

# Triangles in ROC space: History and theory of “nonparametric” measures of sensitivity and response bias

NEIL A. MACMILLAN

Brooklyn College, City University of New York, New York

and

C. DOUGLAS CREELMAN

University of Toronto, Toronto, Canada

Can accuracy and response bias in two-stimulus, two-response recognition or detection experiments be measured nonparametrically? Pollack and Norman (1964) answered this question affirmatively for sensitivity, Hodos (1970) for bias: Both proposed measures based on triangular areas in receiver-operating characteristic space. Their papers, and especially a paper by Grier (1971) that provided computing formulas for the measures, continue to be heavily cited in a wide range of content areas. In our sample of articles, most authors described triangle-based measures as making fewer assumptions than measures associated with detection theory. However, we show that statistics based on products or ratios of right triangle areas, including a recently proposed bias index and a not-yet-proposed but apparently plausible sensitivity index, are consistent with a decision process based on logistic distributions. Even the Pollack and Norman measure, which is based on non-right triangles, is approximately logistic for low values of sensitivity. Simple geometric models for sensitivity and bias are not nonparametric, even if their implications are not acknowledged in the defining publications.

In many experiments in cognitive science, observers try to assign distinct labels to stimuli chosen from different classes. In the simplest case, there are two stimulus classes (signal and noise in a detection experiment, old and new items in a recognition memory study) and two corresponding responses. The experimenter wishes to abstract from the results a measure of accuracy, or sensitivity, that reflects the subject's ability to distinguish the stimulus classes, as well as a measure of response bias, that is, the tendency to choose one response over the other.

Signal detection theory (SDT; Green & Swets, 1966; Macmillan & Creelman, 1991) is a framework for generating and evaluating both kinds of indices. The best known SDT measures,  $d'$  (for sensitivity) and  $\beta$  (for bias), are consistent with a decision model in which stimulus classes lead to equal-variance normal distributions of a decision variable. This model can be tested by examining *receiver-operating characteristic* (ROC) curves—functions that relate the proportion of hits (*yes* responses to signal presentations) to the proportion of false alarms (*yes* responses to noise presentations) as response bias is either manipulated

(usually through instructions or payoffs) or measured (by a confidence rating judgment). Early work in auditory detection, much of it summarized by Green and Swets, found that empirical ROC curves yielded constant values of  $d'$  in many cases. In other cases, ROCs were consistent with a model in which the underlying distributions were normal, but had unequal variance. By collecting an ROC curve, the investigator who is willing to assume normality can estimate the ratio of the variances of the underlying distributions, as well as their mean separation.

An important result that avoids *any* assumption about distributions is due to Green (1964): The area under the *yes-no* ROC equals the predicted proportion correct by an unbiased observer in a two-alternative forced-choice (2AFC) task. Because unbiased performance in 2AFC can be defined and measured without any theoretical assumptions, *yes-no* ROC area is also assumption free. Thus detection theory justifies two distinct types of sensitivity measures, those (like  $d'$ ) that can be calculated from a single hit/false-alarm pair but that require an assumption about distributions; and area under the ROC, which requires a full ROC curve, but makes no assumptions.

The ideal measure, it appears, would require a single ROC point *and* be assumption free. Pollack and Norman (1964), building on Green's (1964) finding, proposed just such a measure. Their area measure, later dubbed  $A'$ , is calculated from triangular areas in ROC space, as shown in Figure 1. The figure shows a single hit/false-alarm pair<sup>1</sup> ( $H, F$ ). Because ROC curves are expected to

This research was supported by National Science Foundation Grant DBS 92-12043 and a PSC-CUNY award to N.A.M. We thank R. Duncan Luce, Irwin Pollack, an anonymous reader, and Associate Editor Richard Schweickert for helpful comments during the review process. Correspondence should be addressed to N. A. Macmillan, Department of Psychology, Brooklyn College, Brooklyn, NY 11210 (e-mail: nmacmill@broadway.gc.cuny.edu).

