Treatments of bias in detection and recognition models: A review

A. E. DUSOIR

City of London Polytechnic, Whitechapel High Street, London E1 7PF, England

The review considers various proposals about the form of isobias curves and about the relation between bias and bias conditions (payoff, presentation probabiiity, and instructions). Though the proposals differ sharply, none of them proves to be adequately supported by existing evidence. There has been a curious tendency to appeal to a priori arguments, to consider only a very restricted set of alternatives from among the published proposals, to contrast proposals at levels of performance where their predictions are minimally different, and to ignore existing evidence. The review tries to provide a solid basis for more adequate experimental work in the future.

The current approach to detection and recognition performance is to define separate *discriminability* and *bias* parameters and to try to find two disjoint sets of experimental variables, each affecting only one type of parameter. Discriminability parameters have been extensively discussed; much effort has been spent in trying to find a discriminability parameter which does indeed remain invariant under change in presentation probability, payoff, and instructions, and in trying to relate such a parameter to objective stimulus conditions. Bias parameters have been neglected, by comparison; much less effort, that is, has been spent in trying to find a bias parameter which does indeed remain invariant under change in objective discriminability, or in trying to relate such a parameter to presentation probability, payoff, and instructions. Nevertheless, these last two issues have by now attracted many quite distinct proposals, and these are reviewed here.

PRELIMINARIES

The review considers detection and recognition situations in which one of two stimuli, s_1 and s_2, is presented on each trial and followed by one of two responses, r_1, r_2, or by one of m rating responses. (For the only relevant work outside this area, see Atkinson, Bower, and Crothers (1965), Atkinson and Kinchla (1965), and Markowitz and Swets (1967); these papers consider bias in forced choice tasks.) Where it is natural to treat one stimulus as "noise," the other as "signal," s_1 will be used to denote "noise"; otherwise assignment of the labels to the stimuli will be arbitrary. γ will be used to denote the probability that s_2 will be presented. Where some payoff is made

contingent on the joint occurrence of the ith stimulus and the jth response, it will be denoted o_{ij}.

We will be concerned with the conditional probabilities, $p(r \mid s)$, of response r given stimulus s. In the rating response case, it will be assumed that the m response categories have been collapsed, by the well-known method, to yield m - 1 separate definitions of r_1 and r_2 (Egan, Schulman, & Greenberg, 1959).

For brevity, let $p(r_2 \mid s_1) = q$, $p(r_2 \mid s_2) = p$. It is usual to write q and p as functions of two separate parameters

$$q = f(\Theta_s, \Theta_b) \qquad (1)$$

$$p = g(\Theta_s, \Theta_b) \qquad (2)$$

where Θ_s is interpreted as a sensitivity parameter, Θ_b as a bias parameter. If one parameter is held constant and the other varied, an isoparameter curve is traced out. It will be convenient to use the brackets [ ] to denote a family of isoparameter curves. Thus $[\Theta_b]$, for example, will denote the family of isobias curves traced out by keeping Θ_b constant and varying Θ_s.

Suppose that there are two different parameterizations

$$q = f(\Theta_s, \Theta_b) = \bar{f}(\bar{\Theta}_s, \bar{\Theta}_b) \qquad (3)$$

$$p = g(\Theta_s, \Theta_b) = \bar{g}(\bar{\Theta}_s, \bar{\Theta}_b). \qquad (4)$$

Equations 3 and 4 implicitly define $\bar{\Theta}_s, \bar{\Theta}_b$ as functions of Θ_s, Θ_b. The relation may be put in explicit form as

$$\bar{\Theta}_s = u(\Theta_s, \Theta_b) \qquad (5)$$

$$\bar{\Theta}_b = v(\Theta_s, \Theta_b) \qquad (6)$$

where u, v depend on f, $\bar{f}$, g, and $\bar{g}$. Equations 5 and 6

determine whether the two parameterizations agree as to their isoparameter curves. They yield identical isosensitivity curves. $[\Theta_s] = [\overline{\Theta}_s]$, if in fact $\overline{\Theta}_s$ is a function of the first variable only

$$\overline{\Theta}_s = u(\Theta_s). \qquad (7)$$

Similarly, they yield identical isobias curves, $[\Theta_b] = [\overline{\Theta}_b]$, if $\overline{\Theta}_b$ is a function of the second variable only

$$\overline{\Theta}_b = v(\Theta_b). \qquad (8)$$

A parameterization may considerably simplify the relation between experimental variables and data. Given a proposed parameterization (Θ_s, Θ_b), and a list $(x_1, x_2, \ldots, x_n)$ of experimental variables, we may partition the x_i into the sets $X_{s\bar{b}}$, $X_{\bar{s}b}$, X_{sb}, $X_{\bar{s}\bar{b}}$ according as they affect Θ_s alone, Θ_b alone, both, or neither. The trouble with the raw data parameterization $(q = \Theta_b, p = \Theta_s)$ is that most, possibly all, variables fall into X_{sb} or $X_{\bar{s}\bar{b}}$. What we want is a parameterization where most, if possible all, the variables fall into $X_{\bar{s}b}$ or $X_{s\bar{b}}$; than Θ_s, Θ_b can serve as a direct source of inference about the x_i.

The uniqueness of any such parameterization depends on which of Equations 5-8 hold, since

$$x_i \in X_{\bar{s}b} \rightarrow \frac{\delta\theta_s}{\delta x_i} = 0, \frac{\delta\theta_b}{\delta x_i} \neq 0 \rightarrow \frac{\delta u}{\delta\theta_s} \cdot \frac{\delta\theta_s}{\delta x_i} + \frac{\delta u}{\delta\theta_b} \cdot \frac{\delta\theta_b}{\delta x_i} \begin{cases} = 0 \\ \neq 0 \end{cases}$$

$$\text{as } \frac{\delta u}{\delta\theta_b} \begin{cases} = 0 \\ \neq 0 \end{cases} \qquad (9)$$

$$x_i \in X_{s\bar{b}} \rightarrow \frac{\delta\theta_s}{\delta x_i} \neq 0, \frac{\delta\theta_b}{\delta x_i} = 0 \rightarrow \frac{\delta v}{\delta\theta_s} \cdot \frac{\delta\theta_s}{\delta x_i} + \frac{\delta v}{\delta\theta_b} \cdot \frac{\delta\theta_b}{\delta x_i} \begin{cases} = 0 \\ \neq 0 \end{cases}$$

$$\text{as } \frac{\delta v}{\delta\theta_s} \begin{cases} = 0 \\ \neq 0 \end{cases}. \qquad (10)$$

So if parameterizations (Θ_s, Θ_b), $(\overline{\Theta}_s, \overline{\Theta}_b)$ determine $X_{\bar{s}b}, \ldots, X_{s\bar{b}}$ and $\overline{X}_{\bar{s}b}, \ldots, \overline{X}_{s\bar{b}}$ then $X_{\bar{s}b} = \overline{X}_{\bar{s}b}$ if Equation 7 holds, $X_{s\bar{b}} = \overline{X}_{s\bar{b}}$ if Equation 8 holds, otherwise $X_{\bar{s}b}$ and $\overline{X}_{\bar{s}b}$ are disjoint, as are $X_{s\bar{b}}$ and $\overline{X}_{s\bar{b}}$. Note, however, that two parameterizations may each make $X_{\bar{s}b}$, $X_{s\bar{b}}$ nonempty, provided they disagree totally about their composition. Thus, for example, there is nothing to prevent Θ_b remaining invariant under change in one variable and $\overline{\Theta}_b = v(\Theta_s, \Theta_b)$ remaining invariant under change in another. Both parameterizations would then have some point.

So far, empirical work has established some degree of consensus, but only about sensitivity. The main, and best known, example is the parameterization

$(d', \cdot)$ in signal detection theory (Green & Swets, 1966), where any of a large number of bias parameters may complete the parameterization. Because these bias parameters are related by Equation 6, not Equation 8, there is almost no agreement about the measurement of bias.

