Psychological Bulletin
1989, Vol. 106, No. 2, 338–340

0033-2909/89/$00.75

Response Bias in Below-Chance Performance: Computation of the Parametric Measure β

Michael R. Waldmann
University of Frankfurt, Frankfurt/Main,
Federal Republic of Germany

Reinhard Göttert
Max Planck Institute for Psychological Research,
Munich, Federal Republic of Germany

We argue that conceptual problems arise with the parametric measure β when above- and below-chance data are aggregated or compared. Depending on the interpretation of β as a likelihood-ratio measure or an indicator of strictness or bias toward signal or noise, the original formula for β should be retained or modified. The response bias measure β can be retained only if it is interpreted as a likelihood-ratio measure. If it is interpreted as an indicator of strictness or bias toward signal or noise, the original formula has to be modified. One possible modified formula is suggested here.

Aaronson and Watts (1987) recently presented formulas for the nonparametric indices A′ and B″ applicable to below-chance performance. They argued that in computing Hodos's (1970) response bias measure B″ for below-chance performance, hit rate and false-alarm rate must be substituted for one another when using Grier's (1971) formula. Otherwise, B″ can yield bizarre values: Points to the left of the equal-bias diagonal may be negative rather than positive. As Aaronson and Watts recognized, parametric indices of discrimination work well with the below-chance case: The accuracy measure d′, for example, yields negative values, indicating that the mean of the probability–density distribution of noise events is located above the distribution of signal events on the evidence axis.

Unfortunately, the same is not true for the parametric response bias measure β when it is computed for individual subjects with below-chance discrimination. For example, take Point b in Figure 1. Point b is located on the left of the negative diagonal, which implies the use of a strict criterion. While $B''_{(BCh)}$ (i.e., the modified B″ for the below-chance case) in fact yields a positive value indicating a strict criterion (i.e., Prob (yes) < Prob (no)), β has a value less than 1, which usually is interpreted as an indicator of a lax criterion (i.e., Prob (yes) > Prob (no)). It should be emphasized, however, that the analogy between B″ and β is not precise. Although in the case of B″, Grier's (1971) formula yields wrong values for below-chance performance, problems with β do not primarily arise on the computational rather than conceptual level.

Measure β for Above- and Below-Chance Performance

The measure β is defined as the likelihood ratio of the densities of the signal and the noise distribution at the decision point. Figure 2 exemplifies this by locating the different decision points shown in Figure 1 on the evidence axis (i.e., the abscissa) in relation to the underlying signal and noise distributions. The measure β for Point a, for example, is computed as the ratio of the probability density values on the signal and noise distributions (i.e., y(signal)/y(noise)). As Green and Swets (1966) proved, the slope of the receiver operating characteristic (ROC) curve at any point is numerically equal to the likelihood-ratio criterion that generates that point. Given the assumption that the signal and the noise distribution are Gaussian with equal variance (as in Figure 2), the slope of the ROC curve (i.e., β) decreases monotonically with increasing hit and false-alarm probability in the above-chance case (see Figure 1). Because the likelihood ratio is monotonically related to the location of the decision criterion on the evidence axis, β can also be interpreted in terms of strictness of the criterion or bias toward signal or noise. For Point a, for example, the probability density of a signal is less than the probability density of noise, yielding a likelihood ratio (i.e., β) below 1, a value that indicates a lax criterion (see Figure 2). This can also be seen by analyzing the bias toward a yes or a no response: For Point a the probability of a yes response is greater than the probability of a no response, which also indicates a lax criterion.

This simple relationship between a likelihood-ratio interpretation and a strictness interpretation is reversed in the below-chance case. As can be seen in Figure 1, the slope actually increases with increasing hit and false-alarm rates (i.e., increasing laxity) in the below-chance case. Consequently, points with equal slope lie on different sides of the negative diagonal (e.g., Points a and d or Points b and c in Figure 1). This means that a specific likelihood ratio that is equivalent to a lax criterion in the above-chance case (e.g., Point a in Figure 1) represents a strict criterion in a symmetric below-chance ROC (Point d in Figure 1).

This relationship can also be seen in Figure 2. In the above-chance case, Point a, which represents a lax criterion, yields a β value less than 1, whereas Point c, which has an equivalent location on the evidence axis, now has a β value greater than 1. The same inverse relationship can be seen for Points b and d, which represent strict criteria.

