Perception & Psychophysics
2003, 65 (6), 844-860

An assessment of response bias for the *same–different* task: Implications for the single-interval task

MICHAEL J. HAUTUS and STEPHEN COLLINS
University of Auckland, Auckland, New Zealand

Two auditory amplitude discrimination experiments were conducted using the *same–different* experimental design. Observer bias was manipulated, in the first experiment, by varying payoff matrices and, in the second, by varying prior probabilities of signal presentation. Five levels of bias manipulation and four levels of difficulty were employed in each experiment. Each observer received all combinations of bias manipulation and difficulty, but with only one of these combinations within each block of trials. Nine indices of bias were assessed by simultaneously fitting isosensitivity and isobias functions to the data and by fitting isobias functions only. Although none of the indices tested provided an exceptionally good fit to the data, two indices stood out from the rest. These were c^*_{sd} and c_{i}, indices with isobias contours similar in shape to those for the c index derived from the *yes–no* task.

Biasing variables play an important role in the process of perceptual judgment. These variables relate to such factors as motivation, expectation, and anticipated payoffs and are not part of the sensory stimulus. Nevertheless, they have an inescapable effect on the judge's response. For example, a wine taster must reduce any propensity toward giving a particular wine an award because of personal preference or friendship with the vintner. In the same way, bidders at auctions need to decipher the fast stream of auditory input they receive from the auctioneer. In doing this, they must ensure that they do not get swept away in the moment and lose track of the criterion they set as a maximum purchase price. Both of these tasks are typically performed against a background of competing stimuli and are subject to bias when a sensory judgment is uncertain. In understanding and quantifying judgments made in such practical situations, as well as those made in an experimental setting, independent measures are required that can reflect the separate contributions of sensory capability and biasing factors to the final decision outcome (Simpson & Fitter, 1973).

Several detection-theoretic bias indices have been proposed, and it may be beneficial to develop a taxonomy of these measures. Macmillan and Creelman (1991, Chapter 2) provide a good starting point, which we will formalize further now. Four basic types of bias index can be defined on the basis of the relationship between the index and the structure of the decision space: decision variable criterion, likelihood ratio, criterion location, and relative criterion location indices. Figure 1 illustrates the relationship for each type of bias index. *Decision variable* criterion indices (e.g., x_c) are simply the value of the decision axis at the location of the criterion. *Likelihood ratio* criterion indices (e.g., β) are based on the ratio of the heights of the two density functions at the location of the criterion. *Criterion location* indices (e.g., c) are equal to the distance in the units of the decision axis between the criterion and some reference point, often that corresponding to $\beta = 1$. The location on the decision axis of the reference point changes with sensitivity. Finally, *relative criterion location* indices (e.g., c') typically scale criterion location bias indices by the sensitivity of the observer. This taxonomy is illustrated for the single-interval case in the left-hand part of Table 1.

The validity of several of the bias indices in Table 1 has been investigated from both theoretical and empirical perspectives. From a theoretical standpoint, there are two conditions that are often required of candidate indices. First, a shift in bias toward a positive response should be reflected in a concomitant increase in both the hit rate and the false alarm rate. Conversely, a shift in bias toward a negative response should yield a decrease in both of these rates. This monotonicity condition is required because bias is directly related to the propensity for making a particular kind of response. The bias indices β and c' have been shown to fail the monotonicity condition (Macmillan & Creelman, 1990).

Second, if bias and sensitivity represent independent processes, the bias index should maintain independence from the index of sensitivity; otherwise, it could be argued that the two indices are influenced by a common factor—that is, that they are, at least in part, measuring the same thing. However, although indices of sensitivity

This research was funded by a grant from the University of Auckland Research Committee. We thank John Irwin for the valuable comments he made on a draft of this manuscript. Correspondence concerning this article should be addressed to M. J. Hautus, Department of Psychology, University of Auckland, Private Bag 92019, Auckland, New Zealand (e-mail: m.hautus@auckland.ac.nz).

Copyright 2003 Psychonomic Society, Inc.

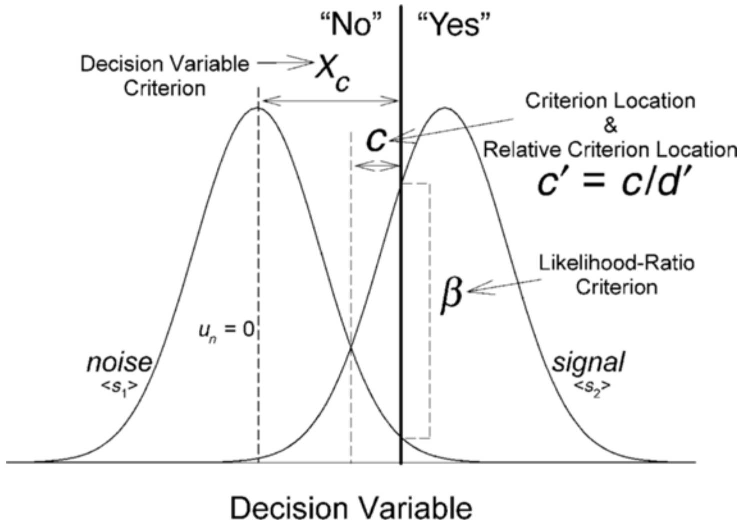

Figure 1. The relationship between the four categories of bias indices and the decision axis for the single-interval (*yes–no*) case.

are invariant under changes in bias, the reverse may not be true (Gardner, 1997; Ingham, 1970; Irwin, Hautus, & Francis, 2001; Macmillan & Creelman, 1991, pp. 46–47). Furthermore, Dusoir (1975) suggested that it may be necessary to adopt more than one bias index to explain response behavior.

Although there are numerous theoretical considerations for the selection of a good bias index, it is clear that an ideal index must account for empirically observed variation. Indeed, if an index describes the data well, it is likely that it will also fulfill the theoretical requirements. After all, these requirements have been inferred from observation as well. We now present some of the ways in which researchers have sought to evaluate the validity of various bias indices.

Ingham (1970) first proposed the c bias index as a valid alternative to the traditional detection-theoretic index, β. Ingham assessed c and β on the data collected from an auditory intensity discrimination task and found that c demonstrated less variation than did β across experimental sessions and that c demonstrated greater independence from d' than did β.

Snodgrass and Corwin (1988) attempted to determine, among other things, whether lack of theoretical independence between d' and β would produce problems when this model was fitted to data or whether the model based on β would perform as adequately as a model that demonstrated independence between the parameters, such as one based on d' and c. That is, would the parameters demonstrate dependencies when fitted to data collected over a range of difficulty levels and bias manipulations? Normal and memory-impaired observers participated in a *yes–no* single-interval recognition memory task, in which bias was manipulated by using payoff matrices. The parameter estimates for the best-fitting models based on β demonstrated dependencies that reflected the mathematical relationship between the parameters. They found that estimates of β derived from their data tended

Table 1
A Taxonomy of Bias Indices Based on the Relationship Between the Index and the Decision Axis

Index	Single Interval (*Yes–No*)	*Same–Different*: Differencing	*Same–Different*: Likelihood Ratio
Decision variable	x_c	k	
Likelihood ratio	β	β_d	β_i
Criterion location	c	c_d	c_i
		c_{sd}	
		c^*_{sd}	
Relative criterion location	c'	c'_d	c'_i

Note—The four most common bias indices for the single-interval task each belong to a different category. Nine bias indices for the *same–different* task are also grouped into their appropriate categories.

to increase with sensitivity. On the other hand, the parameter estimates for the best-fitting models based on c did not demonstrate significant dependencies.

Hoshino (1991) applied the model based on the c index to the data of a *yes–no* recognition memory test that employed words as stimuli. A later analysis was conducted, using the model based on β, to compare estimates of the two bias indices. Despite reports that β was not a useful bias index, Hoshino found similar patterns of results for both β and c.

See, Warm, Dember, and Howe (1997) conducted three vigilance experiments to systematically investigate a number of bias indices. The two detection-theoretic indices they tested were again β and c. Indices were compared using five analytical techniques: analysis of variance, intercorrelations between bias and sensitivity, intercorrelations between bias indices, analysis of residual bias indices, and comparisons of averaged and collapsed bias indices. They found that β was less sensitive than c to manipulations of response bias, that β, but not c, was relatively insensitive to manipulations of bias when those manipulations lead to conservative response behavior, that although both indices demonstrated significant correlations with the sensitivity index, the correlations for β were much larger than those for c, and that estimates of c obtained by averaging the estimates of c obtained for individual subjects were similar to those obtained by first pooling the data of all the subjects and then estimating c, whereas estimates of β obtained by these two methods were quite dissimilar. On the basis of these comparisons, c was taken as the better measure of bias for vigilance tasks, when compared with β.

There are two major approaches to the manipulation of bias: (1) to use a payoff matrix to differentially influence the consequences of behavior and (2) to adjust the a priori probability that the signal is presented on a trial—here, called the signal presentation probability (SPP). Holding both of these things constant as the difficulty of a task is manipulated should encourage the observer to maintain a constant bias. Green and Swets (1966, pp. 21–23) demonstrated that the optimal likelihood ratio criterion for an observer who is maximizing his or her return when faced with differential payoffs and a specific SPP is

$$\beta = \frac{R(\mathrm{H}) - R(\mathrm{M})}{R(\mathrm{CR}) - R(\mathrm{F})} \cdot \frac{\mathrm{SPP}}{1 - \mathrm{SPP}}, \tag{1}$$

where $R(.)$ is the reward function for correct rejections (CR), false alarms (F), hits (H), and misses (M). When there are no differential rewards, the first term on the right side of Equation 1 equals one. Similarly, when SPP = .5, the second term on the right side of Equation 1 equals one. The optimal criterion is unbiased (i.e., $\beta = 1$) when both components equal one or when they hold a reciprocal relationship to one another. For research purposes, it is useful to set one of these components equal to one and systematically vary the other component to manipulate observer bias.

