Nonparametric Relationships between Single-Interval and Two-Interval Forced-Choice Tasks in the Theory of Signal Detectability

Judi A. Lapsley Miller,[1] Brian K. Scurfield, Vit Drga,[2]
Susan J. Galvin,[3] and John Whitmore

Victoria University of Wellington, New Zealand
E-mail: judi@psychophysics.org, vit.drga@psychophysics.org, sue.galvin@psychophysics.org

Green's relationship, $\mathscr{A}_{SI} = P(C)_{2I}$, which equates the area, $\mathscr{A}_{SI}$, under the receiver operating characteristic (ROC) curve in the single-interval forced-choice (SIFC) task with the proportion correct, $P(C)_{2I}$, in the two-interval forced-choice (2IFC) task, is rederived using the cross-correlation functions of the SIFC evidence distributions. The relationship is generalized to include discrete random variables, unidimensional decision axes that do not need to be monotonic with likelihood ratio, and arbitrary prior and guessing probabilities. A 2IFC difference decision rule is assumed. Further nonparametric relationships, including an equality between an entropy transform of $\mathscr{A}_{SI}$ and the 2IFC channel capacity, nonparametric bounds on the area under the 2IFC ROC curve in terms of $\mathscr{A}_{SI}$, and methods for estimating 2IFC ROC curves based on information from the SIFC task, are developed. These relationships are investigated experimentally. Experiment I is a frequency-discrimination task where the evidence is known to be distributed as a discrete random variable. Experiment II is an amplitude-discrimination task where the theoretical evidence distributions are continuously distributed. The problem of observer inconsistency is addressed by repeating the experiments multiple times, using the same stimuli, then using group operating characteristic (GOC) analysis to remove unique noise. Results from Experiment I show excellent support for all the theoretical relationships, and results from Experiment II show partial support for the theoretical relationships.   © 2002 Elsevier Science

1. INTRODUCTION

In modern psychophysics, the experimental tasks most widely used are the single-interval forced-choice (SIFC) task and the two-interval forced-choice (2IFC) task. The analysis of these tasks is based on the theory of signal detectability (TSD) (Marill, 1956; Peterson, Birdsall, & Fox, 1954). With particular assumptions, the two tasks are connected theoretically by Green's relationship, which shows that the area, $\mathscr{A}_{\text{SI}}$, under the receiver operating characteristic (ROC) curve in the SIFC task is equal to the proportion of correct decisions, $P(\text{C})_{\text{2I}}$, in the 2IFC task (Green, 1964b; Green & Moses, 1966; Green & Swets, 1966). This relationship is important because it allows prediction of psychophysical performance independently of any one psychophysical task (Green, 1964b; Green & Swets, 1966; Swets, 1959), which was previously impossible with classical psychophysical methods (Swets & Green, 1961). The purpose here is to clarify and extend the underlying theory for $\mathscr{A}_{\text{SI}} = P(\text{C})_{\text{2I}}$, to derive new nonparametric relationships between the SIFC and 2IFC tasks based on this theory, and to evaluate these new results experimentally.

SIFC and 2IFC experimental tasks defined. The SIFC experiment embodies the simplest nontrivial detection task (Egan, 1975). The experiment consists of a set of independent trials, each containing an observation interval, a decision interval, and an optional payoff interval. During the observation interval either the $\mathscr{S}$ event (signal plus noise) occurs or the $\mathscr{N}$ event (noise alone) occurs. The observer never has direct access to knowledge of these events, only to uncertain evidence associated with the events. During the decision interval, the observer's task is to decide if the $\mathscr{S}$ event occurred, based on this evidence.

The evidence for the two events is modeled as conditional probability functions, $X_{\mathscr{S}}$ and $X_{\mathscr{N}}$, on a univariate random variable, X, where X is the decision axis. To make a decision, the observer partitions X by using a criterion c. If the evidence x is greater than c the observer decides a signal occurred and responds "Yes"; if the evidence is less than c the observer decides no signal occurred and responds "No." If X is discrete and x is equal to c then the observer guesses "Yes" with probability γ and "No" with probability $1 - \gamma$. The observer may vary the criterion and guessing probability depending on the prior probabilities of the events and the payoffs associated with each of the four possible outcomes (hit, miss, false alarm, and correct-rejection).

A trial in a 2IFC experiment consists of two observation intervals, a decision interval, and an optional payoff interval. In one observation interval the $\mathscr{S}$ event occurs, and in the other the $\mathscr{N}$ event occurs. During the decision interval, the observer's task is to decide in which *interval* the $\mathscr{S}$ event occurred or, equivalently, in which *order* the two events occurred. The evidence, δ, the observer uses to make a 2IFC decision is the result of a comparison between the evidence from the first interval, x_{I}, and the evidence from the second interval, x_{II}. There are many ways of making this comparison, the most common being to take the difference $\delta = x_{\text{I}} - x_{\text{II}}$ (Egan, 1975) or the ratio $\delta = x_{\text{I}}/x_{\text{II}}$ (Green & McGill, 1970). The 2IFC observer makes a decision by comparing δ to a criterion, k. If $\delta > k$ then the observer decides the signal was in the first interval and responds "One"; if $\delta < k$ then the observer decides the signal was in the second interval and responds "Two"; and if $\delta = k$ then

the observer guesses "One" with probability ξ and "Two" with probability $1 - \xi$. Assuming a difference decision rule, the 2IFC decision axis is defined as the random variable $\varDelta = X_{\mathrm{I}} - X_{\mathrm{II}}$, with $\varDelta_{\langle \mathscr{S}, \mathscr{N} \rangle}$ and $\varDelta_{\langle \mathscr{N}, \mathscr{S} \rangle}$ denoting the conditional forms of $\varDelta$ associated with the event orderings $\langle \mathscr{S}, \mathscr{N} \rangle$ and $\langle \mathscr{N}, \mathscr{S} \rangle$, respectively. The 2IFC experiment can be analyzed in a similar way to an SIFC experiment, because the event orderings of the 2IFC task can, themselves, be regarded as events.

SIFC and 2IFC ROC curves.   One of the more important concepts to come out of TSD is the ROC curve, which summarizes the observer's ability to discriminate between two events. The ROC curve is a graph of the observer's hit rate against its false-alarm rate, for all possible criteria (Egan, 1975; Green & Swets, 1966; Peterson *et al.*, 1954). Both SIFC and 2IFC ROC curves can be generated either by repeating a binary-decision experiment with the observer using a different criterion each time or by using a rating scale where the observer employs multiple criteria within an experiment. Some authors make a distinction between the binary-decision task and the rating-scale task (e.g., Bamber, 1975; Green, 1964b), but this is unnecessary because the binary-decision task is actually a two-point rating-scale task. Green and Swets (1966) showed that the theoretical SIFC ROC curve is the same for either method, and there is evidence that this is also true for empirical SIFC ROC curves (e.g., Emmerich, 1968a, 1968b; Hanley & McNeil, 1982; Watson, Rilling, & Bourbon, 1964). In principle, there should also be no difference between 2IFC ROC curves obtained by binary decisions or by ratings, but this has not been investigated.

There are, in fact, only a few examples of theoretical or empirical 2IFC ROC curves in the literature (Egan & Clarke, 1966; Falmagne, 1985; Friedman & Carterette, 1964; Leshowitz, 1969; Schulman & Mitchell, 1966). The use of 2IFC ROC curves has not been pursued because, as Luce explained, "… no one seems ever to collect 2-alternative, forced-choice ROC curves, there being a myth to the effect that this procedure, unlike the yes-no one, is unbiased" (Luce, 1997, p. 85). It is assumed that both the prior probabilities of the 2IFC event orderings and the payoffs for the 2IFC outcomes are symmetric. Therefore an ideal 2IFC observer should display no preference for one decision over the other. A 2IFC observer is considered biased if there is a tendency to decide that an event ordering occurred more often than determined by the prior probability of the event ordering. Rather than obtaining a 2IFC ROC curve, $P(\mathrm{C})_{2\mathrm{I}}$ is used as a measure of detectability, along with methods for "correcting" any bias (Green & Swets, 1966). Green and Swets (1966) argued that any biases are usually small and that the economies of deriving just $P(\mathrm{C})_{2\mathrm{I}}$ outweigh the need for deriving the entire 2IFC ROC curve. By obtaining the complete 2IFC ROC curve, however, performance can be evaluated for any given bias, and measures of detectability calculated from the ROC curve may also be estimated. By using a rating scale it is just as fast to obtain an entire 2IFC ROC curve as it is to obtain one point using a binary decision.