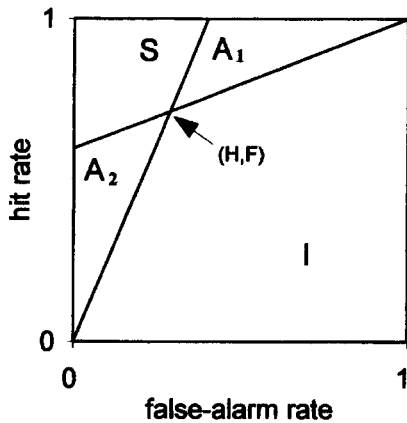


Figure 1. Geometric derivation of  $A'$  and  $B''$  for a single (hit, false-alarm) pair in ROC space. Pollack and Norman (1964) estimated the area under the receiver-operating characteristic (ROC) by subarea  $I$  plus half of  $A_1$  plus  $A_2$ . Hodos's (1970) and Donaldson's (1992) bias measures depend on the difference between  $A_1$  and  $A_2$ .

be monotonic increasing with nonincreasing slope,<sup>2</sup> the area under the curve includes the subarea marked  $I$ . For the same reason, it cannot include any of the subarea marked  $S$ . The triangles  $A_1$  and  $A_2$  may lie under the curve (in whole or in part), depending on the ROC's shape. Pragmatically, Pollack and Norman proposed to estimate the area under the ROC containing  $(H, F)$  by  $I$  plus one half of the ambiguous triangular area:

$$A' = I + \frac{1}{2} (A_1 + A_2). \quad (1)$$

Pollack and Norman (1964) asserted that  $A'$  measured sensitivity nonparametrically,<sup>3</sup> and their analysis soon inspired a triangle-based response bias measure claiming the same status (Hodos, 1970).<sup>4</sup> The statistic  $B''$  is a function of the areas of the right triangles in Figure 1 that include  $A_1$  and  $A_2$  and contain the ROC point  $(1,0)$ . These areas are  $A_1 + S$  and  $A_2 + S$ , and  $B''$  is the difference between them divided by their sum:

$$B'' = \frac{A_1 - A_2}{A_1 + A_2 + 2S}. \quad (2)$$

This index varies from  $-1$  to  $+1$  as response bias varies from all *yes* to all *no* responses.

The new measures did not immediately catch on. From 1964 to 1969, Pollack and Norman's paper was cited only twice, once by Pollack himself. But in 1971, Grier published computing formulas for both  $A'$  and  $B''$ , making these indices much more accessible, and their use soared. The popularity of  $A'$  and  $B''$ , which has continued unabated for 25 years, led us to undertake both a survey of their use and an evaluation of their theoretical status.

In the next section, we review the history of application of triangle-based sensitivity and bias measures, documenting the extent to which they are used, the substan-

tive areas in which they have most often been applied, and the justifications offered by experimenters for choosing them over alternative measures. Many authors choose these measures because they believe them to be free of assumptions entailed by other, competing indices. The most common rationale is that  $A'$  and  $B''$  make no "distributional" assumptions, as do measures derived from detection theory.

In fact, however, the statistic  $B''$  can be derived from a detection-theory model in which the underlying distributions are logistic in form (Macmillan & Creelman, 1990). In the theoretical section of the present paper, we show that logistic distributions arise naturally for indices based on triangular areas, in particular for a recently proposed bias measure (Donaldson, 1992) and a novel but plausible sensitivity measure. The Pollack and Norman (1964) sensitivity measure  $A'$  entails distributions that morph from the logistic (when accuracy is low) to the rectangular (when accuracy is high).

#### APPLICATIONS OF TRIANGLE-BASED MEASURES, 1970-1994

We examined every citation of Pollack and Norman (1964), Hodos (1970), and Grier (1971) listed in either the Science Citation Index or the Social Science Citation Index for the 25-year period 1970 through 1994. Figure 2 shows the pattern of citations over this period. For the first decade, each paper was cited with about equal frequency (5 articles per year); since then, the Pollack and Norman and the Hodos articles have maintained that pace, but the Grier article has been cited in about 15 articles per year.

One or another of these papers has been cited in 403 articles; of course, some papers cite more than one, so the total number of citations is greater. Table 1 gives a Venn diagram analysis of the citations, and can be used to show the relative impact of the Grier (1971) article in another way: Of 290 articles citing the Grier paper, only 74 cited either of the others.