The success of the $(d', \cdot)$ parameterization rests on the following: (i) Under this parameterization, there exist some $X_{\bar{s}b}$ variables. (ii) All these $X_{\bar{s}b}$ variables are ones which might have been expected a priori to affect "bias" but not "sensitivity," namely payoff, instructions, and the response category definitions. Another such variable, γ, probably has to go into X_{sb} (Markowitz & Swets, 1967; Wickelgren, 1968), unless the subject's memory for s_1, s_2 is controlled. (iii) It seems a reasonable requirement of a sensitivity parameter that it should not, for example, take the same value for points (q, p), (q, p'), $p \neq p'$. In fact, for Θ_s to be a sensitivity parameter, there should be a function, ψ, such that

$$\Theta_s = \psi(q, p) \qquad (11)$$

$$\begin{array}{c} q \leqslant q' \ \& \ p > p' \\ \text{or} \\ q < q' \ \& \ p \geqslant p' \end{array} \rightarrow \psi(q, p) > \psi(q', p'). \qquad (12)$$

This is satisfied by the d' parameter. Note, parenthetically, that Equations 11 and 12 make p a strictly increasing function of q given fixed sensitivity. Consequently any second parameter, Θ_b, in Equations 1 and 2, must necessarily index the extent to which the subject chooses r_2 rather than r_1, given fixed sensitivity. This makes it reasonable to call any Θ_b a *bias* parameter. (iv) There seems to be no alternative parameterization satisfying (i), (ii), and (iii).

It should be noted that (i)-(iv) support the d' parameter in quite different ways. (i) gives the parameter some point. (ii) and (iii) justify it being called a *sensitivity* parameter. (iv) gives the parameter greater importance than it would otherwise have.

What still remains is to choose a bias parameter. As will appear, there is no shortage of candidates. What their proponents seem to be in search of is a straightforward analogue of *all* of the points (i), (ii), (iv) above; that is, they want to assume that some measure Θ_b can be found for which $X_{s\bar{b}}$ variables exist, that Θ_b is unique in satisfying this, and that the $X_{s\bar{b}}$ variables include all the a priori "sensitivity" variables such as signal energy, noise level, presence or absence of signal pedestal, ... (Note that any list of such variables must be very long and heterogeneous, given the variety of experimental situations to which detection models have been applied.) It is important to see that the three parts of this assumption are really quite distinct and separable. No bias measure may exist which satisfies

all three requirements together.

Given this (perhaps overambitious) demand that the three requirements should be satisfied simultaneously, to reject some candidate, Θ_b, it is sufficient merely to show that Θ_b varies with some one a priori sensitivity variable, e.g., signal energy. From a more circumspect point of view, this result shows merely that signal energy is not an $X_{s\bar{b}}$ variable for that parameterization. Several important $X_{s\bar{b}}$ variables for that parameterization may nonetheless exist, in which case the parameterization ought to be preserved. For brevity, we will write below of "evidence against $[\Theta_b]$" whenever Θ_b varies with some particular a priori sensitivity variable. This is deliberately noncommittal, and the reader is left to draw the qualified or unqualified conclusion as he thinks fit.

To avoid confusion, something should be said about Hodos's (1970) paper, since its title—"Nonparametric index of response bias for use in detection and recognition experiments"—might be taken to render this review redundant. A nonparametric comparison relation, relative to a set S of parameter-specifying functions $\psi_1(q,p), \ldots \psi_i(q,p), \ldots \psi_n(q,p)$, is a transitive, asymmetric binary relation, $*$, defined on some subset P of pairs of points $[(q,p),(q',p')]$ in the unit square, such that

$$(q,p) * (q',p') \rightarrow \psi_i(q,p) > \psi_i(q',p')$$

$$\text{for all } \psi_i \text{ in } S, \text{ all } [(q,p),(q',p')] \text{ in } P. \quad (13)$$

Thus P includes just those pairs about which all the ψ_i agree. This is interesting, since it allows inference to be extended beyond parameterizations related via Equations 7 and 8 into parameterizations related via Equations 5 and 6. Norman (1964) provides a *discriminability* comparison of exactly this sort, relative to sensitivity parameters which yield monotonic isosensitivity curves of nonincreasing slope. Hodos's title suggests that something similar is about to be provided for *bias*. This is quite misleading. What is in fact provided is a new (arbitrary) bias parameter which in its formulation makes no reference to any specific sensitivity parameter. Hodos's parameter is in this respect no different from several bias parameters already in existence (see below). It is, in any case, a trivial matter to deparameterize Equations 1 and 2 to give a direct isobias relation between p and q (the relation need not be a function, of course). Thus, it is no more than a matter of presentation whether or not a bias measure is nonparametric in Hodos's greatly reduced sense.

Certainly it would be most useful to have some bias relation, $*$, satisfying Equation 13, if S were to include most proposed bias measures and P were to be reasonably full. Unfortunately a glance at the isobias curves so far proposed (Figures 1 and 2) suggests that their disagreement is rather too general to allow these conditions to be satisfied. Of course, this may cease to be true if empirical work considerably reduces the number of plausible bias parameters; a genuine nonparametric approach might then be feasible.

TREATMENTS OF BIAS

Let $p(r_2 \mid s_1) = F_1(z)$, $p(r_2 \mid s_2) = F_2(z)$, where F_1 and F_2 are continuous functions of a real variable and increase monotonically from 0 to 1. Let change of *bias* be represented as change in z and change of *discriminability* as change in F_1, F_2. For fixed F_1 and F_2, any degree of bias, $z = z_c$ say, can be characterized in a number of different ways: in terms of $F_1(z_c)$, $F_2(z_c)$, or some function of the two; in terms of the local behavior of F_1 and F_2 around $z = z_c$; or in terms of the distance $(z_c - r)$ between z_c and some reference point r (where r must be fixed relative to one or both of F_1, F_2 since the origin and scale of the z-axis can only be fixed arbitrarily). Each of these three types of characterization is found among the bias measures currently in use.

Examples of the first type are:

(1) $p(r_2 \mid s_1)$

(2) $p(r_2) = \gamma \cdot p(r_2 \mid s_2) + (1 - \gamma) \cdot p(r_2 \mid s_1)$

(3) $\dfrac{\gamma}{(1 - \gamma)} \cdot \dfrac{p(r_2 \mid s_2)}{p(r_2 \mid s_1)}$

(1) gives the conditional probability of a false r_2, (2) the absolute probability of an r_2, and (3) the absolute probability of a correct r_2 divided by the absolute probability of an incorrect r_2. (1) came to psychology from normative statistical decision theory (Neyman & Pearson, 1933; Swets, Tanner, & Bridsall, Note 1). (2) and (3) arose from more intuitive and psychological considerations (Healy & Jones, 1973; Parks, 1966; Thomas & Legge, 1970).