Bias in the *Same–Different* Task

Can the findings reported so far be generalized to detection-theoretic tasks other than the *yes–no* task? This is an important question, because the detection-theoretic model for the *same–different* task is an extension of that for the single-interval task. The same index of sensitivity is used in the model for each task. Similarly, the same bias index should apply. Irwin et al. (2001) motivated a number of indices of bias that could be adopted in a *same–different* task. They also demonstrated, with a small data set, a useful method for assessing the efficacy of these bias indices. Their analyses suggest further experimental work: to create a larger data set to determine which bias indices perform the best.

The *same–different* task may be a suitable candidate for the investigation of bias, because the task may be easier for observers to undertake than either *yes–no* or two-alternative forced-choice tasks. This is because the observer need not know anything about the identity of the stimuli in order to take part; that is, the observer does not need to learn which stimulus is signal and which is noise. Rather, the observer need decide only whether the two stimuli presented on a trial are the same or different. Consequently, it may be relatively easy for observers to maintain a constant level of bias as task difficulty changes. This argument may not be strong for the simple stimuli used in our experiments but would be more pertinent for more complex stimuli, such as foods and beverages, or auditory profile stimuli.

Recently, the *same–different* task has been used in its own right for the investigation of observer bias. For example, Gauthier, Behrmann, and Tarr (1999) have investigated differences in observer bias for prosopagnosics making judgments about different classes of objects, including faces. It is important to ensure that the bias index employed is valid, before such comparisons are made.

The *same–different* task involves the presentation of two stimuli, and the observer must decide whether those stimuli were drawn from the same or different events. There are two events, S_1 and S_2, and either event has an equal chance of presentation as the stimulus. This allows four possible event combinations: two resulting in *same* trials $\langle S_1, S_1\rangle$ or $\langle S_2, S_2\rangle$ and two resulting in *different* trials $\langle S_1, S_2\rangle$ or $\langle S_2, S_1\rangle$ (Noreen, 1981). In the standard version of this task, all the combinations are equally likely.

A complicating factor for this task is that there are two distinct decision strategies that can be adopted by the observer. According to Sorkin's (1962) difference strategy, the decision variable is the difference between the sensory experiences produced by the two stimuli presented on a trial. To use this strategy, the observer judges the magnitude of the difference between the evidence arising from each observation on a trial and responds *same* only if this magnitude is less than some criterion value; otherwise, the observer responds *different.* (Ennis, Palen, & Mullen, 1988, have extended the difference model of the *same–different* task to multiple perceptual dimensions.)

However, for Noreen's (1981) likelihood-ratio strategy, the observer makes use of more than just the relative difference information employed in the difference strategy. The two observation intervals are independent; therefore, the likelihood that the evidence arising from an observation interval arose from a particular event can be assessed, and the ratio of the likelihoods associated with each observation interval can be compared with a criterion. This strategy makes better use of the information available to the observer on a trial. Consequently, performance is higher when the observer uses the likelihood-ratio strategy.

Given that there are two independent sources of information in the *same–different* task, the decision space for this task can best be represented by two orthogonal dimensions. These dimensions represent the sensory evidence that arises from each of the observation intervals on a trial. The distribution of evidence on the decision space is given by four bivariate-normal distributions, one for each of the possible stimulus sequences (see the bot-

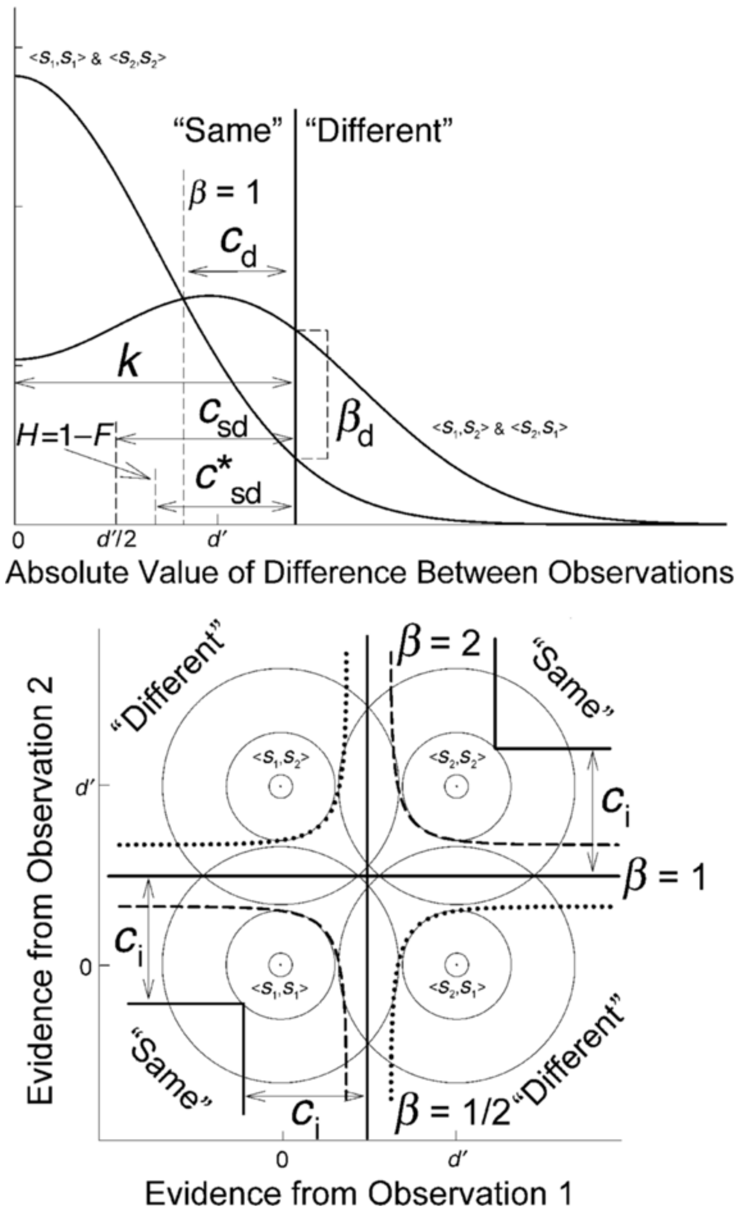

Figure 2. The relationship between the four categories of bias index and the decision axis for the *same–different* task. The top and bottom panels illustrate the decision space for the difference and the likelihood-ratio strategies, respectively. The relative criterion indices c'_d and c'_i are not illustrated.

tom panel of Figure 2). The shape of the decision boundaries on the decision space is determined by the decision strategy that is assumed, and the location of those boundaries is determined by the bias of the observer. The difference strategy gives rise to boundaries that represent a fixed absolute difference between the two dimensions. These boundaries therefore take the form of straight lines. A useful simplifying feature for the difference strategy is that the multidimensional decision space can be reduced to a one-dimensional one (Sorkin, 1962; see the top panel of Figure 2). This is possible because the absolute value of the difference between the evidence arising from the two observations is itself a one-dimensional decision variable. The boundaries for the likelihood-ratio strategy are typically curved (except for the case of an unbiased observer) and coincide with equal likelihood-ratio contours for same and different events. An interesting feature of the *same–different* task is that the shape of the isosensitivity curve is dependent on the decision strategy that is assumed. The difference and likelihood-ratio strategies give rise to asymmetrical and symmetrical isosensitivity curves, respectively (Hautus, Irwin, & Sutherland, 1994; Noreen, 1981, note 4).

Dai, Versfeld, and Green (1996) demonstrated that the difference strategy and the likelihood-ratio strategy are extreme special cases of a more general representation of the *same–different* task. They showed that the method of stimulus presentation can place limits on the type of decision strategy an observer can adopt. For roving experiments, where stimuli are drawn from a larger set on a trial-by-trial basis, the best an observer can possibly do is adopt the difference decision strategy. When the task is not *roving* and only two possible stimuli can be presented to the observer on a trial, the likelihood-ratio decision strategy can be adopted. When used, the likelihood-ratio strategy provides better overall performance than does the difference strategy (Noreen, 1981). Consequently, the likelihood-ratio strategy has often been called the optimal strategy for the *same–different* task. However, Dai et al. demonstrated that the difference strategy is the optimal strategy when a roving design is used. Although they did not shed light on the decision strategy that will be adopted by an observer when performing in a task that does not have roving stimuli, they did demonstrate that the likelihood-ratio strategy is available for observers to use under these conditions.

We now return to our taxonomy of bias indices to incorporate those proposed for the *same–different* task. The right-hand part of Table 1 presents the various bias indices for the difference model and the likelihood-ratio model in the *same–different* task. Where a similar bias index can be defined for the two models, a subscript is employed to indicate the appropriate model for that index. What follows is a description of each of these bias indices in terms of their relationship to the decision space. The derivations of these bias indices are detailed elsewhere (Irwin et al., 2001).

Figure 2 illustrates the relationship between each of the bias indices for the *same–different* task and the decision space for each strategy in that task. The top panel in Figure 2 illustrates the decision space for the difference strategy for the case in which $d' = 2$. The evidence variable for this strategy is the absolute difference between the evidence stemming from the two observations on a trial. Consequently the origin is at zero. The bias index, k, is the distance of the criterion from the origin or, more simply, the value of the absolute difference. The c_d index is the distance of the criterion from $\beta_d = 1$, with β_d itself defined as the ratio of the height at the criterion's location of the distribution arising from *same* events to the height of the distribution arising from *different* events. c_d (not illustrated) is simply c_d/d'. The c_{sd} index is the distance of the criterion from $d'/2$. Note that d' is the mode of the *different* distribution, which is a fact that can be used to motivate the c_{sd} index.

We now introduce yet another bias index for the difference strategy in the *same–different* task. This index, which we denote c^*_{sd}, is based on the idea that an unbiased observer will apportion responses equally between the response alternatives unless a biasing factor is present. For this to be true, irrespective of the sensitivity of the observer, the reference point on the decision space for zero bias must be located where the decision variable leads to equal hit and correct-rejection rates;[1] that is, the hit rate must equal one minus the false alarm rate. Any point on the receiver-operating characteristic (ROC) that originates from a criterion filling this requirement will be located on the negative diagonal of the ROC square. The c^*_{sd} index is the distance that the observer's criterion is from this reference point on the decision space. The strength of the c^*_{sd} index is that it can produce an isobias curve that falls on the minor diagonal of the ROC square. None of the bias indices for the difference strategy discussed earlier can produce an isobias curve that lies along this diagonal.

