Distributional Assumptions and Observed Conservatism in the Theory of Signal Detectability

LAURENCE T. MALONEY

Department of Psychology and Department of Electrical Engineering and Computer Science, University of Michigan

AND

EWART A. C. THOMAS

Department of Psychology, Stanford University

The theory of signal detectability typically fits data from Yes–No detection experiments by assuming a particular form for the noise and signal plus noise distributions of the Observer. Previous work suggests that estimates of the Observer's sensitivity are little affected by small discrepancies between the assumed distributions (usually Gaussian) and the Observer's true underlying distributions. Possibly for this reason, estimates of the Observer's choice of criterion or likelihood ratio suggesting suboptimal performance have also been taken at face value. It is, for example, commonly accepted that human Observers are *conservative*: They are said to choose criteria corresponding to likelihood ratios that are closer to 1 than the ratios produced by optimal criteria. We demonstrate that estimates of likelihood ratio can be markedly biased when the distributions assumed in estimation are not the Observer's true distributions. We derive necessary and sufficient conditions for an optimal Observer to appear conservative when fitted by distributions different from those governing his choices.

These results raise a fundamental question: What information about the Observer's underlying noise and signal plus noise distributions does the Observer's performance in a Yes–No detection task provide? We demonstrate that a small number of isosensitivity (ROC) curves completely determines the Observer's underlying noise and signal plus noise distributions for many familiar forms of the theory of signal detectability. These results open up the possibility of a semiparametric theory of signal detectability.   © 1991 Academic Press, Inc.

443

1. INTRODUCTION

The theory of signal detectability (TSD), in its simplest form, summarizes an Observer's decision-making under uncertainty via two parameters. One parameter, d', measures the Observer's sensitivity or capacity, and the other parameter, β, is a criterion that indexes the effect of motivational factors on the Observer's performance. Since the original presentation of TSD by Tanner and Swets (1954), various aspects of the theory have been examined, including (a) optimal statistical methods for estimating the two parameters (e.g., Altham, 1973; Dorfman, Beavers, & Saslow, 1973; Grey & Morgan, 1972; Ogilvie & Creelman, 1968), (b) the effects of varying the form of the probability distributions of noise and signal plus noise (Egan, 1975; Green & Swets, 1966/1974; Lockhart & Murdock, 1970; Morgan, 1976; Pollack & Hsieh, 1969; Swets, 1986a, 1986b), and (c) the effects of the subject's use of sub-optimal decision rules on the parameter estimates (Creelman & Donaldson, 1968; Thomas & Legge, 1970).

A general conclusion from studies under (a) and (b) is that estimates of the sensitivity parameter are quite robust under various distributional assumptions (e.g., Morgan, 1976). Possibly for this reason, estimates of the Observer's criterion or likelihood ratio have also been taken at face value. These estimates, generally obtained under the assumption of Gaussian equal-variance distributions of noise and signal plus noise, constitute the focus of many studies under (c). The estimated criterion is often found to be less extreme than the value that is optimal for the decision, a type of suboptimality that has been labeled "conservatism". One of the issues considered in the present article is the robustness of such conservatism under various distributional assumptions.

Human judgment and decision-making have been found to be suboptimal in a variety of contexts. While these findings are important as descriptions of behavior, they also provide the impetus for studies of the nature and causes of suboptimal behavior. For example, studies showing that judgments of event likelihood some-times violate the axioms of probability theory (Kahneman, Slovic, & Tversky, 1982) are invoked to explain why judgments between complex gambles can be suboptimal.

Other studies attempt to account for suboptimality by identifying a rule or heuristic to which the data seem to conform. In the case of conservatism, sub-optimality has been attributed to (a) a decision-making heuristic that fails to maximize expected payoff, or (b) an unstable criterion. Much attention has been given to probability matching as a heuristic leading to conservatism (Creelman & Donaldson, 1968; Parks, 1966; Thomas & Legge, 1970). Or, if the Observer varies the criterion from trial to trial but the data is fitted with the assumption that the criterion is fixed, the estimated criterion will differ from the optimal for many choices of sensory distributions. Attention has centered on whether there are plausible rules for this trial-by-trial variation such that the Observer's overall data conform to probability matching or some other heuristic. (See, for example, Dusoir (1974) and Thomas (1975) for discussions of the adequacy of the probability matching heuristic and other rules for criterion adjustment.)

In all these discussions, it is assumed that the experimenter knows the sensory distributions of the Observer. Laming (1986), however, has raised doubts concerning the use of Gaussian distributions for certain kinds of tasks. Studies of detectability, based on stimulus energy (or amplitude) in vision and audition lead to a variety of distribution forms: non-central χ^2, non-central F, and others. While Swets (1986b) finds that Gaussian TSD fits data from several different kinds of discriminative and diagnostic tasks well, at least one of his data sets seems to exhibit patterned deviations from the Gaussian model (his Fig. 6c, d).

If the distributional assumptions underlying data reduction are incorrect, it is possible to construct another account of observed suboptimal behavior. According to this alternative account, behavior might be optimal in reality but *appear* to be suboptimal only because the experimenter has incomplete information about the Observer. Such an account, originally suggested by Laming (1973, p. 95), would be a corrective to the prevailing view that human Observers are far from ideal, even in simple perceptual tasks (compare Healy & Kubovy, 1981; Kubovy, 1977).

In this article, we present two sets of mathematical results concerning the distributional assumptions commonly made in fitting empirical data via TSD. We first characterize precisely how choice of distribution influences parameter estimation for a broad class of TSD models. These results permit us to state necessary and sufficient conditions for an optimal Observer erroneously to be labeled conservative as a consequence of incorrect assumptions concerning the Observer's underlying distributions.

Second, we demonstrate that much weaker distributional assumptions than those commonly used suffice, in principle, to permit estimation of the Observer's sensitivity and criterion. We do so by characterizing what information about the Observer is available in the Observer's performance summarized as a small number of isosensitivy (ROC) curves.

In combination our results indicate that conventional distributional assumptions made in fitting Observer's data in signal detection tasks are (a) potential sources of error, and (b) unnecessary.

2. THE TSD OBSERVER: DEFINITIONS

The TSD Model. TSD provides a model of the Observer's signal detection behavior based on two assumptions. The first is that the sensory information available to the TSD Observer is a random variable X. On trials where the signal is absent, X has the cumulative distribution function (cdf) $F_N(x)$, called the *noise distribution*. On trials where the signal is present, X has the cdf $F_{SN}(x)$, the *signal plus noise distribution*. These two distributions are the *sensory distributions*. When the corresponding probability density functions (pdf's) exist, they will be denoted f_N and f_{SN}, respectively.

The second assumption is that the Observer has a *decision rule* used to decide whether noise alone or signal plus noise was present on a given trial. A decision

rule is a function $D(X)$ that specifies "Yes" or "No" for each value of X. For example (iff is an abbreviation for "if and only if"),

$$D(X, c) = \text{YES} \qquad \text{iff} \quad X > c$$

is a decision rule based on a sensory criterion: The Observer establishes the sensory criterion, c, and responds Yes (signal) or No (no signal) accordingly as the sensory information X is greater than or less than the criterion. Each possible choice of F_N, F_{SN}, and D is a candidate model for a TSD Observer. We will sometimes refer to such a model as a "TSD model," or as a "TSD," or as a "model Observer."

The goal of estimation is to *identify* the empirical Observer, to match the Observer's behavior to a specific model Observer. This goal is not achievable without drastic restrictions on the possible distributions F_N, F_{SN} (see Laming, 1973, pp. 100–101; Marley, 1971). The conventional approach to estimation is to assume that we know the Observer's distributions except for a small number of parameter values. F_N and F_{SN} are commonly assumed to be Gaussian distributions with the same standard deviation σ and with means at 0 and d, respectively (See Fig. 1).

Location TSDs and S3. We next define two subsets of all the possible TSD model Observers which we refer to as location TSDs and S3. Both classes are natural generalizations of familiar equal-variance Gaussian TSD where the Gaussian is replaced by some other distribution. S3 is a subset of the location TSDs which we use to explore the distributional robustness of TSD in the first part of this paper. We use the broader class of location TSDs when we determine what information about the Observer is contained in the Observer's behavior in a signal detection task.

DEFINITION 2.1.   A TSD model is a *location TSD* if and only if $F_{SN}(x) = F_N(x - d)$, $d \geqslant 0$: the signal plus noise distribution is a copy of the noise distribution shifted by d.
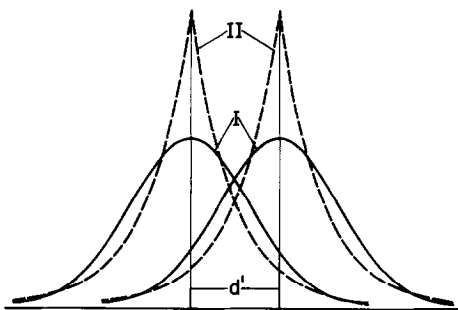

FIG. 1.   The sensory distributions for two location TSDs, the first based on the Gaussian (solid lines), the second based on the Laplacian (dashed lines).