Measures of detectability.   The relationships between the SIFC and 2IFC tasks are illustrated by relationships between their respective measures of detectability. A measure of detectability is used to summarize an observer's performance in a psychophysical task. Ideally, a measure of detectability should be free from

confounding by processes not associated with stimulus properties, such as the prior probabilities of the events, the payoffs, and the decision criterion. A measure of detectability is also more useful if it is nonparametric, for comparisons may then be made among different observers and different detection tasks, regardless of the nature of the evidence distributions. Scurfield (1995, 1996) suggested that a measure of detectability should also be a metric; that is, the measure should have a true zero, be nonnegative, be invariant to labeling of the events and decisions, and obey the triangle inequality. The more common measures of detectability in use are d', which is parametric because it is based on normal distributions; $P(C)_{SI}$ and $P(C)_{2I}$, which are both dependent on the prior probabilities of the events and the decision criterion; and $\mathscr{A}_{SI}$, which although nonparametric, and independent of priors, payoffs, and criterion, is not a metric. These limitations restrict the extent to which the SIFC and 2IFC tasks may be related. The goal is to reduce these limitations, including developing relationships between the SIFC and 2IFC tasks using new measures of detectability, which have the properties Scurfield described as desirable.

Theoretical relationships between SIFC and 2IFC tasks in the literature. Attempts to relate SIFC and 2IFC tasks to each other have been restricted in their generality by the assumptions made. For example, parametric relationships based on normal distributions have only limited usefulness, because many detection tasks do not involve normal distributions. There is, unfortunately, a belief that TSD is applicable only to detection tasks where the evidence is normally distributed (e.g., Hodos, 1970; Simpson & Fitter, 1973). Hanley and McNeil (1982) suggest that many researchers have been turned off TSD because of this mistaken belief. *TSD does not need the assumption that the evidence is distributed as a normal density.* Green's demonstration that $\mathscr{A}_{SI} = P(C)_{2I}$ is more general, because both $\mathscr{A}_{SI}$ and $P(C)_{2I}$ are nonparametric measures of detectability (Green, 1964b; Green & Moses, 1966; Green & Swets, 1966).[4] Green's relationship has been extended by Bamber (1975), Egan (1975), Falmagne (1985), Green and Moses (1966), Iverson and Sheu (1992), and McNicol (1972), but none have explicitly considered the combined effects of the decision axis, decision and guessing rules, prior probabilities, and payoffs.

The derivations of $\mathscr{A}_{SI} = P(C)_{2I}$ by Green and Swets (1966) and Egan (1975) both assume that the decision axis is the likelihood ratio,[5] $\ell(X)$, or is monotonic with it, and that it is continuous and well behaved. There are, however, a number of interesting detection tasks best modeled by discrete random variables. The discrete version of $\mathscr{A}_{SI} = P(C)_{2I}$ was mentioned by Green and Moses (1966), and a worked example appeared in McNicol (1972). Bamber (1975) did show that $\mathscr{A}_{SI} = P(C)_{2I}$ for

---

[4] Green and Swets (1966) also presented one other nonparametric relationship, which related points on the 2IFC ROC curve to the SIFC ROC curve.

[5] In the SIFC task, the likelihood ratio is $\ell(x) = P(x \mid \mathscr{S})/P(x \mid \mathscr{N})$, which is the ratio of the event-conditional probability mass functions for discrete random variables, or $\ell(x) = f(x \mid \mathscr{S})/f(x \mid \mathscr{N})$, which is the ratio of the event-conditional probability density functions for continuous random variables (Egan, 1975). In the 2IFC task an observer could either use X or $\ell(X)$ to form the 2IFC decision axis ($\delta = x_I - x_{II}$ or $\delta = \ell(x_I) - \ell(x_{II})$, respectively), either of which could then also be converted to the 2IFC likelihood ratio, $\ell(\Delta)$.

finitely discrete random variables, but did not consider the effect of varying the prior probabilities or guessing probabilities. The guessing probabilities provide a justification for using linear interpolation between the points in a discrete ROC curve (Egan, 1975; Green, 1960b). Only Green and Moses (1966) mention guessing probabilities in their derivation.

Even though it can be shown that observers can achieve most decision goals by using the likelihood ratio, X is, nonetheless, the appropriate decision axis for some tasks. Moreover, a proof for X as the decision axis is more general, because it automatically includes the case for the likelihood ratio.[6] Falmagne (1985), Green and Moses (1966), and Iverson and Sheu (1992) have generalized $\mathscr{A}_{SI} = P(C)_{2I}$ for X as the decision axis as well as $\ell(X)$, but only for continuous random variables.

In spite of the fact that $\mathscr{A}_{SI}$ is independent of criterion, prior probabilities, and payoffs, $P(C)_{2I}$ does depend on all three to achieve a variety of specific decision goals. Therefore, most derivations of $\mathscr{A}_{SI} = P(C)_{2I}$ assume the prior probabilities of the events to be equal and that an observer sets the criterion at zero for the 2IFC task (assuming a difference decision rule). Usually, the payoffs are not considered, because they are implicitly assumed to be symmetric. If they were not symmetric, then an observer would need to consider changing the criterion and the guessing probabilities in order to achieve a particular decision goal.

These examples demonstrate the need for a general theory relating $\mathscr{A}_{SI}$ to $P(C)_{2I}$ that covers both discrete and continuous random variables and takes into account suboptimal decision strategies and goals for both tasks as well as the effects of the prior and guessing probabilities.

Experimental comparisons of the SIFC and 2IFC tasks. It is important to evaluate the relationships between the SIFC and 2IFC tasks experimentally. There have been a number of experimental studies that have evaluated the relationships between the two tasks. Early studies focused on evaluating parametric relationships, especially those based on normal evidence distributions, but the results did not consistently support the theories. Wickelgren (1968) found it amazing that any of the experiments showed a relationship, given the number of assumptions required. Many reasons have been put forward to explain the variability of the results, the most popular being noise processes that degrade observer performance (e.g., Green, 1964a; Wickelgren, 1968). Although observer inconsistency is a major problem, and a likely explanation of variable results, it is still surprising that many researchers did not consider the possibility that the underlying distributions were not normal.

There have been three attempts to verify $\mathscr{A}_{SI} = P(C)_{2I}$ experimentally. Green and Moses (1966) reported a weak empirical relationship between $\mathscr{A}_{SI}$ and $P(C)_{2I}$ in a short-term memory test using a 6-point rating-scale SIFC task and a binary-decision 2IFC task. The lack of agreement between the two measures of detectability was attributed to variability caused by the method used for measuring $\mathscr{A}_{SI}$ and to the small number of stimuli used. The experiment was repeated with more stimuli, and the relationship was stronger, but still too variable to support an equality between $\mathscr{A}_{SI}$ and $P(C)_{2I}$.

---

[6] Consider $Y = \ell(X)$ as a new decision axis, then apply the proofs to Y instead of X (see Section 2.2).

Emmerich (1968a) found in a binaural tone-in-noise experiment that the "agreement is quite good" between $\mathscr{A}_{SI}$ and $P(C)_{2I}$. He presented psychometric functions of $\mathscr{A}_{SI}$ and $P(C)_{2I}$, but not the numerical difference between the two measures. D. Emmerich (personal communication, March 1996) has kindly given us these values. On average, across all observers and conditions, $\mathscr{A}_{SI}$ approximately equaled $P(C)_{2I}$, but the difference between the two measures for individual observers ranged from -0.111 to 0.062, where the difference could range between -0.5 and 0.5 (assuming $\mathscr{A}_{SI}$ and $P(C)_{2I}$ are both $\geqslant 0.5$).

Whitmore, Williams, and Ermey (1968), in a study on envelope detection, considered both tone-in-noise and noise-in-noise detection using a rating-scale SIFC task and a binary-decision 2IFC task. Because different signal-to-noise ratios were used in the two conditions, $\mathscr{A}_{SI}$ and $P(C)_{2I}$ could not be directly compared. Instead, seven-point psychometric functions were obtained, then an attenuated theoretical psychometric function was fitted to the points. If $\mathscr{A}_{SI} = P(C)_{2I}$, the psychometric functions should have fallen on top of one another. Although the shape of the theoretical psychometric functions fitted the data well, the attenuations were not constant across the SIFC and 2IFC conditions (differing by approximately 0–2.5 dB), therefore an equality between $\mathscr{A}_{SI}$ and $P(C)_{2I}$ was not strongly supported.

The fact that these experiments did not strongly support an equality between $\mathscr{A}_{SI}$ and $P(C)_{2I}$ may be due to observer inconsistency, which can produce different results from the same observer even when it is presented with exactly the same stimulus set (Drga, 1999; Taylor, Boven, & Whitmore, 1991). If this error is not reduced, differences (or similarities) may be found for $\mathscr{A}_{SI}$ and $P(C)_{2I}$ that are due merely to differences in observer inconsistency between the SIFC and 2IFC tasks.