The bottom panel in Figure 2 illustrates the decision space for the likelihood-ratio strategy, again for the case of $d' = 2$. Each dimension on this decision space represents the evidence arising from an observation interval, and the circles denote regions of equal likelihood for each of the four bivariate-normal distributions that represent the evidence that arises from the four possible types of trial ($<S_1, S_1>$, $<S_2, S_2>$, $<S_1, S_2>$, and $<S_2, S_1>$). Because of the multidimensional nature of this decision space, it is more difficult to clearly illustrate the structure of the bias indices associated with it. The likelihood-ratio index β_i is the ratio of the height of the compound distribution for same events to the height of the compound distribution for different events. Criteria for three likelihood ratios are illustrated ($\beta_i = ½$, $\beta_i = 1$, and $\beta_i = 2$). Note that the criterion for $\beta_i = 1$ is defined by the axes of the decision space. This criterion signifies a special case of the likelihood-ratio strategy called the independent-observations strategy (Noreen, 1981). The c_i index is therefore defined as the distance of the criterion from these axes. Note that the criteria for c_i are straight lines, unlike the curves for likelihood-ratio criteria when $\beta_i \neq 1$. If this were not the case, the distance from $\beta_i = 1$

could not be a constant. For cases in which a criterion based on c_i intersects the asymptote of a likelihood-ratio criterion, $c_i = \ln(\beta_i/d')$ (Irwin & Hautus, 1997). Finally, the c'_i index (not illustrated) is given by $c'_i = c_i/d'$.

Each of the bias indices produces a different family of isobias curves (see Irwin et al., 2001, Figures 2 and 4). The most unlikely family of isobias curves is generated by the k index. For this index, isobias curves take the form of horizontal lines that extend from the left axis of the ROC square to the major diagonal. Thus, an observer who adopts a constant value of k in effect holds their hit rate constant. The decision variable index, x_c, for the *yes–no* task exhibits similar behavior, and amounts to the observer's holding the false alarm rate constant as sensitivity changes. No isobias curves have been reported that exhibit this behavior. Consequently, it is unlikely that observers adopt a criterion based on these decision variable indices.

It seems reasonable to expect a successful bias index to describe data that has been obtained by manipulating bias either through the use of a payoff matrix or by varying SPP. We assess the ability of the bias indices listed in Table 1 to account for the variation in the data obtained from two groups of observers: one group whose bias is manipulated by systematically adjusting the SPP while holding a neutral payoff matrix and another group whose bias is manipulated by systematically adjusting the payoff matrix while holding the SPP constant at .5. The set of payoff matrices and SPPs employed are carefully tailored to yield the same optimal criteria for each group.

METHOD

Observers

Eight observers participated in the research. Seven were male, and their ages ranged from 19 to 55 years (mean age of 29). Four of the observers had previous experience on the *same–different* task (O4, O6, O7, and O8), whereas the rest were naive. All the observers were assessed for hearing loss, using pure-tone audiometry (Bruel & Kjaer, Model 1800). Hearing loss was never in excess of 15 dB (re ISO standard, 1975). The observers were paid $20 for their participation. In addition, 2 observers received an additional $50 "bonus" based on their performance in the task.

Apparatus and Stimuli

The stimuli consisted of 1000-Hz tones generated by a multifunction synthesizer (Hewlett Packard, Model 8904A) that were gated (TTES, SW1) to a duration of 100 msec with 10-msec $\cos^2$ ramps. The amplitudes of the tones were set, using a programmable attenuator (TDT, PA2), to one of five levels: 75.0, 75.5, 76.0, 76.5, or 77.0 dB SPL. The tones were presented to the observer, who was seated in a sound-attenuating chamber (Amplaid, Model E), via monaural earphones (TDH-49 with MX-41/AR cushions). Responses were made on a two-button response panel, and information was conveyed to the observer during the experiment via a computer terminal (Hewlett Packard, Model 700/41) and three LEDs arranged on a small panel.

Procedure

A standard *same–different* task was employed in which each trial contained two observation intervals. The observers were required to judge whether the two observations arose from the same or different stimuli. For any block of trials, there were two stimuli, one of which was always the 75.0-dB SPL standard tone and the other of which was selected from the four other amplitudes available. Thus, there were four different levels of difficulty in the experiment.

The observers were assigned to one of two bias manipulation conditions. The first condition utilized payoff matrices to influence observer bias. Points were earned for correct trials only. The points awarded for hits and correct rejections for the five payoff matrices were 1:9, 3:7, 5:5, 7:3, and 9:1. The second condition employed SPP to influence observer bias. The probability that the two stimuli on a trial were the same was set to one of the following five values: 0.1, 0.3, 0.5, 0.7, and 0.9. Payoffs and SPPs were determined using Equation 1 so that the criteria adopted by the observers in the two bias manipulation groups would be approximately the same.

Each trial commenced with a 150-msec warning light. After a further interval of 500 msec, the first tone was presented, followed by a 100-msec pause before the presentation of the second tone. A light flashed in synchrony with the presentation of the stimuli to help demarcate the two observation intervals. The observer made a response, *same* or *different*, on the two-button response panel. Feedback, in the form of a light, was presented 200 msec after the observer made a response. In addition, the total number of points earned in the current block of trials was displayed on the computer terminal. After a further 300 msec, the next trial commenced. Each block consisted of 110 trials.

The observers in the payoff and SPP conditions each undertook 20 different combinations of difficulty and bias manipulation. Each block of trials was at a fixed level of difficulty and bias manipulation. Consequently, this was not a roving experimental design. Each observer first undertook one block of trials with each of the 20 bias/difficulty combinations as practice—a total of 2,200 practice trials. They then undertook two additional blocks of trials in succession with each combination. The first 20 trials were discarded as warm-up trials, leaving 200 experimental trials for analysis from each combination—a total of 4,000 experimental trials per observer. Both the practice and the experimental blocks were undertaken in a different random order by each observer.

To encourage optimal performance, the observers were told that an additional $50 would be paid to the observer with the highest point score in each of the two bias conditions. To make this possible and to ensure that the observers received the same form of feedback in each group, one point was awarded for each correct response in the SPP condition.

RESULTS

The data for each observer consist of 20 points that can be plotted in the ROC square. Each point is based on 200 trials and can be considered as a random sample taken at the point of intersection of one of the four isosensitivity curves and one of the five isobias curves investigated in this experiment. In addition, the data of the 4 observers in each condition can be pooled, so that each point is based on 800 trials. One observer was chosen from each condition to provide a fine-grained view of the data. The observers selected were not atypical in any way. Figure 3 illustrates the data of O1 (SPP), O8 (payoff), and the group for both conditions.

Isosensitivity Curves

For each decision strategy, an estimate of sensitivity was obtained for each isosensitivity curve independently of the other three. To obtain this estimate, the five bias parameters were also estimated. However, the values of these parameters were unique to the isosensitivity curve

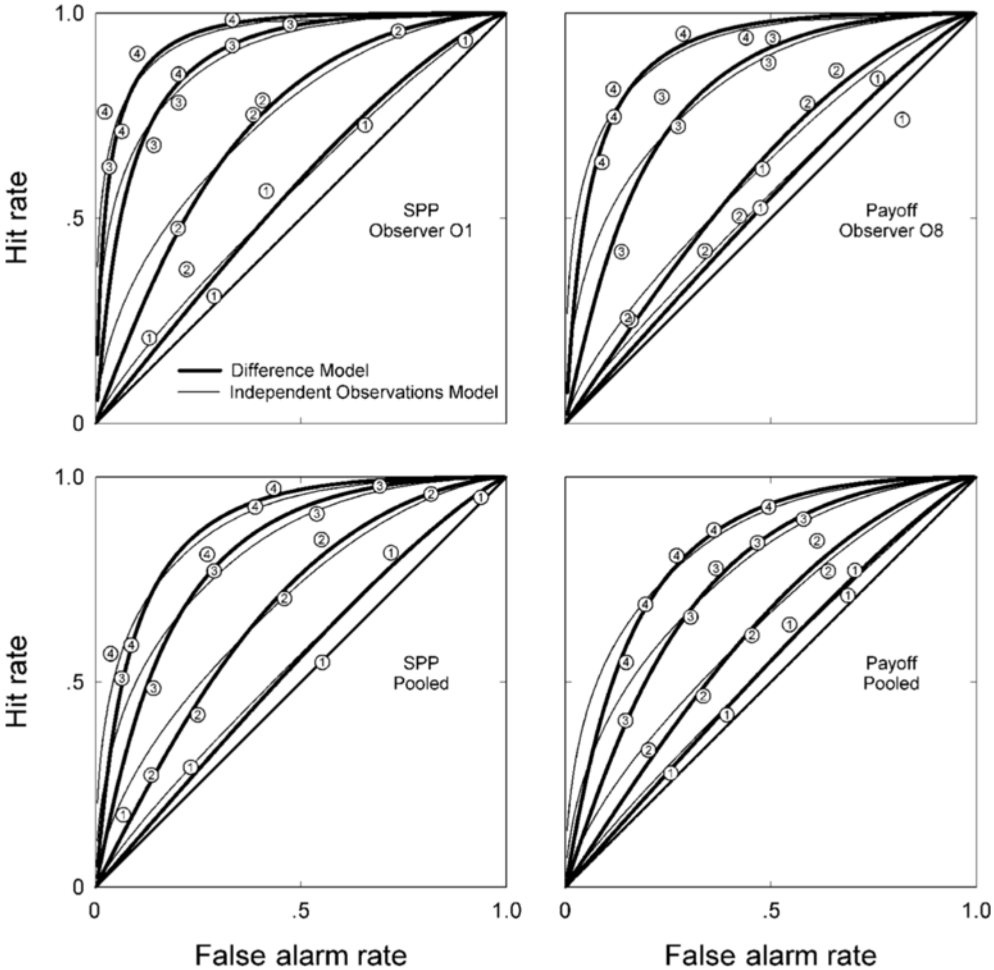

Figure 3. Data collected for two observers (O1 and O8) and the pooled data for all the observers within each condition. For the top two panels (individual observers), each point is based on 200 trials, and for the bottom two panels (pooled data), each point is based on 800 trials. Points with the same numerical symbol fall on the same isosensitivity curve. The thick and thin curves are the best-fitting isosensitivity curves based on the difference and the likelihood-ratio models, respectively. The parameter estimates used to plot these curves and the associated goodness-of-fit statistics are given in Table 2.

under investigation. Therefore, the bias index adopted in this process had no effect on the estimate of sensitivity. These bias parameter estimates were discarded. The fitting procedure involved an iterative search, using an adaptation of the downhill simplex method (Press, Teukolsky, Vetterling, & Flannery, 1992, pp. 408–412), for the maximum likelihood solution. Iterative techniques are notorious for convergence to a local minimum. Consequently, multiple fits were performed from different starting locations to ensure that global minima were obtained. There were very few cases of convergence to local minima. Once the best-fitting parameter estimates were obtained, a χ^2 fit statistic was also calculated.