As noted above, Fig. 1 shows the pdf's of the sensory distributions for two examples of location TSDs. The first is equal-variance Gaussian shown in solid lines. The second is similar, but the Gaussian distribution has been replaced by the Laplacian distribution shown in dashed outline. The sensory distributions of a location TSD are always identical except that F_{SN} is shifted by a distance d. To specify a location TSD we need specify F_N, the parameter d, and the parameters specifying the decision rule. In the case of the sensory criterion rule above, there is a single parameter c. A location TSD with a sensory criterion is completely specified by F_N, d, and c.

For convenience, we shall usually assume that the noise distribution F_N is normalized so that the median of F_N is at 0, and the upper quartile (75th percentile) of F_N is at 1. This normalization is the analogue of using the unit-variance, zero mean Gaussian as the noise distribution in fitting equal-variance Gaussian TSD. Since at least one of the distributions that interests us does not have a finite mean or variance, we normalize to the median and upper quartile. Also, for convenience, we sometimes restrict attention to those cdf's that have corresponding pdf's. This assumption reduces the size of the class of location TSDs, but not in any empirically significant way, since any cdf without a pdf can be approximated arbitrarily closely by cdf's that do have pdf's.

Next we define S3. S3 is a subset of the location TSDs chosen because of its convenient mathematical properties. The assumptions made in delimiting this class are technical in nature and some of them could easily be dropped without affecting the conclusions drawn. However, the results of the next section would not be appreciably strengthened by increasing the size of S3, and doing so would reduce the readability of the proofs offered.

DEFINITION 2.2.   A TSD model is in S3 if it is a location TSD whose sensory distributions further satisfy

(a)   the cdf, $F(x)$, can be differentiated (except possibly at a finite number of points) to yield a probability density function $f(x)$ that is unimodal (has a single maximum), and

(b)   the pdf's $f_N(x) \equiv f(x)$ and $f_{SN}(x) \equiv f(x - d)$ have support everywhere; i.e., $f_N(x) > 0$, and $f_{SN}(x) > 0$, for all x.

(c)   the pdf, $f(x)$, is symmetric about its median.

The initial letters of the three conditions (Support, Smoothness, Symmetry) give us the notation S3. Only the last assumption, symmetry, and the assumption of unimodality, are empirically significant restrictions. Any locations TSD can be arbitrarily well approximated by location TSDs that are smooth and have support everywhere. We speak of model Observers (or TSDs) as being in S3 and we will also refer to distributions as being in S3. The Gaussian and Laplace distributions (Fig. 1) are members of S3.

Figure 2 summarizes the classes of TSD models we use. In addition to S3 and the location TSDs, we have diagrammed the class of location-scale TSDs which we
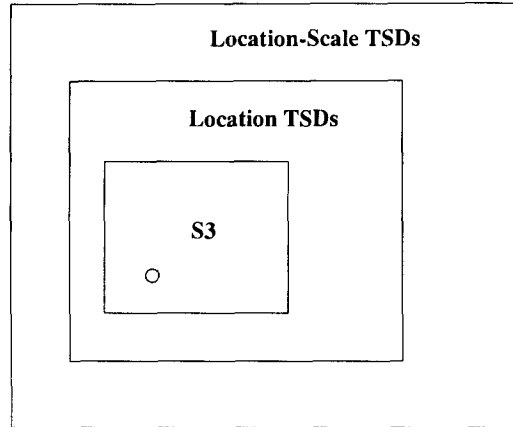
All TSDs

FIG. 2.   The classes of model TSD Observers discussed. The small circle represents the equal-variance Gaussian TSD Observer.

touch on briefly in a later section. They are model TSDs that generalize Gaussian TSD where the assumption of equal variance has been relaxed.

The Sensory Criterion Decision Rule.   If the distributions of a TSD are smooth enough to have probability density functions, we may define the likelihood ratio β corresponding to a given sensory criterion c to be $L(c) = f_{SN}(c)/f_N(c)$. For a location TSD, this simplifies to $L(c) = f(c - d)/f(c)$ where $f = f_N$. For many location TSDs, including the Gaussian, the likelihood ratio, β, is a monotonic function of c. For this class, the x-axis is monotonic with the likelihood ratio axis and the isosensitivity curve is concave everywhere. Our Observer, who uses the x-axis as the decision axis, would then be the ideal Observer of TSD for whom the decision axis is the likelihood ratio axis.

Throughout the analyses following we assume that the model Observer uses a sensory criterion rule, $X > c$. This assumption accords with observed human performance in signal detection tasks. When the TSD has the *monotone likelihood ratio property* (that the likelihood ratio β is a monotone increasing function of the sensory criterion c), this rule coincides with the likelihood criterion rule $L(X) > \beta$. Because of our interest in the biasing effects of distributional assumptions, we did not want to restrict the location TSDs considered to have the monotone likelihood ratio property. We have chosen instead to consider, at least initially, distributions "resembling" the Gaussian, in the sense that the corresponding isosensitivity curves are symmetric about the negative diagonal and are concave in the middle region where most task outcomes lie. These are the distributions that can be distinguished from the Gaussian family only with much difficulty. They include some, like the Cauchy family, that are not monotone likelihood ratio families. All of the proofs

below are valid when we restrict ourselves to the members of S3 and location TSDs exhibiting the monotone likelihood ratio property.

The Isosensitivity Curve. Last, we define the isosensitivity curve (receiver operating characteristic) representing the part of the TSD model that is observable empirically. The outcome of a Yes–No task is usually characterized by a false alarm rate, p_{FA}, and a hit rate, p_H, determined by equations $p_{FA} = 1 - F_N(c)$, and $p_H = 1 - F_{SN}(c)$. If the Observer is induced to vary c, with d held constant, across experimental conditions (say, by varying the payoffs for responses), the point (p_{FA}, p_H) traces out the isosensitivity curve introduced above.

Conservatism. For any combination of payoffs and prior odds on the presence or absence of the signal, there is an optimal value of β maximizing payoffs to the Observer. This value can be computed without knowledge of the sensory distributions. An optimal Observer will select a point along his or her isosensitivity curve corresponding to this optimal β. A conservative Observer shows a characteristic departure from optimal β:

DEFINITION 2.3. An Observer is *conservative* iff he or she consistently chooses values of β closer to 1 than the optimal value of β for the given payoffs and prior odds.

Green & Swets (1966/1974) exhibit data where optimal likelihood ratio ranges from 1/9 to 9/1 yet Observers' estimated β_e's range only from 1/2.20 to 2.17/1. Green & Swets conclude that: "[T]he observer tends to avoid extreme criteria: when the optimal β is relatively large, his actual criterion is not so high as the optimal criterion, and when the optimal β is relatively small, his criterion is not so low as the optimal criterion (p. 91)."
We introduce the notion of a radical Observer as well:

DEFINITION 2.4. An Observer is *radical* iff he or she consistently chooses values of β that are further from 1 than the optimal value of β for the given payoffs and prior odds.

Note that the conservative, radical, and optimal Observers all coincide in the special case where the optimal value of β is in fact 1.

3. CONSERVATISM: AN EXAMPLE

Overview. Can an experimenter's assumptions concerning the Observer's distributions lead him or her incorrectly to label an optimal Observer conservative? In this section, we demonstrate, by example, that failure to recognize the Observer's

actual distributions can produce biased estimates of β. Apparent deviations from optimal performance are large, and very similar to the deviations exhibited by human Observers.

Suppose that the noise and signal plus noise distributions that are used by the Observer are $f_o(x)$ and $f_o(x - d_o)$, respectively, and that the Observer's sensory criterion is c_o, with corresponding likelihood ratio $\beta_o = f_o(c_o - d_o)/f_o(c_o)$. An experimenter wishes to use the Observer's false alarm and hit rates to estimate d_o and β_o and, for this purpose, assumes that $f_e(x)$ is the Observer's noise distribution. The triple, $[p_{FA}, p_H, f_e(\cdot)]$, would then yield estimates, d_e and β_e, of d_o and β_o, respectively. If $f_e(\cdot) = f_o(\cdot)$, then the experimenter's estimates would be accurate. If, however, $f_e(\cdot) \neq f_o(\cdot)$, the question arises as to whether there may be systematic deviations of β_e and d_e from β_0 and d_o, respectively. In the next section, we offer a complete characterization of the patterns of failures to be expected. In this section, we demonstrate, by example, that such systematic deviations can produce the appearance of conservatism in β_e when the Observer is in fact choosing optimal β_o.

Assume that the Observer's true noise distribution is the Laplace distribution, $f_o(x) = \frac{1}{2} e^{-|x|}$ (Johnson & Kotz, 1970, Chap. 23), but that the distribution, $f_e(x)$, used in fitting the data is the familiar, unit-variance Gaussian distribution. Let us assume also that the signal plus noise distribution (for both the true distribution and the distribution used in fitting) differs from the corresponding noise distribution by only a location transformation. Figure 1 contrast the Observer's actual distributions and the experimenter's assumptions concerning the Observer. *We do not intend to recommend the use of a TSD based on the Laplacian distribution.* The latter is instead a device for illustrating how estimates of the criterion parameter c_o and likelihood ratio β_o can be affected by choice of distribution. Such demonstrations motivate the results of the next section where we derive necessary and sufficient conditions that an incorrect choice of distribution mimic conservative behavior.