An example of the second type is:

$$(4) \quad \beta = \frac{F_2'(z_c)}{F_1'(z_c)} = \left. \frac{dp(r_2 \mid s_2)/dz}{dp(r_2 \mid s_1)/dz} \right|_{z=z_c}$$

(4), like (1), came from normative decision theory; the rationale for its use in detection and recognition experiments is discussed extensively by Peterson, Birdsall, and Fox (1954) and by Green and Swets (1966). In models of the signal detection theory type, which assume that the subject takes an observation from one of two probability density functions and
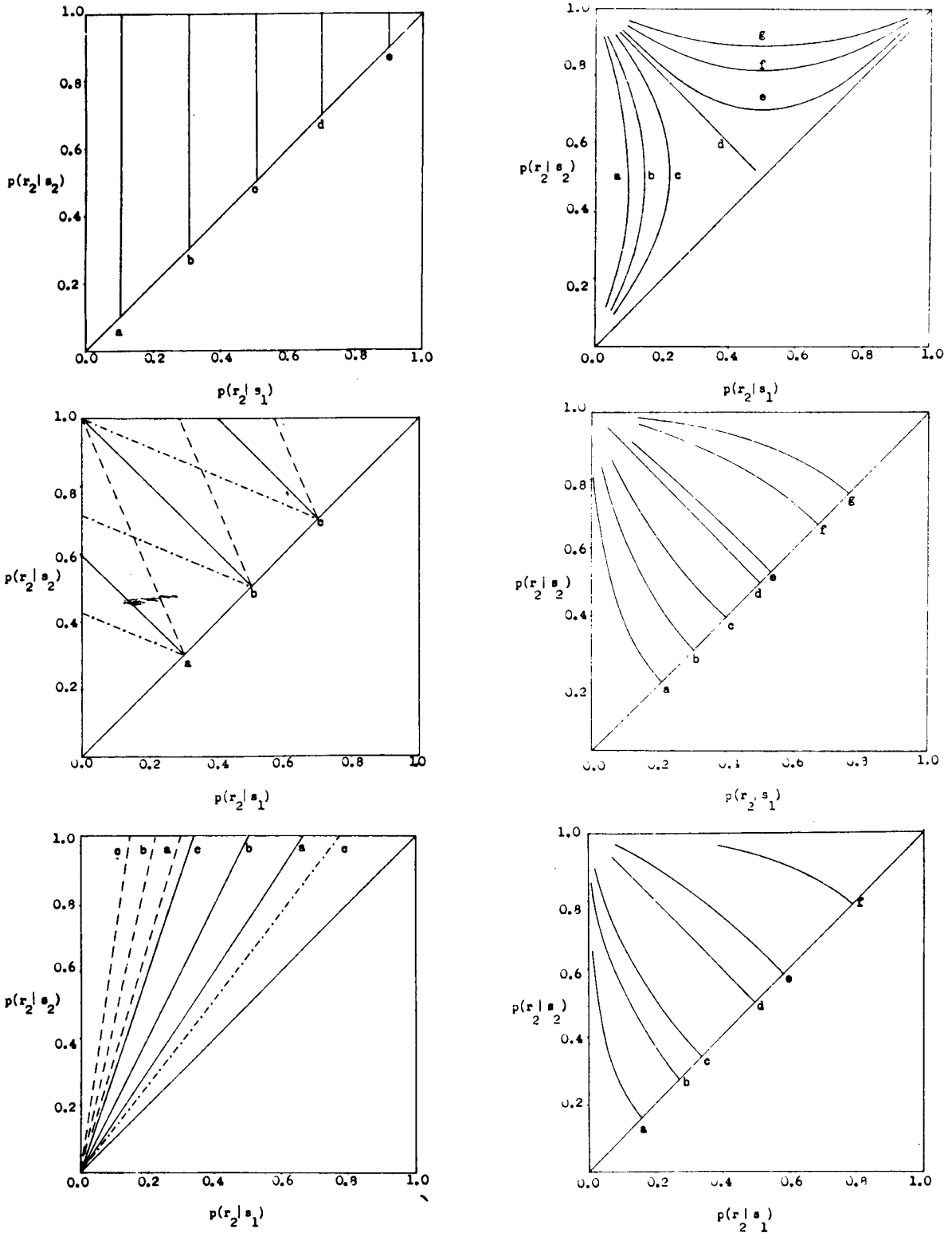
Figure 1. Illustrative isobias curves.

(a) [p(r₂ | s₁)]. ṗ(r₂ | s₁) = 0.1 (a), 0.3 (b), 0.5 (c), 0.7 (d), 0.9 (e).

(b) [p(r₂)]. p(r₂) = 0.3 (a), 0.5 (b), 0.7 (c). ———γ = 0.3, ———γ = 0.5, —·—·—γ = 0.7.

(c) [(γ/(1 - γ)) · (p(r₂ | s₂)/p(r₂ | s₁))]. (γ/(1 - γ)) · (p(r₂ | s₂)/ p(r₂ | s₁)) = 1.5 (a), 2.0 (b), 3.0 (c). ———γ = 0.3, ———γ = 0.5, —·—·—γ = 0.7.

(d) [β], given that F₁(z), F₂(z) are the normal distribution function, σ = 1. β = 2.3 (a), 1.7 (b), 1.3 (c), 1.0 (d), 0.9 (e), 0.7 (f), 0.5 (g).

(e) [b]. b = 3.9 (a), 2.3 (b), 1.5 (c), 1.0 (d), 0.9 (e), 0.5 (f), 0.3 (g).

(f) [C]. C = 1.0 (a), 0.6 (b), 0.4 (c), 0.0 (d), 0.2 (e), -0.8 (f).

compares it with a fixed criterion, (4) gives likelihood ratio at criterion; but an analogous quantity appears with a different interpretation within the choice theory of Luce (1963), so a general definition, such as (4), seems preferable.

Examples of the third type are:

(5) log b

(6) C

Both Luce's (1963) choice theory and signal detection theory in its commonest form assume $F_1(z) = F(z)$, $F_2(z) = F(z + d)$, where d is a discriminability parameter. In the former, F is the logistic distribution function ($\sigma^2 = \pi^2/3$) and d is $2 \log\eta$. In the latter, F is the normal distribution function ($\sigma^2 = 1$) and d is the familiar d'. (5), within choice theory, and (6), within signal detection theory, measure a distance $(z_c - z_0)$ in σ units. In each case, z_0, a point of zero bias, is chosen so that $F_2(z_0) = 1 - F_1(z_0)$; thus at $z_c = z_0$ $p(r_2 \mid s_2) = p(r_1 \mid s_1)$.

Depending on which bias measure the subject keeps constant as discriminability varies, very different isobias curves are traced out (Figure 1). There is one exception: (5) and (6) generate almost identical curves because of their similarity of definition and the relation between the logistic and the normal distributions (Bush, 1963). Clearly the form of $[\beta]$ depends on the particular $F_1(z)$, $F_2(z)$ assumed, while [log b] and [C] lack any rationale unless $F_1(z)$, $F_2(z)$ are as their parent theories suppose.

More complex treatments of bias have been put forward by Broadbent and by Smith. Broadbent's proposal (Broadbent, 1965, 1971) assumes that for each s and r there is a real number A_{sr} such that $p(r \mid s) = A_{sr}/\Sigma_{r'}A_{sr'}$. Log A_{sr} is then supposed to be a weighted average of the expected utility of r given knowledge that s has occurred and its expected utility given no information about s. Thus