Overview of paper. This paper evaluates nonparametric relationships between the SIFC and 2IFC tasks both theoretically and experimentally. $\mathscr{A}_{SI} = P(C)_{2I}$ is derived for both discrete and continuous random variables, based on the cross-correlation functions of $X_{\mathscr{S}}$ and $X_{\mathscr{N}}$. Three relationships implied by $\mathscr{A}_{SI} = P(C)_{2I}$ are then developed: (a) a relationship between Scurfield's (1995, 1996) detectability measure $\mathscr{D}_{SI}$ and the channel capacity of the 2IFC observer, (b) bounds on the area under the 2IFC ROC curve, $\mathscr{A}_{2I}$, derived from $\mathscr{A}_{SI}$, and (c) methods for deriving a 2IFC ROC curve from information in the associated SIFC task. To evaluate these relationships, two experiments were conducted: one where the evidence was discretely distributed, the other where the evidence was continuously distributed. Previous experimental results have been inconclusive because of the variability in the results, therefore group operating characteristic (GOC) analysis was used to reduce the effects of observer inconsistency (Drga, 1999; Taylor *et al.*, 1991).

2. THEORY

The key to deriving relationships between the SIFC and 2IFC tasks is to define the 2IFC evidence distributions in terms of the SIFC evidence distributions. First, the SIFC and 2IFC evidence distributions are defined, along with their ROC curves, for both discrete and continuous random variables. Green's relationship is

then derived for both discrete and continuous random variables. Similar relationships are also derived for Scurfield's measures of detectability. Finally, 2IFC ROC analysis is considered in more detail, and some relationships relating SIFC and 2IFC ROC curves are developed.

In the SIFC task, it is assumed that $X_{\mathscr{S}}$ and $X_{\mathscr{N}}$ are distributed as independent, stationary, univariate random variables. When X is discrete, probability is assumed to be massed on the integers, for convenience. In general, for either discrete or continuous X, $P(C)_{2I}$ may be derived from a simple comparison between the evidence values in each observation interval without having to derive any 2IFC decision axis. 2IFC decision axes are needed for 2IFC ROC analysis, however, and a difference decision rule is assumed here, partly for ease of exposition and partly because of its popularity as a decision rule (Drga, 1999, reviewed other rules such as the ratio decision rule). $P(C)_{2I}$ may then be found for any 2IFC criterion.

For continuous X, the SIFC ROC curve is determined by the hit rate

$$\mathrm{HR}(c) = P(X_{\mathscr{S}} > c) = \int_{c}^{\infty} f_X(x \mid \mathscr{S}) \, dx \tag{1}$$

and the false-alarm rate

$$\mathrm{FAR}(c) = P(X_{\mathscr{N}} > c) = \int_{c}^{\infty} f_X(x \mid \mathscr{N}) \, dx. \tag{2}$$

For discrete X, the definitions of the hit rate and the false-alarm rate must take account of the probability of guessing when the evidence is tied with the criterion,

$$\mathrm{HR}(c, \gamma) = P(X_{\mathscr{S}} > c) + \gamma P(X_{\mathscr{S}} = c)$$

$$= \sum_{x=c+1}^{\infty} P(X_{\mathscr{S}} = x) + \gamma P(X_{\mathscr{S}} = c), \tag{3}$$

$$\mathrm{FAR}(c, \gamma) = P(X_{\mathscr{N}} > c) + \gamma P(X_{\mathscr{N}} = c)$$

$$= \sum_{x=c+1}^{\infty} P(X_{\mathscr{N}} = x) + \gamma P(X_{\mathscr{N}} = c). \tag{4}$$

The discrete ROC curve consists of a set of distinct points (defined when either $\gamma = 0$ or $\gamma = 1$), joined by line segments, because Eqs. (3) and (4) are linear functions of γ.

In the 2IFC task, the evidence from each interval constitutes a vector $\mathbf{x} = (x_I, x_{II})^T$ that is an instance of a random vector $\mathbf{X} = (X_I, X_{II})^T$. It is assumed that in any 2IFC trial X_I and X_{II} are independent and the same as either $X_{\mathscr{S}}$ and $X_{\mathscr{N}}$, respectively, or $X_{\mathscr{N}}$ and $X_{\mathscr{S}}$, respectively, from the associated SIFC task. The independence assumption implies that for continuous $\mathbf{X}$

$$f_{\mathbf{X}}(\mathbf{x} \mid \langle \mathscr{S}, \mathscr{N} \rangle) = f_X(x_I \mid \mathscr{S}) \, f_X(x_{II} \mid \mathscr{N}), \tag{5}$$

$$f_{\mathbf{X}}(\mathbf{x} \mid \langle \mathscr{N}, \mathscr{S} \rangle) = f_X(x_I \mid \mathscr{N}) \, f_X(x_{II} \mid \mathscr{S}), \tag{6}$$

and for discrete $\mathbf{X}$

$$P(\mathbf{X}_{\langle \mathscr{S}, \mathscr{N} \rangle} = \mathbf{x}) = P(X_{\mathscr{S}} = x_{\mathrm{I}}) \, P(X_{\mathscr{N}} = x_{\mathrm{II}}), \tag{7}$$

$$P(\mathbf{X}_{\langle \mathscr{N}, \mathscr{S} \rangle} = \mathbf{x}) = P(X_{\mathscr{N}} = x_{\mathrm{I}}) \, P(X_{\mathscr{S}} = x_{\mathrm{II}}). \tag{8}$$

Assuming a difference decision rule, the probability density functions of $\varDelta_{\langle \mathscr{S}, \mathscr{N} \rangle}$ and $\varDelta_{\langle \mathscr{N}, \mathscr{S} \rangle}$, for continuous $\varDelta$, are obtained by substituting $x_{\mathrm{I}} = x_{\mathrm{II}} + \delta$ into Eqs. (5) and (6) and then integrating over all $\mathbf{x}$,

$$f_{\varDelta}(\delta \,|\, \langle \mathscr{S}, \mathscr{N} \rangle) = \int_{-\infty}^{\infty} f_X(x+\delta \,|\, \mathscr{S}) \, f_X(x \,|\, \mathscr{N}) \, \mathrm{d}x, \tag{9}$$

$$f_{\varDelta}(\delta \,|\, \langle \mathscr{N}, \mathscr{S} \rangle) = \int_{-\infty}^{\infty} f_X(x+\delta \,|\, \mathscr{N}) \, f_X(x \,|\, \mathscr{S}) \, \mathrm{d}x, \tag{10}$$

where the dummy variable x_{II} has been replaced with x. Similarly, for discrete $\varDelta$,

$$P(\varDelta_{\langle \mathscr{S}, \mathscr{N} \rangle} = \delta) = \sum_{x = -\infty}^{\infty} P(X_{\mathscr{S}} = x+\delta) \, P(X_{\mathscr{N}} = x), \tag{11}$$

$$P(\varDelta_{\langle \mathscr{N}, \mathscr{S} \rangle} = \delta) = \sum_{x = -\infty}^{\infty} P(X_{\mathscr{N}} = x+\delta) \, P(X_{\mathscr{S}} = x). \tag{12}$$

Equations (9)–(12) are the statistical cross-correlation functions of the probability functions of $X_{\mathscr{S}}$ and $X_{\mathscr{N}}$. Note that $f_{\varDelta}(\delta \,|\, \langle \mathscr{S}, \mathscr{N} \rangle)$ is a reflection of $f_{\varDelta}(\delta \,|\, \langle \mathscr{N}, \mathscr{S} \rangle)$ in the line $\delta = 0$; that is, $f_{\varDelta}(\delta \,|\, \langle \mathscr{S}, \mathscr{N} \rangle) = f_{\varDelta}(-\delta \,|\, \langle \mathscr{N}, \mathscr{S} \rangle)$. Similarly, $P(\varDelta_{\langle \mathscr{S}, \mathscr{N} \rangle} = \delta) = P(\varDelta_{\langle \mathscr{N}, \mathscr{S} \rangle} = -\delta)$.