For a given value of d_o, say $d_o = 1$, each choice of c_0 generates values for p_{FA} and p_H. These values, in turn, can be inserted into the formulae for the Gaussian, unit-variance model yielding estimates, β_e and d_e (usually called d').

Figures 3a and 3c show plots on logarithmic axes of the estimated criterion, β_e, versus the true criterion, β_o, for $d_o = 1$ and 3. (Figures 3e and 3f are discussed in a later section.) The estimates, β_e, lie above the 45° line for $\beta_o < 1$ and below it for $\beta_o > 1$. Estimates depart less from 1 than the true criterion values, a pattern indicating conservatism if the true values are optimal. Compare Fig. 4-3 in Green & Swets (1966/1974, p. 92).

It is interesting to note that, in an experimental situation, it would be difficult to determine that the Laplacian and not the Gaussian is the true underlying distribution. Figures 3b and 3d plot the residual estimated location parameter (the normalized difference between each d_e and the mean d_e) versus true $\log \beta_o$. Note that, in comparision with the magnitude of the mean d_e, the estimates vary only slightly, though systematically, as the value of β_o varies. It would be difficult to reject Gaussian theory TSD on the grounds that the estimated separation parameter varies with the choice of β_o.
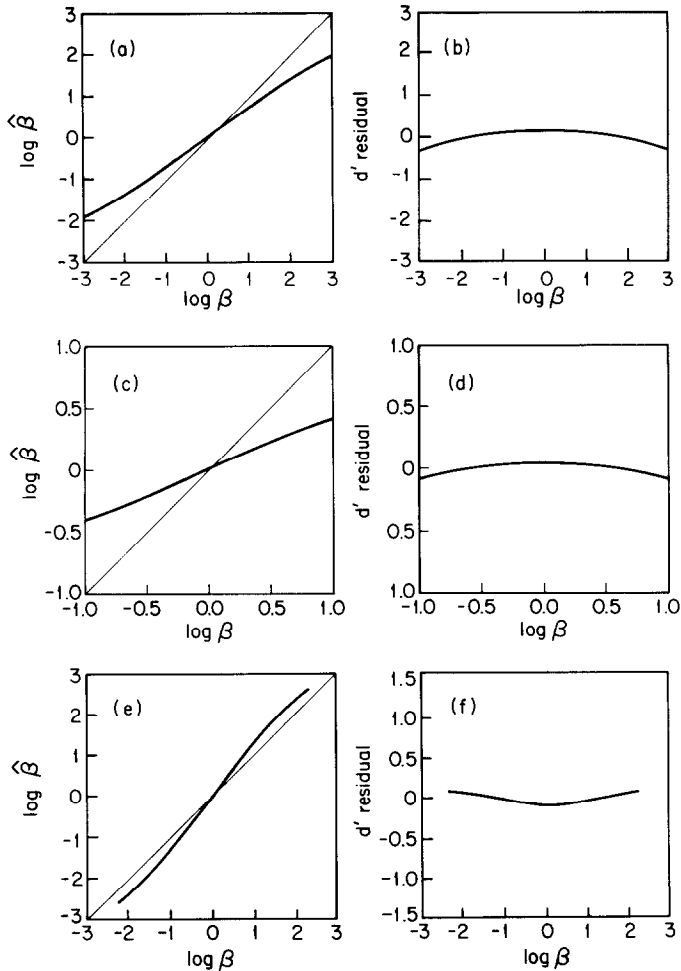
FIG. 3. (a) and (c): Plots of log β_e estimated vs true log β_o for $d_o = 1$ and 3, with true distribution the Laplacian. (b) and (d): Plots of residual d_e estimated vs true log β_o for the same two distributions as in (a) and (c). (e) and (f): The same as (a) and (b), except that the true distribution is the hyper-normal. The distribution assumed in fitting is always the Gaussian.

Figure 4 plots p_{FA} versus p_H as isosensitivity curves when f_o is Gaussian, Laplace, and Cauchy. They are fairly similar across the middle range of outcome values, and the shape of the contour provides few clues as to which of them best describes experimental data. Swets (1986a) makes a similar point concerning the isosensitivity curves of Gaussian TSD and TSDs based on other distributions.

DEFINITION 3.1.   We call the distribution f_o *locally conservative* with respect to the distribution f_e if Observers having the first distribution appear conservative for particular values of d_o and criterion c_o when responses are fit using f_e.
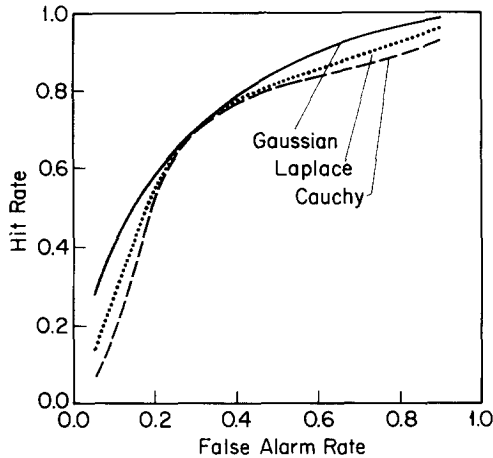
FIG. 4. Isosensitivity (ROC) curves for Gaussian, Laplace, and Cauchy location TSD models, constructed to pass through (0.3, 0.7): for the Gaussian, $d = 1.048$, for the Laplace, $d = 1.022$, and for the Cauchy, $d = 1.453$.

DEFINITION 3.2. We call the distribution f_o *everywhere conservative* with respect to the distribution f_e if Observers having the first distribution appear conservative for all values of d_o and criterion c_o when responses are fit using f_e. The meaning of the term conservative used without qualification should be clear from context as in "f_o is conservative with respect to f_e when d_o is greater than 2."

In the next section, we derive necessary and sufficient conditions that a distribution f_o be conservative with respect to a second distribution f_e either over the entire real line or any specified subinterval. We restrict attention to location TSDs with distributions drawn from S3 for the duration of the following section.

4. EVERYWHERE CONSERVATIVE RELATIONS

Overview. In this section we introduce the quantile map ψ and demonstrate that it completely characterizes biases introduced by the fitting process when the distribution assumed is not the Observer's true distribution. In particular, we state in terms of ψ the condition under which an optimal Observer will erroneously be judged conservative, as in the example of the last section. We demonstrate that the results of the last section are not a special property of the Laplacian equal-variance TSD. In fact, the results of this section permit us to "manufacture," using ψ, an indefinitely large number of model Observers who behave normatively and optimally in actuality, but who will be judged conservative by an experimenter assuming the Gaussian equal-variance Observer. The same construction allows us to generate Observers exhibiting radical (anti-conservative) behavior.

The conclusion we draw is that estimates of likelihood β and, to a lesser extent, estimates of sensitivity d, are not distributionally robust. Without knowledge of the Observer's true distributions, estimates of β should not be used to label an Observer conservative.

Throughout this section, we assume that all distributions are drawn from S3, as defined above. We intend to show that, within the class S3, there exist model Observers that will be diagnosed as behaving conservatively when in fact they are in fact behaving optimally. If we were to expand the class of model Observers to all location TSDs or even larger classes of TSDs, this problem of misdiagnosis would only worsen. Consequently, it suffices to develop the results below in the constricted, but mathematically tractable, class S3.

DEFINITION 4.1. Define $\psi = F_e^{-1} \circ F_o$ to be the *quantile map* between F_o and F_e.

Let f_* refer to either f_o or f_e, and let us adopt a similar notation for other quantities. Then since we are working in S3, $p_{FA} = 1 - F_*(c_*)$ and $p_H = 1 - F_*(c_* - d_*)$. Therefore, $F_e(c_e) = F_o(c_o)$. F_e is invertible (it is a cdf in S3). Hence we have

$$c_e = F_e^{-1} \circ F_o(c_o) = \psi(c_o). \tag{1}$$

A similar argument gives us

$$c_e - d_e = F_e^{-1} \circ F_o(c_o - d_o) = \psi(c_o - d_o), \tag{2}$$

expressing c_e and d_e as functions of the true parameters of the Observer.

The map ψ, the quantile map, is readily interpretable. It maps the quantiles of the distribution f_o to the corresponding quantiles of the distribution f_e. The median of f_o is mapped to the median of f_e, and so forth. (Q–Q plots, used to test whether a given sample of data is drawn from a hypothesized distribution, are related to the quantile map. See Gnanadesikan, 1977, pp. 198–199). For example, the map $\psi(x) = ax + b$ transforms a distribution f_o to f_e, which is just a shifted, scaled copy of itself. Since we required above that all noise distributions be normalized to have median 0 and upper quartile 1, we can easily show

PROPOSITION 4.2. *If ψ is a linear quantile map between normalized distributions, then $\psi(x) = x$.*

Proof. Substitute $\psi(0) = 0$ and $\psi(1) = 1$ in $\psi(x) = ax + b$. ∎

That is, the only linear map ψ between normalized distributions is the identity x. Nex we summarize some properties of ψ:

PROPOSITION 4.3. *Suppose that F_o and F_e are both drawn from S3. Assume both distributions to be normalized with medians at 0 and upper quartiles at 1. Let $\psi = F_e^{-1} \circ F_o$. Then*

(a) ψ *is continuous.*

(b) ψ *is strictly increasing (i.e.,* $\psi' > 0$)*, and therefore invertible.*

(c) ψ *is an odd function:* $\psi(-x) = -\psi(x);\ \psi(0) = 0.$

(d) $\psi'(x) = f_o(x)/f_e \circ \psi(x)$

(e) ψ' *is an even function:* $\psi'(x) = \psi'(-x).$

Proof.   (a)   We defined **S3** so that F_0 and F_e are differentiable and therefore continuous. Their composition, ψ, is also continuous.