$$\log A_{ij} = \Phi u_{ij} + (1 - \Phi)((1 - \gamma)u_{1j} + \gamma u_{2j})$$

$$0 < \Phi < 1$$

where Φ is a discriminability parameter and the u_{ij} are unknown transforms of the o_{ij}, to be estimated from the data. Isobias curves for this model are shown in Figure 2. Smith's (1968) model assumes that each decision in a psychophysical task may involve some *cost* (formally a nonnegative real number) and that the subject takes average cost into account in choosing his bias. Together with some plausible assumptions about the determination of cost and the nature of the detection process, this is shown to predict discontinuous isobias curves; that is, above a certain critical level of discriminability the subject should
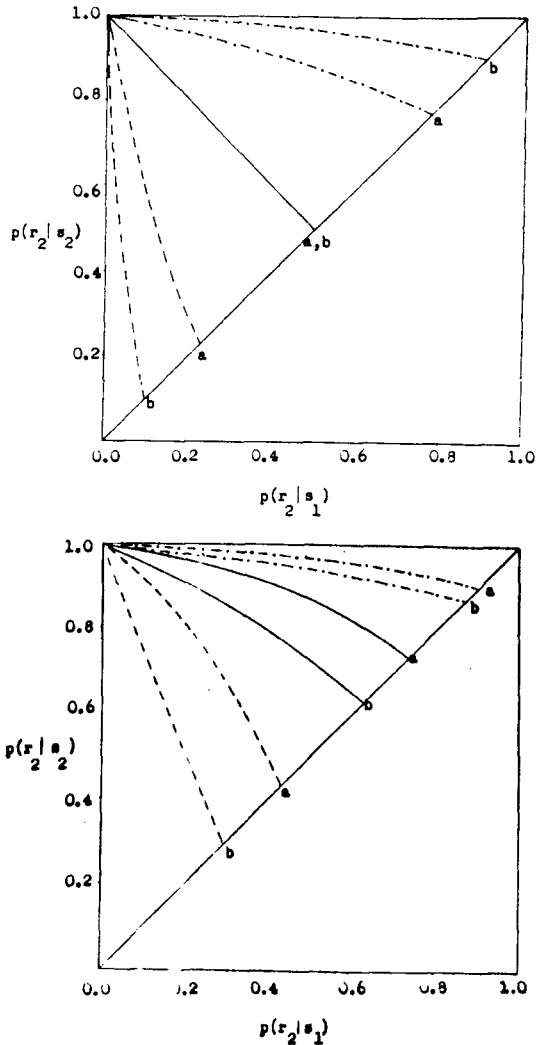
Figure 2. Illustrative isobias curves for Broadbent's model.
(a) $u_{11} = u_{22} = 5$, $u_{12} = u_{21} = -1$ (a); $u_{11} = u_{22} = 5$, $u_{12} = u_{21} = -5$ (b); — — — $\gamma = 0.4$, ———— $\gamma = 0.5$, — · — · —$\gamma = 0.6$.
(b) $u_1 = u_{22} = 4$, $u_{12} = -1$, $u_{21} = -3$ (a); $u_{11} = 5$, $u_{22} = 4$, $u_{12} = -1$, $u_{21} = -3$ (b); — — — $\gamma = 0.4$, ———— $\gamma = 0.5$, — · — · —$\gamma = 0.6$.

show negligible bias ($p(r_2 \mid s_2) \cong p(r_1 \mid s_1)$) even if bias is considerable at lower levels of discriminability. Smith's isobias predictions are much more complex than any described above, since a whole family of two-parameter isobias curves pass through any (q,p).

ISOBIAS CURVES: EXPERIMENTAL EVIDENCE

Some general points should be borne in mind throughout this section. First: the isobias functions so far taken seriously have arisen mostly from stationary independent trial models of detection and recognition. Certainly it is *possible*, despite the

existence in typical data of sequential effects, that one of these functions should describe empirical isobias curves (for example, see Bush, Luce, & Rose, 1964). It is to be hoped, however, that more attention will be given in the future to isobias functions within dependent trials models or in models which are careful to avoid assumptions about trial by trial events (e.g., Thomas & Legge, 1970). Second: existing work assumes that isobias curves have the same form across different subjects, different tasks, different ways of varying discriminability, and different types of bias conditions; there have been no tests of this (optimistic) assumption. Third: a common procedure has been to run a group of subjects at two or more discriminability levels under fixed bias conditions, and to compare the distribution of bias change with some chosen null hypothesis distribution. Even if isobias curves *are* identical across subjects, this is a very dubious procedure because of the way the proposed isobias curves are related. Thus, for example, if the subject holds $p(r_2 \mid s_1)$ constant, calculated β may show an increase, a decrease, or no change, depending on the starting point and the size of discriminability change (see Figure 1); so the fact that β change is approximately symmetrical about zero in a group of subjects does not, by itself, reveal much about the underlying isobias function. (Conversely, there is no reason to expect symmetry about zero, even if [β] holds, without some specific assumptions about variance in the unit square.) Fourth: the dissimilarity between the proposed isobias functions varies crucially with the level of bias and range of discriminability involved; there is little point in testing an isobias prediction in a region of the unit square where dissimilarities are minimal.

[$p(r_2 \mid s_1)$]

Treisman (1964) and Ingham (1970) have argued for [$p(r_2 \mid s_1)$] on the grounds that this requires less knowledge on the part of the subject than does [β]; they suggest that everyday life provides ample opportunity for the subject to learn to generate a given $p(r_2 \mid s_1)$. This latter point seems dubious; the experimenter normally takes elaborate steps to make s_1 more controlled than everyday-life stimuli, and this seems misguided unless the internal representation of s_1 is commensurately altered.

It is of course open to direct test whether subjects can generate given false-alarm rates when given instructions to do so. Swets, Tanner, and Birdsall (1961) instructed four experienced subjects to maintain $p(r_2 \mid s_1)$ within the ranges 0.0-0.07, 0.21-0.28, 0.43-0.50, 0.64-0.71, in a visual detection task. There was no trial by trial feedback, but at the end of each block of 50 trials they were told obtained $p(r_2 \mid s_1)$. Averaged over 18 blocks, obtained values were close to target values, the largest deviation

being 0.04; and this was achieved without drop in d' relative to more normal viewing conditions. Swets et al. do not point out, however, that errors were very systematic. Of the 10 points which lie outside their target ranges, nine lie below, despite one range being bounded below at zero. The argument is presumably that any adjustment process based on the feedback must provide a source of variance *additional* to any involved in normal performance; therefore d' should fall (Thomas & Myers, 1972) unless the adjustment process is very efficient. Against this, it seems possible that one source of variance has been *substituted* for another. (Suppose, for example, that under normal conditions the subject adjusts bias at the end of each block, but on the bais of $p(r_2)$ rather than $p(r_2 \mid s_1)$. There ceases to be any simple reason to expect a d' decrement in the Swets et al. condition.)

Direct evidence for [$p(r_2 \mid s_1)$] was put forward by Treisman et al. (Howarth & Treisman, 1958; Treisman, 1964; Treisman & Howarth, 1959). They found that given an accessory stimulus at a fixed interval from s_2 subjects increased $p(r_2 \mid s_2)$ but kept $p(r_2 \mid s_1)$ constant. A fundamental weakness, however, is that $p(r_2 \mid s_1)$ was low throughout the experiments (average value 0.041); in this region [β], [b], [C] and Broadbent's model each predict approximately constant $p(r_2 \mid s_1)$ (Figures 1 and 2). Treisman et al. give no evidence that [$p(r_2 \mid s_1)$] provides a better fit than these alternatives.

Evidence against [$p(r_2 \mid s_1)$] is to be found in some experiments discussed in later sections and in the data of Markowitz and Swets (1967), who obtained six-point isobias plots from three subjects under three values of γ in an auditory detection task; $p(r_2 \mid s_1)$ was negatively related to discriminability throughout. Markowitz and Swets, in fact, provide the best defined individual-subject isobias plots available. Regrettably, their binary response, forced choice, and rating response data suggest quite different assumptions about $F_1(z)$, $F_2(z)$, and so cannot be used to test the theory dependent functions [β], [b], or [C].

[$p(r_2)$]

The curves [$p(r_2)$] have the general form

$$p = k - ((1 - \gamma)/\gamma)q$$

$$0 < \gamma < 1$$

$$k - 1 \leqslant \frac{(1-\gamma)}{\gamma} q \leqslant k$$

The slope, $(-(1 - \gamma)/\gamma)$, is determined by the presentation probabilities alone; k is chosen by the subject. If $k = 1$, the subject matches $p(r_2)$ to γ; otherwise $p(r_2) \gtrless \gamma$ as $k \gtrless 1$. If k is assumed to depend on the outcome matrix and instructions, but not on γ, the model can be tested in more than one

way: its isobias predictions can be tested as usual; in addition, a single (q,p) is sufficient to estimate k, and this determines the isobias functions on which all other (q,p) must lie given other γ. The first procedure is discussed here, the second in the next section.

