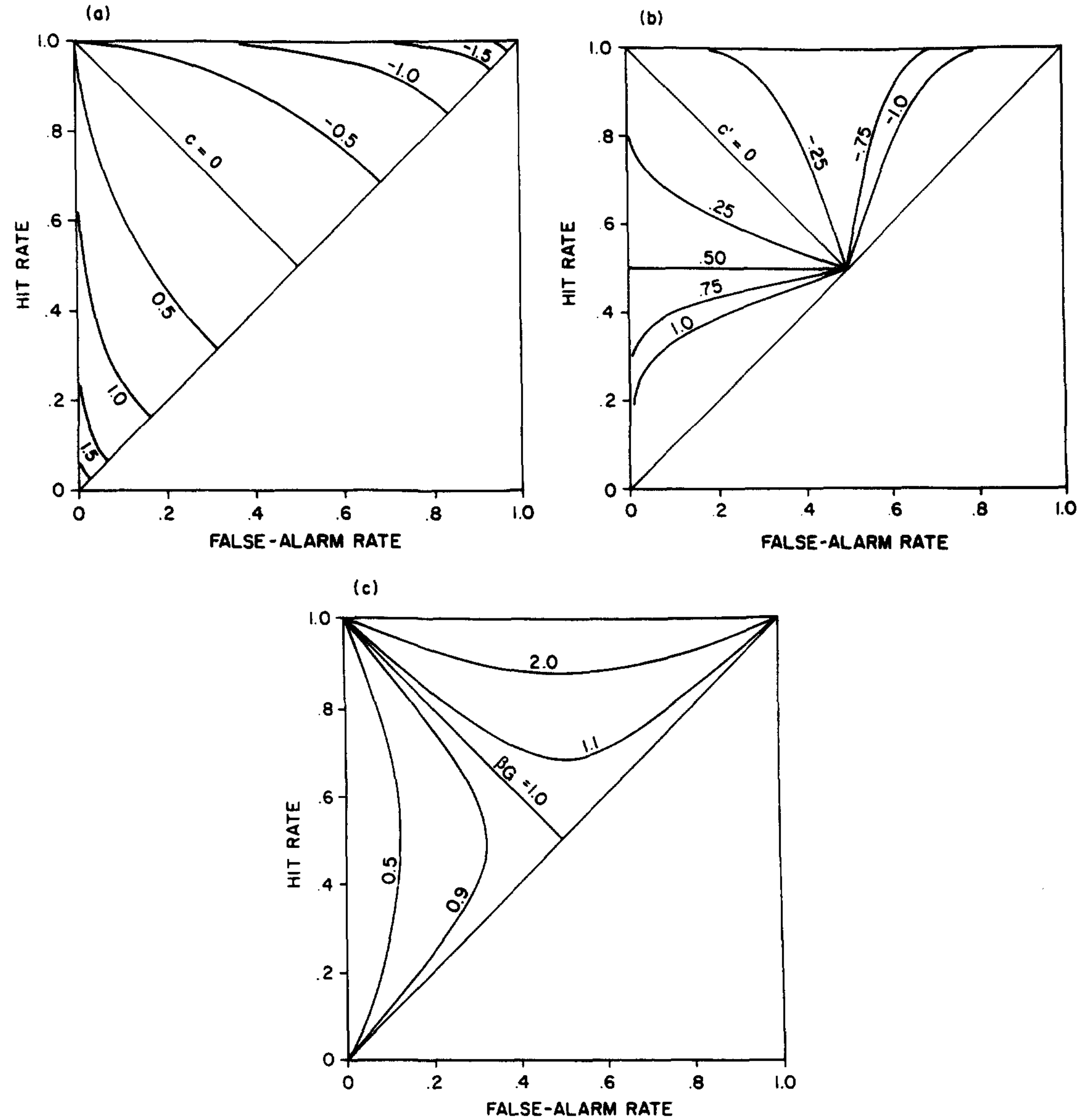


Figure 3. Isobias curves, assuming Gaussian distributions, for (a) criterion location, (b) criterion location relative to sensitivity, and (c) likelihood ratio.

1963a, 1963b) predicts an ROC that more nearly agrees with existing discrimination data than does high-threshold theory, it has an important practical disadvantage: Two parameters are needed to specify sensitivity and two to specify bias. Perhaps because of this complexity, the theory is little used.

Single High-Threshold Theory

The first model we consider is also the oldest (e.g., Blackwell, 1953). Numerous authors (Egan, 1958; Green & Swets, 1966; Luce, 1963a) have pointed out that the theory's predicted ROC is contrary to data, and it is widely recognized that the single high-threshold sensitivity index is inappropriate. The theory's bias statistic, however, is the false-alarm rate F. The connection of F to threshold theory is less generally appreciated, and this measure continues to be used by psychologists to assess bias.

State diagram and sensitivity measure. Single high-threshold theory proposes that sensitivity be measured by the adjusted ("true") hit rate:

$$q = (H - F)/(1 - F). \tag{13}$$

This relation can be derived from a state diagram (Figure 5a) containing only two internal states, D_1 and D_2. The true hit rate q is the probability that stimulus S_2 leads to state D_2. (Notice that S_1 never leads to D_2.) Observers respond "yes" whenever in state D_2 and with probability u when in state D_1; it is these latter contaminating "guesses" that make the correction recommended by Equation 13 necessary.

The dependence of H and F on the adjusted hit rate and the guessing rate can be calculated directly from the state diagram:

$$H = P(\text{"yes"} \mid S_2) = q + u(1 - q)$$
$$F = P(\text{"yes"} \mid S_1) = u. \tag{14}$$

Eliminating the guessing parameter u leads to Equation 13.

The ROC implied by the model is obtained by solving Equation 13 for H in terms of F. As shown in Figure 5b, it is a straight line from $(q, 0)$ to $(1, 1)$.[3] It is this prediction of the theory that has led to its rejection.

Underlying distributions. Now let us infer the underlying distributions (see Figure 6). To allow for the point $(q, 0)$, there must be a region on the decision axis where events only occur due to S_2—otherwise the false-alarm rate would not be zero. In the rest of the decision space, corresponding to the ROC segment $(q, 0)$ to $(1, 1)$, the S_1 and S_2 distributions could have any shape. They must be proportional to each other, however, because the ratio of their heights—the likelihood ratio—is constant when the ROC has constant slope. For simplicity, the underlying distributions are represented in Figure 6 as rectangles. The boundary between these two areas is the "threshold," the decision-axis value above which only S_2 events occur. The sensitivity parameter q is the proportion of the S_2 distribution that does not overlap the S_1 distribution.