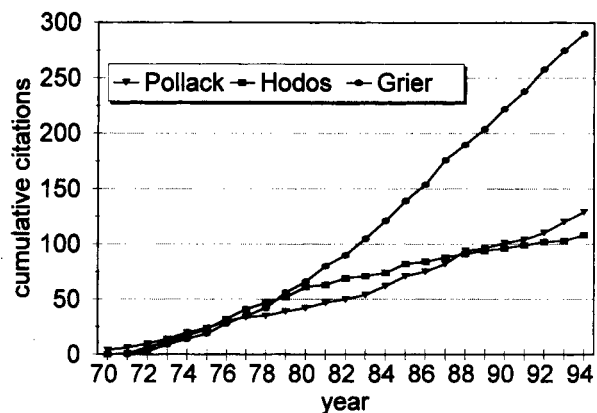


Figure 2. Number of citations for Pollack and Norman (1964), Hodos (1970), and Grier (1971) from 1970 through 1994.

**Table 1**  
**Venn Diagram of Citations**

	Cited Pollack			
	Yes		No	
	Cited Hodos		Cited Hodos	
	Yes	No	Yes	No
Cited Grier				
Yes	33	24	17	216
No	19	55	39	

Note—Pollack, Pollack & Norman (1964); Hodos, Hodos (1970); Grier, Grier (1971). Citations tabulated are all those listed in the Science Citation Index or the Social Science Citation Index from 1970 through 1994.

Pollack and Norman (1964) presented their original measure as useful in recognition memory experiments and referred also to simple detection tasks. Their measure and Hodos's (1970) continue to be used in memory and perception research, and have also been popular with workers in many other fields. Table 2 shows that journals in which articles using these area measures are published divide into three approximately equal categories: physiological/animal, perception/cognition, and a pot-pourri of other subdisciplines.

We sampled the 403 papers to find out why authors used triangle-based measures.<sup>5</sup> To do this, we listed all the articles in chronological order, then randomly selected 1 article of every successive set of 10. Table 3 summarizes these appearances. Most articles (28, or 70%) reported data that were analyzed using *A'* or *B''*; of these, 89% cited the supposed distribution-free or nonparametric nature of these statistics as the reason for using them. Six papers (15%) mentioned the measures in passing, for example, to note use by another investigator. Only one of these references was at all critical. The final 6 articles (15%) were theoretical, and 4 of these were critical, questioning the nonparametric status of the measures.

According to our survey, then, triangle-based measures of sensitivity and bias are heavily used by behav-

ioral scientists in a wide variety of content areas. Users choose these measures because they believe them to make few assumptions, particularly about underlying distributions. Theoretical evaluation has been less enthusiastic, for reasons we consider next.

**TRIANGLE-BASED MEASURES,  
ODDS RATIOS, AND  
THE LOGISTIC DISTRIBUTION**

**The Algebra of Triangular Areas**

One of the appeals of triangle-based indices is their geometric simplicity. Unsurprisingly, such measures are also simple algebraically, and expressing them in terms of the coordinates of ROC space reveals similarities among the indices that might lead us to expect similarities in their entailments. Consider first the small triangles in Figure 1, with areas *A*<sub>1</sub> and *A*<sub>2</sub>. These areas are as follows:<sup>6</sup>

$$A_1 = \frac{1}{2} \frac{(1 - H)}{H} (H - F) \tag{3}$$

$$A_2 = \frac{1}{2} \frac{F}{(1 - F)} (H - F). \tag{4}$$

It is convenient to know the area of *S*, the region in which the ROC cannot fall:

$$S = \frac{1}{2} \frac{F}{(1 - F)} \frac{(1 - H)}{H} (H + 1 - F). \tag{5}$$

Adding the area of *S* to the areas of the small triangles gives the areas of the large triangles:

$$A_1 + S = \frac{1}{2} \frac{(1 - H)}{(1 - F)}, \tag{6}$$

$$A_2 + S = \frac{1}{2} \frac{F}{H}. \tag{7}$$

**Table 2**  
**Journals in Which Target Articles Were Cited**

Content Area	No. Citations	No. Journals	Journal With Most Citations (No.)
Physiological/animal	141	55	
Animal learning, behavior	41	9	<i>Journal of the Experimental Analysis of Behavior</i> (21)
Medical	15	11	(four with 2 each)
Neuroscience	60	28	<i>Neuropsychology</i> (8)
Pharmacology	25	7	<i>Psychopharmacology</i> (13)
Perception/cognition	143	44	
Cognition	26	10	<i>Journal of Memory &amp; Language</i> (7)*
Perception	51	12	<i>Perception &amp; Psychophysics</i> (16)
General†	66	22	<i>Bulletin of the Psychonomic Society</i> (13)‡
Other	119	53	
Personality/clinical	41	19	<i>Journal of Abnormal Psychology</i> (6)
Developmental	23	6	<i>Journal of Gerontology</i> (9)
Applied	36	23	<i>Human Factors</i> (8)
Quantitative	19	5	<i>Psychological Bulletin</i> (11)

\*Includes articles in journal under previous name, *Journal of Verbal Learning & Verbal Behavior*. †General category includes many psychology journals, and probably contains more perception and cognition articles than anything else. Citations in five books were also placed in this category. ‡Includes articles in journal under previous name, *Psychonomic Science*.

**Table 3**  
**Reasons for Citation of Target Articles in the Sample**

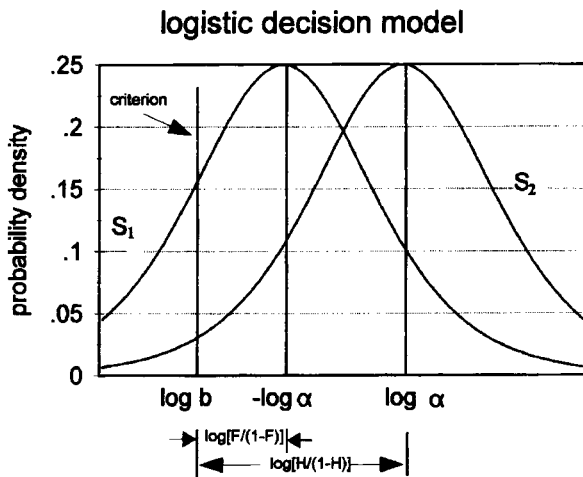
Reasons for Citations	Frequency
Empirical use	
No distributional assumptions	12
Nonparametric	13
No rationale	3
Passing reference (all but one positive)	6
Theoretical analysis (all but two negative)	6
Total	40

Each of these triangular areas depends on the product or ratio of  $H$  and  $F$ , their complements  $1 - H$  and  $1 - F$ , and/or their difference. This dependence arises because the operating point in ROC space has coordinates  $H$  and  $F$ , and triangle areas involve the product of triangle sides. These areas combine to produce sensitivity and bias measures in a natural way: Bias indices depend on a difference or ratio of  $A_1$  and  $A_2$ , whereas sensitivity indices depend on a sum or product. An examination of Equations 3–7 makes it clear that ratios and products of triangular areas contain *odds ratios*, that is, ratios of probabilities such as  $F/H$  and  $(1 - H)/(1 - F)$ . Odds ratios in turn are closely related to the logistic distribution.

**Logistic Model**

Figure 3 shows the decision space for the one-interval experiment, assuming that the underlying distributions are logistic. The stimuli  $S_1$  and  $S_2$  give rise to distributions with means of  $-\log(\alpha)$  and  $+\log(\alpha)$ . The criterion is located at  $\log(b)$ . The hit rate is the area above the criterion under the  $S_2$  distribution, and the false-alarm rate is the area above the criterion under the  $S_1$  distribution. As the figure shows, the distance from the criterion to the mean of the  $S_2$  distribution is  $\log[H/(1 - H)]$  and its distance from the mean of the  $S_1$  distribution is  $\log[F/(1 - F)]$ .

Sensitivity in detection-theory models is the distance between the means, which in this case equals



**Figure 3.** Decision space for the standard one-interval experiment, assuming underlying logistic distributions.

$$2 \log(\alpha) = [\log(\alpha) - \log(b)] - [-\log(\alpha) - \log(b)]$$

$$= \log \frac{H}{(1 - H)} - \log \frac{F}{(1 - F)} \tag{8}$$

Expressed in nonlogarithmic terms,

$$\alpha^2 = \frac{H}{(1 - H)} \frac{(1 - F)}{F} \tag{9}$$

Response bias can be measured in a number of ways, two of which turn out to be important here. First, the location of the criterion is  $\log[F/(1 - F)]$  below the  $S_1$  mean:

$$\log(b) = -\log(\alpha) - \log \frac{F}{(1 - F)}$$

$$= -\frac{1}{2} \left[ \log \frac{H}{(1 - H)} + \log \frac{F}{(1 - F)} \right] \tag{10}$$