If $\varDelta$ is continuous and a hit is defined to be the correct selection of the event-ordering $\langle \mathscr{S}, \mathscr{N} \rangle$, then the 2IFC ROC curve is defined by the 2IFC hit rate

$$\mathrm{HR}(k)_{2\mathrm{I}} = P(\varDelta_{\langle \mathscr{S}, \mathscr{N} \rangle} > k) = \int_{k}^{\infty} f_{\varDelta}(\delta \,|\, \langle \mathscr{S}, \mathscr{N} \rangle) \, \mathrm{d}\delta \tag{13}$$

and the 2IFC false-alarm rate

$$\mathrm{FAR}(k)_{2\mathrm{I}} = P(\varDelta_{\langle \mathscr{N}, \mathscr{S} \rangle} > k) = \int_{k}^{\infty} f_{\varDelta}(\delta \,|\, \langle \mathscr{N}, \mathscr{S} \rangle) \, \mathrm{d}\delta. \tag{14}$$

If $\varDelta$ is discrete and ξ is the 2IFC guessing probability,

$$\mathrm{HR}(k, \xi)_{2\mathrm{I}} = P(\varDelta_{\langle \mathscr{S}, \mathscr{N} \rangle} > k) + \xi P(\varDelta_{\langle \mathscr{S}, \mathscr{N} \rangle} = k)$$

$$= \sum_{\delta = k+1}^{\infty} P(\varDelta_{\langle \mathscr{S}, \mathscr{N} \rangle} = \delta) + \xi P(\varDelta_{\langle \mathscr{S}, \mathscr{N} \rangle} = k), \tag{15}$$

$$\mathrm{FAR}(k, \xi)_{2\mathrm{I}} = P(\varDelta_{\langle \mathscr{N}, \mathscr{S} \rangle} > k) + \xi P(\varDelta_{\langle \mathscr{N}, \mathscr{S} \rangle} = k)$$

$$= \sum_{\delta = k+1}^{\infty} P(\varDelta_{\langle \mathscr{N}, \mathscr{S} \rangle} = \delta) + \xi P(\varDelta_{\langle \mathscr{N}, \mathscr{S} \rangle} = k). \tag{16}$$

With a difference decision rule, the 2IFC ROC curve is necessarily symmetrical in the negative diagonal of the ROC space (Egan, 1975); that is,

$$\text{HR}(k, \xi)_{2\text{I}} = 1 - \text{FAR}(-k, 1 - \xi)_{2\text{I}}. \tag{17}$$

THEOREM 2.1 (The Area under the SIFC ROC Curve). *The area under the SIFC ROC curve is related to the random variables $X_{\mathscr{S}}$ and $X_{\mathscr{N}}$ by the equation*

$$\mathscr{A}_{\text{SI}} = P(X_{\mathscr{N}} < X_{\mathscr{S}}) + \tfrac{1}{2} P(X_{\mathscr{N}} = X_{\mathscr{S}}). \tag{18}$$

COROLLARY 2.1 (The Area over the SIFC ROC Curve). *Because the area of the ROC space is unity, the area over the ROC curve is*

$$1 - \mathscr{A}_{\text{SI}} = P(X_{\mathscr{N}} > X_{\mathscr{S}}) + \tfrac{1}{2} P(X_{\mathscr{N}} = X_{\mathscr{S}}). \tag{19}$$

If X is continuous, then $P(X_{\mathscr{N}} = X_{\mathscr{S}}) = 0$ and the right-hand sides of Eqs. (18) and (19) reduce to $P(X_{\mathscr{N}} < X_{\mathscr{S}})$ and $P(X_{\mathscr{N}} > X_{\mathscr{S}})$, respectively.

Theorem 2.1 was described by Wilcox (1968) for continuous X, and Bamber (1975) provided a proof for continuous and finitely discrete X. Here, an alternative proof, based on the cross-correlation function of $X_{\mathscr{S}}$ and $X_{\mathscr{N}}$, is presented.

Proof (of Theorem 2.1 *for the continuous case).* For continuous X, $\mathscr{A}_{\text{SI}}$ is obtained by integrating the hit rate with respect to the false-alarm rate; that is,

$$\mathscr{A}_{\text{SI}} = \int_0^1 \text{HR}(c) \, d(\text{FAR}(c)). \tag{20}$$

The differential $d(\text{FAR}(c))$ is obtained by using the second fundamental theorem of calculus to differentiate Eq. (2): that is

$$d(\text{FAR}(c)) = -f_X(c \mid \mathscr{N}) \, dc. \tag{21}$$

The lower limit of zero in Eq. (20) occurs when $c = \infty$ and the upper limit of one occurs when $c = -\infty$. Therefore, after substitution for $\text{HR}(c)$ and $d(\text{FAR}(c))$,

$$\mathscr{A}_{\text{SI}} = \int_{-\infty}^{\infty} \left[\int_c^{\infty} f_X(x \mid \mathscr{S}) \, dx \right] f_X(c \mid \mathscr{N}) \, dc. \tag{22}$$

By setting $x = \delta + c$, Eq. (22) can be recast as

$$\mathscr{A}_{\text{SI}} = \int_0^{\infty} h(\delta) \, d\delta, \tag{23}$$

where

$$h(\delta) = \int_{-\infty}^{\infty} f_X(c + \delta \mid \mathscr{S}) \, f_X(c \mid \mathscr{N}) \, dc. \tag{24}$$

Here, the order of integration has been reversed. This is permissible because the integrand is a probability density function and therefore uniformly convergent.

By Eq. (9), $h(\delta)$ represents the same cross-correlation function as $f_{\Delta}(\delta \,|\, \langle \mathscr{S}, \mathscr{N} \rangle)$. Thus, $h(\delta)$ is the density of $\Delta_{\langle \mathscr{S}, \mathscr{N} \rangle} = X_{\mathscr{S}} - X_{\mathscr{N}}$. It then follows from Eq. (23) that $\mathscr{A}_{\mathrm{SI}} = P(\Delta_{\langle \mathscr{S}, \mathscr{N} \rangle} > 0) = P(X_{\mathscr{S}} - X_{\mathscr{N}} > 0) = P(X_{\mathscr{N}} < X_{\mathscr{S}})$, which completes the proof of Theorem 2.1 for continuous X.   ∎

Proof (of Theorem 2.1 *for the discrete case).*   Assume, now, that X is discrete. In Fig. 1, the shaded area, $a(c)$, below the ROC curve between $\mathrm{FAR}(c, 0)$ and $\mathrm{FAR}(c-1, 0)$ is

$$a(c) = \tfrac{1}{2}\left[\mathrm{HR}(c, 0) + \mathrm{HR}(c-1, 0)\right]\left[\mathrm{FAR}(c-1, 0) - \mathrm{FAR}(c, 0)\right]$$

$$= \tfrac{1}{2}\left[\sum_{x=c+1}^{\infty} P(X_{\mathscr{S}} = x) + \sum_{x=c}^{\infty} P(X_{\mathscr{S}} = x)\right]$$

$$\times \left[\sum_{x=c}^{\infty} P(X_{\mathscr{N}} = x) - \sum_{x=c+1}^{\infty} P(X_{\mathscr{N}} = x)\right]$$

$$= \tfrac{1}{2}\left[P(X_{\mathscr{S}} = c) + 2\sum_{x=c+1}^{\infty} P(X_{\mathscr{S}} = x)\right] P(X_{\mathscr{N}} = c)$$

$$= \tfrac{1}{2} P(X_{\mathscr{S}} = c)\, P(X_{\mathscr{N}} = c) + \sum_{x=c+1}^{\infty} P(X_{\mathscr{S}} = x)\, P(X_{\mathscr{N}} = c).$$

Summing over all $a(c)$ gives the area under the discrete SIFC ROC curve,

$$\mathscr{A}_{\mathrm{SI}} = \sum_{c=-\infty}^{\infty} \sum_{x=c+1}^{\infty} P(X_{\mathscr{S}} = x)\, P(X_{\mathscr{N}} = c)$$

$$+ \tfrac{1}{2} \sum_{c=-\infty}^{\infty} P(X_{\mathscr{S}} = c)\, P(X_{\mathscr{N}} = c). \tag{25}$$

By setting $x = \delta + c$ and then reversing the order of summation, Eq. (25) can be recast as

$$\mathscr{A}_{\mathrm{SI}} = \sum_{\delta=1}^{\infty} h(\delta) + \tfrac{1}{2} h(0), \tag{26}$$

where

$$h(\delta) = \sum_{c=-\infty}^{\infty} P(X_{\mathscr{S}} = c+\delta)\, P(X_{\mathscr{N}} = c). \tag{27}$$

By Eq. (11), $h(\delta)$ represents the same cross-correlation function as $P(\Delta_{\langle \mathscr{S}, \mathscr{N} \rangle} = \delta)$. Thus, $h(\delta)$ is the mass function of $\Delta_{\langle \mathscr{S}, \mathscr{N} \rangle} = X_{\mathscr{S}} - X_{\mathscr{N}}$. It then follows from Eq. (26) that

$$\mathscr{A}_{\mathrm{SI}} = P(\Delta_{\langle \mathscr{S}, \mathscr{N} \rangle} > 0) + \tfrac{1}{2} P(\Delta_{\langle \mathscr{S}, \mathscr{N} \rangle} = 0)$$

$$= P(X_{\mathscr{S}} - X_{\mathscr{N}} > 0) + \tfrac{1}{2} P(X_{\mathscr{S}} - X_{\mathscr{N}} = 0)$$

$$= P(X_{\mathscr{N}} < X_{\mathscr{S}}) + \tfrac{1}{2} P(X_{\mathscr{N}} = X_{\mathscr{S}}),$$

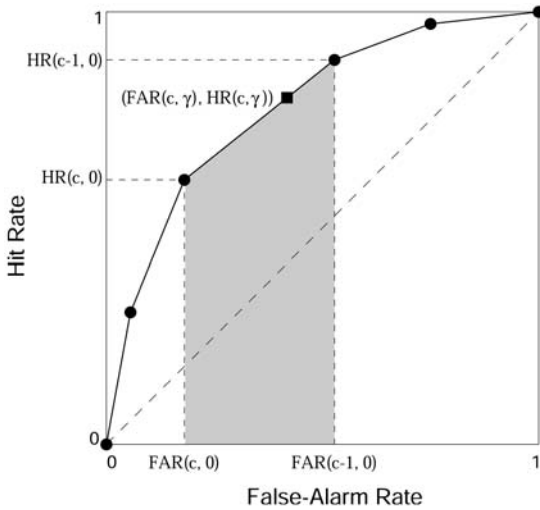which completes the proof of Theorem 2.1 for discrete X.   ∎

FIG. 1. An illustrative SIFC ROC curve generated from two discrete probability functions.