Table 2 gives the maximum-likelihood estimates of the sensitivity parameter, d', for each isosensitivity curve, together with the χ^2 fit statistic ($df = 4$). One observer in each condition was not performing as well as the others (O3 for SPP and O7 for payoff). These 2 observers demonstrated inferior sensitivity at every level of task difficulty, when compared with the other observers in the same condition. However, the data from these 2 observers did not provide a bad fit to the best-fitting model, except for the case of O3 at the two easiest levels of difficulty, and then only for the difference strategy fits. Therefore, these observers remain in the analyses that follow. Figure 3 includes the best-fitting isosensitivity curves based on each decision strategy.

Pooling the data of all of the observers in a condition and then fitting the detection-theoretic model to this amalgamated data can provide an overall estimate of sensitivity for each model and condition. Such overall estimates are given in Table 2, together with their χ^2 fit statistics ($df = 4$).

An alternative method for assessing how well the best-fitting models account for the data is to cumulate the χ^2 values for the observers in each condition. Since the data

Table 2
Estimates of Sensitivity for Both the Signal Presentation Probability (SPP) and the Payoff Conditions

	0.5 dB		1.0 dB		1.5 dB		2.0 dB	
Observer	d'	χ^2	d'	χ^2	d'	χ^2	d'	χ^2
Difference Strategy: SPP Condition								
1	0.94	1.7	1.96	1.5	3.12	4.9	3.78	8.4
2	0.70	31.5*	1.87	3.6	2.16	1.8	3.17	5.7
3	−0.29	6.4	0.34	7.8	1.70	12.7*	1.96	10.5*
4	0.98	2.4	1.99	6.9	3.01	16.0*	3.64	2.5
Cumulated		42.0*		19.8		35.4*		27.1*
Pooled	0.75	15.3*	1.64	2.7	2.51	17.6*	3.06	20.0*
Difference Strategy: Payoff Condition								
5	0.65	2.5	1.45	8.5	2.20	5.0	2.72	12.3*
6	1.00	2.8	1.71	3.3	2.42	7.2	3.07	0.1
7	0.27	3.6	0.81	6.0	1.20	3.6	1.51	4.6
8	0.68	8.1	1.25	2.6	2.45	7.9	3.30	4.1
Cumulated		17.0		20.4		23.7		21.1
Pooled	0.69	2.8	1.31	6.9	2.09	1.2	2.61	1.5
Likelihood-Ratio Strategy: SPP Condition								
1	0.76	1.6	1.60	3.7	2.46	3.0	2.94	7.1
2	0.77	22.6*	1.51	1.7	1.71	4.2	2.50	2.9
3	0.02	6.2	0.45	6.2	1.35	6.5	1.52	4.7
4	0.83	1.7	1.65	9.8*	2.43	14.1*	2.90	3.6
Cumulated		32.1*		21.4		27.8*		18.3
Pooled	0.67	10.8*	1.35	9.7*	2.00	9.8*	2.41	11.65*
Likelihood-Ratio Strategy: Payoff Condition								
5	0.53	2.6	1.19	23.1*	1.81	7.6	2.19	16.0*
6	0.78	4.5	1.41	1.6	1.97	10.1*	2.48	1.5
7	0.14	3.7	0.62	7.5	1.03	1.5	1.27	4.7
8	0.63	7.0	1.00	4.1	1.99	4.1	2.62	4.4
Cumulated		17.8		36.3*		23.3		26.6*
Pooled	0.57	3.64	1.09	9.3	1.72	6.9	2.13	7.6

Note—Estimates were obtained for each observer at each of the four levels of task difficulty. The χ^2 ($df = 4$) fit statistics are also shown. Group estimates of sensitivity are the average sensitivities for the observers in the group, and group χ^2 ($df = 16$) fit statistics are the sum of those for the observers in the group. *$p < .05$.

obtained from each observer are independent, the cumulated χ^2 values given in Table 2 have 16 degrees of freedom.

A comparison of the sensitivity estimates obtained for the two different decision strategies indicates that those for the difference strategy are generally higher than those for the likelihood-ratio strategy. At first glance, this may appear anomalous, but it must be remembered that the likelihood-ratio model predicts better performance than does the difference model, given the same stimulus conditions. Thus, given the same empirical isosensitivity curve, the likelihood-ratio model will yield lower estimates of sensitivity than will the difference model.

A comparison of the sensitivity of the observers in each condition indicates better performance from those in the SPP condition. Although this is an interesting observation, it is not possible to attribute this effect to the type of bias manipulation; the effect could be due to individual differences. The small number of observers in each condition means that this effect does not reach significance.

There are two distinct ways to determine from Table 2 which strategy provides a better fit to the data. First, the number of model rejections could indicate better performance for one model. For individual observers, 5 of 32 curves for the difference strategy and 6 of the 32 curves for the likelihood-ratio strategy provided a poor fit to the data ($p < .05$). For the group data, 3 and 4 curves, of the 8 available for each strategy, provided a poor fit to the data for the difference and likelihood-ratio strategies, respectively. It is clear, then, that neither model stands out as superior when judged by the number of model rejections. The second approach is to investigate the cumulated χ^2 values for each model. These cumulated values are 207 and 204 for the difference and the likelihood-ratio strategies, respectively. Given the considerable variation in the χ^2 values in Table 2, it is clear that these cumulated χ^2 values also do not permit the selection of a superior model. Consequently, it is not possible to reject the bias indices for an entire decision strategy on the basis of the analysis of the isosensitivity curves. The following analyses therefore compare the ability of all nine indices to account for the variation in the data.

Isobias Curves

Although there is plenty of evidence that sensitivity is relatively invariant when stimulus conditions remain constant, there is considerably less evidence either way for the notion that bias is invariant when the utility of a decision remains constant. Furthermore, detection theory posits that sensitivity is independent of cognitive factors; it is determined primarily by perceptual factors. These tend to change very slowly. On the other hand, bias is a cognitive factor and is much more likely to change over time. Therefore, a useful approach to fitting our detection-theoretic models would be first to estimate the four sensitivity parameters and then to estimate the five bias parameters with the sensitivity parameters fixed. This approach attributes most of the variability in the data to the bias factor and forces this factor to account for that variability. The result is a very stringent test of the various bias indices.

The estimates of d' given in Table 2 were used to constrain the model fitted to the data. Thus, the points corresponding to the same isosensitivity curve, on each isobias curve, would be assigned the same value of d' throughout the fitting process. The χ^2 statistics ($df = 35$) for the best-fitting isobias curves based on each bias index are presented in Table 3 for individual observers and the pooled data of all the observers within each condition. Since this is a very strong test of the ability of the bias indices to account for the data, most of the χ^2 statistics reach significance ($p < .05$). Consequently, contrary to convention, the best-fitting models from which the data do not differ significantly are indicated.

A more discerning measure of the efficacy of the various bias indices is the cumulative value of χ^2 across the observers in each condition ($df = 140$). Table 4 lists the nine indices with their associated cumulative χ^2 values separately for each condition. The indices obtained for each condition are ordered on the basis of the cumulative χ^2 value across all the observers in the condition. These cumulative values vary considerably across the indices. The data from the two conditions are independent, and in many respects these two conditions constitute an experimental replication. They demonstrate reasonable consistency in the ability of the indices to account for the variability observed in the data.

If we focus now on the rankings, the c^*_{sd}, c_{d}, c_{i}, and c_{sd} indices have all performed reasonably well. The c_{i} index performed best in the SPP condition, whereas c^*_{sd} was better in the payoff condition. Thus, it does not seem reasonable to proclaim either as the best index. All four of the criterion location indices have a similar cumulative χ^2, and it is difficult to choose between them on the basis of this analysis. There is a considerable drop in performance to the fifth ranked index, which for both conditions is k, and an even larger drop to the remaining indices. It is interesting that this analysis has caused the indices to fall reasonably well into their taxonomic groupings: The criterion location indices perform best (c^*_{sd}, c_{d}, c_{i}, and c_{sd}), followed by the decision variable index (k), and the relative criterion location indices (c'_{i} and c'_{d}), and the likelihood-ratio indices (β_{i} and β_{d}) collectively perform the poorest. Likelihood-ratio indices, despite their theoretical importance, are clearly not very good at accounting for these data.

An alternative form of analysis is to pool the data for all of the observers within each condition and then fit the detection-theoretic models to those amalgamated data. Table 4 also contains the resulting χ^2 statistics for the best-fitting models obtained in this way. The χ^2 values for the pooled data are understandably different from those obtained by accumulation, the former having 35 degrees of freedom and the latter having 140. The rankings of the indices are reasonably consistent with those obtained by cumulating the χ^2 across the observers in each condition. The main feature in this analysis is that the criterion location indices again stand out above the rest.