In nonlogarithmic terms,

$$b^2 = \frac{HF}{(1 - H)(1 - F)} \tag{11}$$

A second measure of response bias is the *likelihood ratio*, the ratio of the height of the  $S_2$  density to the  $S_1$  density at the criterion value. It can be shown (Macmillan & Creelman, 1990) that the likelihood ratio  $\beta_L$  equals

$$\beta_L = \frac{H(1 - H)}{F(1 - F)} \tag{12}$$

The logistic decision space is implied by Luce’s (1963) choice theory, which is discussed in McNicol (1972) and Macmillan and Creelman (1990, 1991). Note the importance that odds ratios—which we have seen arise naturally from products and ratios of triangle areas—play in expressions for the parameters of this model.

**DISTRIBUTIONAL ANALYSIS OF “NONPARAMETRIC” MEASURES**

We return now to area-based measures of sensitivity and bias that purport to be nonparametric. One sense in which this claim is true is that these indices are derived without any reference to underlying distributions. This is not a strong foundation on which to build such a claim, however, because it depends on the intent of the person presenting the statistic rather than its mathematical properties. For example, it is not hard to imagine someone proposing Equation 9 as a sensitivity measure on the heuristic grounds that it is an increasing function of  $H$  and  $(1 - F)$ , which quantify successful responding, and a decreasing function of  $(1 - H)$  and  $F$ , which quantify unsuccessful responding. This rationale does not mention the logistic distribution, but the omission does not render  $\alpha^2$  distribution free.

Detection theory offers a quantitative method for determining, or at least delimiting, the underlying distribu-

tions consistent with any measure of sensitivity or bias. Every sensitivity index has an *implied ROC* (Swets, 1986b), the relation between the hit rate and the false-alarm rate when sensitivity (as measured by that statistic) is constant. For example, the implied ROC for  $\alpha$  is found by solving Equation 9 for  $H$ ; implied ROCs for  $A'$  were first presented by Pollack and Norman (1964). Similarly, every bias index has an *implied isobias curve*, the relation between  $H$  and  $F$  when bias (as measured by that statistic) is constant. For example, the implied isobias curve for  $b$  is found by solving Equation 11 for  $H$ ; implied isobias curves for  $B''$  were first presented by Hodos (1970).

The connection between ROC and isobias curves on the one hand and underlying distributions on the other is not one-to-one. Suppose that Figure 3 were drawn on a sheet of rubber. If this sheet were stretched horizontally, more in some regions than in others, the distributions would no longer have a logistic shape, but the hit and false-alarm rates, which correspond to relative areas above the criterion, would be unchanged. ROC curves, isobias curves, and sensitivity and bias statistics, all of which depend on  $H$  and  $F$ , would also remain the same. Thus although  $\alpha^2$  is consistent with a logistic-distributions model, it is also consistent with models obtained by stretching the logistic decision space horizontally. Detection-theory statistics such as  $\alpha^2$  and  $d'$  are derived from distributional assumptions, but do not uniquely imply them.<sup>7</sup>

ROC and isobias data cannot distinguish between two measures that are monotonically related. Thus if ROC data are consistent with a constant value of  $\alpha$ , they are also consistent with a constant value of  $\alpha^2$  or  $\log(\alpha)$ . To stress this connection, we call two measures that are monotonically related *equivalent*. In the following paragraphs, we demonstrate equivalence between triangle-based measures of sensitivity and bias on the one hand, and logistic parameters on the other.

**Hodos's Bias Measure,  $B''$**

Hodos's (1970) measure was defined above (Equation 2) as the normalized difference between the areas of the larger triangles  $A_1 + S$  and  $A_2 + S$ . Substituting Equations 3–7 into Equation 2 leads to a computing formula for  $B''$ :

$$B'' = \frac{H(1 - H) - F(1 - F)}{H(1 - H) + F(1 - F)} \tag{13}$$

Dividing numerator and denominator of Equation 13 by  $F(1 - F)$  leads to

$$B'' = \frac{\beta_L - 1}{\beta_L + 1} \tag{14}$$

This relation is monotonic because the slope,  $2/(\beta_L + 1)^2$ , is always positive, and  $B''$  is thus equivalent to the logistic likelihood ratio  $\beta_L$ . Hodos's measure therefore derives from the logistic model.