(b) F_o and F_e are both strictly increasing because the corresponding pdf's have support everywhere. As F_e is continuous and strictly increasing, so is F_e^{-1}. Then ψ is the composition of strictly increasing functions and therefore also strictly increasing.

(c)   The symmetry assumption (evenness) in the definition of **S3** permits us to write

$$\psi(-x) = F_e^{-1} \circ F_o(-x) = F_e^{-1}(1 - F_o(x)) = -F_e^{-1}(F_o(x)) = -\psi(x). \qquad (3)$$

(d)   This result follows from differentiation of ψ using the chain rule and inverse function theorem (See Apostol, 1967, pp. 174–176, 252–253).

(e)   Apply the chain rule to $\psi(-x)$ to get $(d/dx)\,\psi(-x) = -\psi'(-x)$. Then differentiate both sides of (c) to get $(d/dx)\,\psi(-x) = -\psi'(x)$. Equate the two expressions. ∎

Next we express conservatism in terms of ψ. The likelihood ratio is

$$\beta_* = \frac{f_*(c_* - d_*)}{f_*(c_*)}. \qquad (4)$$

We have the following series of equivalences:

$$\beta_e < \beta_o$$

iff

$$\frac{f_e(c_e - d_e)}{f_o(c_o - d_o)} < \frac{f_e(c_e)}{f_o(c_o)} \qquad (6)$$

iff substituting from Eqs. (1) to (2),

$$\theta(c_o - d_o) < \theta(c_o), \qquad (7)$$

where

$$\theta(x) = \frac{f_e \circ \psi(x)}{f_o(x)} = \frac{1}{\psi'(x)}.$$

Equation (7) is a necessary and sufficient condition that $\beta_e < \beta_o$. Rewriting Eq. (7) in terms of ψ' (recalling that $\psi' > 0$),

$$\beta_e < \beta_o \qquad \text{iff} \quad \psi'(c_o) < \psi'(c_o - d_o). \tag{8}$$

Because of the iff condition in Eq. (8), our argument also gives necessary and sufficient conditions that $\beta_e = \beta_o$ and that $\beta_e < \beta_o$; i.e., the series of iff's (Eqs. (5)–(8)) above go through with the inequalities reversed (or replaced by equalities). Consequently, we have proven

PROPOSITION 4.4. *For TSDs drawn from* S3, *the quantile map* ψ *determines the relation between true and estimated* β.

We next formulate conservatism in terms of ψ.
Since the pdf's f_* drawn from S3 are unimodal and even, we can prove

LEMMA 4.5.

$$\beta_* > 1 \qquad \textit{iff} \quad c_* > \frac{d_*}{2}. \tag{9}$$

Proof. (We suppress the subscript $*$ in the subsequent proof for clarity's sake.) Since each of the noise and signal plus noise pdf's, $f(x)$ and $f(x - d)$, has a single mode at 0 and d, respectively, the difference $f(x) - f(x - d)$ is positive when $x = 0$, and negative when $x = d$. Therefore, since f is continuous, there is a value x_0, with $0 < x_0 < d$, such that $f(x_0) = f(x_0 - d)$. Therefore, $\beta = 1$ when $c = x_0$. We now show that $\beta \geqslant 1$ iff $c \geqslant x_0$.
By the unimodality of f, for $x_0 < c < x_0 + d, f(c - d) > f(x_0) > f(c)$, which implies that $\beta > 1$. For $c > x_0 + d, c - d > x_0 > 0$, and $f(c - d) > f(c)$, again implying that $\beta > 1$. A similar argument shows that $\beta < 1$ if $c < x_0$. This concludes the proof that $\beta \geqslant 1$ iff $c \geqslant x_0$. Since f is even, $x_0 = d/2$. ∎

The definition of local conservatism for a particular c_o and d_o can be restated as

$$\beta_e < \beta_o \qquad \text{iff} \quad c_o > \frac{d_o}{2}.$$

Recall that ψ is strictly increasing and, therefore, $\psi' > 0$.
Merging Eqs. (8) and (10), we get

$$\psi'(c_o) < \psi'(c_o - d_o) \qquad \text{iff} \quad c_o > \frac{d_o}{2}. \tag{10}$$

Equation (10) is a necessary and sufficient condition for local conservatism framed only in terms of ψ, c_o, and d_o. In particular, local conservatism depends on the underlying distributions f_* only through the quantile mapping ψ.

The following theorem gives a necessary and sufficient condition on ψ guaranteeing that f_o be everywhere conservative with respect to f_e. Before turning to the theorem, however, we rewrite Eq. (10) in a more convenient form. Making the substitutions $x = c_o - d_o/2$ and $\Delta = d_o/2$, we have the following characterization of local conservatism:

$$\psi'(x + \Delta) < \psi'(x - \Delta) \qquad \text{iff} \quad x > 0. \tag{11}$$

Note that $\Delta > 0$ always since $d_o > 0$. Therefore, f_o will be everywhere conservative with respect to f_e precisely when Eq. (11) holds for all x and all $\Delta > 0$.

Begin with a brief lemma that says, in effect, a smooth non-increasing function whose derivative is non-zero almost everywhere is actually a strictly decreasing function. This result is a technical result needed to state the subsequent theorem in its strongest possible form.

LEMMA 4.6. *Let $h: \mathbf{R} \to \mathbf{R}$ be a continuous, differentiable function and let A be an open interval of $\mathbf{R}$. Suppose that, for $x \in A$, $h'(x) \leqslant 0$ with $h'(x) = 0$ on a subset of A with Lebesgue measure 0. Then $x, y \in A$ with $x < y$ implies $h(x) > h(y)$.*

Proof. If $h(x) < h(y)$ then, by the mean value theorem, there is $x < \zeta < y$ with $h'(\zeta) = (h(y) - h(x))/(y - x) > 0$ contrary to the given. So $h(x) \geqslant h(y)$. Suppose $h(x) = h(y)$. Then $h(x) = h(\zeta) = h(y)$ for any $\zeta \in (x, y)$. Consequently $h'(\zeta) = 0$ for $\zeta \in (x, y)$. But then h' is zero across an open interval (x, y), a set of positive Lebesgue measure. Contradiction. Therefore, $h(x) > h(y)$. ∎

THEOREM 4.7. *Assume that the quantile mapping, ψ, is twice differentiable, i.e., that $\psi^{(2)}$ exists everywhere. Then f_o is everywhere conservative with respect to f_e iff $\psi^{(2)}(x) \leqslant 0$ for $x > 0$, with equality occurring on a set of Lebesgue measure 0.*

Proof. (The second derivative of ψ need not be negative everywhere in order to be conservative. If $\psi^{(2)}$ is negative except at a few points where it takes on the value 0, conservatism still holds.) Assume, first, that $\psi^{(2)}(x) \leqslant 0$ for $x > 0$, with equality occurring on a set of Lebesgue measure 0. We show that f_0 is everywhere conservative with respect to f_c by showing that Eq. (11) holds for all values of Δ and x.

Recall that since f_* is an even function, ψ is odd, ψ' is even, and $\psi^{(2)}$ is odd. Therefore, $\psi^{(2)}(x) \geqslant 0$ for $x < 0$ iff $\psi^{(2)}(x) \leqslant 0$ for $x > 0$. It suffices to derive the condition of Eq. (11) for $x > 0$. (As mentioned above, Δ is positive.) Then, for any $x > 0$, $\Delta > 0$, we wish to show that Eq. (11) holds. We distinguish two cases:

(a) $0 < x < \Delta$. Then $0 < \Delta - x < \Delta$ and

$$\psi'(x - \Delta) = \psi'(\Delta - x) > \psi'(\Delta) > \psi'(x + \Delta), \tag{12}$$

the last two inequalities following from two applications of the lemma with ψ' replacing h.

(b) $0 < \Delta < x$. Then $x - \Delta > 0$ and $\psi'(x - \Delta) > \psi'(x + \Delta)$ by the lemma.

We have shown that $\psi^{(2)}(x) \leqslant 0$ for $x > 0$, with equality only on a set of measure 0, implies that f_0 is everywhere apparently conservative with respect to f_e. Next we prove the converse.

If f_o is apparently conservative with respect to f_e, then Eq. (11) is valid for any x and $\Delta > 0$, and it can be shown that $\psi^{(2)}(x) \leqslant 0$ for $x > 0$ by letting $\Delta \to 0$. We need only shown that $\psi^{(2)}$ is not zero across some open interval. But then ψ' would be constant across an open interval, which contradicts Eq. (11). This proves the converse. ∎

A simple consequence of Theorem 4.7 is

COROLLARY 4.8. *Assume that the quantile mapping, ψ, is twice differentiable, i.e., that $\psi^{(2)}$ exists everywhere. Then if $\psi^{(2)}(x) < 0$ for $x > 0$, then f_o is everywhere conservative with respect to f_e.*

Failure to identify the correct distribution underlying the Observer's performance results in biased estimates of β_o that can mimic conservatism. The quantile mapping ψ determines whether or not conservatism appears. If the relevant properties of ψ can be computed from the Observer's performance, this potential source of apparent conservatism can be identified and then removed. One obvious place to look for information about ψ is in the estimates d_e of a fixed d_o, as c_o varies.

