The effects of response bias in two-alternative-forced-choice discrimination matrices¹

C. M. HOLLOWAY, MEDICAL RESEARCH COUNCIL, APPLIED PSYCHOLOGY RESEARCH UNIT, Cambridge, England

Traditionally the two-alternative forced-choice technique has been employed in the investigation of stimulus discriminability in cases where it has been desirable to limit the effects of response bias on the discrimination performance. The efficacy of this method of response bias control is assessed in relation to a number of perceptual models of the receiver-operating characteristic.

The two-alternative forced-choice (2 Alt.FC) technique has been commonly employed in perceptual discrimination experiments in order to minimize the effect of S's response preferences on his discrimination performance. In the detection paradigm, the signal is presented on each trial in one of two discrete time intervals and the S is asked to state in which interval the signal occurred. In the recognition paradigm the signal is presented on each trial in a single observation interval and is one of two alternatives, and the S must say which.

Two advantages have been claimed for the 2 Alt.FC procedure. First, the response categories provided (Interval 1 vs Interval 2, or Alternative 1 vs Alternative 2) seem to differ little in cost of selection, so that the S should have no a priori reason for preferring any one response. Given this assumption, the articulation score² calculated from the stimulus-response (S-R) confusion matrix may be taken to reflect directly the S's discrimination performance; no response-bias-free statistic need be calculated. A second advantage is that when response bias does exist, its effect on the articulation measure is likely to be minimal, since any preference for a particular response will tend to raise the hit rate for one stimulus while tending to depress the hit rate for the other stimulus. The rise in hit rate for the one stimulus may not be exactly matched with the fall in the other hit rate, but a balance is approximated with small amounts of response bias. The articulation measure may thus be noted as relatively insensitive to the effects of response bias.

The assumption of lack of appreciable response bias will be more or less valid depending on the S's previous familiarity and expectations with regard to the response categories offered. In some instances—with the recognition of spoken phonemes, for example—the assumption may be unreasonable (Holloway, 1968). Furthermore, if a significant amount of bias does exist, the articulation measure may be sensitive to its effects. However, the extent to which the articulation measure is affected can only be assessed by deriving response-bias-free measures of performance from the raw data and is thus dependent upon which of a number of conceptualizations of the perceptual process is used in the application of the response bias correction procedures. The efficiency of the articulation score from the 2 Alt.FC S-R matrix as a discrimination measure is a complex function of the amount of response bias present and the model of the perceptual process assumed to be an appropriate reflection of discrimination behavior.

Green and Swets (1966, pp. 408-411) report one correction procedure, due to Egan and Jeffress, which is based on the theory of signal detectability. The description offered, however, assumes that the two alternatives presented to the S are equally discriminable, and response bias is defined in such a way as to be

Table 1
A general notation for the stimulus-response matrix. All cell entries represent probabilities.

	Stimulus		
	A	B	NIL
Response	A	1 - q	r
	B	q	1 - r

p, q: hit-rates for stimuli A, B,

r: response bias

The articulation score is given by $\frac{1}{2}(p + q)$

dependent upon the discriminability of the alternatives. In this article an attempt is made to assess the effects of response bias on articulation measures with respect to an independent definition of response bias and more than one perceptual model.

In the first section the models are elaborated. In the second section a simulation is performed to explore systematically the size of error caused by failure to correct for response bias for a wide range of independent bias values and the full range of discrimination performance, i.e., from chance performance to perfect discrimination.

RESPONSE BIAS CORRECTION PROCEDURES

General Considerations

The notation for a S-R confusion matrix is shown in Table 1. Entries in the matrix represent probability values. The right-hand column represents a no-stimulus situation and its entries depict response preferences.

By the application of rules specifying the interaction between response preference and discriminability,³ the entries in the six cells of Table 1 can be changed to give an infinitely large series of matrices, each corresponding to the expected performance given a constant level of discriminability and different amounts of response bias (as specified in the right-hand column of the matrix).