Parks (1966) argued for $[p(r)_2]$ on the basis of the recognition memory data of Strong (1912). (For the application of signal detection models to recognition memory, see Green and Swets (1966).) Strong held γ at 0.5 and varied discriminability via the size of the exposure set (5, 10, 25, 50, 100, 150 items). Items were magazine advertisements. The subject respond-ed with a 4-point confidence rating. Parks cumulated the data to give a single $p(r_2 \mid s_1)$ and $p(r_2 \mid r_2)$ for each condition. Since the subject was not told γ, Parks writes $p = k$ -Φq, where Φ represents the subject's estimate of $((1 - \gamma)/\gamma)$ and is assumed constant over groups. Two of the groups were used to estimate k and Φ, and this was sufficient to predict q given p for the other groups. But since γ was 0.5, and estimated k was 0.971, the isobias curve under test consisted, approximately, of the unit square negative diagonal (see Figure 1). Thus the success of the prediction argues only against $[p(r_2 \mid s_1)]$ and $[(\gamma/(1 - \gamma)) \cdot (p(r_2 \mid s_2)/p(r_2 \mid s_1))]$.

A similar objection applies to Thomas and Legge's more recent case for $[p(r_2)]$ (Thomas & Legge, 1970). They quote the results of Swets (1959), who ran three subjects at each of four discriminability levels in an auditory detection task, with γ throughout at 0.5. $p(r_2)$ was close to γ in all cases, but again this argues against only $[p(r_2 \mid s_1)]$ and $[(\gamma/(1 - \gamma)) \cdot [p(r_2 \mid s_2)/p(r_2 \mid s_1)]]$. Thomas and Legge's support for $[p(r_2)]$ is in part motivated by their unwillingness to assume that the subject has the information required to compute β, especially when he is inexperienced and has no trial by trial feedback. It should be noted that even if the subject does set his bias so as to generate a target value of $p(r_2)$, that in no way implies the isobias function $[p(r_2)]$, since the subject might still take task difficulty into account in choosing his target $p(r_2)$.

$$[(\gamma/(1 - \gamma)) \cdot (p(r_2 \mid s_2)/p(r_2 \mid s_1))]$$

This isobias function preserves the ratio of correct to incorrect r_2 responses. It was recently suggested by Healy and Jones (1973) on grounds that can only be described as intuitive. Like the proposal just discussed, it makes predictions about the effects of varying γ as well as about the shape of isobias curves; only the latter are dealt with here. A crucial prediction is that isobias curves have positive slope. This is clearly falsified by the data of Swets (1959), Markowitz and Swets (1967), and Ingleby (1969), for auditory detection, and by that of Strong (1912) for recognition memory tasks. It is falsified, for intelligibility (word recognition) experiments, by data

which Healy and Jones themselves quote (Clarke, 1960; Decker & Pollack, 1959; Pollack & Decker, 1958). Thus, Healy and Jones's attempt to draw substantive conclusions about recall on the basis of this assumed isobias function seems misguided.

[β]

Broadbent (1971, pp. 193-196, 237-239) has recently lent his weight to [β], drawing on data from three different fields.

(1) Auditory Selective Attention

A crucial experiment, which has never been performed, would be to require subjects to dis-criminate between s_1 and s_2, under two conditions, first with s_1, s_2 occurring on an unattended channel, second with s_1, s_2 occurring on an attended channel, and to trace out isobias curves in each condition by varying the objective discriminability of s_1 and s_2; it might then be seen directly whether the isobias curves were of the same family (they need not be) and if so whether they were coincidental. Unfortunately, the existing studies (Broadbent & Gregory, 1963; Moray & O'Brien, 1967; Treisman & Geffen, 1967) consider only one objective discriminability level, under attended and unattended conditions. All apply the classical signal detection theory model and find change in d' without significant change in β. This, by itself, implies nothing about isobias curves, since it is unclear whether the data points lie on one isobias curve or on two. Broadbent, however, argues that the latter interpretation is in detail implausible. He makes the crucial assumption that channel rejection, if it affects bias at all, must increase bias against r_2. The subject cannot then be moving between members of $[p(r_2 \mid s_1)]$; for $p(r_2 \mid s_1)$ in fact increases given channel rejection. Nor can he be moving between members of $[p(r_2)]$ or $[C]$; for this would be compatible with unchanged β only if $p(r_2 \mid s_2) < p(r_1 \mid s_1)$ (see Figure 1), which is contradicted by the Broadbent and Gregory (1963) rating data.

The argument is weak, for three reasons. First: the justification for Broadbent's crucial assumption seems obscure. It would be reasonable enough if one were trying to account for selective attention solely in terms of bias change. But, given that discriminability change certainly occurs, whatever else, it needs to be decided by experiment, not fiat, whether in addition bias changes against r_2, in favor of r_2, or in a direction which depends on the initial bias level. Second: the argument assumes that selective attention proceeds by channel attenuation rather than by time-sharing between channels; it seems uncertain, at present, whether this is correct (Broadbent, 1971; Moray, 1969). Third: the data on which the argument rests are far from ideal. Only in the Broadbent and Gregory (1963) experiment is there any evidence for the

correctness of the distributional assumptions on which β is based. In addition, Treisman and Geffen's results are based on pooled data and estimated values of $p(r_2 \mid s_1)$; subjects were trying to detect target (s_2) words among nontarget words, but the authors did not count a nontarget word as an s_1 unless it was phonemically similar to the target by their fairly arbitrary definition of similarity. Moray and O'Brien's data include zero false-alarm rates, which have made necessary lower-bound estimates of β in some cases.

(2) Vigilance

In the modified vigilance situation studied by Ingleby (1969), subjects detected sinusoids in noise, with trial by trial feedback. In some conditions, before the warning which preceded each trial a "cue" light gave probabilistic information about whether a signal would be presented on that trial. γ was always 0.5. Subjects responded with a four-category confidence rating with differential payoff attached to each category. Separate experiments varied objective discriminability and the information given by the cue light; within each experiment, a condition was run in which the cue light was dispensed with.

Isosensitivity functions were compatible with $F_1(z)$ and $F_2(z)$ being cumulative normal, but with $\sigma_2 = 1.25\sigma_1$. Calculating β on this assumption, Ingleby argues for [β] on three grounds. First: as discriminability varied, the central criterion remained on the negative diagonal. Second: criteria lay further apart (in terms of the F_1 standard deviation, σ_1) when they were lax than when they were strict. This would follow if subjects maintained the criteria at a similar distance in terms of log β, because of the relation between log β and z/σ_1 when $\sigma_1 < \sigma_2$; this relation is illustrated in Figure 3, the crucial point being that the function is positively accelerating. Third: at low discriminability, the cue light had a greater effect on criteria, in terms of z/σ_1, than at high discriminability, and in conditions when the cue light was omitted criteria lay closer together in terms of z/σ_1 at high discriminability.