Bias measure. A changing value of the parameter u (the proportion of D_1 trials on which the observer responds "yes") is modeled in this decision space by a shift in the criterion but not by a change in likelihood ratio. The criterion is located at a proportion $1 - u$ of the distance spanned by the S_1 distribution, for u is the false-alarm rate. Accordingly, the criterion location, relative to the S_1 distribution, is equivalent to F. Isobias curves for the parameter u (or for F) are vertical lines in ROC space, so u does not depend on H. As a bias statistic, F is doubly flawed: It fails to satisfy the monotonicity condition and it implicates a model that has been rejected on other grounds.

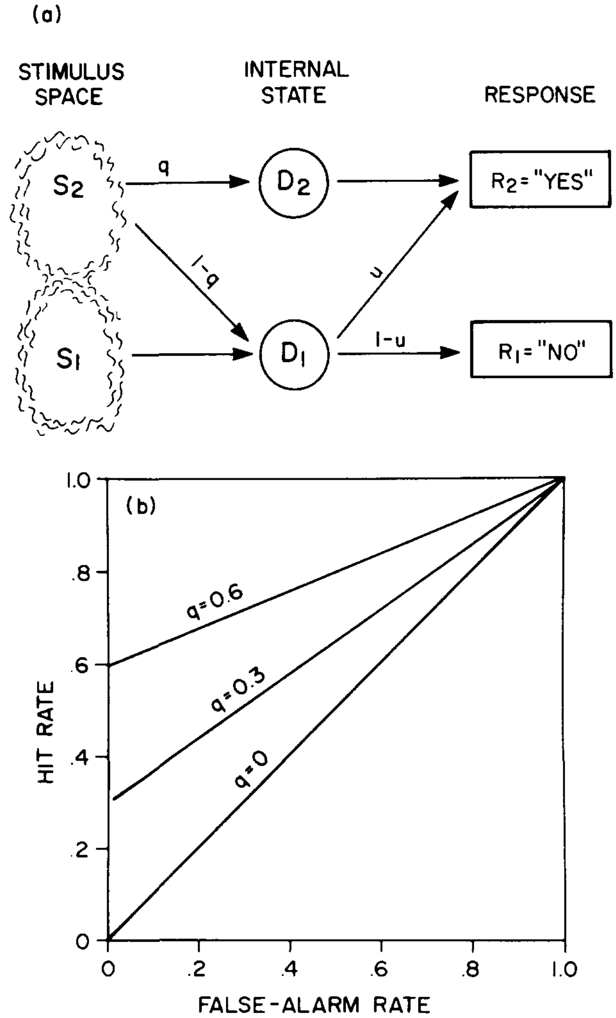


Figure 5. Single high-threshold theory: (a) state diagram, with two internal states; and (b) receiver operating characteristic curve.

Double High-Threshold Theory

Double high-threshold theory has a long history.[4] Woodworth (1938) presented a method for correcting recognition memory data for guessing that is equivalent to it. Egan (1958; summarized in Green & Swets, 1966, pp. 337–341) was the first to plot its ROC; he also pointed out that single and double high-threshold theories each call for a different correction for guessing and are special cases of the same general model. The double high-threshold theory has been explicitly discussed by Macmillan and Kaplan (1985), Swets (1986b), Snodgrass and Corwin (1988), and Morgan (1976). It often arises implicitly via its sensitivity parameter, for this model has proportion correct as its constant performance measure.

State diagram. The state diagram of the underlying model is shown in Figure 7a. There are three discrete states: D_1 arises only when S_1 occurs, D_2 can be triggered only by S_2, and an intermediate state, $D_?$, can occur for either stimulus. The model assumes two "high" thresholds, each insurmountable by one of the stimuli.

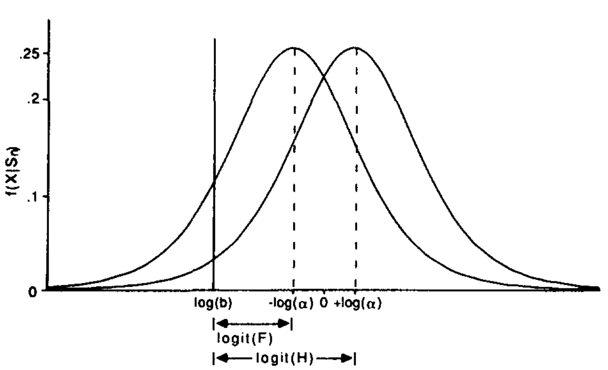


Figure 4. Decision space under logistic assumptions. (Distances on the decision axis are measured in logit units of log $[p/(1 - p)]$.)

[3] We denote points in ROC space as (hit, false-alarm) pairs, the reverse of the usual graphical convention, in which the order (x, y) is used.

[4] We thank our anonymous reviewer for pointing this out to us.

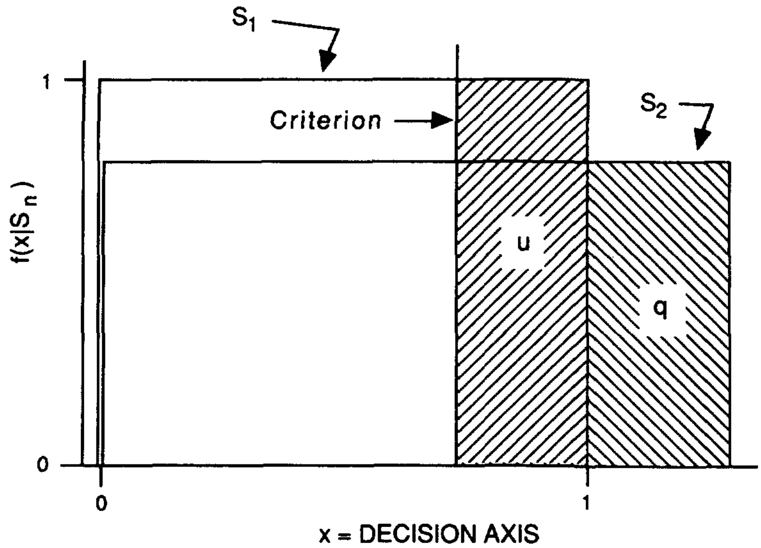


Figure 6. Rectangular underlying distributions consistent with single high-threshold theory.