THEOREM 2.2 (The Area under the 2IFC ROC Curve). *Using the same method of derivation as in the SIFC case, it can be shown that in terms of the 2IFC random variables $\Delta_{\langle \mathscr{S}, \mathscr{N} \rangle}$ and $\Delta_{\langle \mathscr{N}, \mathscr{S} \rangle}$,*

$$\mathscr{A}_{2I} = P(\Delta_{\langle \mathscr{N}, \mathscr{S} \rangle} < \Delta_{\langle \mathscr{S}, \mathscr{N} \rangle}) + \tfrac{1}{2} P(\Delta_{\langle \mathscr{N}, \mathscr{S} \rangle} = \Delta_{\langle \mathscr{S}, \mathscr{N} \rangle}). \tag{28}$$

COROLLARY 2.2 (The Area over the 2IFC ROC Curve). *Like in the SIFC case,*

$$1 - \mathscr{A}_{2I} = P(\Delta_{\langle \mathscr{N}, \mathscr{S} \rangle} > \Delta_{\langle \mathscr{S}, \mathscr{N} \rangle}) + \tfrac{1}{2} P(\Delta_{\langle \mathscr{N}, \mathscr{S} \rangle} = \Delta_{\langle \mathscr{S}, \mathscr{N} \rangle}). \tag{29}$$

$\mathscr{A}_{2I}$ is independent of strictly monotonically increasing transformations of the 2IFC decision axis Δ. $\mathscr{A}_{2I}$ is, however, dependent on strictly monotonically increasing transformations of the SIFC decision axis X, which is used to form Δ, because the 2IFC ROC curve is dependent on such transformations. This point is returned to in Section 2.2.

THEOREM 2.3 (Green's Relationship: $\mathscr{A}_{SI} = P(C)_{2I}$). *The area under the SIFC ROC curve, $\mathscr{A}_{SI}$, is equal to the proportion correct in the 2IFC task, $P(C)_{2I}$, for both continuous and discrete random variables. For discrete random variables, this equality holds if and only if $P(\langle \mathscr{S}, \mathscr{N} \rangle) = P(\langle \mathscr{N}, \mathscr{S} \rangle) = 1/2$, or $\xi = 1/2$, or both.*

Proof. First, $P(C)_{2I}$ needs to be defined, and its dependence on the 2IFC prior probabilities and guessing probabilities noted. From there, the relationship $\mathscr{A}_{SI} = P(C)_{2I}$ follows from Theorem 2.1.

For any criterion, k, and any guessing probability, ξ, the general formula for $P(C)_{2I}$ is

$$P(C, k, \xi)_{2I} = P(\langle \mathscr{S}, \mathscr{N} \rangle) \, P(\Delta_{\langle \mathscr{S}, \mathscr{N} \rangle} > k) + \xi P(\langle \mathscr{S}, \mathscr{N} \rangle) \, P(\Delta_{\langle \mathscr{S}, \mathscr{N} \rangle} = k)$$
$$+ P(\langle \mathscr{N}, \mathscr{S} \rangle) \, P(\Delta_{\langle \mathscr{N}, \mathscr{S} \rangle} < k) + (1 - \xi) \, P(\langle \mathscr{N}, \mathscr{S} \rangle) \, P(\Delta_{\langle \mathscr{N}, \mathscr{S} \rangle} = k). \tag{30}$$

Like in the SIFC case, observers may choose to set the 2IFC criterion based on the prior probabilities of the event orderings and the payoffs associated with the 2IFC outcomes. It is assumed that an unbiased observer sets a criterion at $k = 0$, because the priors and payoffs are usually designed to be symmetrical in 2IFC experiments and because the difference distributions associated with the two event orderings in the 2IFC task are mirror-symmetrical around $k = 0$. In this case, $P(C)_{2I}$ can be expressed as

$$
\begin{aligned}
P(C, 0, \xi)_{2I} &= P(\langle \mathscr{S}, \mathscr{N} \rangle)\, P(X_{\mathscr{N}} < X_{\mathscr{S}}) + \xi P(\langle \mathscr{S}, \mathscr{N} \rangle)\, P(X_{\mathscr{N}} = X_{\mathscr{S}}) \\
&\quad + P(\langle \mathscr{N}, \mathscr{S} \rangle)\, P(X_{\mathscr{N}} < X_{\mathscr{S}}) + (1 - \xi)\, P(\langle \mathscr{N}, \mathscr{S} \rangle)\, P(X_{\mathscr{N}} = X_{\mathscr{S}}) \\
&= P(X_{\mathscr{N}} < X_{\mathscr{S}}) + P(X_{\mathscr{N}} = X_{\mathscr{S}})[\xi P(\langle \mathscr{S}, \mathscr{N} \rangle) + (1 - \xi)\, P(\langle \mathscr{N}, \mathscr{S} \rangle)].
\end{aligned}
\tag{31}
$$

Comparing this equation to Theorem 2.1 shows that in the continuous case $\mathscr{A}_{SI} = P(C)_{2I}$ for any prior probability since $P(X_{\mathscr{N}} = X_{\mathscr{S}}) = 0$. If X is discrete, however, the role of the 2IFC priors and 2IFC guessing probability needs to be taken into consideration, because $\mathscr{A}_{SI} = P(C)_{2I}$ if and only if $P(\langle \mathscr{S}, \mathscr{N} \rangle) = P(\langle \mathscr{N}, \mathscr{S} \rangle) = 1/2$, or $\xi = 1/2$, or both. ∎

COROLLARY 2.3. *The area under the SIFC ROC curve is equal to the hit rate of an unbiased 2IFC observer.*

Proof. By setting $k = 0$ and $\xi = 1/2$ in Eqs. (13) and (15) and expanding $\varDelta_{\langle \mathscr{S}, \mathscr{N} \rangle} > 0$ to $X_{\mathscr{N}} < X_{\mathscr{S}}$, it is easily seen by comparison with Eq. (31) that

$$
HR(0, \tfrac{1}{2})_{2I} = P(C, 0, \tfrac{1}{2})_{2I}.
\tag{32}
$$

It follows from Theorem 2.3 that

$$
\mathscr{A}_{SI} = HR(0, \tfrac{1}{2})_{2I}. \quad \blacksquare
\tag{33}
$$

The form of the 2IFC decision rule and, consequently, the 2IFC decision axis do not change the value of $P(C)_{2I}$ because all that is required is a method for evaluating the inequality $X_{\mathscr{N}} < X_{\mathscr{S}}$. It follows that Green's relationship also is not affected by the form of the comparison. Results that depend on the 2IFC ROC curve in its entirety, however, are dependent on the 2IFC decision rule and decision axis.

2.1. Relationships Implied by $\mathscr{A}_{SI} = P(C)_{2I}$

Bamber (1975) used Theorem 2.1 to interpret $\mathscr{A}_{SI}$ as a measure of the extent to which one distribution lies above (or below) the other distribution. This interpretation is not correct, because it implies that $\mathscr{A}_{SI}$ is a distance measure or metric. $\mathscr{A}_{SI}$ cannot be considered a metric because it does not have a true zero and it varies with respect to the labeling of the events and decisions (Scurfield, 1996).

Although $\mathscr{A}_{SI}$ itself is not a metric, it is possible to transform it into a metric. Definition 2.1 shows one such transformation (Scurfield, 1996).