Of these three points, the first in fact argues only against $[p(r_2 \mid s_1)]$ and $[(\gamma/(1 - \gamma)) \cdot (p(r_2 \mid s_2)/p(r_2 \mid s_1))]$. The second fails to argue against $[p(r_2)]$, because the relation between $p(r_2)$ and z/σ_1 when $\sigma_1 < \sigma_2$ is rather similar to that between log β and z/σ_1 (Figure 3). The third is potentially a stronger argument for [β]. Unfortunately, as Ingleby himself points out, it is not at all clear from his data whether the effect was in fact large enough to keep β constant across discriminability. Ingleby tried to detect change in log β using both analysis of variance and Wilcoxon's test; the latter, but not the former, gave a significant result. There seems little theoretical basis for either of the null hypotheses assumed; both
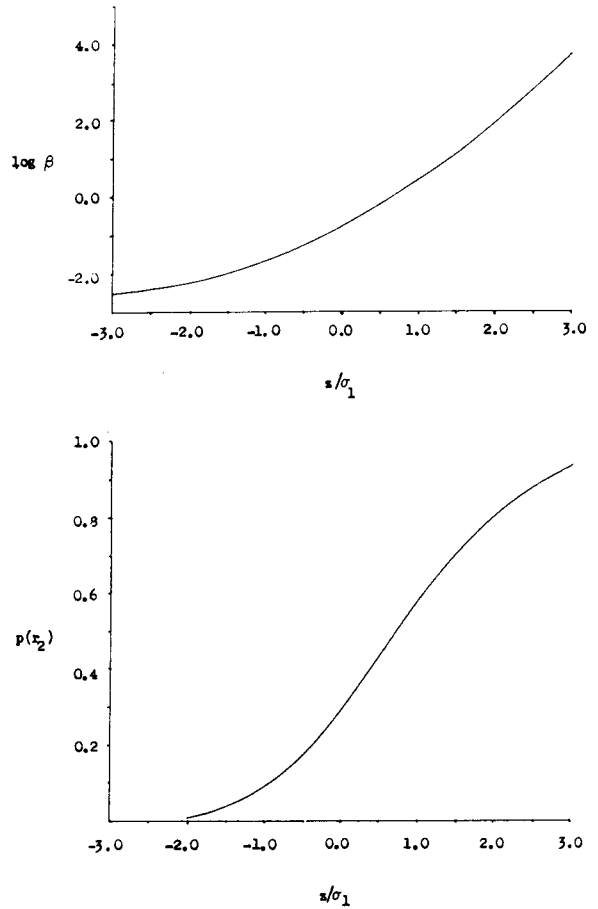

Figure 3. The relations between log β and z/σ_1, and between $p(r_2)$ and z/σ_1, under the illustrative assumption

$$F_1(z) = F(z) = \int_{-z/\sigma_1}^{+\infty} N(0,1)\, dx,$$

$F_2(z) = F(0.8z + 1.2\sigma_1)$. For Figure 3(b), it is assumed that $\gamma = 0.5$.

depend on unstated assumptions about variance in the unit square. This, and the disagreement between the tests, make Broadbent's unqualified claim that Ingleby's subjects kept β constant (Broadbent, 1971, p. 195) seem rather overstated.

(3) Scaling

Broadbent and Gregory, in an unpublished experiment reported briefly in Broadbent (1971, pp. 237-239), required subjects to discriminate among collections of ball-bearings according as they were more or less numerous than each of three different standard collections; there were four possible responses at each trial. Two groups were run with the stimuli equally distributed over the four categories, but with the range of stimuli differing between the groups. This allowed three values of β to

be calculated for each subject (using choice theory assumptions about $F_1(z)$, $F_2(z)$). Discriminability did indeed vary between groups, but the β values did not. Unfortunately, however, there was no change either in log b, C, $p(r_2)$, $p(r_2 \mid s_1)$, or even $p(r_1 \mid s_2)$, using the same Mann-Whitney U statistic that was used to assess β change.[2] Thus Broadbent's claim that the experiment supports $[\beta]$ seems very puzzling.

In summary, then, the case for $[\beta]$ seems most equivocal. It should be noted that Altham (1973) has recently assumed $[\beta]$ in formulating her nonparametric index of discriminability. There seems little point in trying to avoid assumptions about isosensitivity curves (about which there is much evidence) at the cost of making assumptions about isobias curves (about which much less is known).

[log b] and [C]

Hardy and Legge (1968) argue for [C] on the basis of an experiment which investigated the effect of concurrent visual presentation of "emotional" and "neutral" material on performance in an auditory signal detection task. The basic task was to detect a sinusoid in continuous white noise. The observation intervals were marked by illumination of a rectangular area containing a faint image of a word (subjects were unaware of its presence) which was on some trials "emotional" and on some trials "neutral." γ was 0.5. Subjects responded on a 4-point rating scale. Fourteen subjects were run for 64 trials, after minimal practice (six trials). Hardy and Legge show that if the effect of the "emotional" words is assumed to be on discriminability alone then [C] must be postulated rather than $[\beta]$. Against this, there seems no reason to make this assumption. Even if it is made, it seems likely that $[p(r_2)]$ would provide much the same sort of fit given that γ was 0.5 (see Figure 1).

Smith's Cost Model

Smith has made two attempts to find the predicted discontinuities in isobias curves. In Smith (1969), a recognition memory design was used: the material was three-letter nonsense syllables, γ was 0.5, and the subjects responded "yes" (r_2) or "no" (r_1), feedback being given at the end of each list. Separate groups of subjects were assigned to three payoff matrices, to produce zero bias, bias towards r_1, and bias towards r_2.

There was no experimental variation of discriminability; rather, the data from the zero bias condition were used to give a partial ordering of the items with respect to discriminability and bias was compared across different levels of the partial ordering. In Smith (1970), subjects were presented with lists of names and were required to respond "yes" (r_2) or "no" (r_1) to each one according as they thought the person concerned was or was not born

before 1910 (s_2 and s_1 items, respectively). Again, γ was 0.5 and three payoff matrices were used, exactly as in the previous experiment.

In each case, there was some evidence for the predicted discontinuities, though in neither case is there any evidence that they occurred at the particular critical level of discriminability stipulated by the model. Three criticisms should be noted. First: within the psychophysical context, the data would have been considered pre-asymptotic; the subject is being required to establish consistent behavior to items scattered over the full range of discriminability, and it is quite unclear whether he can do this with such minimal practice. Second: the stimuli were complex and could be perfectly identified prior to the subject's decision about them. It is possible, therefore, that subjects adopt different biases for different subsets of items and that these subsets may not be random samples with respect to discriminability. (An arbitrary example: the subject might be biased to respond "no" to items in which the letter x appears, but such items may be more or less difficult than the average.) This, too, would lead to interactions between discriminability and bias, but for reasons quite foreign to Smith's model. Third: even if, as the model suggests, the subject makes a preliminary estimate of the discriminability of an item before choosing an appropriate level of bias, it need not be for the reason given by the model. An example: the subject's own initials are presented as a member of the combined set in a recognition memory experiment. He reflects that if they had been presented in the exposure set he would almost certainly have noticed it and remembered it very clearly, so he responds "no" unless he is very sure that the item was presented. In theoretical terms, he has adjusted his bias for an item of high discriminability. But this has nothing whatever to do with cost as it appears in Smith's model.

BIAS AND THE PRESENTATION PROBABILITIES

The next two sections deal with some attempts to describe the relation between bias and the variables which determine bias. Clearly, if two parameterizations are related via Equation 6, not Equation 8, then there is no possibility of their agreeing as to this further issue, since they must necessarily write bias as a function of different variables.

This section deals with the effects of varying γ when the payoff matrix and discriminability are held fixed. It is natural to suspect that the effects may differ depending on whether the subject knows γ is being varied, whether the subject knows γ, and whether feedback is presented; this is strongly borne out by the results of Kinchla (Note 3), Tanner, Haller, and

Atkinson (1967), and Tanner, Rauk, and Atkinson (1970). To accommodate these results, any model must allow $p(r_2)$ to increase or decrease as γ increases, depending on the information conditions. The models to be considered in this section and the next completely fail to do this. At best, therefore, they may hold for some restricted range of conditions not yet made explicit.