Sensitivity measure p(c). As for single high-threshold theory, the sensitivity parameter in this model is a true hit rate q. For simplicity, this parameter is assumed to equal both the proportion of S_2 presentations leading to the D_2 state and the proportion of S_1 presentations leading to the D_1 state. The hit and false-alarm rates are

$$H = q + (1 - q)v \text{ and}$$
$$F = (1 - q)v. \quad (15)$$

It is easy to see that this model predicts an ROC of the form $H = q + F$; some examples are shown in Figure 7b.

For equal presentation probabilities, the proportion correct is the average of the hit and correct rejection rates, and

$$q = H - F = 2[0.5(H + 1 - F)] - 1 = 2p(c) - 1. \quad (16)$$

Proportion correct is monotonically related to q and is therefore an equivalent sensitivity parameter of the model. If $p(c)$ equals 0.8, for example, the proportion q of trials falling in the "sure" D_1 and D_2 states is 0.6. Other trials lead to the uncertain state $D_?$ and elicit "yes" and "no" responses according the observer's response bias v.

A decision space consistent with the double high-threshold ROC is shown in Figure 8. The sensitivity parameter q equals the distance between the means of two rectangular distributions and also equals the proportion of each distribution that does not overlap the other. The S_1 and S_2 distributions each cover a region corresponding to two states: S_1 presentations can lead to either D_1 or $D_?$, S_2 presentations to either D_2 or $D_?$. Unlike the distributions of single high-threshold theory, those of double high-threshold theory are symmetric. There are three values of likelihood ratio: zero, one, and infinity.

Bias measures. Double-high threshold theory has two alternative bias measures (both discussed by Dusoir, 1975, 1983), each of which parallels one from detection theory. We first calculate the criterion location relative to the equal-bias point. In Figure 8, the center of the region of overlap is set to zero, and the criterion k is measured with respect to this origin. Then $H = p(c) - k$ and $F = 1 - p(c) - k$; solving these equations yields

$$k = 0.5[1 - (F + H)]. \quad (17)$$

The criterion is thus a simple linear transformation of $0.5(F + H)$, the *yes rate* when presentation probabilities are equal. Like the c of SDT and $\log(b)$ of Choice Theory, the yes rate reflects the location of the criterion relative to the halfway point between the S_1 and S_2 distributions. The relation between k and $p(c)$ is suggested by the similarity between Equation 17 and the corresponding expression for sensitivity in the equal presentation probability case:

$$p(c) = 0.5[1 - (F - H)]. \quad (18)$$

The false-alarm and hit rates are added in Equation 17 and subtracted in Equation 18, and the same transformation is applied to the result. We encountered a similar parallel for detection-theory models (compare, for example, Equations 3 and 4).

Like c and $\log(b)$, the yes rate measures the same distance along the decision axis whether the sensitivity measure is large or small. If we linearly transform k into a new variable k' that varies from 0 to 1, no matter what $p(c)$ is, we obtain

$$k' = 1/\{1 + [F/(1 - H)]\}, \quad (19)$$

that is, something that depends only on $(1 - H)/F$, the *error ratio.* The parameter k' is, in fact, equal to $1 - v$ and is therefore equivalent as a bias measure to v, which was nominated for this role by Snodgrass and Corwin (1988). Atkinson (1963) proposed a different transformation of the same statistic.

Isobias curves for the yes rate and the error ratio are shown in Figure 9. As might be expected from their interpretation in terms of criterion location, curves for the yes rate resemble SDT curves for criterion location, and curves for the error ratio resemble SDT curves for criterion location relative to sensitivity.

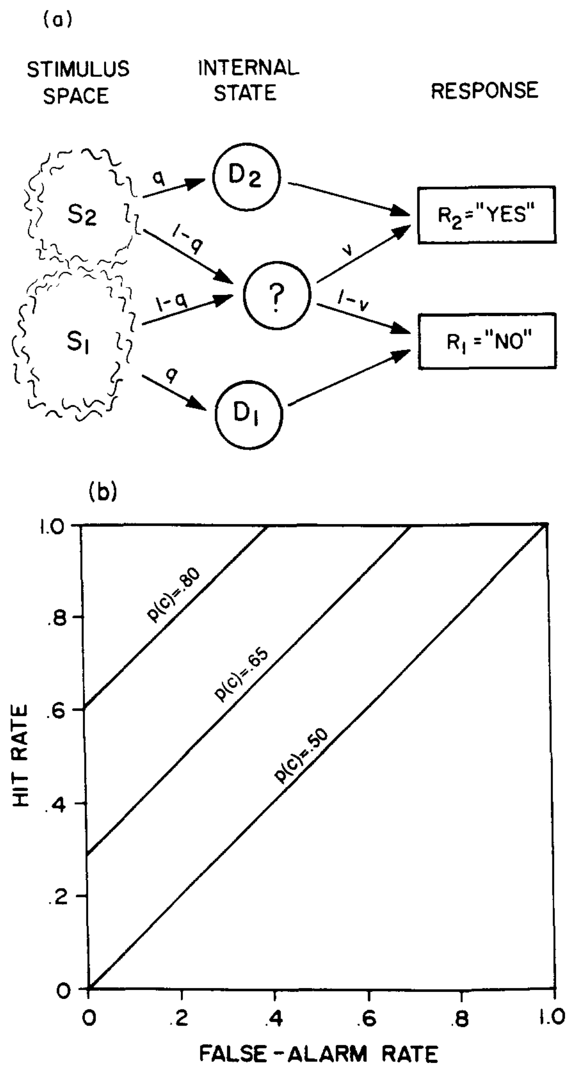


Figure 7. Double high-threshold theory: (a) state diagram, with three internal states; and (b) receiver operating characteristic curve.

Measures Based on Areas in ROC Space

Sensitivity Measure A′

The better an observer performs in a discrimination task, the more closely the ROC approaches the upper and left limits of ROC space. An appealing measure of sensitivity, therefore, is the area under the ROC; like proportion correct with equal presentation probabilities, this area increases from 0.5 for chance performance to 1.0 for perfect responding. This area equals the proportion correct by an unbiased observer in a two-interval forced-choice experiment (Green & Swets, 1966), giving it some credibility as a natural, nonparametric statistic.