Table 3
χ^2 Fit Statistics for the Best-Fitting Isobias Curves Based on Each Bias Index

Observer	β_{d}	c_{d}	c_{sd}	c'_{d}	c^*_{sd}	k	β_{i}	c_{i}	c'_{i}
				SPP Condition					
1	194.8	56.5	74.7	249.8	42.6*	251.4	169.7	47.1*	162.7
2	681.5	97.6	157.7	450.1	90.3	243.8	323.1	68.1	237.0
3	980.1	109.7	318.1	1,050.2	131.9	154.7	1,293.3	111.1	787.2
4	157.0	89.8	49.6*	253.5	50.7	318.3	154.4	61.9	165.6
Pooled	1,641.4	136.1	292.1	1,293.8	96.8	612.4	1,127.5	92.3	749.0
				Payoff Condition					
5	260.2	122.0	165.3	160.6	108.9	227.4	209.6	110.9	77.9
6	213.8	97.5	89.4	288.6	76.5	178.7	254.2	149.5	252.7
7	801.3	259.3	265.6	449.5	227.4	407.0	1,094.5	544.5	801.4
8	384.6	54.6	110.3	280.2	39.0*	253.4	214.5	43.7*	160.5
Pooled	1,159.7	131.3	208.4	684.7	54.7	543.7	784.5	67.0	305.1

Note—All χ^2 values have 35 degrees of freedom. Contrary to convention, statistics for which the models did *not* differ significantly from the model are indicated. SPP, signal presentation probability. $*p > .05$.

Table 4
Bias Indices Ranked in Order of the Cumulated Value of χ^2 Across Observers in Each Condition and in Order of the χ^2 Obtained When Each Model was Fitted to the Data Pooled Across the Observers Within Each Condition

	Cumulated χ^2 ($df = 140$)				Pooled Data ($df = 35$)			
	SPP		Payoff		SPP		Payoff	
Rank	Index	χ^2	Index	χ^2	Index	χ^2	Index	χ^2
1	c_i	288	c^*_{sd}	452	c_i	92	c^*_{sd}	55
2	c^*_{sd}	316	c_d	533	c^*_{sd}	97	c_i	67
3	c_d	354	c_{sd}	631	c_d	136	c_d	131
4	c_{sd}	600	c_i	849	c_{sd}	292	c_{sd}	208
5	k	968	k	1,067	k	612	c'_i	305
6	c'_i	1,353	c'_d	1,179	c'_i	749	k	544
7	β_i	1,940	c'_i	1,293	β_i	1,128	c'_d	685
8	c'_d	2,003	β_d	1,660	c'_d	1,294	β_i	785
9	β_d	2,013	β_i	1,773	β_d	1,641	β_d	1,160

Note—The criterion location indices perform the best in each analysis. All χ^2 values are significant ($p < .05$). SPP, signal presentation probability.

Figure 4 illustrates the best-fitting models, based on the c_d, c^*_{sd}, and c_i indices, for O1 (SPP condition) and O8 (payoff condition). The estimates of the bias index and the χ^2 statistic for each curve are provided together with the overall χ^2 statistic for each panel. For O8, the isobias curve for the 7:3 payoff condition has been continued below the major diagonal. Knowledge of the shape of the isobias curves in this region is important for situations in which sampling variability has produced a datum that estimates sensitivity to be below chance level (Irwin et al., 2001). Figure 5 is analogous to Figure 4, except that the fits are to the data pooled across the observers in each condition. Although the values of the indices for the best-fitting models based on c_d, c^*_{sd}, and c_i are given in Figure 5, the values for all nine indices are provided in Table 5.

For the individual isobias curves illustrated in Figures 4 and 5, many of the χ^2 values do not reach significance ($p > .05$), indicating a satisfactory fit to the model. The c_i and c^*_{sd} indices each provide good fits to four of the five curves for O1, whereas c_d provides a good fit to three curves. The overall values of χ^2 for c_i and c^*_{sd} do not reach significance for O1. The c^*_{sd} index has the lowest overall χ^2 (42.7), followed by c_i (47.1) and c_d (56.5). For O8, all isobias curves based on c_d, c^*_{sd}, and c_i provide a good fit to the data. The overall χ^2 for each of the three indices for O8 do not reach significance. The c^*_{sd} index has the lowest overall χ^2 (39.2), followed by c_i (43.7) and c_d (54.6). Thus, the evidence from isobias curves for individual observers indicates that the indices *fall out* in the order c^*_{sd}, c_i, and then c_d.

For the pooled data of the payoff condition (Figure 5), there are two, four, and three good fits to the data for c_d, c^*_{sd}, and c_i, respectively. For the SPP condition, there are two, two, and three good fits, using these indices. Although the overall values of χ^2 are all significant, the c_d index has the highest value in both conditions. In the payoff condition, the c^*_{sd} index has the lowest value of χ^2 (54.7), and in the SPP condition, c_i has the lowest value (92.3). Thus, the evidence from isobias curves for the pooled data indicates that the c_i and c^*_{sd} indices clearly outperform c_d in accounting for these data.

The data from the two conditions are independent and, therefore, can be combined to provide an overarching test of the effectiveness of the various bias indices. The left half of Table 6 (labeled "Constrained") gives the cumulated χ^2 values for the best fits based on each of the nine bias indices. It is clear that the four criterion location indices outperform the others, with c^*_{sd} providing the best overall performance. Figure 6 illustrates the combined data. Each point is based on 1,600 trials. The best-fitting isosensitivity curves for the difference model (top and middle panels on left; $d' = 1.04, 1.59, 2.35$, and 2.87; $\chi^2 = 71.3$) and the likelihood-ratio model (bottom left panel; $d' = 0.87, 1.30, 1.90$, and 2.29; $\chi^2 = 73.7$) are illustrated together with the best-fitting isobias curves based on the c_d, c^*_{sd}, and c_i indices. Again, the fit statistics for the isosensitivity curves are very similar, thus concealing the underlying decision model. Both the isosensitivity and the isobias curves have been plotted on the same panels to highlight the nature of the analyses conducted on the bias indices, which have considered both bias and sensitivity. The distance of each point from the best-fitting isobias curve is not the sole determinant of the fit statistic. Rather, the distance between the point and the intersection of the best-fitting isobias curve and the isosensitivity curve determines the fit statistic. This is illustrated in the bottom left panel of Figure 6 by the third point from the left on the top isobias curve (triangles). Although the point lies on top of the isobias curve, it is some distance from the second isosensitivity curve. The χ^2 for this isobias curve, based on the c_i, index is 44.5, the largest fit statistic on this panel. The magnitude of this χ^2 value stems primarily from the third point. The other four isobias curves on this panel provide a good fit to the data ($p > .05$).

The performance of the c^*_{sd} and c_i indices is clearly better than that of the other indices that have been investigated—albeit, even if the χ^2 values are fairly large.

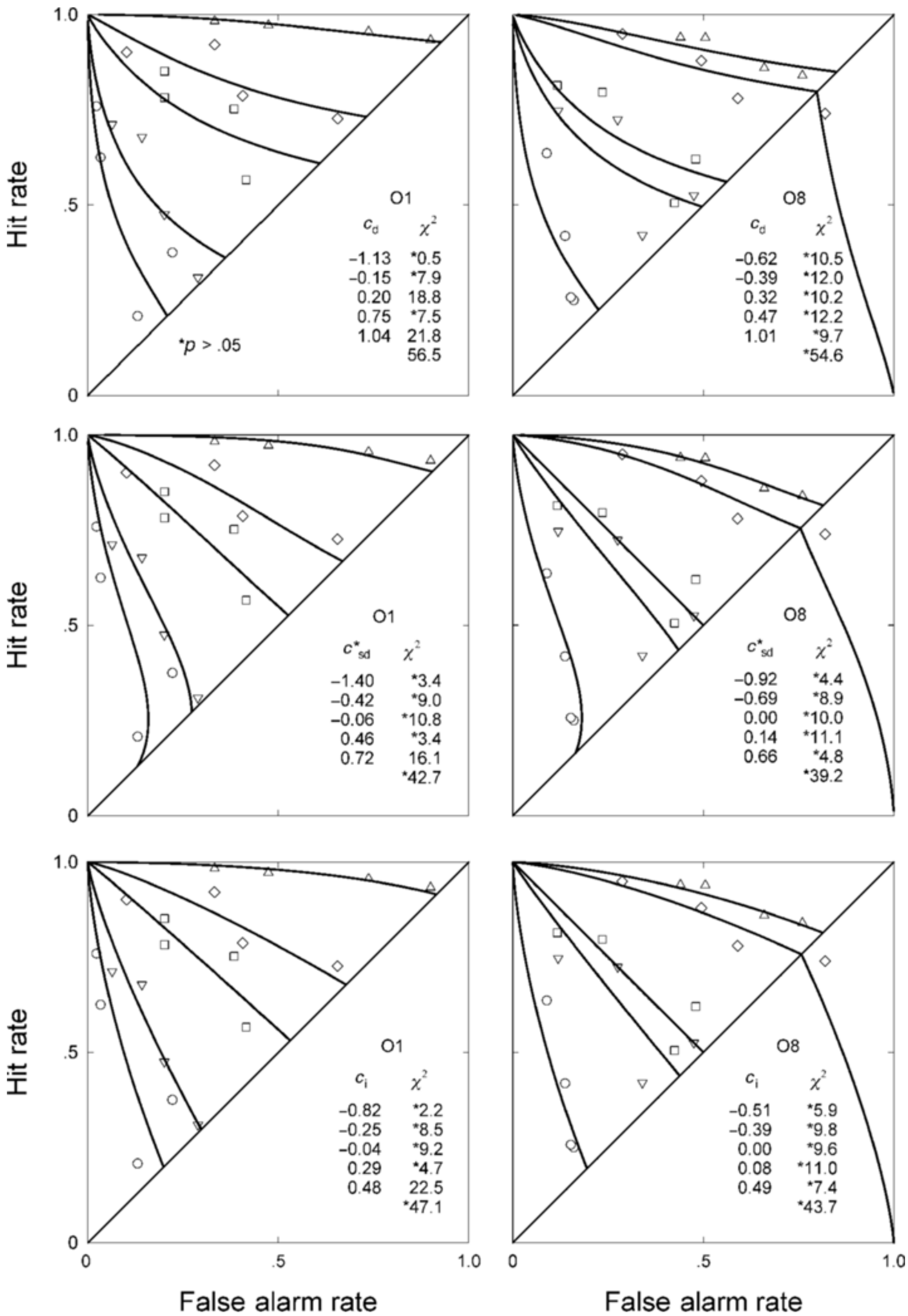

Figure 4. Best-fitting isobias curves based on the three criterion location indices, c_d (top panels), c^*_{sd} (middle panels), and c_i (bottom panels), that are shown in Table 4 to account best for the variation in the data. The panels on the left show the best fits to the data of O1 (signal presentation probability condition), and those on the right are for O8 (payoff condition). Goodness-of-fit indices for individual curves are provided next to the index value (df = 7; cumulated χ^2 df = 35). The isobias curve for the 7:3 payoff condition is illustrated for the region below the major diagonal in which one point falls.