In the application of response-bias correction procedures the entries in the initial matrix, 1, are based on observed performance; another matrix, 2, is then derived by specifying absence of response bias ($r = 1 - r = .5$). Thus the initial matrix, 1, corresponds with an "uncorrected" matrix and the derived matrix, 2, corresponds with a "corrected" matrix in which the articulation score is a bias-free estimate of discrimination performance. The difference between the articulation scores calculated from the two matrices reflects directly the error produced by using the articulation score from the uncorrected matrix as a discrimination index, and is termed the correct error.

Three response-bias correction procedures are considered, and these may now be described.

The Correction Procedures

(a) *The simple threshold model.* The simple threshold model assumes that a stimulus is either discriminated perfectly or is not discriminated at all. If it is assumed that discrimination occurs with a probability x and y for each of the stimuli presented, then the entries in the confusion matrix may be conceptualized as in Table 2.

Thus, given Stimulus A, the probability of being correct, p , is

Table 2
Simple threshold model conceptualization of the stimulus-response matrix.

		Stimulus	
		A	B
Response	A	$x + (1 - x)r$	$(1 - y)r$
	B	$(1 - x)(1 - r)$	$y + (1 - y)(1 - r)$

the sum of the probability of correct discrimination, x , and the probability that, given no discrimination, the S will choose Alternative A as a response, $(1 - x)r$. The probability of an incorrect response, $(1 - p)$, is the probability that, given no discrimination, $(1 - x)$, the S chooses Alternative B as a response, $(1 - x)(1 - r)$. Similar arguments apply for Stimulus B.

It can then be shown that:

$$x = \frac{p - r}{1 - r} \quad (1a)$$

and

$$y = \frac{q + r - 1}{r} \quad (1b)$$

Since we assume that x and y remain constant from Matrix 1 to Matrix 2, then

$$p_2 = \frac{p_1(1 - r_2) - (r_1 - r_2)}{(1 - r_1)} \quad (2a)$$

and

$$q_2 = 1 - \frac{r_2(1 - q_1)}{r_1} \quad (2b)$$

where the subscripts denote the relevant matrix.

(b) *The choice model.* The choice model assumes a multiplicative relationship between the confusion parameter, α , and the response preference parameter, t . The S-R matrix is conceptualized as in Table 3, and is derived from expositions by Luce (1959, pp. 62, 66; 1963, pp. 103-189) and Shipley (1960). From this conceptualization it can be shown that:

$$\alpha_a = \frac{(1 - p)r}{p(1 - r)} \quad (3a)$$

and

$$\alpha_b = \frac{(1 - q)(1 - r)}{qr} \quad (3b)$$

If we assume that α_a and α_b remain constant from Matrix 1 to Matrix 2 then:

$$p_2 = \frac{p_1(1 - r_1)r_2}{p_1(r_2 - r_1) + r_1(1 - r_2)} \quad (4a)$$

and

$$q_2 = \frac{q_1 r_1 (1 - r_2)}{q_1 (r_1 - r_2) + (1 - r_1) r_2} \quad (4b)$$

Table 3
Choice model conceptualization of the stimulus-response matrix.

		Stimulus		
		A	B	nil
Response	A	1	α_b	1
	B	$\alpha_a t$	t	t

(c) *The signal detectability model.* The signal detectability analysis of the 2 Alt.FC situation assumes the operation of the integration model for multiple observations. It is assumed that each interval, or each alternative, is monitored independently, and that there is no loss of information due to the necessity for multiple observation. Under the assumptions stated it can be shown that:

$$d'_{fc} = \sqrt{(d'_1)^2 + (d'_2)^2} \quad (5)$$

where d'_{fc} is the index of discrimination for the 2 Alt.FC situation and d'_1 and d'_2 are the discrimination indices for each of the alternatives. Given the response bias, r , the discrimination function, d' , is found for each stimulus by entering the yes/no tables with the given values of hit rate and false-alarm rate for each stimulus. In the present notation the false-alarm rates for Stimuli A and B are (r) and $(1 - r)$, respectively. The predicted hit rate for a stimulus is found by leaving the tables with the previously obtained values of d' and the defined false-alarm rate [i.e., $r = (1 - r) = .5$, for the no-bias matrix].