From Eqs. (1) and (2), d_e can be written as a function of c_o (d_o fixed):

$$d_e = \psi(c_o) - \psi(c_o - d_o), \tag{13}$$

which, after differentiating with respect to c_o and changing variables, yields

$$d'_e = \psi'(x + \Delta) - \psi'(x - \Delta).$$

Comparing this last equation with Eq. (11), we see that, for fixed d_o, conservatism occurs precisely when

$$d'_e < 0 \quad \text{iff} \quad c_o > (d_o/2). \tag{14}$$

But since f_e and f_o are even and unimodal, β_e and $\beta_o \geqslant 1$ for $c_o \geqslant (d_o/2)$.

THEOREM 4.9. *The distribution f_o is everywhere conservative with respect to the distribution f_e iff, as β_o varies, the estimate, d_e, of a fixed d_o is a maximum when $\beta_e = \beta_o = 1$, and decreases monotonically on both sides of this maximum.*

The slight bow in Fig. 3b and 3d illustrates Eq. (14) and this theorem. In principle, the systematic variation in d_e as β_o varies is perfectly diagnostic of the conservatism that is introduced by a failure to choose the proper distribution for the Observer. These figures also suggest that the variation in d_e may be slight and correspondingly difficult to detect in the data.

Further, to vary β_o in practice requires altering the conditions of the experiment (vary payoffs, prior odds). It is commonly assumed that such manipulations do not

affect d_o but perhaps they do. Any systematic change in d_e, then, may diagnose a failure of the distributional assumptions—or may diagnose a failure of the assumption that d_o is independent of payoffs, etc.

The conditions derived from Theorems 4.7 and 4.9 can be readily modified to predict what we called radicalism above, i.e., pairs of distributions, f_o and f_e, where the fitted β_e is consistently more extreme than the true β_o. A cursory examination of several commonly encountered distributions (e.g., the Laplace, Luce, and Cauchy distributions) discloses that all are conservative with respect to the Gaussian over the range of parameter values examined. Analogues of Theorems 4.7 and 4.9 of the previous section are easily proven for radicalism:

THEOREM 4.10. *Assume that the quantile mapping, ψ, is twice differentiable, i.e., that $\psi^{(2)}$ exists everywhere. Then f_o is everywhere radical with respect to f_e iff $\psi^{(2)}(x) \geqslant 0$ for $x > 0$, with equality occurring on a set of Lebesgue measure 0.*

THEOREM 4.11. *The distribution f_o is everywhere radical with respect to the distribution f_e iff, as β_o varies, the estimate, d_e, of a fixed d_o is a minimum when $\beta_e = \beta_o = 1$, and increases monotonically on both sides of this minimum.*

The results of this section are strong enough to permit computing a distribution that is everywhere conservative or everywhere radical with respect to any cumulative distribution function in S3. We need only pick a odd, increasing, continuous function ψ with $\psi^{(2)}(x) < 0$ for $x > 0$. Since, by definition, $\psi = F_e^{-1} \circ F_o$, we have $F_o = F_e \circ \psi$. Applying Theorem 4.7, we have shown

PROPOSITION 4.12. *Let F_e be any cdf in S3. Let ψ be any odd monotone increasing function with $\psi^{(2)}(x) < 0$ for $x > 0$. Then $F_o = F_e \circ \psi$ is a cumulative distribution function that is everywhere conservative with respect to F_e. When F_e is the Gaussian cumulative distribution function, denoted Φ, then $F_o = \Phi \circ \psi$ is a cumulative distribution function that is everywhere conservative with respect to the Gaussian. $\psi(x) = x^{1/3}$ satisfies the proposition, and therefore $\Phi(x^{1/3})$ is a distribution everywhere conservative with respect to the Gaussian.*

The implication of the proposition is that, from the experimenter's point of view, there exist an infinitude of possible Observers F_o that are optimal but appear conservative, as many as there are strictly increasing, odd functions concave on the positive reals. The example of the previous section is not an isolated property of the Laplace distribution. Given the results above, we can manufacture any number of examples by choice of ψ.

Similarly, to create a distribution that is everywhere radical with respect to the Gaussian, we need only pick a odd, continuous function ψ with $\psi^{(2)}(x) > 0$ for $x > 0$ such as x^3.

5. A Necessary Condition for Conservatism/Radicalism

Overview. In this section, we obtain a simple necessary condition for apparent conservatism based on the curvature at 0 of the two distributions. This condition permits rapid screening of distributions for possible radicalism/conservatism and allows us to exhibit a distribution that is locally radical with respect to the Gaussian. It also gives some insight into which distributions are not conservative with respect to the Gaussian. The conditions on ψ above are currently the only necessary and sufficient conditions for radicalism/conservatism.

We also demonstrate that an index of the curvature of f_o at 0 must be greater than that of f_e if f_o is to be everywhere apparently conservative with respect to f_e. The reverse inequality is a necessary condition for radicalism. As above, we are concerned with location TSDs for which the pdf's are drawn from S3. In addition, we now assume that the second derivatives, $f_*^{(2)}$, are defined in a neighborhood of 0.

The index we use is

$$\kappa_0(f) = \frac{|f^{(2)}(0)|}{f(0)^3}, \tag{15}$$

an inverse measure of the "flatness" of the pdf, $f(x)$, near 0. A low value of $\kappa_0(f)$ indicates that $f(x)$ is relatively flat near its mode, and a high value of $\kappa_0(f)$ indicates that $f(x)$ is sharply peaked. We continue the notation of the previous section.

Suppose that the Observer sets the decision criterion c_o at the median of the (symmetric) signal plus noise distribution. Then $c_o = d_o$, p_{FA} is $1 - F_o(d_o)$, and the hit rate p_H is p.5. The likelihood ratio at criterion is

$$\beta_o = \frac{f_o(0)}{f_o(d_o)}. \tag{16}$$

The Observer's responses are fitted by assuming a second even distribution f_e. By symmetry, the decision criterion c_e must be placed at the median of the signal distribution since $p_H = 0.5$; also, the estimated likelihood at criterion would be

$$\beta_e = \frac{f_e(0)}{f_e(c_e)}. \tag{17}$$

Since the modes of f_* are at 0, both likelihood criteria, β_*, exceed 1, and conservatism would not be observed if $\beta_e > \beta_o$. Let

$$y_e(d_o) = \frac{f_e(c_e(d_o))}{f_e(0)},$$

the reciprocal of β_e. (Note that we have made explicit the dependence of c_e on d_o.)

Let

$$y_o(d_o) = \frac{f_o(d_o)}{f_o(0)},$$

i.e., the reciprocal of β_o. Since zero is a mode of both distributions, f_*, $y_*^{(2)}$ are both negative at 0. Further $y_*(0) = 1$, $y_*'(0) = 0$, and

$$y_o^{(2)}(0) = \frac{f_o^{(2)}(0)}{f_o(0)}, \tag{18}$$

$$y_e^{(2)}(0) = \frac{f_e^{(2)}(0)[f_o(0)]^2}{[f_e(0)]^3}. \tag{19}$$

Consequently, y_e would be less than y_o for some range of d about 0, if $y_e^{(2)}(0) < y_o^{(2)}(0) \leqslant 0$. This last inequality implies (substituting from Eqs. (18) and (19)):

$$\frac{f_e^{(2)}(0)}{[f_e(0)]^3} < \frac{f_o^{(2)}(0)}{[f_o(0)]^3}; \tag{20}$$

that is,

$$\kappa_0(f_e) > \kappa_0(f_o). \tag{21}$$

We have proven

THEOREM 5.1. *If κ_0 for the true model is less than κ_0 for the fitted model, the estimated likelihood criterion would be* more *extreme than the true likelihood criterion for at least some (p_{FA}, p_H) pairs.*

Consequently,

COROLLARY 5.2. *If f_o is everywhere conservative with respect to f_e, then $\kappa_0(f_0) > \kappa_0(f_e)$, a necessary condition for conservatism.*

EXAMPLES. Let f_e be the unit-variance Gaussian distribution. It can be shown that the value of κ_0 for the Gaussian is 2π. κ_0 for the Laplace distribution discussed previously is $\infty > 2\pi$, consistent with our conclusion above that the Laplace distribution is apparently conservative with respect to the Gaussian. κ_0 for the Logistic distribution is 8 and for the Cauchy distribution is $2\pi^2$. Hence these two distributions cannot be everywhere radical with respect to the Gaussian. Analyses similar to those performed for the Laplace distribution (not reported here) indicate that the Logistic and Cauchy are conservative with respect to the Gaussian. We next use the κ_0 criterion to provide an example of a pdf that is radical with respect to the Gaussian over at least part of its range.