The right half of Table 6 (labeled "Unconstrained") provides the outcome of a less rigorous approach to testing the efficacy of the nine bias indices. In this analysis, one bias parameter and four sensitivity parameters have been estimated for each isobias curve. This allows the four sensitivity parameters for each isobias curve to take on unique values unrelated to the sensitivity parameters estimated for the other isobias curves. In this case, the pro-

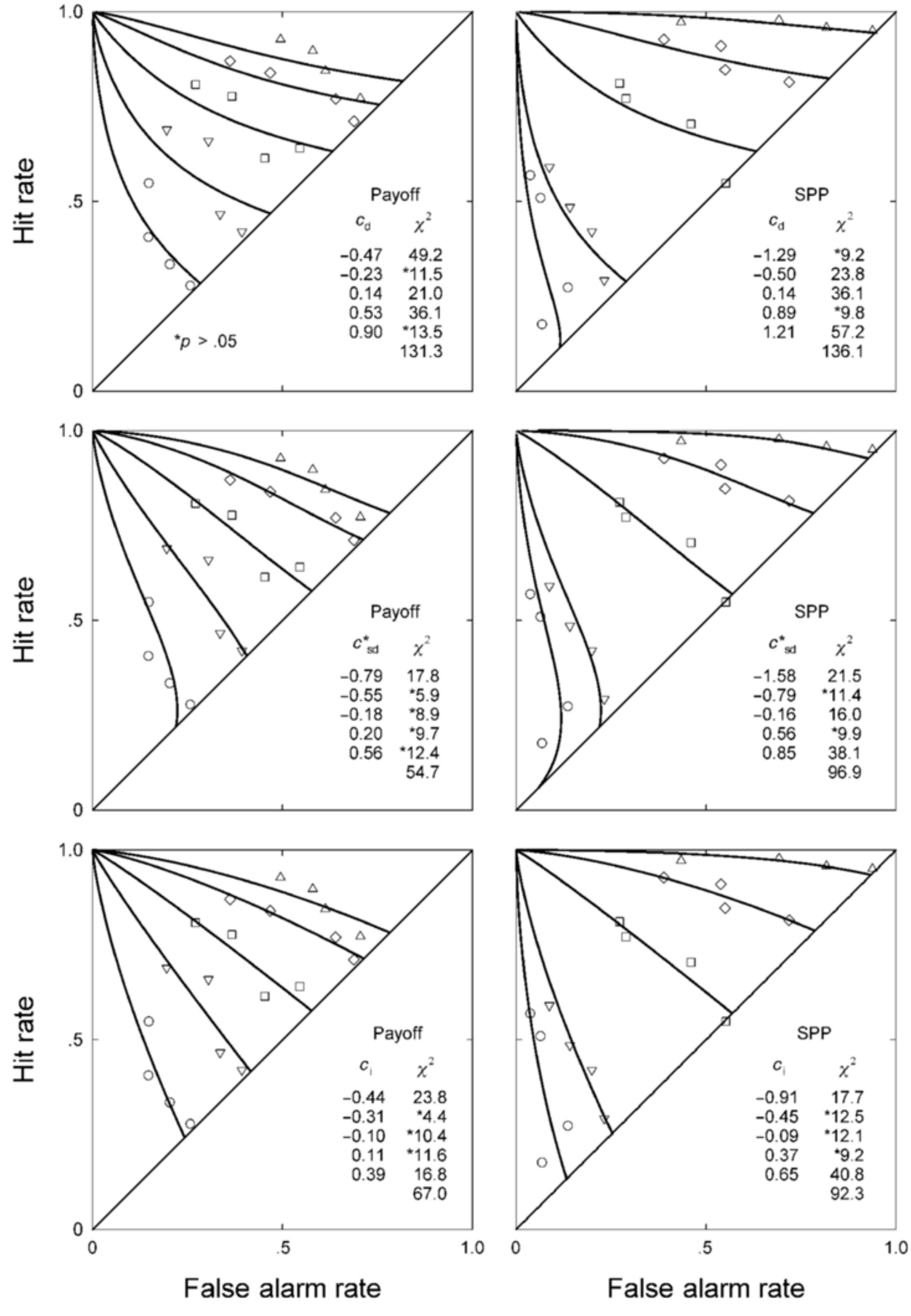

Figure 5. Best-fitting isobias curves based on the three criterion location indices, c_d (top panels), c^*_{sd} (middle panels), and c_i (bottom panels), that are shown in Table 4 to account best for the variation in the data. The panels on the left show the best fits to the pooled data for the payoff condition, whereas those on the right are for the signal presentation probability condition. Goodness-of-fit indices for individual curves are provided next to the index value (*df* = 7; cumulated χ^2 *df* = 35).

cess involves the fitting of 25 parameters in total (20 sensitivity and 5 bias parameters) for the combined data. This process is akin to the usual approach taken by researchers fitting isosensitivity curves to data for the purpose of gaining an estimate of sensitivity. In that case, the estimates of the bias indices are usually discarded, and in the present analysis, the estimates of the sensitivity parameters are discarded. Figure 6 (right-hand panels) illustrates the best-fitting isobias curves based on c_d, c^*_{sd}, and c_i. The c^*_{sd} and c_i indices provide particularly good fits to

Table 5
Estimates of Each of the Bias Indices Obtained for the Pooled Data of All Observers Within Each Condition

	Bias Measure								
	c_d	c_{sd}	c'_d	c^*_{sd}	k	β_d	c_i	c'_i	β_i
SPP									
0.1	1.21	.26	0.52	0.85	0.41	1.15	0.65	0.38	1.46
0.3	0.89	0.02	0.38	0.56	0.76	1.14	0.37	0.23	1.14
0.5	0.14	−0.64	0.04	0.16	1.54	1.05	−0.09	−0.06	0.99
0.7	−0.50	−1.26	−0.25	0.79	2.18	0.88	−0.45	−0.27	0.84
0.9	−1.29	−2.06	−0.60	1.58	2.98	0.52	−0.91	−0.55	0.53
Payoff									
1:9	0.90	0.01	0.44	0.56	0.70	1.12	0.39	0.26	1.11
3:7	0.53	−0.32	0.24	0.20	1.07	1.09	0.11	0.07	1.02
5:5	0.14	−0.69	0.05	−0.18	1.46	1.04	−0.10	−0.07	0.98
7:3	−0.23	−1.05	−0.14	−0.55	1.84	0.96	−0.31	−0.20	0.93
9:1	−0.47	−1.29	−0.30	−0.79	2.08	0.91	−0.44	−0.31	0.90

Note—The isobias contours for the three best-performing indices, c_d, c^*_{sd}, and c_i are illustrated in Figure 5. SPP, signal presentation probability.

the data, and in these cases χ^2 does not reach significance. These are the only two indices, out of the original nine, whose overall fit cannot be rejected statistically.

An interesting comparison can be made between the constrained and the unconstrained fits in Figure 6. It is expected that if the correct model is fitted to the data, the difference between the parameter estimates obtained from the two fitting procedures will be very similar. If the model were not representative of the data, the removal of constraints would lead to a considerable change in the parameter estimates. Figure 6 indicates that the corresponding bias parameter estimates obtained by the two fitting procedures are almost identical for the c^*_{sd} and the c_i indices. Changes in parameter estimates are larger for c_d. This provides further evidence in support of the c^*_{sd} and c_i indices.

DISCUSSION

Variability in the Data

The constrained detection-theoretic models were relatively poor at accounting for the data, as indicated by a large proportion of significant χ^2 statistics. Observation of the data (both group and individual), however, indicates that no model that upheld the requirement of a monotonically increasing isosensitivity curve could do well for these data. This suggests that there is more variation in each data point than would be expected of a binomial sampling distribution, and this would necessarily inflate the obtained fit statistics.

Whereas sensitivity is expected to be relatively stable over time, it is probable that bias is not. This is primarily because the location of the criterion is under the observer's control. Possible reasons for shifting the criterion during an experimental session are diverse. For example, the observer may wish to optimize his or her performance with respect to some goal, and the nature of the feedback given over successive trials may encourage the observer to adjust his or her response allocation, or bias, on the basis of how he or she assesses the likelihood of achieving that goal. Alternatively, an observer may decide that he or she is using a particular response too frequently and may decide to adjust the ratio of *same* to *different* responses. Movement of the decision criterion during an experimental session implies that the points in the unit square are an average of performance with the same level of sensitivity but slightly different biases. The point itself would have an inflated standard error, when compared with what would be expected for a fixed criterion. This would result in a disproportionate number of poor fits, as determined by a test statistic such as χ^2. It is likely that this problem would be more severe for binary responses than for rating responses. This is because the degree to which the decision criterion can be shifted in the rating task is limited by the presence of other criteria. These boundary limits are not present in the binary task.