Parks (1966), Creelman and Donaldson (1968), and Thomas and Legge (1970) assume $[p(r_2)]$ and take k in $p = k - ((1 - \gamma)/\gamma)q$ to depend on payoff but not on γ. All test the special case $k = 1$, under which assumption the subject matches $p(r_2)$ to γ (the matching hypothesis). Parks (1966) uses his own recognition memory data and that of Davis, Sutherland, and Judd (1961). In both cases, γ was varied by keeping the size of the exposure set constant while increasing the size of the confusion set. In neither case did the subject have trial by trial feedback, but in the Parks experiment he was clearly told γ. The fit seems satisfactory for both experiments, but Parks considers only group data. This seems very regrettable, given the recent finding in the related field of probability learning that response rates for groups are not representative of response rates for individual subjects (Norman, 1971). Creelman and Donaldson (1968), on the other hand, apply the prediction to individual subjects. Two subjects were run in a line-length recognition task under two discriminability conditions, with γ taking 11 values in the range 0.05-0.95. Payoff was fixed and symmetric. The subjects were later run under five levels of γ with all payoff entries multiplied by 10. No formal test of fit is presented, but at an informal level both subjects seem to show systematic tendencies to give $p(r_2 \lessgtr \gamma$ as $\gamma \gtrless 0.5$ in the low discriminability, low payoff condition (Creelman & Donaldson, 1968, Figure 2), whereas under high payoff one subject shows $p(r_2) < \gamma$ for all values of γ. The data are very equivocal therefore. Thomas and Legge (1970) quote the data of Tanner, Swets, and Green (Note 4). Two subjects were run in an auditory detection task, under a constant symmetric payoff matrix ($o_{11} = o_{22}$, $o_{12} = o_{21}$), with γ varied through 0.1, 0.3, 0.5, 0.7, 0.9. Trial by trial feedback was always given. Each subject was run for two sessions of 300 trials under each γ. Thomas and Legge assert that both subjects conform to the matching hypothesis. Both subjects in fact seem aberrant, however, but in different ways (see Thomas & Legge, 1970, p. 68). One subject shows large differences between conditions run under the same γ (thus there is a difference in p of 0.2, in one case, and a difference in q of 0.2, in another). The other subject seems consistently more biased than the model demands (7 out of 8 points err in this direction). Thus, the fit does not seem very impressive.

A second proposal (Healy & Jones, 1973) assumes $[(\gamma/(1 - \gamma)) \cdot (p(r_2 \mid s_2)/p(r_2 \mid s_1))]$ and takes k in $(\gamma/(1 - \gamma)) \cdot (p(r_2 \mid s_2)/p(r_2 \mid s_1)) = k$ to depend on payoff and instructions but not on γ. This equation predicts bizarre isosensitivity functions, a fact ignored by Healy and Jones. If for $\gamma = 0.5$ the subject generates the point (q',p') then $((1 - \gamma)/\gamma) < (q'/p')$ implies $p(r_2 \mid s_2) < p(r_2 \mid s_1)$, which is worse than chance performance. For example, if the subject generates $(0.4, 0.6)$ for $\gamma = 0.5$, then he must do worse than chance for $\gamma > 0.6$. Thus, against the facts (Green & Swets, 1966), isosensitivity curves should always be grotesquely skewed if they are traced out by varying γ, except when discriminability is negligible. It is not surprising, therefore, that the proposal conflicts wildly with the results of Galanter and Holman (1967) and Tanner, Swets, and Green (Note 4), who obtained symmetric (unskewed) isosensitivity curves by varying γ under constant payoff. The central points of their functions may be used to estimate k, and $p(r_2 \mid s_2)$ can then be predicted from the other $p(r_2 \mid s_1)$. The predictions are sometimes embarrassingly low $(p(r_2 \mid s_2) < p(r_2 \mid s_1))$ and sometimes embarrasingly high $(p(r_2 \mid s_2) > 1.0)$. Thus, this proposal is clearly inadequate as it stands.

BIAS, PRESENTATION PROBABILITY, AND PAYOFF

This section considers attempts to predict bias as a function of γ and the o_{ij} conjointly. As is well known, an early approach was to assume that subjects set their bias so as to maximize expected payoff. The reasons for rejecting the model are equally well known, and are discussed extensively in Green (1960), Luce (1963), and Broadbent (1971). This review considers two later proposals, which were prompted by the failure of the maximization of expected value hypothesis, and which have not been adequately discussed.

Thomas and Legge (1970) assume $[p(r_2)]$ and for payoff matrices with $o_{11} = -o_{12}$ and $o_{22} = -o_{21}$ take k in $p = k - ((1 - \gamma)/\gamma)q$ to be given by $k = 2 \cdot o_{22}/(o_{11} + o_{22})$. Because bias is a simple function of the o_{ij} alone, no room is left for individual differences; this seems a little implausible. They apply the suggestion to the data of Green and Swets (1966, p. 90), who ran one subject over five payoff conditions in an auditory detection task, with $\gamma = 0.5$ throughout; the fit seems satisfactory. On the other hand, they ignore the rather fuller data of Galanter and Holman (1967), with which their proposal is clearly incompatible. Observed and predicted values of k are given in Table 1. The deviations are often large and are perfectly systematic: no subject shows sufficient bias under any condition. (The data have been read off double-probability plots; hence the use

Table 1
Observed and Predicted Values of k for the Data of Galanter and Holman (1967)

Matrix	1	2	3	4
o_{11}	.1	.1	1.5	2.5
o_{22}	2.5	1.5	.1	.1
Expected k	1.923	1.875	.125	.077
Observed k				
s_1	< 1.75	< 1.50	> .40	> .25
s_2	< 1.80	< 1.60	> .50	> .35
s_3		< 1.30	> .60	
s_4		< 1.40	> .70	

Note—The o_{ij} are in units of 1 cent. In all matrices $o_{12} = -o_{11}$ and $o_{21} = -o_{22}$.

of inequality signs. In all cases, ambiguities have been resolved so as to favor Thomas and Legge's prediction; thus Table 1 gives a lower bound to the deviations between expected and observed values.) Problematically, discriminability is quite similar between experiments. Galanter and Holman used greater ratios of o_{11} to o_{22} than the maximum used by Green and Swets (9:1), but the former's data for matrices 2 and 3 show less bias than would be predicted even if the ratio 9:1 had been used.

Broadbent's model, summarized in the second section of this review, has never been seriously tested. It cannot apply to all detection and recognition tasks, since it predicts that $p(r_2)$ will always increase as γ increases; as pointed out previously, this is incorrect (Tanner, Rauk, & Atkinson, 1970). Broadbent has suggested (personal communication) that a plausible modification would be to replace γ in the model's equations with γ' (say), representing the subject's estimate of γ. Since, however, for constant γ, γ' can clearly not be the same for zero and perfect discriminability, it becomes necessary to allow that between these two extremes γ' may vary with discriminability. Thus the isobias predictions of the original model (Figure 1) have to be withdrawn (since they are traced out by keeping γ and the u_{ij} constant while varying Φ) and none can be put in their place until it is known how γ' varies as a function of γ and discriminability conjointly.

CONCLUSION

The review has failed to find any adequate experimental foundation for current treatments of bias. There has been a long continued tendency (e.g., Treisman (1964, p. 26) and Healy and Jones (1973, passim)) to choose measures of bias on the basis of a priori argument and intuition. This contrasts oddly with the detection theory approach to discriminability, which advanced from a realization that a discriminability measure could not be chosen arbitrarily, or on the basis of intuition (as enshrined in the traditional psychophysicist's guessing correction).

It is to be hoped that something of the empirical fervor which once went into the proliferation of isosensitivity curves may soon be transferred to questions about bias.