The pdf, $f_o(x) = A \exp[-(|x|/B)^3]$, $A \approx 0.315$, $B \approx 1.68$, which we call the *hyper-*
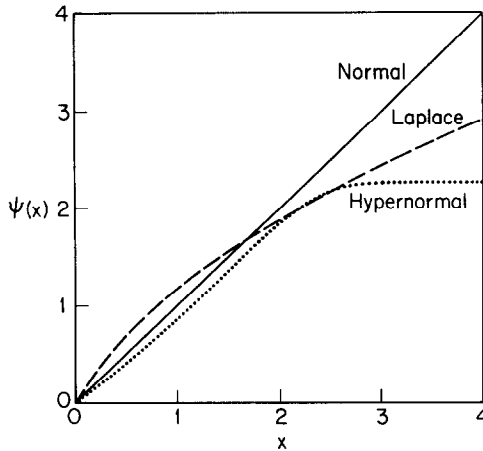
FIG. 5. Plots of the quantile functions ψ between three distributions (Gaussian, Laplacian, and hypernormal) and the Gaussian distribution. See text for details.

normal distribution, has a mean of 0 and a variance of 1. For this pdf, κ_0 is 0, implying that this pdf would not be everywhere conservative with respect to the Gaussian pdf. Indeed, Fig. 3e shows the radicalism of the distribution f_o with respect to the distribution f_e for this choice of distributions. Figure 3f shows that the estimate, d_e, has a local minimum when $\beta_o = 1$, consistent with Theorem 2.

For the purpose of comparison, we have plotted in Fig. 5 the quantile mapping, ψ, for f_o Gaussian, Laplacian, and hypernormal, and f_e Gaussian. As expected from Theorem 1, the first function is linear, the second is concave, and the third is not concave for all x. Other calculations show that when, for example, f_e is the Cauchy pdf and f_o is, in turn, Gaussian, Laplacian, and hypernormal, ψ is always convex increasing, suggesting that these choices of f_o are radical with respect to the Cauchy pdf. This convexity of ψ, coupled with the fact that κ_0 is greater for the Laplacian pdf than for the Cauchy pdf, demonstrates that the above condition on κ_0 is not sufficient even for local conservatism. The condition on κ_0 is necessary for ψ to be concave as we prove below. The usefulness of this condition stems from the fact that the shape of the pdf is more easily interpretable than the curvature of ψ. So far we have been unable to obtain a sufficient condition that is as easily interpretable as the condition on κ_0.

PROPOSITION 5.3. *If ψ is concave, then $\kappa_0(f_e) > \kappa_0(f_o)$.*

Proof. From the definition of ψ, $F_e \circ \psi = F_o$. On differentiating this equation twice, we get $\psi'(x) = f_o(x)/f_e \circ \psi(x)$, and

$$\psi^{(2)}(x) = \frac{f_o(x)^2}{f_e \circ \psi(x)} \left[\frac{f_o'(x)}{f_o(x)^2} - \frac{f_e'(x)}{f_e(x)^2} \right].$$

Since 0 is the median of the pdf's, $\psi(0) = 0$. Therefore, for x close to 0,

$$\psi(x) \approx \psi'(0)x = x\,\frac{f_o(0)}{f_e(0)}.$$

Also, since f_* are even and unimodal, $f'_*(0) = 0$, and $f^{(2)}(0) \leqslant 0$. Now $\psi^{(2)} < 0$ implies that

$$\frac{f'_o(x)}{f_o(x)^2} < \frac{f'_e(x)}{f_e(x)^2}.$$

On expanding each side of the inequality in a Taylor series about $x = 0$, substituting for $\psi(x)$, and $f'_*(0)$, and noting that $f^{(2)}_* \leqslant 0$, we get

$$\kappa_0(f_o) > \kappa_0(f_e),$$

as required.   ∎

6. DISTRIBUTIONAL INFORMATION IN THE OBSERVER'S ISOSENSITIVITY CURVES

In the preceding sections, attention is confined to the subset S3 of the full set of location TSDs. We now drop that restriction and consider all location TSDs.

Biases introduced by incorrectly specifying the Observer's underlying distribution have led us to examine what can be inferred about the Observer's underlying distribution from detection data. It is often stated or implied that an isosensitivity curve completely characterizes the Observer's performance. (See, for example, Coombs, Dawes, & Tversky, 1970, p. 175). We believe that the meaning of this statement is that, once a given f_o is assumed, the isosensitivity curve is uniquely determined by d_o. Each isosensitivity curve corresponds to exactly one value of d_o. However, one can inquire after a deeper meaning to the statement and ask, for example, "What information about the underlying pdf f_o is available in the subject's performance as summarized by the isosensitivity curve?"

It is known that a given isosensitivity curve can be generated by infinitely many pairs of stimulus distributions, $F_N(x)$, and $F_{SN}(x)$ (Laming, 1973, pp. 100–101; Marley, 1971). Marley (1971) also provides necessary conditions for a set of isosensitivity curves to be generated by a location TSD. In this section, we consider the simpler question of whether two distinct location TSDs can generate the same set of isosensitivity curves. The stimulus distributions are *not* assumed to be even (symmetric about the median). We show, after some intermediate results, that, given two isosensitivity contours, there is at most one location TSD that could generate them both. We also show that the isosensitivity contour generated by a given location TSD can be generated also by at least one other location TSD. This last result implies that the minimum number of isosensitivity contours required for the characterization of a location TSD is two.

Levine (1970) provides necessary and sufficient conditions for the existence of transformations that render curves parallel. Location TSDs have cumulative distribution functions that are parallel curves in Levine's sense and the requirement that there be only one location TSD generating a given set of isosensitivity curves can be construed withing Levine's framework as an assertion concerning the uniqueness of the transformation rendering certain cumulative distribution functions parallel. Levine's necessary and sufficient conditions must, of course, be equivalent to the necessary and sufficient conditions we derive below, but we have found no direct path from his result to ours. Since Levine's results may extend to the location-scale TSD case discussed in the dnext section, while our results, so far, cannot, it would be desirable to express Levine's results in terms appropriate to TSD.

An isosensitivity curve for a location TSD is determined by specifying the underlying distribution, f_1, and the location parameter, d_1, separating the noise and signal plus noise distributions. Each value of d_1 generates a distinct isosensitivity curve. Let us denote the set of isosensitivity curves generated by a particular location TSD based on f_1 as $I(f_1)$. Let f_2 be a second distribution such that $I(f_1) = I(f_2)$. That is, for any isosensitivity curve generated by f_1 with location parameter d_1, there is d_2 such that f_2 with location parameter d_2 sweeps out the same isosensitivity curve. Let F_1 and F_2 be the cdfs of f_1 and f_2, respectively, and suppose that both distributions have been normalized to have a median of 0 and a semi-interquartile range of 1. We next show that f_1 and f_2 are identical: The full set of isosensitivity curves produced from a location TSD determines the underlying family distribution.

THEOREM 6.1. $I(f_1) = I(f_2)$ iff $f_1 = f_2$.

Proof. The isosensitivity curves can be ordered monotonically with respect to d_1 (for the location TSD based on f_1) and with respect to d_2 (for the location TSD based on f_2). Therefore, two distinct values of d_1 correspond to distinct isosensitivity curves and, therefore, to distinct values of d_2. This shows that d_2 is a well-defined function of d_1. But these two variables can be interchanged in the argument to show that d_1 is a continuous surjective (onto) function of d_2. This implies that there is a monotonic function m such that $d_2 = m(d_1)$. Then Eqs. (1) and (2) imply

$$\psi(x - d_1) = \psi(x) - m(d_1), \tag{22}$$

where $\psi = F_2^{-1} \circ F_1$. Setting $d_1 = 0$ in the previous equation gives us $m(0) = 0$, and it can be shown that $\psi(0) = 0$. On putting $x = 0$ in Eq. (22), we have

$$\psi(d_1) = -m(-d_1), \tag{23}$$

so that $m(x) = -\psi(-x)$. Therefore, substituting $y = -d_1$ and substituting for m in Eq. (22), we obtain

$$\psi(x + y) = \psi(x) + \psi(y), \tag{24}$$

and, consequently, $\psi(x) = ux$ for some constant u (Aczel, 1966, pp. 31, 32). Since f_1 and f_2 are normalized to have the same semi-interquartile range, $u = 1$ (shown above), ψ is the identity, and $f_1 \equiv f_2$ as desired. ∎

This result can be strengthened considerably. Next we demonstrate that as few as two isosensitivity curves, corresponding to appropriate choices of the parameter d for a location TSD Observer, completely specify that Observer. (In the next section, we demonstrate that three isosensitivity curves suffice to determine a particular type of location-scale TSD Observer.) In practice, then, a small number of isosensitivity curves, corresponding to different location parameters d_o for a TSD Observer, determine the Observer. All location TSDs share the degenerate isosensitivity contour corresponding to $d = 0$. For convenience in the presentation below we ignore this contour: "isosensitivity contour" means "non-degenerate isosensitivity contour."

Also, in stating the theorems below, it is convenient for us to refer to the zero function as a "periodic function."

First we prove a simple lemma.