Table 6
Bias Indices Ranked in Order of the χ^2 Obtained When Each Model Was Fitted to the Data Pooled Across All Observers

	Cumulated χ^2			
	Constrained		Unconstrained	
Rank	Index	χ^2 (df = 35)	Index	χ^2 (df = 15)
1	c^*_{sd}	78	c^*_{sd}	18*
2	c_i	82	c_i	20*
3	c_d	178	c_{sd}	109
4	c_{sd}	261	c_d	134
5	c'_i	651	c'_i	273
6	k	974	β_d	318
7	β_i	1,191	β_i	341
8	c'_d	1,306	c'_d	364
9	β_d	2,000	k	518

Note—The constrained model forced the four values of d' estimated for each isobias curve to be the same for every curve (as in all previous model fits), whereas the unconstrained model allowed these estimates of d' to take on unique values for each isobias curve. The criterion location indices perform the best in each analysis. All χ^2 values are significant ($p < .05$), except for the fits based on c^*_{sd} and c_i in the unconstrained model. *$p > .05$.

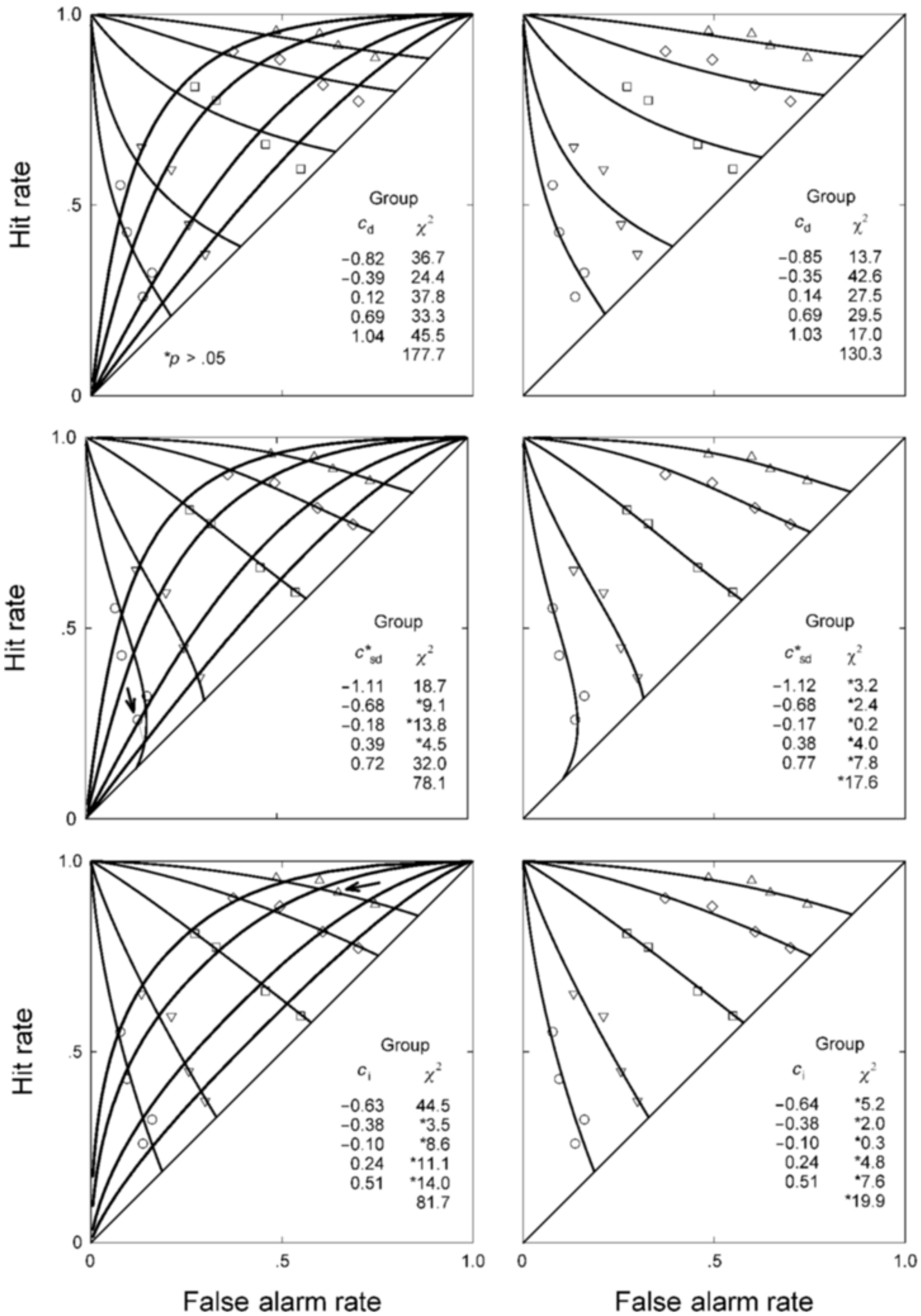

Figure 6. Best-fitting models based on the bias indices c_d, c^*_{sd}, and c_i. Data have been pooled across all the observers (both conditions). Each point is based on 1,600 trials. The panels on the left illustrate the constrained fits, whereas those on the right show the unconstrained fits. Goodness-of-fit statistics for individual isobias curves are provided next to the index value (see the text for sensitivity parameter estimates and fit statistics for the constrained fits). For the constrained model, $df = 7$ (cumulated χ^2 $df = 35$) for each isobias curve, whereas for the unconstrained model $df = 3$ (cumulated χ^2 $df = 15$).

General Findings and Taxonomy

Tables 4 and 6 indicate that the class of bias indices referred to as criterion location indices account for the variation in the data considerably better than do any of the other indices. In particular, the c^*_{sd} and c_i indices have, in most analyses, performed particularly well. Indeed these two indices were the only ones that provided fits that did not significantly deviate from the data when parameters were estimated independently for each isobias curve (see Table 6 and Figure 6). Irwin et al. (2001) obtained similar results for the c_i index in a small experiment on face discrimination. Although they did not provide any unconstrained fits for their data, c_i was found to provide the best fit when the models were constrained. The detection-theoretic model based on c_i was the only one examined that provided a statistically good fit to their data (Irwin et al. did not investigate the ability of c^*_{sd} to fit their data). For reasons outlined above, all of the constrained models in the present study provided statistically poor fits to the data.

As for the remaining indices, there is a weak pattern evident in Tables 4 and 6. For all cases of constrained modeling, the rank of the k index is either five or six. This places this decision variable index second behind the criterion location indices, c^*_{sd}, c_i, c_d, and c_{sd}. As for the other two taxonomic classes (likelihood ratio and relative criterion location), the ordering is not that clear. The likelihood-ratio indices always have a rank of seven or greater, and in two of five instances they hold the top two ranks. This may justify giving the likelihood-ratio indices the prize for worst performance. However, this demarcation is rather mute, since we are seeking the best-performing indices and the precise ordering of those that perform particularly poorly is not important. What is interesting is that the various indices within a taxonomic group tentatively demonstrate similar performance in accounting for the data.

Decision Strategies

The finding that the fit statistics for the isosensitivity curves did not clearly favor either of the available decision strategies encouraged us to fit the bias indices available for both strategies. The fit statistics for the isobias curves demonstrated a clear superiority of two bias indices, c^*_{sd} and c_i. In many ways, this is a good finding. If the isosensitivity curves did not clearly indicate a decision strategy, there is no reason that isobias curves should.

There is evidence that the discrimination of simple sinusoidal auditory stimuli in the *same–different* task is usually accompanied by the adoption of a difference decision strategy (Hautus et al., 1994). The data from the present study do not comment strongly on this earlier finding. First, the fit statistics for the isosensitivity curves did not favor either model. Second, there were two best-fitting bias indices—again, one from each decision strategy.

There are two differences between this research and previous attempts to determine the decision strategy adopted by observers on the *same–different* task. Arguably the most important difference is the use of a binary response task instead of a rating task. There is more variation in the distribution of points in the unit square when a binary response task is used. This is largely because each point is independent of the others. In the rating procedure, all the points share information because of the method used to calculate the frequencies upon which the false alarm and the hit rates are based. The method adopted guarantees that a point associated with a relatively lax criterion will always have a false alarm rate and a hit rate that are greater than or equal to these probabilities for any point with a relatively strict criterion. No such organization is imposed on binary response data.

The second difference is the provision of quantitative feedback. Previous research informed the observer only whether his or her response on a trial was correct or not. It is likely that this subtle difference in feedback would cause changes in bias. There is no evidence that it would cause a change in sensitivity.

There is another issue that may have an effect on the isosensitivity curve. It has been suggested that the use of SPP to induce changes in bias may also affect sensitivity (e.g., Markowitz & Swets, 1967). Laming (1986, pp. 95–100) illustrated this by showing that rating task isosensitivity curves for auditory discrimination provide estimates of sensitivity that are proportional to SPP. Although we have no direct evidence to support this finding, the χ^2 fit statistics given for the pooled data in Table 2 are larger for the SPP condition for three out of four fits based on the difference strategy and all fits based on the likelihood-ratio strategy. One explanation for this is that each point in the SPP condition lies on a slightly different isosensitivity curve. The attempt to fit a single isosensitivity curve to these data would lead to the inflated χ^2 values observed for this condition.

Range Effects

Irwin et al. (2001) highlighted some range effects that can influence the desirability of the various bias indices proposed for the *same–different* task. The problem is that, for some indices, the range of values that can be assumed is dependent on sensitivity. For the difference model of the *same–different* task, all bias indices except for k are limited in their range by the level of sensitivity, whereas none of the indices for the likelihood-ratio model exhibit range effects. For example, β_d must be less than $\exp(d'^2/4)$, c_d must be less than $2/d' \operatorname{arcosh}[\exp(d'^2/4)]$, c'_d must be less than $2/d'^2 \operatorname{arcosh}[\exp(d'^2/4)]$, and c_{sd} must be less than $d'/2$ (see Irwin et al., 2001, Appendix). There is no closed-form expression for the limiting value of c^*_{sd} as a function of sensitivity. Although these range effects do not preclude the calculation of a value of sensitivity and bias for any location in the unit square, they could pose some difficulties. For example, the estimated value of any of these bias indices for a given isobias curve is limited by the smallest estimate of sensitivity for a point on that curve.