**Donaldson's Bias Measure,  $B''_D$**

Recently, Donaldson (1992) proposed a measure that is analogous to  $B''$ . He argued that a nonparametric bias statistic should use the smaller triangles in Figure 1, not the larger ones that include the area  $S$ . Donaldson suggested the statistic

$$B''_D = \frac{A_1 - A_2}{A_1 + A_2} \tag{15}$$

Substituting Equations 3 and 4 into Equation 15 gives a computing formula for  $B''_D$ :

$$B''_D = \frac{(1 - H)(1 - F) - HF}{(1 - H)(1 - F) + HF} \tag{16}$$

It is easy to show that

$$B''_D = \frac{1 - b^2}{1 + b^2} \tag{17}$$

which is monotonic with (and thus equivalent to)  $b$ , the logistic criterion location. Donaldson's bias measure therefore derives from the logistic model.

**The New Sensitivity Measure  $A_{\text{right}}$**

As a complement to Hodos's (1970) bias measure, we introduce the *product* of the larger triangles in Figure 1 as a natural (inverse) measure of sensitivity. For points near the upper left corner of ROC space, where sensitivity is high,  $A_1 + S$  and  $A_2 + S$  will be small; near the ROC diagonal, where sensitivity is low, they will be large. This index, which we call  $A_{\text{right}}$  to indicate its dependence on right triangles, can be found from:

$$A_{\text{right}} = \frac{1}{4} \frac{(1 - H)}{H} \frac{F}{(1 - F)} \tag{18}$$

It is easy to show that

$$A_{\text{right}} = \frac{1}{4\alpha^2} \tag{19}$$

which is monotonic with  $\alpha$ . Our new index, therefore, derives from the logistic model. Our purpose in suggesting this measure is preemptive: Because it is geometrically similar to the popular  $B''$ , we fear that someone may seriously offer it as a nonparametric measure of sensitivity. It is not one.

**Pollack and Norman's Sensitivity Measure  $A'$**

As Pollack and Norman (1964) defined the measure that was later denoted  $A'$ , it depends on the sum of the smaller triangles  $A_1$  and  $A_2$  (see Equation 1). The computing formula is:

$$A' = \frac{1}{2} + \frac{1}{4} \frac{(H - F)(1 + H - F)}{H(1 - F)} \tag{20}$$

Unlike the previous three measures,  $A'$  is not monotonic with a parameter of the logistic model. Instead, it depends on two different measures of sensitivity. One is  $\alpha$ ,

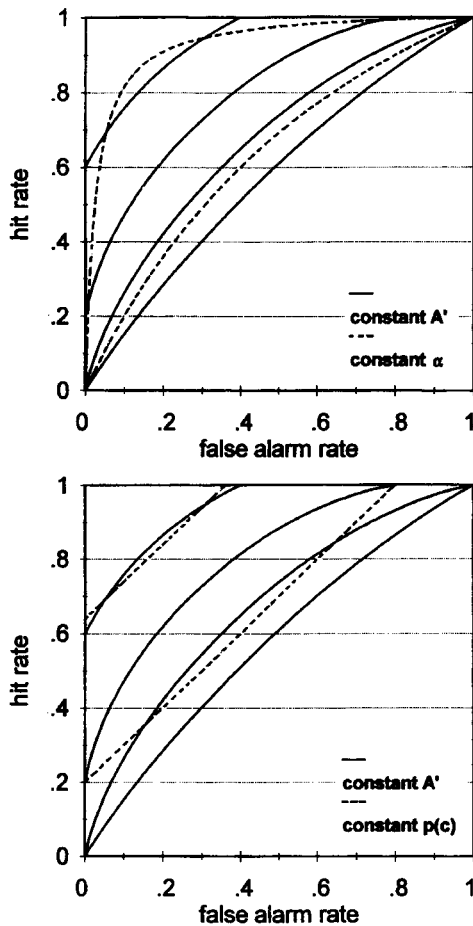


Figure 4. Each panel shows a family of receiver-operating characteristic (ROC) curves implied by  $A'$ . In the upper panel, ROC curves for two levels of  $\alpha$  (logistic sensitivity) are also shown; in the lower panel, ROC curves for two levels of  $p(c)$  (rectangular sensitivity) are shown. The comparison shows that  $A'$  is approximately consistent with logistic distributions at low levels, and rectangular distributions at high levels.