THE SIMULATION

By generating the matrix represented in Table 1, an Elliott 803 computer was used to sample the correction error for various values of p , q , and r , and for each correction procedure. Data were collated in two ways.

In one condition the constraint was applied that the two stimuli should be equally discriminable ($x = y$, $\alpha_a = \alpha_b$, or $d'_1 = d'_2$). In the second condition, those two values of stimulus discriminability were found that generated the maximum possible correction error for a given response bias value. Values of r ranged from .5 to .05.

Results

A summary of the results for the equal discriminability condition is presented graphically in Fig. 1. The axes are the correction error (corrected articulation percentage minus uncorrected articulation percentage), and the corrected articulation percentage, corresponding to the "true" discrimination performance of the S. It will be noted that the correction errors predicted on both the signal detectability model and the choice model are similar in both extent and point of maximum on the corrected articulation range. The consistent trend is for the uncorrected score to be less than the corrected score, this effect being greater for the more extreme levels of the response bias. The correction error may exceed 1% when response bias values are in excess of a level of 1:2. The discrepancy in correction errors between the models is not large. Below bias values of the order of $r = .25$ the maximum difference between the models is 1%.

Under equal discriminability conditions a zero correction error is predicted by the simple threshold model. Thus for Stimulus A the correction error is:

$$p_2 - p_1 = [x + (1 - x)r_2] - [x + (1 - x)r_1] \quad (6a)$$

and for Stimulus B:

$$q_2 - q_1 = [y + (1 - y)(1 - r_2)] - [y + (1 - y)(1 - r_1)] \quad (6b)$$

Since it is assumed that $x = y$, then

$$[x + (1 - x)] = [y + (1 - y)] = k \quad (7)$$

and since

Fig. 1. The correction error (corrected articulation percentage minus uncorrected articulation percentage) is plotted as a function of the corrected articulation percentage. The main parameter is the assumed level of response bias. The pairs of curves represent the derivations from the choice model (curve with highest peak) and the signal detectability model (curve with lowest peak). The derivation from the simple threshold model is the abscissa line.

$$r_2 = 1 - r_1 = \frac{1}{2} \quad (8)$$

then from Eqs. 6, 7, and 8:

$$p_2 - p_1 = k(r_2 - r_1) \quad (9a)$$

and

$$q_2 - q_1 = k[r_2 - (1 - r_1)] = -k(r_2 - r_1) \quad (9b)$$

Since $p_2 = q_2$ (because the corrected matrix is symmetrical) the correction error for Stimulus A, $(p_2 - p_1)$, is of equal magnitude but opposite sign to the error for Stimulus B, $(q_2 - q_1)$.

Sizable correction errors can occur when the stimuli are not assumed to be equally discriminable (Fig. 2). In such cases, the correction error depends on which particular stimulus is the more discriminable, so that two maximum correction errors occur, one that overestimates the corrected articulation score and one that underestimates it. The maximum negative error, i.e., error due to the uncorrected articulation score being smaller than the corrected articulation score, is dependent on the model employed. For the simple threshold model the maximum error is obtained when the stimulus associated with the most preferred response has a corrected hit rate of 100% (perfect discrimination) and the other stimulus has a corrected hit rate of 50% (is not

Fig. 2. The maximum negative error is plotted as a function of the level of response bias. The parameter represents the model used in deriving the predictions: (a) the choice model (b) the signal detectability model, and (c) the simple threshold model. The maximum positive error obtainable on all three models corresponds to Curve c.

discriminated). Under these conditions the maximum error is simply described by Eq. 10:

$$\text{Max negative error} = |r_2 - r_1| \quad (10)$$

With the other two models, the maximum negative error may exceed the value for the simple threshold model. This will occur when the stimulus associated with the most preferred response is fully discriminated, but the other stimulus assumes various values depending on the particular response bias value considered. The corrected hit rate for the nonpreferred stimulus that yields the maximum negative error is given in Table 4. The values shown are for the choice model, but are within 1% of those derived from the signal detectability model.