DEFINITION 2.1 (Scurfield's SIFC Detectability Measure).

$$\mathscr{D}_{\mathrm{SI}} = \log 2 - \mathscr{H}_{\mathrm{SI}}, \tag{34}$$

where $\mathscr{H}_{\mathrm{SI}} = -\mathscr{A}_{\mathrm{SI}} \log(\mathscr{A}_{\mathrm{SI}}) - (1 - \mathscr{A}_{\mathrm{SI}}) \log(1 - \mathscr{A}_{\mathrm{SI}})$.

The result of this transformation of $\mathscr{A}_{\mathrm{SI}}$ is a measure of information, because $\mathscr{H}_{\mathrm{SI}}$ represents the *Shannon entropy* (Shannon, 1949) of the areas above and below the ROC curve. If base-two logarithms are used, then $\mathscr{D}_{\mathrm{SI}}$ ranges from zero to one in bits.

Scurfield (1996) used Theorem 2.1 to show that $\mathscr{D}_{\mathrm{SI}}$ is a pseudo-metric for X and a true metric for $\ell(X)$ (or its equivalent). This leads to the important conclusion that $\mathscr{D}_{\mathrm{SI}}$ is a measure of the distance between the random variables $X_{\mathscr{G}}$ and $X_{\mathscr{N}}$.[7] If $X_{\mathscr{G}}$ and $X_{\mathscr{N}}$ do not overlap then $\mathscr{D}_{\mathrm{SI}} = 1$ bit, and if $X_{\mathscr{G}}$ and $X_{\mathscr{N}}$ are identically distributed then $\mathscr{D}_{\mathrm{SI}} = 0$ bits. $\mathscr{D}_{\mathrm{SI}}$ is also nonparametric and independent of prior probabilities, payoffs, and the decision criterion. It takes account of both the shape and the location of the evidence distributions, unlike measures such as d', which is the Euclidean distance between the means of two normal evidence distributions.

DEFINITION 2.2 (Scurfield's 2IFC Detectability Measure). Similarly, $\mathscr{D}_{\mathrm{2I}}$ is the 2IFC task analog of $\mathscr{D}_{\mathrm{SI}}$.

$$\mathscr{D}_{\mathrm{2I}} = \log 2 - \mathscr{H}_{\mathrm{2I}}, \tag{35}$$

where $\mathscr{H}_{\mathrm{2I}} = -\mathscr{A}_{\mathrm{2I}} \log(\mathscr{A}_{\mathrm{2I}}) - (1 - \mathscr{A}_{\mathrm{2I}}) \log(1 - \mathscr{A}_{\mathrm{2I}})$.

An equality similar to Green's relationship, which relates $\mathscr{D}_{\mathrm{SI}}$ to the channel capacity of an unbiased 2IFC observer, $\mathscr{C}_{\mathrm{2I}}$ may be derived (Scurfield, 1996).

THEOREM 2.4. *$\mathscr{D}_{\mathrm{SI}}$ is equal to the channel capacity, $\mathscr{C}_{\mathrm{2I}}$, of the unbiased 2IFC observer*.

Proof. In general, for fixed k and fixed ξ, the channel capacity is defined as the maximum value of the *mutual information* (Abramson, 1963) between the event orderings and the decisions as the prior probabilities of the event orderings are varied for the unbiased observer. When $k = 0$ and $\xi = 1/2$, the matrix of event orderings and decisions has one degree of freedom because, by Eq. (17), $\mathrm{HR}(0, \tfrac{1}{2})_{\mathrm{2I}} = 1 - \mathrm{FAR}(0, \tfrac{1}{2})_{\mathrm{2I}}$. In this case, the observer is said to be a *uniform channel*. Applying the definition for the channel capacity of a uniform channel (Abramson, 1963) gives

$$\mathscr{C}_{\mathrm{2I}} = \log 2 + \mathrm{HR}(0, \tfrac{1}{2})_{\mathrm{2I}} \log[\mathrm{HR}(0, \tfrac{1}{2})_{\mathrm{2I}}] + [1 - \mathrm{HR}(0, \tfrac{1}{2})_{\mathrm{2I}}] \log[1 - \mathrm{HR}(0, \tfrac{1}{2})_{\mathrm{2I}}]; \tag{36}$$

[7] With a likelihood ratio decision axis, the only occasion where the distance is zero is when $X_{\mathscr{G}} = X_{\mathscr{N}}$, which results in an ROC curve that falls along the chance line. A pseudo-metric allows for a distance of zero even when $X_{\mathscr{G}} \neq X_{\mathscr{N}}$. This can occur when the evidence distributions result in an ROC curve that goes equally above and below the chance line (e.g., uniform distributions of equal mean and unequal variance).

and this maximum value of mutual information occurs if $P(\langle \mathscr{S}, \mathscr{N} \rangle) = P(\langle \mathscr{N}, \mathscr{S} \rangle)$. From Corollary 2.3,

$$\mathscr{C}_{2\mathrm{I}} = \log 2 + \mathscr{A}_{\mathrm{SI}} \log \mathscr{A}_{\mathrm{SI}} + (1 - \mathscr{A}_{\mathrm{SI}}) \log(1 - \mathscr{A}_{\mathrm{SI}}). \tag{37}$$

By the definition of $\mathscr{D}_{\mathrm{SI}}$, it follows that $\mathscr{C}_{2\mathrm{I}} = \mathscr{D}_{\mathrm{SI}}$. Therefore $\mathscr{D}_{\mathrm{SI}}$ is equal to the channel capacity of an unbiased 2IFC observer. ∎

2.2. Transformed Decision Axes

In both the SIFC and 2IFC tasks, the evidence variables (X in the SIFC task or X_{I} and X_{II} in the 2IFC task) can be transformed using some function, $g(X)$, before they are used to make a decision.

One case of interest is when the SIFC likelihood ratio is strictly monotonically increasing with X; that is, $g(X) = \ell(X)$. A second case relates difference-based decision axes to ratio-based decision axes when X is positive-valued. If $g(X) = \log(X)$, then the difference decision axis based on the transform $\varDelta_g = g(X_{\mathrm{I}}) - g(X_{\mathrm{II}}) = \log(X_{\mathrm{I}}/X_{\mathrm{II}})$ is a strictly monotonically increasing transform of a 2IFC decision axis based on the ratio $X_{\mathrm{I}}/X_{\mathrm{II}}$.

In the SIFC case, the ROC curve is not affected by strictly monotonically increasing transforms of X (Egan, 1975). Similarly, such transforms on the 2IFC decision axis do not affect the form of the 2IFC ROC curve. A 2IFC ROC curve derived from $g(X_{\mathrm{I}})$ and $g(X_{\mathrm{II}})$, however, may be different from the 2IFC ROC curve derived from X_{I} and X_{II}, even if $g(X)$ is strictly monotonically increasing with X. This is because $g(X_{\mathrm{I}}) - g(X_{\mathrm{II}})$ is not necessarily strictly monotonically increasing with $X_{\mathrm{I}} - X_{\mathrm{II}}$.

For all g, including the identity $g(X) = X$, the 2IFC ROC curves based on $\varDelta_g = g(X_{\mathrm{I}}) - g(X_{\mathrm{II}})$ are constrained to pass through the same point on the negative diagonal, namely the point associated with $P(\mathrm{C})_{2\mathrm{I}}$, where $\varDelta_g = 0$.

Proof. Since $g(X)$ is a strictly monotonically increasing transform defined over the domain of X_{I} and X_{II}, then $P(g(X_{\mathrm{I}}) > g(X_{\mathrm{II}})) = P(X_{\mathrm{I}} > X_{\mathrm{II}})$, $P(g(X_{\mathrm{I}}) = g(X_{\mathrm{II}})) = P(X_{\mathrm{I}} = X_{\mathrm{II}})$, and $P(g(X_{\mathrm{I}}) < g(X_{\mathrm{II}})) = P(X_{\mathrm{I}} < X_{\mathrm{II}})$. Hence the value of $P(\mathrm{C})_{2\mathrm{I}}$ in Eq. (31) can be also expressed as

$$\begin{aligned} P(\mathrm{C})_{2\mathrm{I}} = {} & P(g(X_{\mathscr{N}}) < g(X_{\mathscr{S}})) \\ & + P(g(X_{\mathscr{N}}) = g(X_{\mathscr{S}}))[\xi P(\langle \mathscr{S}, \mathscr{N} \rangle) + (1 - \xi) P(\langle \mathscr{N}, \mathscr{S} \rangle)]. \end{aligned} \tag{38}$$