As an example, consider the constrained fits of the c^*_{sd} index to the group data (see Figure 6). The estimate of sensitivity for the most difficult condition is $d' = 1.0$, and hence, c^*_{sd} must be less than 1.1. This must be true for all of the fits reported. Three of the four points lie to the left of the first isobias curve fitted to these data (the left-hand curve with $c^*_{sd} = 0.72$), suggesting that a limitation has been placed on the value of the bias index by range effects. But for this case, or any other case illustrated in this manuscript, this is not the cause of the apparent improper fit. The data point on this isobias curve that corresponds to the most difficult condition actually lies almost above the point of intersection between the best-fitting isobias and isosensitivity curves. Increasing the value of c^*_{sd} would move the curve further to the left, increasing the already large distance between that intersection and the associated data point. Indeed, the only data point that the isobias curve would move closer to, if the curve were moved to the left, is the second point from the top. To reinforce the premise that this odd-looking fit is not a range effect, consider the unconstrained fits in Figure 6. Here, without constraints placed on sensitivity, the best-fitting isobias curve has $c^*_{sd} = 0.77$, still considerably less than 1.1.

A theoretical appraisal of this issue is worth considering. Assume that the estimate of sensitivity for one of the isosensitivity curves, collected in a similar experiment to this one, is 0.1. This places a severe limitation on the values that these bias indices can assume *for any* isobias curve: β_d, c_d, c'_d, c^*_{sd}, and c_{sd} must be less than 1.003, 1.415, 14.15, 0.955, and 0.050, respectively. This appears limiting. However, if the data do conform to any of these models, the range effects will be irrelevant for that model; if the model reflects underlying structure, the data will conform. Turning this around, it is apparent that a powerful test of an index with range limitations to account for the data would be to assess the performance of the index for stimuli that are almost undetectable—that is, as d' approaches zero.

This example raises an important question. Why not assess the maximum value of these indices when d' is zero? As has been pointed out by others, the likelihood ratio bias index for the single-interval task is undefined when $d' = 0$, because the underlying normal density functions for signal and noise are identical in this case (Macmillan & Creelman, 1991, p. 47). This is equally true for the *same–different* indices β_d and β_i. This is not a range effect as defined above (however, β_d, but not β_i, suffers from such an effect). It is more a mathematical oddity caused by a singularity that occurs only when d' is exactly zero. For any value of d' in the vicinity of zero, β_d and β_i can take on any value in their permissible range.

A Superior Bias Index

The results of this experiment showed c^*_{sd} and c_i to be the best bias indices for the two independent conditions in this auditory *same–different* experiment. This concurs with the outcome obtained by Irwin et al. (2001) for a visual *same–different* experiment. This provides support for the claims (e.g., See et al., 1997; Snodgrass & Corwin, 1988) that c is a good index to use for the single-interval task, because c^*_{sd}, c_i, c_d, c_{sd}, and c are defined in a similar manner for the two psychophysical tasks—they all belong to the criterion location class of indices.

Also bearing on the issue of index performance in the single-interval task is our finding that the likelihood-ratio indices performed particularly poorly, if not the worst of all the indices, in the *same–different* task. The performance of β has on occasion been compared with c in the single-interval task—occasionally, with mixed results. For example, Hoshino (1991) found similar performance for both indices on a memory recognition task. The present results suggest that further investigation is warranted to contrast the efficacy of the c and β indices for the single-interval task.

The Manipulation of Bias

Although Table 4 indicates a great deal of similarity in the underlying structure of the data collected for the SPP and payoff conditions—which motivated the accumulation of the data across both conditions for the final analyses—there are some obvious differences. These occur primarily in the magnitude of the best-fitting parameters. For example, it appears that the SPP condition yielded more control over observer bias, as indicated by a wider spacing between the points along each isosensitivity curve. However, as was mentioned earlier with respect to the apparent reduced observer sensitivity in the payoff condition relative to the SPP condition, this study was not designed to investigate these types of intergroup differences. For example, the small number of observers in each condition leads to low statistical power, even to detect what appear to be large differences. Neither of the two intergroup differences mentioned reached significance.

Summary and Recommendation

This study showed that the c^*_{sd} and c_i bias indices provided the best-fitting isobias curves to these data collected in a *same–different* task. This is in agreement with an earlier study (Irwin et al., 2001), which found that the c_i index provided the best description of the data when incorporated into a standard detection-theoretic model for the *same–different* task (c^*_{sd} was not investigated in that study). The present study had a built-in replication and employed auditory stimuli, whereas Irwin et al. employed visual stimuli. Both studies agree that the likelihood-ratio indices provide a poor description of the data. This finding contrasts with some of the reports on this index for the single-interval task (Hoshino, 1991). However, there are theoretical reasons to be cautious about the likelihood-ratio indices; for example, they do not conform to the monotonicity requirement (Macmillan & Creelman, 1990). Further research using the single-interval design is required on this class of indices.

The c^*_{sd} index performed slightly better overall than the c_i index; however, the difference in performance was not large (c^*_{sd} outperformed c_i for the data combined across the two conditions, both when the fits were constrained and when they were not constrained). Although the fit of the isosensitivity curves to the data provided little indication of the decision strategy adopted by the observers, this difference in performance between c^*_{sd} and c_i gives tentative evidence that the difference decision strategy was adopted. This is consistent with earlier research that showed that observers adopt a difference decision rule when making *same–different* judgments about the amplitude of tones (Hautus et al., 1994).

The traditional definition of the c index—the distance of the criterion from the location on the decision axis that has a likelihood ratio of one—has led to the c_d index for the *same–different* task. This index did not perform as well as c^*_{sd}, which we have defined as the distance from the location on the decision axis that gives rise to equal hit and correct-rejection rates (i.e., the hit rate plus the false alarm rate equals one). However, the definition for c^*_{sd} would still give rise to the c index if applied to the single-interval task. It also gives rise to the c_i index for the *same–different* task when the likelihood-ratio strategy is assumed. This suggests that the formal definition for c could be changed so that all three of these indices (c, c^*_{sd}, and c_i) are defined identically. The only difference between these indices would be the nature of the decision space upon which the definition is effected. This idea is supported by the empirical finding that these three indices tend to be superior in accounting for human behavior when compared with other bias indices.

Those who argue that bias should remain constant when sensitivity changes (all other things being equal) may draw comfort from the consistency with which the criterion location indices have performed. This parallels the results supporting the c index that were obtained by Snodgrass and Corwin (1988) and See et al. (1997). Furthermore, it does not necessarily disagree with Hoshino (1991), who found that c and β performed about equally well.

Although evidence appears to be mounting for the criterion location class of indices, particularly in the *same–different* task, we recommend a pragmatic approach to analysis, rather than a prescriptive one. When one compares observer bias for data at the same level of sensitivity, it does not matter which index is adopted, the outcome of comparisons between the bias indices will always be the same. However, when sensitivity varies, it may be prudent to adopt the index that best describes the data—that is, if the nature of the data allows such an assessment to be made. If not, a prescriptive approach may be required, and the best evidence to date supports the adoption of a criterion location index of bias.

REFERENCES

Dai, H., Versfeld, N. J., & Green, D. M. (1996). The optimum decision rules in the *same–different* paradigm. *Perception & Psychophysics*, **58**, 1-9.

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

Ennis, D. M., Palen, J. J., & Mullen, K. (1988). A multidimensional stochastic theory of similarity. *Journal of Mathematical Psychology*, **32**, 449-465.

Gardner, R. M. (1997). Misconceptions about classical psychophysics and the measurement of response bias. *Perceptual & Motor Skills*, **84**, 587-594.

Gauthier, I., Behrmann, M., & Tarr, M. J. (1999). Can face recognition really be dissociated from object recognition? *Journal of Cognitive Neuroscience*, **11**, 349-370.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.

Hautus, M. J., Irwin, R. J., & Sutherland, S. (1994). Relativity of judgements about sound amplitude and the asymmetry of the same–different ROC. *Quarterly Journal of Experimental Psychology*, **47A**, 1035-1045.

Hoshino, Y. (1991). A bias in favor of the positive response to high-frequency words in recognition memory. *Memory & Cognition*, **19**, 607-616.

Ingham, J. G. (1970). Individual differences in signal detection. *Acta Psychologica*, **34**, 39-50.

Irwin, R. J., & Hautus, M. J. (1997). Likelihood-ratio decision strategy for independent observations in the *same–different* task: An approximation to the detection-theoretic model. *Perception & Psychophysics*, **59**, 313-316.

Irwin, R. J., Hautus, M. J., & Francis, M. A. (2001). Indices of response bias in the *same–different* experiment. *Perception & Psychophysics*, **63**, 1091-1100.

Laming, D. (1986). *Sensory analysis*. New York: Academic Press.

Macmillan, N. A., & Creelman, C. D. (1990). Response bias: Characteristics of detection theory, threshold theory, and "nonparametric" indexes. *Psychological Bulletin*, **107**, 401-413.

Macmillan, N. A., & Creelman, C. D. (1991). *Detection theory: A user's guide*. Cambridge: Cambridge University Press.

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

Noreen, D. L. (1981). Optimal decision rules for some common psychophysical paradigms. In S. Grossberg (Ed.), *Mathematical psychology and psychophysiology* (Proceedings of the Symposium in Applied Mathematics of the American Mathematical Society and the Society for Industrial and Applied Mathematics, Vol. 13, pp. 237-280). Providence: RI: American Mathematical Society.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. (1992). *Numerical recipes in C: The art of scientific computing* (2nd ed.). Cambridge: Cambridge University Press.

See, J. E., Warm, J. S., Dember, W. N., & Howe, S. R. (1997). Vigilance and signal detection theory: An empirical evaluation of five measures of response bias. *Human Factors*, **39**, 14-29.

Simpson, A. J., & Fitter, M. J. (1973). What is the best index of detectability? *Psychological Bulletin*, **80**, 481-488.

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

Sorkin, R. D. (1962). Extension of the theory of signal detectability to matching procedures in psychophysics. *Journal of the Acoustical Society of America*, **34**, 1745-1751.

NOTE

1. The hit rate is defined as the probability that the observer responded *same* given that the stimuli were the same, and the false alarm rate is defined as the probability of the response *same* given that the stimuli were different.

(Manuscript received December 31, 2001;
revision accepted for publication January 30, 2003.)