On all three models the maximum positive error, i.e., error due to the uncorrected articulation score being larger than the corrected articulation score, is obtained when the stimulus associated with the most preferred response has a corrected hit rate of 50% (is not discriminated) and the other stimulus is maximally discriminated. Under these conditions the maximum error is defined by Eq. 10, and corresponds to the size of the maximum negative error for the simple threshold model.

It is apparent that differential discriminability of the stimuli will be an important determinant of the size of the correction error and that the direction of this error will depend largely on the relationship between response preferences and the differential discriminability of the stimuli. Response preference associated with the more discriminable stimulus will give an uncorrected score that tends towards underestimation of true discrimination; a tendency to prefer the least discriminable stimulus will give an uncorrected score that tends towards overestimation of true

Table 4
Corrected hit-rate (choice model) for the non-preferred stimulus yielding the maximum negative error. Corrected hit-rate for the preferred stimulus = 100%.

Response Bias	50:50	45:55	40:60	35:65	30:70	25:75	20:80	15:85	10:90	5:95
Hit Rate (%)	50.0	52.5	55.1	57.7	60.5	63.4	67.0	70.4	75.0	81.0

performance. The magnitude of the correction errors that are possible with differential stimulus discriminability underlines the importance of making the correct assumption about stimulus discriminability when deciding whether to apply correction procedures to data.

If it is assumed that the stimuli presented in an experiment are equally discriminable, as in the detection paradigm, the correction procedures outlined in the previous section can be applied without necessitating an independent estimate of response bias.⁴

For the choice model it can be shown that if $\alpha_a = \alpha_b$, then:

$$p_2 = \frac{p_1 q_1}{(1 - p_1)(1 - q_1)} \quad (11)$$

For the detection model, the Egan and Jeffress procedure described by Green and Swets (1966, pp. 408-411) can be directly applied.

CONCLUSION

A marked degree of unanimity is observed between the prediction of the signal detection model and the choice model. Consideration of the simulated effects of all models shows that the correction error is a complex function of the following variables:

- (1) the assumed level of response bias;
- (2) the average discriminability of the stimuli;
- (3) the differential discriminability of the stimuli;
- (4) the association between response preference and differential discriminability.

For those cases in which equal discriminability of stimuli can be assumed, as in the detection paradigm, the generalization can be made that only response biases in excess of the level 1:2 can have an appreciable effect on the articulation score. However, if equal discriminability is not assumed, it becomes difficult to predict the precise effects of response bias on the articulation score, without an exact specification of the parameters involved.

REFERENCES

- GREEN, D. M., & SWETS, J. A. *Signal detection theory and psychophysics*. New York: Wiley, 1966.
- HOLLOWAY, C. M. Decision complexity and speech recognition. Unpublished doctoral thesis, University of Hull, 1968.
- LUCE, R. D. *Individual choice behavior: A theoretical analysis*. New York: Wiley, 1959.
- LUCE, R. D., Detection and recognition. In R. D. Luce, R. R. Bush, and E. Galanter (Eds.), *Handbook of mathematical psychology*. Vol. 1. New York: Wiley, 1963.
- SHIPLEY, E. F. A model for detection and recognition with signal uncertainty. *Psychometrika*, 1960, 25, 273-289.

NOTES

1. This study was carried out at the University of Hull. The experimental work reported in this article was conducted as a part requirement for a PhD thesis.
2. The articulation score is the average probability of a correct response given each of the two stimuli, i.e., $\frac{1}{2}(p + q)$ in the terminology of Table 1.
3. Although the conventional application of the models to be considered may be within either the detection or the recognition paradigm, the index of discrimination for each model is here interpreted as a general index for situations in which the O is requested either to discriminate presence from absence of a signal or to discriminate between alternative stimuli.
4. No correction is necessary on the simple threshold model.