From Corollary 2.3, $P(\mathrm{C})_{2\mathrm{I}}$ is associated with the point in the 2IFC ROC space where the 2IFC ROC curve crosses the negative diagonal. In the discrete case, the 2IFC hit rate in Eq. (15) for this point can be re-expressed as

$$\begin{aligned} \mathrm{HR}(0, \xi)_{2\mathrm{I}} = {} & P(X_{\mathrm{I}} > X_{\mathrm{II}} \mid \langle \mathscr{S}, \mathscr{N} \rangle) + \xi P(X_{\mathrm{I}} = X_{\mathrm{II}} \mid \langle \mathscr{S}, \mathscr{N} \rangle) \\ = {} & P(X_{\mathscr{N}} < X_{\mathscr{S}}) + \xi P(X_{\mathscr{N}} = X_{\mathscr{S}}) \\ = {} & P(g(X_{\mathscr{N}}) < g(X_{\mathscr{S}})) + \xi P(g(X_{\mathscr{N}}) = g(X_{\mathscr{S}})), \end{aligned} \tag{39}$$

which is the hit rate for $\Delta_g = 0$. Similarly, the false-alarm rate is

$$\begin{aligned}
\text{FAR}(0, \xi)_{2\text{I}} &= P(X_\text{I} > X_\text{II} \mid \langle \mathcal{N}, \mathcal{S} \rangle) + \xi P(X_\text{I} = X_\text{II} \mid \langle \mathcal{N}, \mathcal{S} \rangle) \\
&= P(X_\mathcal{N} > X_\mathcal{S}) + \xi P(X_\mathcal{N} = X_\mathcal{S}) \\
&= P(g(X_\mathcal{N}) > g(X_\mathcal{S})) + \xi P(g(X_\mathcal{S}) = g(X_\mathcal{N})).
\end{aligned} \tag{40}$$

Thus, the 2IFC ROC curve based on transformed evidence values $g(X_\text{I})$ and $g(X_\text{II})$ will pass through the same point on the negative diagonal for all such transforms $g(X)$ and give the same value of $P(\text{C})_{2\text{I}}$ in each case. The same result holds in the continuous case when $P(X_\text{I} = X_\text{II}) = 0$. ∎

2.3. Relationship of $\mathscr{A}_{\text{SI}}$ to $\mathscr{A}_{2\text{I}}$

It is possible to draw bounds on the area under the 2IFC ROC curve in terms of the area under the SIFC ROC curve. Here, two cases are considered: (a) the 2IFC ROC curve based on the 2IFC likelihood ratio decision axis, $\ell(\Delta)$, which constrains the slope of the 2IFC ROC curve to be monotonically decreasing; and (b) the general case of the 2IFC ROC curve based on the Δ decision axis. The proofs are based on a geometric argument and apply to both discrete and continuous evidence variables.

THEOREM 2.5. *The bounds on $\mathscr{A}_{2\text{I}}$ in terms of $\mathscr{A}_{\text{SI}}$ for the $\ell(\Delta)$ decision axis are $\mathscr{A}_{\text{SI}} \leqslant \mathscr{A}_{2\text{I}} \leqslant \mathscr{A}_T$, where $\mathscr{A}_T = 1 - 2(1 - \mathscr{A}_{\text{SI}})^2$.*

Like any ROC curve based on a likelihood ratio decision axis, the slope of the 2IFC ROC curve will be monotonically decreasing if the decision axis is monotonic with the 2IFC likelihood ratio decision axis $\ell(\Delta)$. For such decision axes, Theorem 2.5 indicates that the 2IFC observer performs at a level at least equal to that of the SIFC observer.

Proof. To prove Theorem 2.5, consider the 2IFC ROC curve shown in Fig. 2. The area of the shaded square is $\mathscr{A}_{\text{SI}}^2$, because, by Eq. (17) and Corollary 2.3, the hit rate at the point where the 2IFC ROC curve cuts the negative diagonal is $\mathscr{A}_{\text{SI}}$. Note that the area of each shaded triangle is $\frac{1}{2}(1 - \mathscr{A}_{\text{SI}}) \mathscr{A}_{\text{SI}}$. Thus, the area of the region enclosed by the square and the two triangles is $\mathscr{A}_{\text{SI}}^2 + 2[\frac{1}{2}(1 - \mathscr{A}_{\text{SI}}) \mathscr{A}_{\text{SI}}] = \mathscr{A}_{\text{SI}}$. If the slope of the 2IFC ROC curve is monotonically decreasing, then the curve must lie on or outside the region bounded by the square and two triangles. Consequently, $\mathscr{A}_{\text{SI}} \leqslant \mathscr{A}_{2\text{I}}$.

To derive the upper bound on $\mathscr{A}_{2\text{I}}$, consider a line of unit slope drawn through the point where the 2IFC ROC curve cuts the negative diagonal (see the dot–dashed line in Fig. 2). If the ROC curve rises above the line on one side of the negative diagonal, then, by symmetry, it must rise above the line on the other side of the negative diagonal. However, the slope of such an ROC curve cannot be monotonically decreasing, because the ROC curve is constrained to pass through the point where the dot–dashed line cuts the negative diagonal. Hence, the ROC curve can never rise above the dot–dashed line. From Fig. 2, the area of the triangle, defined by the upper left corner of the ROC space and the two points
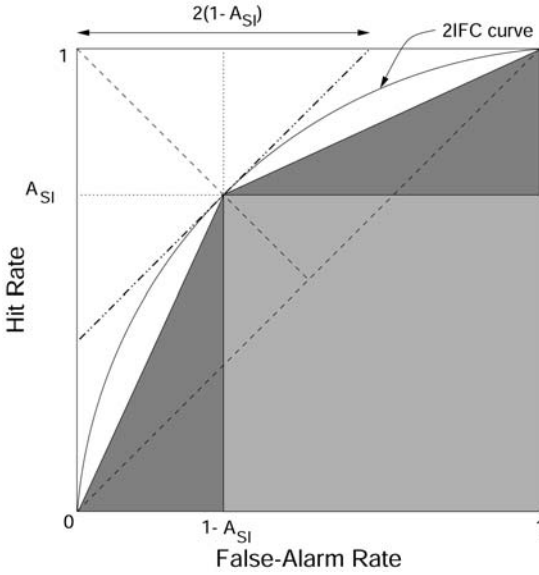
FIG. 2. An illustrative 2IFC ROC curve, indicating the relationship of $\mathscr{A}_{\mathrm{SI}}$ to $\mathscr{A}_{\mathrm{2I}}$.

where the dot–dashed line cuts the boundary of the ROC space, is $2(1-\mathscr{A}_{\mathrm{SI}})^2$. It follows that $\mathscr{A}_{\mathrm{2I}} \leqslant 1 - 2(1-\mathscr{A}_{\mathrm{SI}})^2$. ∎

COROLLARY 2.4 (to Theorem 2.5). $\quad \mathscr{D}_{\mathrm{SI}} \leqslant \mathscr{D}_{\mathrm{2I}} \leqslant \mathscr{D}_{T}$.

The upper bound in Corollary 2.4 is defined as $\mathscr{D}_{T} = \log 2 - \mathscr{H}_{T}$, where $\mathscr{H}_{T} = -T \log T - (1-T) \log(1-T)$ and $T = 2(1-\mathscr{A}_{\mathrm{SI}})^2$. Corollary 2.4 indicates that if the 2IFC decision axis is monotonic with $\ell(\varDelta)$, then the distance between $X_{\mathscr{N}}$ and $X_{\mathscr{G}}$ can be no greater than the distance between $\varDelta_{\langle \mathscr{N},\mathscr{G} \rangle}$ and $\varDelta_{\langle \mathscr{G},\mathscr{N} \rangle}$.

Proof. To prove Corollary 2.4, note that $\mathscr{D}$ is a strictly monotonically increasing transform of $\mathscr{A}$, providing $\mathscr{A}$ is greater than or equal to 0.5. If the slope of the 2IFC ROC curve is constrained to be monotonically decreasing, then $\mathscr{A}_{\mathrm{2I}}$ can never be less than 0.5. Because $\mathscr{A}_{\mathrm{SI}}$ is the point where the 2IFC ROC curve crosses the negative diagonal, it too can never be less than 0.5. Therefore, because $\mathscr{D}$ is a strictly monotonically increasing transform, the relationship among $\mathscr{D}_{\mathrm{SI}}$, $\mathscr{D}_{\mathrm{2I}}$, and $\mathscr{D}_{T}$ must be the same as that among $\mathscr{A}_{\mathrm{SI}}$, $\mathscr{A}_{\mathrm{2I}}$, and $\mathscr{A}_{\mathrm{T}}$. Corollary 2.4 follows immediately from these observations and Theorem 2.5. ∎

THEOREM 2.6. *Bounds on $\mathscr{A}_{\mathrm{2I}}$ in terms of $\mathscr{A}_{\mathrm{SI}}$ for the $\varDelta$ decision axis are* $\mathscr{A}_{\mathrm{SI}}^2 \leqslant \mathscr{A}_{\mathrm{2I}} \leqslant \mathscr{A}_{\mathrm{SI}}(2 - \mathscr{A}_{\mathrm{SI}})$.