If only one point in ROC space is obtained in an experiment, there are many possible ROCs on which it could lie, and some assumptions must be made to estimate the area under the ROC. The measure A' proposed by Pollack and Norman (1964) is a kind of average between minimum and maximum performance and can be calculated (Aaronson & Watts, 1987; Grier, 1971; Snodgrass & Corwin, 1988) by

$$A' = 0.5 + (H - F)(1 + H - F)/[4H(1 - F] \quad \text{if } H \geq F,$$

$$A' = 0.5 + (F - H)(1 + F - H)/[4F(1 - \mathrm{H})] \quad \text{if } H \leq F. \quad (20)$$

The ROCs implied by the use of A' were presented by Pollack and Norman (1964; see also Macmillan & Kaplan, 1985). They resemble detection-theory ROCs at low levels and threshold-theory ROCs when performance is good. Macmillan and Kaplan (1985) also showed that A' cannot be written in the form of Equation 1, so that the underlying distributions do not form a shift family.

Bias Measures

Two bias measures based on the geometry of ROC space have been proposed. In this section, we show that they are equivalent to each other and that they are not nonparametric.

Hodos's (1970) measure B'_H is defined by

$$B'_H = 1 - F(1 - F)/[H(1 - H)] \quad \text{if} \quad H \leq 1 - F,$$

$$B'_H = H(1 - H)/[F(1 - F)] - 1 \quad \text{if} \quad H \geq 1 - F. \quad (21)$$

Grier (1971) suggested the related statistic B''; like A', it must be modified if $H < F$ (Aaronson & Watts, 1987; Snodgrass & Corwin, 1988):

$$B'' = [H(1 - H) - F(1 - F)]/[H(1 - H) + F(1 - F)] \quad \text{if } H \geq F$$

$$B'' = [F(1 - F) - H(1 - H)]/[H(1 - H) + F(1 - F)] \quad (22) \quad \text{if } H \leq F.$$

The two statistics B'' and B'_H are equivalent (have identical iso-

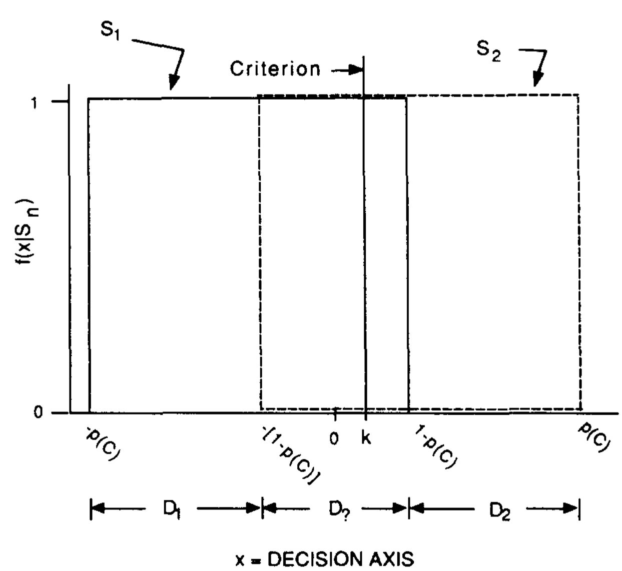
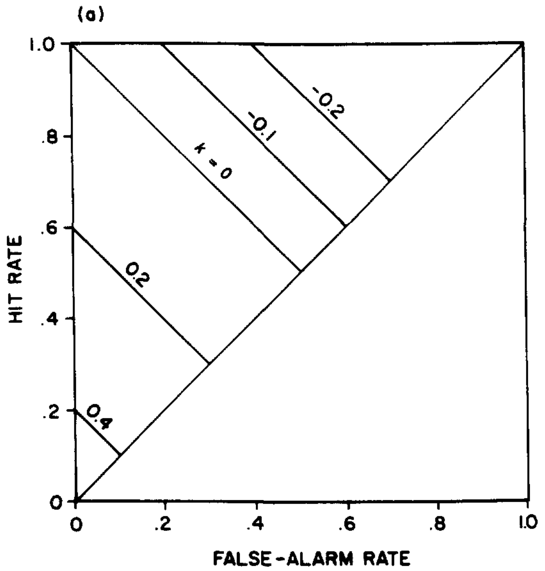


Figure 8. Rectangular underlying distributions consistent with double high-threshold theory.



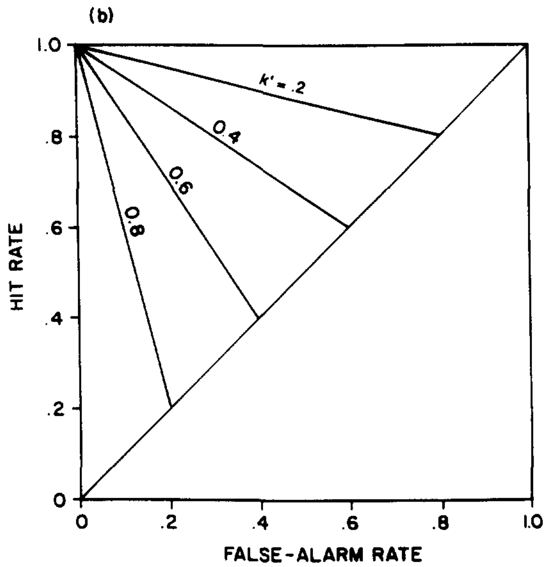


Figure 9. Isobias curves for (a) k (yes rate) and (b) k' (error ratio), the bias measures of double high-threshold theory.

bias curves) above the major diagonal, for one is a monotonic function of the other (see Appendix).

The Hodos (1970) or Grier (1971) statistic is often paired with A', just as β_G and d' are paired by users of Gaussian SDT, or b and α by users of Choice Theory. There is, however, no model that unifies A' and B'', as there is for the others. In fact, the relation of B'' to A' is entirely superficial, but its connection with another sensitivity index is firm: Both B'' and B'_H are monotonic functions of, and are therefore equivalent to, the likelihood ratio statistic β_L in the logistic detection-theory model (see Appendix). Their isobias curve, accordingly, is similar to that shown in Figure 3c. These nonparametric indexes are consistent with a parametric likelihood ratio measure. That the parametric–nonparametric distinction is vacuous has been argued previously[5] (Macmillan & Kaplan, 1985; Swets, 1986b; and with regard to B'_H, by Dusoir, 1975). The equivalence of two measures from opposite sides of this supposed divide serves to make the point concrete.