LEMMA 6.2. *Let g be a real-valued function, and suppose $g(x + a) = g(x) + b$, for all x, where a and b are constants. Then g is the sum of a linear function, g_L, and a periodic function, g_P, either of which may be zero. Further, if h_L and h_P are a linear and a periodic function that also sum to g, then $h_L = g_L$ and $h_P = g_P$: the decomposition is unique.*

Proof. Aczel proves this result in his discussion of Abel's Equation (Aczel, 1966). Let

$$g_L(x) = \frac{b}{a}x + g(0), \tag{25}$$

$$g_P(x) = g(x) - g_L(x). \tag{26}$$

Then g_P is either periodic, with a being a multiple of the period, or zero, because

$$g_P(x + a) = g(x + a) - g_L(x + a) = g_P(x). \tag{27}$$

To prove uniqueness, one can evaluate $g_L + g_P$ and $h_L + h_P$ at a and $2a$. Both periodic functions vanish, $g_L = h_L$ necessarily, and uniqueness follows. ∎

Now suppose that there are two isosensitivity curves that are generated by (a) the sensitivity parameter values d_1 and d_1' within the location TSD, f_1, and also by (b) the sensitivity parameter values d_2 and d_2' within the location TSD, f_2. Let c_1 and c_2 denote criterion values within the models f_1 and f_2, respectively.

Let F_1 and F_2 be the cdf's corresponding to f_1 and f_2, and let $\psi = F_2^{-1} \circ F_1$ be the quantile function mapping f_1 to f_2, where it is assumed that the distributions have been normalized to have median 0 and semi-interquartile range 1. We denote by $I[f, d]$ the isosensitivity curve within a location TSD, f, generated by the sensitivity parameter d.

DEFINITION 6.3.   For two positive real numbers, x and y, we say that x and y are *relatively prime* if there is no real z with x/z and y/z both integers.

Some illustrations of relative primeness: The real numbers 3.2 and 1.6 are not relatively prime; the numbers 1 and π are. If two numbers are picked independently at random from the uniform distribution on $[0, 1]$, they will be relatively prime with probability 1. If two numbers are picked from any smooth joint distribution on $\mathbf{R}^2$, the real plane, they will be relatively prime with probability 1.

Next we show that, if two location TSDs share two isosensitivity curves, then the two TSDs are identical.

THEOREM 6.4.   *If* $\mathbf{I}[f_1, d_1] = \mathbf{I}[f_2, d_2]$ *and* $\mathbf{I}[f_1, d_1'] = \mathbf{I}[f_2, d_2']$, *where* d_1 *and* d_1' *are relatively prime, then* $f_1 \equiv f_2$.

Proof.   As in an earlier section, it can be shown, for the first isosensitivity curve, that

$$\psi(c_1) = c_2, \tag{28}$$

and

$$\psi(c_1 - d_1) = c_2 - d_2, \tag{29}$$

so that

$$\psi(c_1 - d_1) = \psi(c_1) - d_2. \tag{30}$$

Applying the lemma, we see that $\psi = \psi_L + \psi_P$, where either ψ_P is zero or d_1 is an integral multiple of its period. Repeating the above construction for the second isosensitivity curve, we conclude that either ψ_P is zero or else both d_1 and d_1' are integral multiples of its period. Since d_1 and d_1' are relative prime, ψ_P must be zero and ψ must be a linear transformation. Since f_1 and f_2 are noth normalized, ψ must be the identity: $f_1 = f_2$ as desired.   ∎

As part of the proof of Theorem 5, we proved a result required for the next theorem.

LEMMA 6.5.   *If* $\mathbf{I}[f_1, d_1] = \mathbf{I}[f_2, d_2]$, *then the quantile map* ψ *relating* f_1 *and* f_2 *can be decomposed into the sum of a linear and a periodic map of period* d_1.

We next show that the number of isosensitivity curves specified in the conditions of Theorem 6.4 cannot be reduced to one: There are isosensitivity curves generated by two distinct location TSDs.

THEOREM 6.6.   *There are distinct location TSDs generating the same isosensitivity contours.*

Proof.   We need only construct distinct distributions f_1, f_2, with parameters d_1, d_2, respectively, having the same isosensitivity curve: $\mathbf{I}[f_1, d_1] = \mathbf{I}[f_2, d_2]$. Let

$\psi(x) = x + (1/4\pi) \sin(2\pi x)$. Note that ψ is odd, continuous, and increasing. ψ is also the sum of a linear and a periodic function of period 1. For any normalized F_2, let $F_1 = F_2 \circ \psi$. Let $d_1 = d_2 = 1$. Note that F_1 is also normalized. The two distributions are distinct as both are normalized yet $F_1(1/4) \neq F_2(1/4)$.

Now we want to show that if (p_{FA}, p_H) is on the isosensitivity curve $I[f_1, d_1]$, then it is also on the isosensitivity curve $I[f_2, d_2]$ and conversely. If (p_{FA}, p_H) is on the isosensitivity curve $I[f_1, d_1]$, then there is an x with

$$F_1(x) = 1 - p_{FA}$$

and

$$F_1(x - 1) = 1 - p_H.$$

But then, substituting $F_1 = F_2 \circ \psi$,

$$F_2(\psi(x)) = 1 - p_{FA}$$

and

$$F_2(\psi(x - 1)) = 1 - p_H.$$

Since $\psi(x - 1) = \psi(x) - 1$, the second equation becomes

$$F_2(\psi(x) - 1) = 1 - p_H,$$

and, substituting $y = \psi(x)$, we see that (p_{FA}, p_H) is on the isosensitity curve $I[f_2, d_2]$. Since ψ is invertible, we can reverse the roles of the two distributions and prove that every point on $I[f_2, d_2]$ is on $I[f_1, d_1]$ as well. The two isosensitivity curves coincide, yet f_1 is distinct from f_2. ∎

The construction used in the previous theorem proves the stronger result,

THEOREM 6.7. *If* **I** *is an isosensitivity curve generated by a location TSD, then there are infinitely many location TSDs that also generate the contour* **I**.

Proof. The ψ used above may be used to generate a second location TSD. There is a distinct location TSD for each of the infinitely many possible choices of ψ that is the sum of a linear and periodic function of period 1, that is odd and monotone increasing, and that carries normalized distributions into normalized distributions. ∎

7. AN EXTENSION TO LOCATION-SCALE TSDs

The class of location TSDs is a generalization of equal-variance Gaussian TSD relaxing the assumptions on the form of the underlying distributions. We can

similarly generalize unequal-variance Gaussian TSD to the class of *location-scale TSDs* defined by the requirement that

$$F_{SN}(x) = F_N\left(\frac{x - d}{s}\right),$$

where d is as before a location parameter and s is a scaling parameter.

When the range of permissible TSDs is expanded to include a particular type of location-scale TSDs with possibly asymmetric distributions, a result analogous to the previous theorem can be proved given *three* isosensitivity curves. Let d_1'', etc., denote the parameters corresponding to the third isosensitivity curve; otherwise the notation is as in the previous theorem. f_1 and f_2 are now location-scale TSDs with *fixed* scale parameters s_1 and s_2. It should be noted that the more typical location-scale assumption is that the scale parameter *varies* with separation between the stimulus distributions. However, we have not had much success in studying this larger class of location-scale TSDs. The next result concerns a location-scale model that is, therefore, more restrictive than is generally assumed for location-scale models.

THEOREM 7.1.   *Suppose that*

$$I[f_1, s_1, d_1^{(i)}] = I[f_2, s_2, d_2^{(i)}], \qquad i = 1, 2, 3,$$

where $d_1^{(2)} - d_1^{(1)}$ and $d_1^{(3)} - d_1^{(1)}$ are relatively prime. Then $f_1 \equiv f_2$.

Proof.   For the location-scale case,

$$\psi(c_1^{(1)}) = c_2^{(1)}, \tag{31}$$

$$\psi\left(\frac{c_1^{(1)} - d_1^{(1)}}{s_1}\right) = \frac{c_2^{(1)} - d_2^{(1)}}{s_2}, \tag{32}$$

for the first isosensitivity curve. Consequently,

$$s_2\psi\left(\frac{x^{(1)}}{s_1}\right) = \psi(x^{(1)} + d_1^{(1)}) - d_2^{(1)}, \tag{33}$$

where $x^{(1)} = c_1^{(1)} - d_1^{(1)}$. For the second isosensitivity curve we obtain an analogous equation in $x^{(2)} = c_1^{(2)} - d_1^{(2)}$. By varying $c_1^{(2)}$ we can set $x^{(1)} = x^{(2)} = x$, so that

$$s_2\psi\left(\frac{x}{s_1}\right) = \psi(x + d_1') - d_2'. \tag{34}$$

Then, equating (33) and (34) and setting $z = x + d_1^{(1)}$, $a = d_1^{(2)} - d_1^{(1)}$, and $b = d_2^{(2)} - d_2^{(1)}$, we have,

$$\psi(z) = \psi(z + a) + b. \tag{35}$$

The remainder of the proof involves retracing the above steps for the first and third

isosensitivity curves, obtaining a second equation involving ψ analogous to (35), and then applying the lemma as in the previous theorem. ∎

Levine (1970, 1972) examines the groups of transformations rendering curves parallel or rendering curves into a common shape. As noted above, the problem of the uniqueness of the location TSD generating a given set of isosensitivity curves can be expressed in Levine's framework as an assertion concerning the uniqueness of a specific transformation. It may be possible also to state the problem of the uniqueness of the location-scale TSD that generates a given set of isosensitivity curves within this same framework, but we have been unable to do so.