The paper ends with some strategic considerations: (i) It should be remembered that no one bias parameter need remain invariant under change in *all* the a priori sensitivity variables. We may need several bias parameters, not one. This is so even if it is optimistically assumed that there are no differences between tasks, between subjects, and between types of bias conditions. (ii) Attempts to relate bias to presentation probability, or to presentation probability and payoff conjointly, must presuppose an isobias function if they are to apply to more than one discriminability level. There seems little point, therefore, in multiplying such attempts until the form of isobias curves is well established. (iii) Attempts to make inferences from the distribution of bias change in groups of subjects should be avoided, until there is some good evidence that isobias curves are indeed of the same family in different subjects. (iv) The relatively clear-cut differences which do exist between the isobias curves so far proposed should not be wasted. Thus $[\beta]$ is unique in going to (0,0) or (1,1) as discriminability goes to zero. $[p(r_2)]$, unlike $[C]$ or $[b]$, does not go to (0,1) as discriminability improves (provided $\gamma \neq 0.5$), and unlike $[p(r_2 \mid s_1)]$ or $[(\gamma/(1 - \gamma)) \cdot (p(r_2 \mid s_2)/p(r_2 \mid s_1))]$ goes to (0,p), $p < 1.0$ (provided $\gamma > 0.5$, $k < 1$). These examples suggest that the failure of the last decade to resolve disputes about bias is partly due to the inadequate methods used rather than to the difficulty of the problem.

REFERENCES

ALTHAM, P. M. E. A nonparametric measure of signal discriminability. *British Journal of Mathematical and Statistical Psychology*, 1973, 26, 1-12.

ATKINSON, R. C., BOWER, G. H., & CROTHERS, E. J. *An introduction to mathematical learning theory.* New York: Wiley, 1965.

ATKINSON, R. C., & KINCHLA, R. A. A learning model for forced-choice detection experiments. *British Journal of Mathematical and Statistical Psychology*, 1965, 18, 183-206.

BROADBENT, D. E. A reformulation of the Yerkes-Dodson law. *British Journal of Mathematical and Statistical Psychology*, 1965, 18, 145-157.

BROADBENT, D. E. *Decision and stress.* London: Academic Press, 1971.

BROADBENT, D. E., & GREGORY, M. Vigilance considered as a statistical decision. *British Journal of Psychology*, 1963, 54, 309-323.

BUSH, R. R. Estimation and evaluation. In R. D. Luce, R. R. Bush, and E. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. 1). New York: Wiley, 1963.

BUSH, R. R., LUCE, R. D., & ROSE, R. M. Learning models for psychophysics. In R. C. Atkinson (Ed.), *Studies in mathematical psychology.* Stanford: Stanford University Press, 1964.

CLARKE, F. R. Confidence ratings, second-choice responses, and confusion matrices in intelligibility tests. *Journal of the Acoustical Society of America*, 1960, **32**, 35-46.

CREELMAN, C. D., & DONALDSON, W. ROC curves for discrimination of linear extent. *Journal of Experimental Psychology*, 1968, **77**, 514-516.

DAVIS, R., SUTHERLAND, N. S., & JUDD, B. R. Information content in recognition and recall. *Journal of Experimental Psychology*, 1961, **61**, 422-429.

DECKER, L., & POLLACK, I. Confidence ratings and message reception for filtered speech. *Journal of the Acoustical Society of America*, 1959, **30**, 432-434.

EGAN, J. P., SCHULMAN, A. I., & GREENBERG, G. Z. Operating characteristics determined by binary decisions and by ratings. *Journal of the Acoustical Society of America*, 1959, **31**, 768-773.

GALANTER, E., & HOLMAN, G. L. Some invariances of the iso-sensitivity function and their implications for the utility function of money. *Journal of Experimental Psychology*, 1967, **73**, 333-339.

GREEN, D. M. Psychoacoustics and detection theory. *Journal of the Acoustical Society of America*, 1960, **32**, 1189-1203.

GREEN, D. M., & SWETS, J. A. *Signal detection theory and psychophysics*. New York: Wiley, 1966.

HARDY, G. R., & LEGGE, D. Cross-modal induction of changes in sensory thresholds. *Quarterly Journal of Experimental Psychology*, 1968, **20**, 20-29.

HEALY, A. F., & JONES, C. Criterion shifts in recall. *Psychological Bulletin*, 1973, **79**, 335-340.

HODOS, W. Nonparametric index of response bias for use in detection and recognition experiments. *Psychological Bulletin*, 1970, **74**, 351-354.

HOWARTH, C. I., & TREISMAN, M. The effect of warning interval on the electric phosphene and auditory thresholds. *Quarterly Journal of Experimental Psychology*, 1958, **10**, 130-141.

INGHAM, J. B. Individual differences in signal detection. *Acta Psychologica*, 1970, **34**, 39-50.

INGLEBY, J. D. Decision making processes in human perception and memory. PhD thesis, University of Cambridge, 1969.

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.

MARKOWITZ, J., & SWETS, J. A. Factors affecting the slope of empirical ROC curves: Comparison of binary and rating responses. *Perception & Psychophysics*, 1967, **2**, 91-100.

MORAY, N. *Attention: Selective processes in vision and hearing*. London: Hutchinson, 1969.

MORAY, N., & O'BRIEN, T. Signal detection theory applied to selective listening. *Journal of the Acoustical Society of America*, 1967, **42**, 765-772.

NEWMAN, J., & PEARSON, E. S. On the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society [London]*, 1933, **A231**, 289-337.

NORMAN, D. A. A comparison of data obtained with different false-alarm rates. *Psychological Review*, 1964, **71**, 243-246.

NORMAN, M. F. Statistical inference with dependent observations: Extensions of classical procedures. *Journal of Mathematical Psychology*, 1971, **8**, 444-451.

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

PETERSON, W. W., BIRDSALL, T. G., & FOX, W. C. The theory of signal detectability. *IRE Transactions, Professional Group on Information Theory*, 1954, **4**, 171-212.

POLLACK, I., & DECKER, L. Confidence ratings, message reception, and the receiver operating characteristic. *Journal of the Acoustical Society of America*, 1958, **30**, 286-292.

SMITH, P. T. Cost, discriminability, and response bias. *British Journal of Mathematical and Statistical Psychology*, 1968, **21**, 35-60.

SMITH, P. T. The non-monotonicity of the psychometric function in recognition memory. *Perception & Psychophysics*, 1969, **5**, 329-337.

SMITH, P. T. Decision processes in long term memory. *Acta Psychologica*, 1970, **33**, 315-325.

STRONG, E. K. The effect of length of series upon recognition memory. *Psychological Review*, 1912, **19**, 447-462.

SWETS, J. A. Indices of signal detectability obtained with various psychophysical procedures. *Journal of the Acoustical Society of America*, 1959, **31**, 511-513.

SWETS, J. A., TANNER, W. P., & BIRDSALL, T. G. Decision processes in perception. *Psychological Review*, 1961, **68**, 301-340.

TANNER, T. A., JR., HALLER, R. W., & ATKINSON, R. C. Signal recognition as influenced by presentation schedules. *Perception & Psychophysics*, 1967, **2**, 349-358.

TANNER, T. A., JR., RAUK, J. A., & ATKINSON, R. C. Signal recognition as influenced by information feedback. *Journal of Mathematical Psychology*, 1970, **7**, 259-274.

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

THOMAS, E. A. C., & MYERS, J. L. Implications of latency data for threshold and non-threshold models of signal detection. *Journal of Mathematical Psychology*, 1972, **9**, 253-285.

TREISMAN, A. M., & GEFFEN, G. Selective attention: Perception or response. *Quarterly Journal of Experimental Psychology*, 1967, **19**, 1-17.

TREISMAN, M. The effect of one stimulus on the threshold for another: An application of signal detectability theory. *British Journal of Statistical Psychology*, 1964, **17**, 15-35.

TREISMAN, M., & HOWARTH, C. I. Changes in threshold level produced by a signal preceding or following the threshold stimulus. *Quarterly Journal of Experimental Psychology*, 1959, **11**, 129-142.

WICKELGREN, W. A. Unidimensional strength theory and component analysis of noise in absolute and comparative judgements. *Journal of Mathematical Psychology*, 1968, **5**, 102-122.

NOTES