Comparing Response Bias Measures

We turn now to the task of comparing the many competing response-bias indexes. Faced with the analogous problem for sensitivity, Swets (1986b) was able to propose a single compelling touchstone—the form of the empirical ROC curve—by which the quality of a sensitivity statistic might be ascertained. It is natural to ask whether empirical isobias curves might similarly provide an answer to the present problem, especially as the shapes implied by the indexes we have considered are so varied.

Dusoir (1983) essayed this approach. Using data from a series of tone-in-noise detection experiments with stimuli of varying detectability, he plotted isobias curves based on rating responses. No measure was consistently supported; the three indexes rejected least often (for 9 of 21 data sets) were likelihood ratio, criterion location, and relative criterion location, all assuming the Gaussian model.

Dusoir's (1983) ambiguous results probably reflect not just bad luck but an important asymmetry between sensitivity and bias. To demonstrate the invariance of a sensitivity statistic, it is necessary to hold relevant stimulus characteristics constant while varying, say, instructions. In most applications, it is quite clear which stimulus characteristics are relevant. To show the invariance of a bias index, on the other hand, we must hold motivation constant while varying sensitivity. In general, our understanding of even simple aspects of motivation lags far behind our understanding of, for example, psychoacoustics. Experiments like Dusoir's may be too optimistic: It is hard to be sure that changes in stimulus factors do not also affect motivation.

When isobias curves are collected empirically, they are more likely to reflect truly constant bias if *roving* rather than fixed discrimination designs are used, that is, if stimulus pairs differing in sensitivity are intermixed, so that the observer is encouraged as much as possible to assign a constant meaning to an instruction or rating. Unfortunately, this precaution does not guarantee success. Wood (1976) conducted a roving voice-onset time experiment, but found that the Choice Theory parameter (here called b) varied systematically nonetheless. Wood proposed that his observers adjusted their bias from trial to trial on the basis of their estimate of the discriminability of the pair presented. If such strategies occur, they render empirical testing of isobias predictions very difficult.

The present obstacles to testing isobias predictions may eventually be overcome by improved technique or theoretical advance. As an example of the former, Snodgrass and Corwin (1988) collected isobias data for word recognition by varying the imageability of words; Snodgrass (personal communication, March 1989) pointed out that this method of manipulating dis-

[5] The conclusion applies only to single (H, F) pairs. In rating experiments, area under the ROC is truly nonparametric (Green & Swets, 1966). Estimates of this area from single ROC points, however, are not.

Table 1
Classification of Bias Measures

Form of underlying distribution	Bias parameter: Criterion location	Bias parameter: Criterion relative to sensitivity	Bias parameter: Likelihood ratio	Sensitivity parameter
Gaussian	c (4)	c' (5)	β_G (6)	d' (3)
Logistic	b (9)	b' (10)	[β_L (11) / B'' (21) / B'_H]	α (8)
Rectangular	[k (17) / yes rate]	[k' (19) / error ratio]	(NA)	$p(c)$ (18)

Note. Numbers in parentheses refer to equations in which measures are defined. When two or more measures are bracketed, they are equivalent. NA = not applicable.

criminability is relatively opaque to the subject, who is therefore unlikely to adopt strategies like that apparently used by Wood's (1976) observers. A useful theoretical advance would relate response bias to measurable aspects of the experimental situation. Progress on this route has been made, for operant situations using animals, by McCarthy and Davison (1981).

For now, however, the data provide little guidance in evaluating competing bias measures, and we use a different strategy. First, we note that all proposed statistics can be categorized according to two independent attributes of the underlying decision space. The psychophysical parameter estimated may be criterion location, criterion location relative to sensitivity, or likelihood ratio; and the form of the underlying distributions may be Gaussian, logistic, or rectangular. Combining these two partitions generates the taxonomy shown in Table 1.

Theoretical isobias curves for the measures in Table 1 are replotted in Figure 10, this time including the below-chance region of ROC space. The figure makes it clear that the decision characteristics of a bias measure are much more important than the distributional assumptions. Curves for criterion location (or the yes rate) are nonintersecting, curves for criterion relative to sensitivity (or the error ratio) fan out from a single point, and curves for likelihood ratio fan out from ROC space corners. Our comparison of bias measures therefore focuses on decision aspects but considers distributional assumptions where necessary. We discuss five characteristics of our candidate measures: (a) the form of the isobias curve, especially near and below chance; (b) the relation between a bias index's range of values and sensitivity; (c) the distortions introduced by averaging data across observers and the likely need to use this strategy; (d) available strategies for statistical analysis; and (e) the theoretical status of the measure.

Isobias Functions

We have argued that a minimum requirement for an isobias curve is that H be a monotonically decreasing function of F. All measures satisfy this condition in the upper-left quadrant of ROC space, where $H \geq 0.5$ and $F \leq 0.5$, but both relative criterion and likelihood ratio (and therefore B'' and B'_H) violate it elsewhere.

Two other regions of ROC space in which the curves of Figure 10 show different behavior are the chance line $H = F$ and the area below it. When sensitivity is zero, it is still meaningful to talk about bias: An observer for whom $H = F = 0.1$ clearly has a different bias from one operating at $H = F = 0.9$. The criterion location statistics do report different values of bias along the diagonal, but the likelihood ratio does not, because when underlying distributions of likelihood are identical this statistic is undefined.[6] The behavior of the relative criterion depends on the form of the distribution: For the threshold model all is well, but in the normal and logistic cases the index cannot be computed.

Below-chance behavior may seem uninteresting or even illogical, but statistical fluctuations can easily lead to such performance. Two points in ROC space that are symmetrically located across the chance line—(H, F) and (F, H)—should, intuitively, show the same (or very similar) bias. By this test, criterion location is again a superior measure, giving the same value for the two points. The error ratio $(1 - H)/F$ gives similar values near the minor diagonal and more disparate ones near the lines $H = 1.0$ and $F = 0.0$.