Correspondence concerning this article should be addressed to Michael R. Waldmann, Institut für Psychologie, Johann Wolfgang Goethe Universität, Georg-Voigt-Straße 8, D-6000 Frankfurt/Main 11, Federal Republic of Germany, or to Reinhard Göttert, who is now at Städtische Klinik Bogenhausen, Abteilung für Neuropsychologie, Englschalkinger-Straße 77, D-8000 München 81, Federal Republic of Germany.

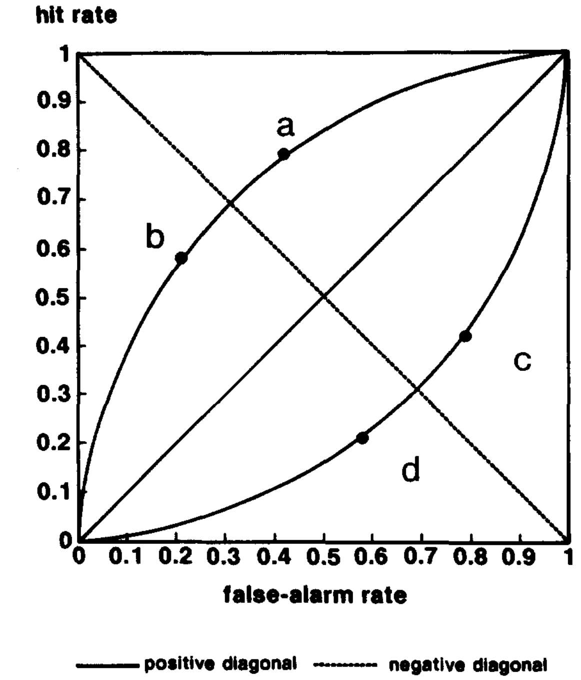

Figure 1. Decision points on symmetric above- and below-chance receiver operating characteristic curves.

This dissociation of a likelihood-ratio interpretation and a strictness or bias interpretation is especially critical when data with above-chance and below-chance performance are aggregated or compared. The resulting value is interpretable only if the aggregated β values are conceived of as a measure for likelihood-ratio criterion setting but not as an indicator of strictness.

Alternative Computation of β in Below-Chance Case

An informal overview of recent work in memory and perception shows that researchers using signal detection measures generally appear to interpret β in terms of strictness and not as a likelihood ratio. Therefore, a modification similar to that of B″ seems appropriate for the below-chance case when signal and noise distributions are Gaussian with equal variance. To extend the computation of β (Equation 1) to below-chance performance, one needs to make an adaptation such that the inverse of β is calculated (Equation 2):

$$\beta = y(\text{signal})/y(\text{noise}), \quad (1)$$

$$\beta_{(\text{BCh})} = y(\text{noise})/y(\text{signal}). \quad (2)$$

Equation 2 is equivalent to

$$\beta_{(\text{BCh})} = 1/\beta, \quad (3)$$

where y(signal) and y(noise) are the probability density values corresponding to the areas under the normal distribution function that represent the hit rate and the false-alarm rate, respectively.

This modification yields criterion values for points on below-chance ROCs that are interpretable in terms of strictness or bias in the same way as is the case for the regular above-chance ROCs. In the example ROCs in Figure 1, Points a and c that according to their location in relation to the negative diagonal are considered equivalent with respect to bias and strictness, are now characterized by the same β value. The same holds true for Points b and d. Note that the $\beta_{(\text{BCh})}$ does not represent a likelihood-ratio criterion: The slopes of Points a and c, or b and d, are clearly different.

Discussion

In summary, if the hypothesis is that subjects base their criterion setting on a likelihood-ratio decision, the classical formula for β should be used in both above- and below-chance performances. If, however, the researcher wants the response criterion measure to represent the strictness of the criterion or bias toward signal or noise response, then the modified formula should be used for individuals with below-chance performance, provided that the signal and noise distributions are assumed to be Gaussian with equal variance.