8. Concluding Remarks

The model of the Observer in a Yes–No detection task provided by the theory of signal detectability consists of three functions: (a) the distribution of the sensory information X available to the Observer on trials where the signal is absent (F_N), (b) the distribution of X on trials where the signal is present (F_{SN}), and (c) a decision rule $D(\cdot)$ that, when applied to X on a given trial, produces a Yes or No response. The experimenter's task is to determine these three functions (F_N, F_{SN}, D), given the Observer's hit and false alarm rates. Without some further restriction on distributions and decision rules, the experimenter cannot succeed in identifying the Observer from the data available.

In practice, then, the missing information specifying the Observer is assumed. In the most familiar case, F_N is assumed to be Gaussian with mean 0 and unit variance, and F_{SN} is assumed to be Gaussian with unknown mean d' and unit variance. The decision rule D is taken to be Yes iff $X > c$, in which the only unknown is the criterion parameter c. The analytical task is to estimate the free parameters remaining in the model of the Observer and use these as indices of the Observer's performance. In particular, estimates of the criterion for human Observers that are less extreme than the value that is optimal for the decision are taken as evidence that human Observers behave conservatively.

The predominant response in the literature to this finding of suboptimal behavior has been a search for heuristics accounting for the behavior. In contrast, the concern in this paper has been to question whether the "finding" might be due to suboptimal model-fitting by the experimenter, rather than suboptimal behavior by the Observer.[1] We have found this to be the case. A discrepancy between assumed and true sensory distributions can result in labelling as conservative behavior that is in fact optimal. The appearance of conservatism is governed by the curvature of the function mapping the quantiles of the true distribution into the quantiles of the assumed distribution.

―――――――――

[1] As noted above, findings of suboptimality in human judgment and decision-making also have prompted an examination of whether suboptimality inheres in the model-fitting rather than in the Observer's behavior. For example, Kahneman *et al.* (1982) suggest that the model of probability theory used by the Observer may be different from the classical model that is typically used by the experimenter.

This result led us to investigate ways in which distributional information might be extracted from the Observer's responses. In the weak form of this investigation, we looked for indications within the data that particular distributional assumptions might be in error. We showed that slight, but systematic variations in the estimate of sensitivity, as the Observer varies c, can be adequately diagnostic of an incorrect distributional assumption. In a similar vein, it is not difficult to show that there are slight, but systematic changes in the shape of the isosensitivity curve as we go from one family of location TSDs to another (Laming, 1973, pp. 122–124). For example, the slope of the isosensitivity curve near the origin depends on the distribution: it is 1 for the Cauchy, finite and greater than 1 for the Laplacian, and infinite for the Gaussian (see Fig. 4). However, since the estimation of the slope of the isosensitivity curve in this region is unreliable, it would be difficult to use this feature of the curve to test the adequacy of a distributional assumption. (Swets (1986a) discusses the problems of diagnosing distributional failures.)

The strong form of our investigation of distributional information focused on the isosensitivity curve and resulted in Theorems 4 and 5. These results show that, in location TSD models, distributional information is fully contained in the isosensitivity curve, and that, in principle, no further assumptions about the Observer's underlying distributions are necessary. Rather surprisingly, the data summarized in as few as two isosensitivity curves from a location TSD specify the underlying distributions completely.

In practice, it is unlikely that distributional information can be extracted from an isosensitivity curve in a transparent or tractable form until a tractable (non-parametric) representation of the isosensitivity curve (or, equivalently, of the sensory distributions) is posited. However, our results suggest that there is a cost associated with constraining the set of possible isosensitivity curves too severely (for example, to the parametric Gaussian or Laplace families of curves). An intermediate approach would be a semi-parametric one (Bloxom, 1985) in which an isosensitivity curve is characterized by (a) a parameter specific to that curve and (b) a function that is the same for all isosensitivity curves but not specified a priori. This invariant function could be drawn from a prespecified class of splines. Then, given a set of empirical isosensitivity curves and the invariant function, the parameters for the curves could be estimated by, e.g., methods used by Bloxom (1985). It is likely that this approach would reduce significantly the bias in the criterion estimates, and further work in this direction seems warranted.

REFERENCES

ACZEL, J. (1966). *Lectures on functional equations and their applications.* New York: Academic Press.

ALTHAM, P. M. E. (1973). A non-parametric measure of signal discriminability. *British Journal of Mathematical and Statistical Psychology,* **26**, 1–12.

APOSTOL, T. M. (1967). *Calculus* (Vol. I). *One-variable calculus with an introduction to linear algebra* (2nd ed.). Lexington, MA: Xerox.

BLOXOM, B. (1985). Considerations in psychometric modeling of response time. *Psychometrika,* **50**, 383–397.

COOMBS, C. H., DAWES, R. M. & TVERSKY, A. (1970). *Mathematical psychology: an elementary introduction.* Englewood Cliffs, NJ: Prentice–Hall.

CREELMAN, C. D., & DONALDSON, W. (1968). ROC curves for discrimination of linear extent. *Journal of Experimantal psychology,* **77**, 514–516.

DORFMAN, D. D., BEAVERS, L. L., & SASLOW, C. (1973). Estimation of signal detection theory parameters from rating-method data: A comparision of the method of scoring and direct search. *Bulletin of the Psychological Society,* **1**, 207–208.

DUSOIR, A. E. (1974). Thomas and Legge's matching hypothesis for detection and recognition tasks: Two tests. *Perception and Psychophysics,* **16**, 466–470.

EGAN, J. P. (1975). *Signal detection theory and ROC analysis.* New York: Academic Press.

GNANADESIKAN, R. (1977). *Methods for statistical analysis of multivariate observations.* New York: Wiley.

GREEN, D. M., & SWETS, J. A. (1966). *Signal detection theory and psychophysics.* New York: Wiley. Reprinted (1974) with corrections. Huntington, NY: Krieger. (References are to 1974 reprint)

GREY, D. R., & MORGAN, B. J. T. (1972). Some aspects of ROC curve-fitting: Normal and logistic models. *Journal of Mathematical Psychology,* **9**, 128–139.

HEALY, A. F., & KUBOVY, M. (1981). Probability matching and the formation of conservative decision rules in a numerical analog of signal detection. *Journal of Experimental Psychology: Human Learning and Memory,* **7**, 344–354.

JOHNSON, N. L., & KOTZ, S. (1970). *Continuous univariate distributions-2.* New York: Wiley.

KAHNEMAN, D., SLOVIC, P., & TVERSKY, A. (1982). *Judgment under uncertainty: Heuristic biases.* Cambridge: Cambridge univ. Press.

KUBOVY, M. (1977). A possible basis for conservatism in signal detection and probabilistic categorization tasks. *Perception and Psychophysics,* **22**, 277–281.

LAMING, D. R. J. (1973). *Mathematical psychology.* New York: Academic Press.

LAMING, D. R. J. (1986). *Sensory analysis.* London: Academic Press.

LEVINE, M. V. (1970). Transformations that render curves parallel. *Journal of Mathematical Psychology,* **7**, 410–443.

LEVINE, M. V. (1972). Transforming curves into curves with the same shape. *Journal of Mathematical Psychology,* **9**, 1–16.

LOCKHART, R. S., & MURDOCK, B. B. (1970). Memory and the theory of signal detection. *Psychological Bulletin,* **74**, 100–109.

MARLEY, A. A. J. (1971). Conditions for the representation of absolute judgment and pair comparison isosensitivity curves by cumulative distributions. *Journal of Mathematical Psychology,* **8**, 554–590.

MORGAN, B. J. T. (1976). The uniform distribution in signal detection theory. *British Journal of Mathematical and Statistical Psychology,* **29**, 81–88.

OGILVIE, J. C., & CREELMAN, C. D. (1968). Maximum-likelihood estimation of receiver operating characteristic curve parameters. *Journal of Mathematical Psychology,* **5**, 377–391.

PARKS, T. E. (1966). Signal-detectability theory of recognition memory performance. *Psychological Review,* **73**, 44–58.

POLLACK, I., & HSIEH, R. (1969). Sampling variability of the area under the ROC curve and of d'_e. *Psychological Bulletin,* **71**, 161–173.

SWETS, J. A. (1986a). Indices of discrimination or diagnostic accuracy: Their ROCs and implied models. *Psychological Bulletin,* **99**, 100–117.

SWETS, J. A. (1986b). Form of empirical ROCs in discrimination and diagnostic tasks: Implications for theory and measurement of performance. *Psychological Bulletin,* **99**, 181–198.

TANNER, W. P., JR., & SWETS, J. A. (1954). A decision-making theory of visual detection. *Psychological Review,* **61**, 401–409.

THOMAS, E. A. C. (1975). Criterion adjustment and probability matching. *Perception and Psychophysics,* **18**, 158–162.

THOMAS, E. A. C., & LEGGE, D. (1970). Probability matching as a basis for detection and recognition decision. *Psychological Review,* **77**, 65–72.