The least appealing measures are the relative criterion location with Gaussian or logistic distributions and the likelihood ratio statistics. For the logistic-distribution index $H(1 - H)/[F(1 - F)]$, for example, the bias of an ROC point is the inverse of the symmetrically opposite one (or, if logarithms are taken, the negative). These measures can be salvaged, however, by adjusting their meaning below the chance line (as has been suggested by Aaronson & Watts, 1987, and Snodgrass & Corwin, 1988, for B''; and by Waldmann & Göttert, 1989, for β_G).

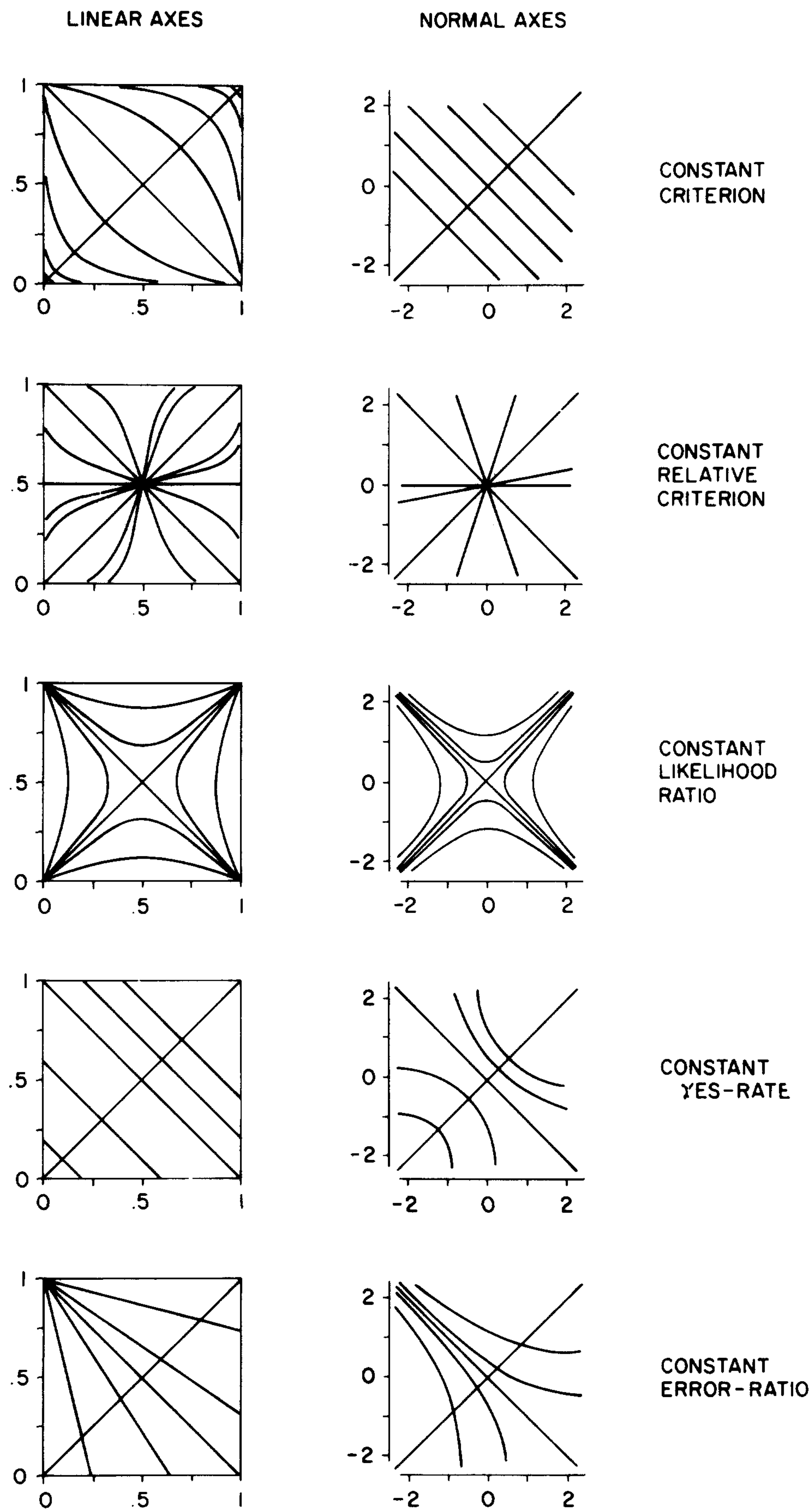
Relation Between Bias and Sensitivity

A desirable characteristic of a bias measure is that it be independent of an associated sensitivity measure. Ingham (1970) (and also Banks, 1970) argued that c had this property with respect to d', provided that the ROC curve had unit slope, and collected data suggesting that c was at least more nearly independent of d' than was the likelihood ratio β_G. (Recall that, according to Equation 7, SDT likelihood ratio is monotonically related to d'.) Snodgrass and Corwin (1988) extended the theoretical argument to other measures, plotting (in their Figure 5) the range of possible bias values at each level of sensitivity for logistic likelihood ratio (β_L), logistic criterion location (log (b)), and the double-threshold relative criterion (error ratio).[7] We add to their treatment by also considering logistic and Gaussian relative criterion and the yes rate.

By the standard of range independence, isobias curves should be parallel on some coordinates. Figure 10 shows that this con-

[6] In one interpretation, the likelihood ratio equals 1 for the zero-sensitivity case. However, in SDT the decision axis itself is considered to be likelihood ratio, so the decision space collapses to a single point.

[7] Snodgrass and Corwin (1988) also considered the range independence of B''. The equivalence of this measure to β_L was disguised in their presentation by the use of different sensitivity measures (effectively, log(α) and A').


LINEAR AXES
NORMAL AXES
CONSTANT CRITERION
CONSTANT RELATIVE CRITERION
CONSTANT LIKELIHOOD RATIO
CONSTANT YES-RATE
CONSTANT ERROR-RATIO
0
.5
1
-2
0
2

dition applies to criterion location with rectangular distributions (for which the curves are parallel on linear coordinates) and criterion location with logistic or Gaussian distributions (for which they are parallel on transformed coordinates). For the error ratio, isobias curves converge only at very high sensitivity. The condition is not met by relative criterion location or likelihood ratio (with logistic or Gaussian distributions), whose curves converge for both very high and very low sensitivities. The issue of range independence is clearly very close to the previously discussed issue of behavior near chance, and our evaluations by these two criteria are, predictably, similar.