the logistic measure; the other is  $p(c)$ , the proportion correct by an unbiased observer, which equals  $\frac{1}{2} \cdot [H + (1 - F)]$ . The relation is:

$$A' = \frac{1}{2} + \frac{1}{2} p(c) (1 - \alpha^2) \tag{21}$$

At low sensitivity, Equation 21 is dominated by the logistic term  $\alpha$ , whereas at high sensitivities  $p(c)$  is more important. This shift is illustrated in Figure 4, which shows

a family of ROC curves implied by  $A'$ , together with those for  $\alpha$  and  $p(c)$ . The upper panel shows that at low levels, but not high levels, a constant- $\alpha$  ROC is very similar to a constant- $A'$  curve. The lower panel shows that at high levels, but not at low levels, a constant- $p(c)$  ROC is very similar to a constant- $A'$  curve. The statistic  $p(c)$  is the sensitivity index of a decision model in which the underlying distributions are *rectangular* (Macmillan & Creelman, 1991). Area under the one-point ROC, as estimated by Pollack and Norman's (1964) method, is consistent with logistic distributions for low sensitivities and rectangular distributions at high levels; in neither case does it deserve the label "nonparametric."

### CONCLUSIONS AND IMPLICATIONS

Table 4 summarizes our findings about the distributional implications of "nonparametric" measures based on triangles in ROC space. Two bias measures are equivalent to natural bias indices of the logistic model. A new measure of accuracy constructed by a rationale typical of area measures is equivalent to logistic sensitivity. The accuracy measure of Pollack and Norman (1964) is approximately equivalent to either a logistic or a rectangular index, depending on performance level.

These measures have been characterized as "nonparametric" for 25 or 30 years because no theory was used in their construction. This history cannot, of course, guarantee that no theory is entailed in applying them. A pitcher who develops a curve ball without knowledge of aerodynamics has not come up with a pitch outside the realm of physics. An implication of this observation is that future indices claiming nonparametric status need to provide positive support for such a claim, not merely the absence of explicit theoretical foundation. New measures based on triangular areas should be treated with particular suspicion: Like the measures we have examined, they are very likely to entail logistic distributions.

Our survey uncovered a rather extensive literature that is based, according to our theoretical analysis, on a misconception. That area measures so often reflect the logistic distribution can be seen as good news, in that ROC data often support distributions of this sort. The rectangular distributions implied by high values of  $A'$ , on the other hand, are not supported by empirical ROC data, and are therefore especially inappropriate for summarizing performance.

If, as we have argued, performance is not to be summarized by triangle-based measures of accuracy, what

Table 4  
Summary of Triangle-Based Measures

Type of Index	"Nonparametric" Index	ROC Triangles Used*	Equivalent Parametric Index
Response bias	$B''$	Large	Logistic likelihood ratio
	$B''_D$	Small	Logistic criterion
Sensitivity	$A'_{right}$	Large	Logistic mean difference
	$A'$	Small	Logistic mean difference <sup>†</sup>
			Rectangular mean difference <sup>‡</sup>

\*"Small" triangles are those with areas  $A_1$  and  $A_2$  in Figure 1; "large" triangles are those with areas  $A_1 + S$  and  $A_2 + S$ . <sup>†</sup>At low performance levels. <sup>‡</sup>At high performance levels.

statistics should be used instead? In 1986, Swets found that detection-theory based indices were most consistent with available data in a wide range of disciplines, and we are aware of no important change in the database since then (Swets, 1986a). In particular, Swets urged the use of an area statistic,  $A_z$ , the area under the best fitting normal ROC curve. This index has most of the advantages that Green (1964) originally claimed for ROC area. Of course, it cannot be computed from a single hit/false-alarm pair, but requires a full ROC. Contrary to the bulk of the literature surveyed in this article, an adequate area-based measure requires this extra effort.