Below-chance performance is not the only case in which β is not simply interpretable as a measure of strictness or bias toward signal or noise. In case of unequal variances of signal and noise probability density distributions, β is not monotonically related to the location of the criterion point on the evidence axis (Green & Swets, 1966). In this case, only a likelihood-ratio interpretation of β is possible. Other bias measures must be con-

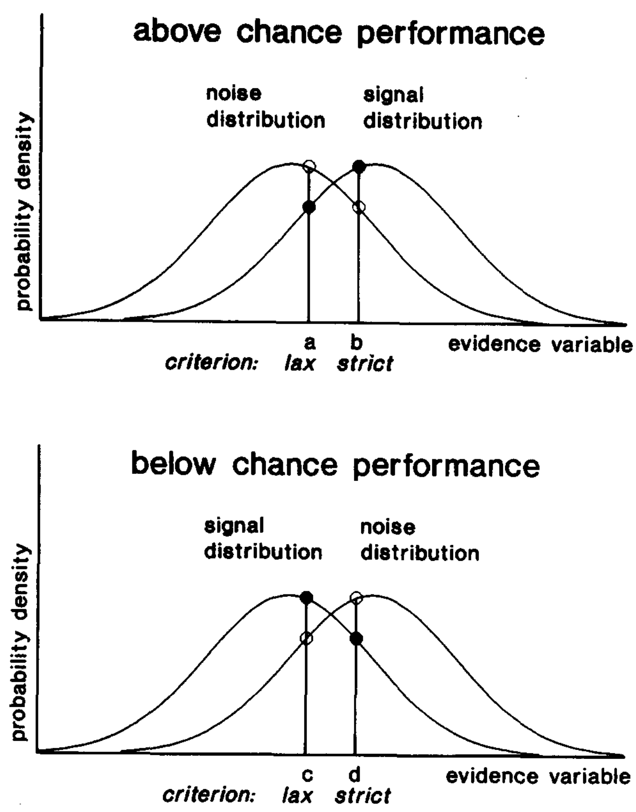

Figure 2. Signal and noise distributions in above- and below-chance performance.

sidered when the researcher is interested in the strictness or bias of the decision criterion (see Dusoir, 1975, 1983; Snodgrass & Corwin, 1988).

Independent of the exact computation of a bias measure in the below-chance case, the question remains concerning what significance below-chance performance has in discrimination tasks. Green and Swets (1966) discussed this "worst-possible behavior" (p. 39) in terms of hypothetical subjects who reverse the decisions that are dictated by the likelihood-ratio criterion. It is unclear if such an irrational behavior can really be found reliably in subjects. Green and Swets argued that such a behavior occasionally may be seen in psychophysical tasks but usually disappears with further practice under conditions with corrective feedback.

Swets and Pickett (1982, p. 23) offered another explanation for occasional below-chance performance: They thought of points or ROC curves below the positive diagonal as chance variations of true discrimination. However, if empirical ROC points are conceived of as chance variations of true values, then estimating the true value will require the aggregation over empirical ROC points, regardless of their location above or below chance level. Consequently, values above and below chance level should represent the same criterion setting type.

However, true negative discrimination is not entirely implausible. Although in the classical psychophysical paradigms, reversal of signal and noise distribution is peculiar indeed—given the basic assumption that the signal is physically added to a noisy background, yielding a signal-plus-noise distribution and a noise distribution—in other signal-detection applications, the classical assumption may be unwarranted (e.g., measurement of memory performance). In these cases it seems more appropriate to assume independent signal events and noise events without any assumption regarding the relative location of their probability density distributions on the evidence axis.

In memory experiments, for example, it seems conceivable that for some individuals and for some item groups, the evidence values of the distractor items surpass the evidence values of the old items, as, for example, when distractor items have a higher familiarity or a higher prototypicality than old items. One plausible case may involve people with certain memory impairments (e.g., amnesic patients) whose recent episodic memory representations are poor but for whom episodes dating back several decades still have strong memory traces. As another example, eyewitnesses may systematically misidentify specific types of persons in lineups because the appearance of these persons fits the schema of a criminal much more closely than does the person the witnesses actually have observed. In such situations, the modified measure suggested here may prove useful.

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.

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

Dusoir, A. E. (1983). Isobias curves in some detection tasks. *Perception and Psychophysics, 33,* 403–412.

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.

Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General, 117,* 34–50.

Swets, J. A., & Pickett, R. M. (1982). *Evaluation of diagnostic systems.* New York: Academic Press.

Received June 6, 1988
Revision received February 27, 1989
Accepted March 9, 1989 ■