Averaging Across Observers

In some applications, frequency data must be averaged across observers before summary statistics are computed. Such averaging can be more or less benign, in the sense that the statistic based on pooled data may or may not be close to the average of the same statistic based on individual data. Macmillan and Kaplan (1985) considered the effects of pooling on sensitivity statistics; which of our bias indexes are best in this sense?

In the simplest situation of this type (and the only one we consider here), two observers have the same bias but different sensitivity. Collapsed bias equals average bias whenever isobias curves are straight lines (of any slope) on linear coordinates. Both measures based on rectangular distributions, the yes rate and the error ratio, have this property. Next best are criterion location and relative criterion location, with logistic or Gaussian distributions. Isobias curves for these indexes are linear on z coordinates, and the error involved in averaging is moderate as long as sensitivities are not too different. (This conclusion is analogous to that of Macmillan and Kaplan, 1985, about sensitivity.)

Likelihood ratio is most susceptible to distortion due to averaging, because the curves for this statistic are most deeply curved. The discontinuities of these curves in going from above to below chance are especially disruptive.

Statistical Analysis

When bias is estimated from experimental data, it is important to know the variability of the estimate so as to construct confidence intervals and test hypotheses. Bias indexes whose statistical properties are simple or well-understood are to be preferred over those whose properties are not.

Criterion location, assuming logistic, Gaussian, or rectangular distributions, is statistically the most tractable conception of bias. For Gaussian distributions, Banks (1970) pointed out that, because $\mathrm{var}(c) = \mathrm{var}(d')/4$ (Equations 3 and 4), the variance of c can be estimated by using the approximation of Gourevitch and Galanter (1967). Assuming logistic distributions, $\log(b)$ can be estimated, and hypotheses about it tested, by using log-linear modeling (Macmillan & Kaplan, 1986), a standard technique. Under the rectangular-distribution model, the yes rate has a binomial distribution (Morgan, 1976).

We are not aware of any statistical studies of likelihood ratio or relative criterion. Certainly it would be valuable to find, by calculation or simulation, the distributions of these indexes.

Theoretical Status

In this final category, we allow theoretical properties of bias measures to intrude on our pragmatic evaluation. Specifically, we consider the historical importance of likelihood ratio and the virtues of bias measures derived from otherwise successful models of discrimination.

In SDT, likelihood ratio plays a central role. Green and Swets (1966, chap. 1) showed that comparing observations with a fixed likelihood ratio was the optimal decision procedure, under various definitions of *optimal.* In fact, the specific optimal level can be calculated from such experimental variables as presentation probabilities and payoffs. Also, optimal models can be specified in terms of likelihood ratio for discrimination designs too complex for a simple criterion construct (e.g., Macmillan, Kaplan, & Creelman, 1977).

Likelihood ratio was able to play such an important role historically, in spite of its imperfections, because of a significant limitation of early work in SDT: In most experiments, the alternative stimuli were fixed, so that sensitivity did not vary. When there is only a single ROC curve to consider, any measure of bias will serve. This history has led to the dutiful use of β_G in cases for which some measure of criterion location would be more instructive.

Second, how important is the assumed form of the underlying distribution in choosing a bias measure? In much of the foregoing discussion, the best measures have come with equal credentials from continuous (Gaussian or logistic) and discrete (or rectangular) distributions. Considering only the bias problem, neither type of distributional assumption is stronger than the other, and neither can be viewed as the default choice. But an argument can be made for preferring c and b, for example, over the yes rate: Other things being equal, one should use sensitivity and bias statistics that arise from the same rather than different underlying models. On the basis of sensitivity predictions (ROCs), normal and logistic models are superior to discrete or rectangular ones.

Considerations of theoretical consistency also have relevance for the many researchers who report A' as a measure of sensitivity and B'' as a measure of bias. Although both statistics can be motivated by examining the geometry of the unit square, we have seen that they have no legitimate relationship. A more principled choice is to pair B'' with logistic sensitivity.

Overall Comparison of Measures

In Table 2, we summarize our evaluation of bias statistics. For each test, a measure is rated "+" if it performs well with

Figure 10. Isobias curves for c, c', β_G, k, and k', on both linear and z coordinates; all curves include the region below the chance diagonal. (Logistic indexes, including B'', are omitted because of their similarity to Gaussian ones.)

Table 2
Comparison of Measures

Property	Bias measure				
	c, log(b)	k, yes rate	c', b'	k', error ratio	$\beta_G, \beta_L, B'', B'_H$
Isobias is monotonic	+	+	−	+	−
Isobias is well-behaved at chance	+	+	−	+	−
Isobias is well-behaved below chance	+	+	−[a]	0	−[a]
Range is independent of sensitivity	+	+	−	0	−
Averaging does not distort	0	+	0	+	−
Statistical analysis is well-known	+	+	0	0	0
Associated with desirable index of sensitivity	+	−	+	−	+
Useful in theory	0	−	0	−	+

Note. + = measure performs well with regard to property; − = measure performs poorly; 0 = marginal success.
[a] If statistic is inverted for below-chance performance, entry is +.

regard to that criterion, "−" if it performs poorly, and "0" for marginal success. As the bases of comparison are neither of equal weight nor wholly independent, we do not attempt to reduce our assessments to single numbers.

Still, one specific comparison seems justified: Criterion location, especially as calculated from normal or logistic models, deserves more attention as a bias index. Contrariwise, the importance of likelihood ratio in theory does not argue for its automatic use in data analysis.

Summary

We have considered eight distinct measures of response bias, in addition to some that are equivalent to one or another of these. All can be classified as reflecting criterion location, criterion location relative to sensitivity, or likelihood ratio and as assuming Gaussian, logistic, or rectangular underlying distributions.

Measures were judged superior if their isobias curves were monotonic, if they measured bias sensibly at and below chance, if their range was independent of sensitivity, if they were not distorted by averaging across observers, if they lent themselves to statistical analysis, and if they were of theoretical importance. By these standards, indexes of criterion location (c, b, and the yes rate) are the best response bias statistics. Detection-theory measures (those assuming logistic or normal distributions) have some advantages over those derived from threshold theory, but nonparametric indexes are actually equivalent to a logistic detection-theory statistic. An important problem for future research is to discover experimental manipulations for which specific bias measures are invariant.