#### 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.
- BRADLEY, J. V. (1968). *Distribution-free statistical tests*. Englewood Cliffs, NJ: Prentice-Hall.
- DONALDSON, W. (1992). Measuring recognition memory. *Journal of Experimental Psychology: General*, **121**, 275-277.
- GREEN, D. M. (1964). General prediction relating yes-no and forced-choice results [Abstract]. *Journal of the Acoustical Society of America*, **36**, 1042.
- 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.
- LUCE, R. D. (1963). 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.
- MACMILLAN, N. A., & CREELMAN, C. D. (1990). Response bias: Characteristics of detection theory, threshold theory, and "nonparametric" measures. *Psychological Bulletin*, **107**, 401-413.
- MACMILLAN, N. A., & CREELMAN, C. D. (1991). *Detection theory: A user's guide*. New York: Cambridge University Press.
- MALONEY, L. T., & THOMAS, E. A. C. (1991). Distributional assumptions and observed conservatism in the theory of signal detectability. *Journal of Mathematical Psychology*, **35**, 443-470.
- MCCNICOL, D. (1972). *A primer of signal detection theory*. Sydney: Allen & Unwin.
- NORMAN, D. A. (1964). A comparison of data obtained with different false-alarm rates. *Psychological Review*, **71**, 243-246.
- POLLACK, I., & NORMAN, D. A. (1964). A nonparametric analysis of recognition experiments. *Psychonomic Science*, **1**, 125-126.
- SWETS, J. A. (1986a). Form of empirical ROCs in discrimination and diagnostic tasks: Implications for theory and measurement of performance. *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.

#### NOTES

1. The ordering ( $H, F$ ) is the reverse of the conventional one, in which the abscissa value is given first. However, in detection and some other contexts, signal trials (on which  $H$  is based) are of more substantive interest than are noise trials (on which  $F$  is based), and we prefer the unusual ordering. (See also Macmillan and Creelman, 1990, 1991.)

2. The rationale for delimiting the region in which an ROC may fall was given by Norman (1964), who argued that an observer operating at point ( $H, F$ ) could also reach any point on the line segments connecting this point with (0,0) or (1,1) by guessing. Thus any point below these segments (and in area  $I$  in Figure 1) is inferior to ( $H, F$ ) and cannot be on the ROC. Analogously, any point in  $S$  is superior to ( $H, F$ ) and cannot be on the ROC. Norman pointed out that the argument is equivalent to requiring that ROCs be monotonic increasing with nonincreasing slope.

3. The term *nonparametric* has been consistently used by the developers and appliers of the measures described in this article, but *distribution free* may be more appropriate. In the context of statistical tests, Bradley (1968, p. 15) defined a nonparametric test as one that takes no stand about the value of a parameter in a distribution and a distribution-free test as one that takes no stand about the form of the distribution. In this article, we for the most part use the latter term, but use *nonparametric* when referring to the claims of others.

4. Hodos (1970) actually proposed a slightly different measure that has different formulas above and below the minor diagonal of ROC space. Grier (1971) proposed  $B''$ , which is equivalent to Hodos's measure in terms of isobias curves (discussed later). Perhaps because its single formula makes it easier to use,  $B''$  has dominated in application.

5. The search for the articles in the sample was conducted at the library system of the University of Massachusetts, Amherst. A total of 57 articles were chosen to locate the sample of 40, the others being in journals to which the library did not subscribe. The distribution across content areas of the sample was quite similar to that of the entire corpus: physiological/animal articles totaled 32% (vs. 35% in the corpus), perception/cognition 38% (vs. 35%), and other areas 30% (vs. 30%).

6. In the following discussion, we consider only the usual case in which  $H \geq F$ . The equations are different for subchance performance; see Aaronson and Watts (1987).

7. Although underlying distributions are not specified by one ROC curve, Maloney and Thomas (1991) have shown that they may be specified by as few as two for models in which the  $S_1$  and  $S_2$  distributions have the same shape. This result (and related findings in the Maloney and Thomas paper) provides a kind of converse to the straightforward implication of an ROC from an underlying model, but the authors have acknowledged (p. 469) that it is difficult to apply.

(Manuscript received May 30, 1995;  
revision accepted for publication January 9, 1996.)