References

Aaronson, D., & Watts, B. (1987). Extensions of Grier's computational formulas for A' and B'' to below-chance performance. *Psychological Bulletin, 102,* 439–442.

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

Banks, W. P. (1970). Signal detection theory and human memory. *Psychological Bulletin, 74,* 81–99.

Blackwell, H. R. (1953). *Psychophysical thresholds: Experimental studies of methods of measurement* (Engineering Research Bulletin 36). Ann Arbor: University of Michigan.

Dusoir, A. E. (1975). Treatments of bias in detection and recognition models: A review. *Perception & Psychophysics, 17,* 167–178.

Dusoir, T. (1983). Isobias curves in some detection tasks. *Perception & Psychophysics, 33,* 403–312.

Egan, J. P. (1958). *Recognition memory and the operating characteristic* (Tech. Note AFCRC-TN-58-51). Bloomington: Indiana University, Hearing and Communication Laboratory.

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

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

Grier, J. B. (1971). Nonparametric indexes for sensitivity and bias: Computing formulas. *Psychological Bulletin, 75,* 424–429.

Hodos, W. (1970). Nonparametric index of response bias for use in detection and recognition experiments. *Psychological Bulletin, 74,* 351–354.

Ingham, J. G. (1970). Individual differences in signal detection. *Acta Psychologica, 34,* 39–50.

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

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

Luce, R. D. (1963a). Detection and recognition. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. 1, pp. 103–189). New York: Wiley.

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

Macmillan, N. A., & Kaplan, H. L. (1985). Detection theory analysis of group data: Estimating sensitivity from average hit and false alarm rates. *Psychological Bulletin, 98,* 185–199.

Macmillan, N. A., & Kaplan, H. L. (1986). Log-linear models provide answers to signal-detection questions (abstract). *Bulletin of the Psychonomic Society, 24,* 342.

Macmillan, N. A., Kaplan, H. L., & Creelman, C. D. (1977). The psychophysics of categorical perception. *Psychological Review, 84,* 452–471.
McCarthy, D., & Davison, M. (1981). Towards a behavioral theory of 'as in signal detection. *Perception & Psychophysics, 29,* 371–382.
McNicol, D. (1972). *A primer of signal detection theory.* London: Allen & Unwin.
Morgan, B. J. T. (1976). The uniform distribution in signal detection theory. *British Journal of Mathematical and Statistical Psychology, 29,* 81–88.
Noreen, D. L. (1977). *Relations among some models of choice.* Paper presented at Mathematical Psychology Meeting, University of California, Los Angeles.
Pollack, I., & Norman, D. A. (1964). A nonparametric analysis of recognition experiments. *Psychonomic Science, 1,* 125–126.
Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General, 117,* 34–50.
Swets, J. A. (1986a). Form of empirical ROCs in discrimination and diagnostic tasks. *Psychological Bulletin, 99,* 181–198.
Swets, J. A. (1986b). Indices of discrimination or diagnostic accuracy: Their ROCs and implied models. *Psychological Bulletin, 99,* 100–117.
Waldmann, M. R., & Göttert, R. (1989). Response bias in below-chance performance: Computation of the parametric measure β. *Psychological Bulletin, 106,* 338–340.
Wood, C. C. (1976). Discriminability, response bias and phoneme categories in discrimination of voice onset time. *Journal of the Acoustical Society of America, 60,* 1381–1389.
Woodworth, R. S. (1938). *Experimental psychology.* New York: Holt.

Appendix

Relations Among Logistic-Distribution Statistics

Likelihood Ratio, Criterion Location, and Sensitivity

We wish to show that, according to Choice Theory, the likelihood ratio is given by Equation 12. Hit and false-alarm rates are as follows (Luce, 1963a, p. 123):

$$H = \alpha/(\alpha + b),$$
$$F = 1/(1 + \alpha b).$$

Substituting these expressions for H and F into Equation 12, we obtain

$$H(1 - H)/[F(1 - F)] = [(1 + \alpha b)/(\alpha + b)]^2. \quad \text{(A1)}$$

We now show that the right side of Equation A1 (and therefore the left side) equals the likelihood ratio. The logistic density function can be written

$$f(x, \mu) = e^{x-\mu}/(1 + e^{x-\mu})^2. \quad \text{(A2)}$$

The Choice Theory yes–no model is consistent with two logistic distributions having means of $\pm\log(\alpha)$ and a criterion located at $\log(b)$. Substituting $\mu = \pm\log(\alpha)$ and $x = \log(b)$ into Equation A2,

$$\begin{aligned}\beta_L &= f[\log(b), \log(\alpha)]/f[\log(b), -\log(\alpha)]\\ &= [(1 + \alpha b)/(\alpha + b)]^2. \quad \text{(A3)}\end{aligned}$$

Thus, combining Equations A1 and A3,

$$\beta_L = H(1 - H)/[F(1 - F)],$$

which is Equation 12.

Likelihood Ratio, B'', and B'_H

We wish to show that β_L, B'', and B'_H are equivalent, that is, have the same isobias curve. To do this, it is sufficient to show that they are monotonic functions of each other.

From Equations 21 and 12,

$$\begin{aligned}B'_H &= (\beta_L - 1)/\beta_L \quad \text{if} \quad \beta_L \geq 1\\ &= \beta_L - 1 \quad \text{if} \quad \beta_L \leq 1. \quad \text{(A4)}\end{aligned}$$

Equation A4 represents B'_H as a monotonic increasing function of β_L over its entire range, so these two measures are equivalent.

The relation between B'' and β_L is also monotonic. Combining Equations 22 and 12 (for above-chance performance),

$$B'' = (\beta_L - 1)/(\beta_L + 1). \quad \text{(A5)}$$

Because B'_H and B'' are each monotonic functions of β_L, they are themselves monotonically related. The actual relation is

$$\begin{aligned}B'' &= B'_H/(2 + B'_H) \quad \text{if} \quad H \geq 1 - F\\ &= B'_H/(2 - B'_H) \quad \text{if} \quad H \leq 1 - F.\end{aligned}$$

Received December 12, 1988
Revision received April 7, 1989
Accepted June 14, 1989 ■