For the $\varDelta$ decision axis, $\mathscr{A}_{\mathrm{2I}}$ may be less than $\mathscr{A}_{\mathrm{SI}}$ and may also result in performance below the chance line, even if SIFC performance was above the chance line.

Proof. If the 2IFC ROC curve is based on the $\varDelta$ decision axis, rather than on $\ell(\varDelta)$, then its slope may not be monotonically decreasing and it may rise above the dot–dashed tangent or fall below the chance line. The curve, however, is still constrained to pass through the same point on the negative diagonal. Therefore the minimum $\mathscr{A}_{\mathrm{2I}}$ is simply the area of the shaded square, $\mathscr{A}_{\mathrm{SI}}^2$, from Fig. 2. The

maximum $\mathscr{A}_{2I}$ is the area of the region described by the shaded square and the two rectangles formed by the horizontal and vertical lines at $HR = \mathscr{A}_{SI}$ and $FAR = (1 - \mathscr{A}_{SI})$, respectively, from Fig. 2. This area is $\mathscr{A}_{SI}^2 + 2\mathscr{A}_{SI}(1 - \mathscr{A}_{SI})$ which, after rearrangement, is equal to $\mathscr{A}_{SI}(2 - \mathscr{A}_{SI})$. Therefore, $\mathscr{A}_{SI}^2 \leqslant \mathscr{A}_{2I} \leqslant \mathscr{A}_{SI}(2 - \mathscr{A}_{SI})$. ■

2.4. Determining 2IFC ROC Curves

It is possible to derive or estimate a 2IFC ROC curve, given information about the associated SIFC task. If the SIFC density or mass functions on X are known, then the 2IFC distributions and ROC curve may be derived by cross-correlating the SIFC distributions using Eqs. (9)–(12) as appropriate. Similarly, if the SIFC ROC curve is empirical, and the observer ratings that were used to generate the curve are known, then the distribution of ratings may be considered to be discrete probability mass functions. From these probability mass functions, a 2IFC ROC curve is estimated by taking the cross-correlation as above. This method was used in the following experiments to see how predictable 2IFC performance was from SIFC performance. It works best for a high-resolution rating scale, otherwise only a few points on the 2IFC ROC curve are estimated. It is assumed that the ratings are monotonic with the decision axis. If the density or mass functions of X are unknown (either theoretical or empirical), then a 2IFC ROC curve can be determined from the SIFC ROC curve itself, as follows.

THEOREM 2.7. *A 2IFC ROC curve can be generated from the SIFC hit rates and false-alarm rates.*

Proof. A generalization of Corollary 2.3 can be used to generate a 2IFC ROC curve from the corresponding SIFC ROC curve. From the SIFC hit rates and false-alarm rates, for each value of c, a curve is plotted with $FAR(c)$ as the abscissa versus $HR(c+k)$ as the ordinate, where k is some constant. If X is continuous, the area under this curve, $A(k)$, is

$$A(k) = \int_0^1 HR(c+k)\, d(FAR(c))$$

$$= \int_{-\infty}^{\infty} \left[\int_{c+k}^{\infty} f_X(x \mid \mathscr{S})\, dx \right] f_X(c \mid \mathscr{N})\, dc.$$

Then with a change of variable, $x = c + \delta$, and reversing the order of integration,

$$A(k) = \int_k^{\infty} \int_{-\infty}^{\infty} f_X(c+\delta \mid \mathscr{S})\, f_X(c \mid \mathscr{N})\, dc\, d\delta$$

$$= \int_k^{\infty} f_A(\delta \mid \langle \mathscr{S}, \mathscr{N} \rangle)\, d\delta \qquad \text{(by Eq. (9))}$$

$$= HR(k)_{2I}.$$

For discrete X, the area under the curve between $FAR(c, 0)$ and $FAR(c-1, 0)$, is

$$a(c, k) = \tfrac{1}{2} \left[HR(c+k, 0) + HR(c+k-1, 0) \right] \left[FAR(c-1, 0) - FAR(c, 0) \right]$$

$$= \tfrac{1}{2} \left[\sum_{x=c+k+1}^{\infty} P(X_{\mathscr{S}} = x) + \sum_{x=c+k}^{\infty} P(X_{\mathscr{S}} = x) \right]$$

$$\times \left[\sum_{x=c}^{\infty} P(X_{\mathscr{N}} = x) - \sum_{x=c+1}^{\infty} P(X_{\mathscr{N}} = x) \right]$$

$$= \tfrac{1}{2} \left[P(X_{\mathscr{S}} = c+k) + 2 \sum_{x=c+k+1}^{\infty} P(X_{\mathscr{S}} = x) \right] P(X_{\mathscr{N}} = c)$$

$$= \tfrac{1}{2} P(X_{\mathscr{S}} = c+k) \, P(X_{\mathscr{N}} = c) + \sum_{x=c+k+1}^{\infty} P(X_{\mathscr{S}} = x) \, P(X_{\mathscr{N}} = c).$$

Summing over all $a(c, k)$ gives

$$A(k) = \sum_{c=-\infty}^{\infty} \sum_{x=c+k+1}^{\infty} P(X_{\mathscr{S}} = x) \, P(X_{\mathscr{N}} = c) + \tfrac{1}{2} \sum_{c=-\infty}^{\infty} P(X_{\mathscr{S}} = c+k) \, P(X_{\mathscr{N}} = c).$$

Then with a change of variable, $x = c+\delta$, and reversing the order of summation,

$$A(k) = \sum_{\delta=k+1}^{\infty} \sum_{c=-\infty}^{\infty} P(X_{\mathscr{S}} = c+\delta) \, P(X_{\mathscr{N}} = c)$$

$$+ \tfrac{1}{2} \sum_{c=-\infty}^{\infty} P(X_{\mathscr{S}} = c+k) \, P(X_{\mathscr{N}} = c)$$

$$= \sum_{\delta=k+1}^{\infty} P(\varDelta_{\langle \mathscr{S}, \mathscr{N} \rangle} = \delta) + \tfrac{1}{2} P(\varDelta_{\langle \mathscr{S}, \mathscr{N} \rangle} = k) \qquad \text{(by Eq. (11))}$$

$$= HR(k, \tfrac{1}{2})_{2I}.$$

Because the 2IFC ROC curve is symmetric about the negative diagonal, the corresponding 2IFC false-alarm rate can be obtained from $FAR(-k)_{2I} = 1 - HR(k)_{2I}$. By varying k between $-\infty$ and ∞ the entire 2IFC ROC curve can be generated. ∎

For the discrete case, the resulting 2IFC ROC curve is an approximation, because only the midpoints of each line segment (where $\xi = 1/2$), rather than the vertices, are defined. For empirical SIFC ROC curves with many points, the resulting 2IFC ROC curve and $\mathscr{A}_{2I}$ will only be slightly underestimated.

A crucial assumption in this proof is that both the SIFC task and the 2IFC task are based on the same random variable X. If a strictly monotonically increasing transform of X rather than X itself is used to derive a 2IFC ROC curve, then the derived curve may differ from the 2IFC curve based on X. The family of 2IFC ROC curves that result from any strictly monotonically increasing transform, however, is constrained in the ROC space, as described by Theorem 2.6 (or by Theorem 2.5 for the special case where $\varDelta$ is monotonic with $\ell(\varDelta)$).

Green and Swets (1966, pp. 48–49) provided a similar method to estimate specific points on the 2IFC ROC curve. Their method makes the assumption that the SIFC decision axis is $\ell(X)$, that $\ell(X)$ is continuous, and that the 2IFC decision rule is based on ratios rather than differences. They did not show how to estimate the

entire ROC curve, however, because the purpose of their proof was to prove 2IFC ROC curve symmetry.

As a final point, note that it is not possible to uniquely derive SIFC ROC curves from 2IFC ROC curves because information is lost with the integrations and summations involved in calculating cross-correlation functions.

3. EXPERIMENTS

Two experiments were conducted to investigate the nonparametric relationships between the SIFC and 2IFC tasks developed here. There are few reported experiments using discrete $X_{\mathscr{G}}$ and $X_{\mathscr{N}}$, (Boven, 1976; Podd, 1982; Taylor, 1984; Taylor et al., 1991), and there are none that specifically compare the SIFC and 2IFC tasks. Therefore, the first experiment used a frequency discrimination task where the evidence for the two events was distributed as overlapping discrete uniform distributions. The observers' performances could then be compared with a known theoretical model. The second experiment was the task of narrow-band Gaussian noise amplitude discrimination. The distributions of $X_{\mathscr{G}}$ and $X_{\mathscr{N}}$ were not known but were assumed to be continuous. The designs for both experiments included ways to minimize stimulus sampling variability and to remove observer inconsistency.

Observer inconsistency is the biggest problem facing experimental psychophysics. As Green and Luce (1974, p. 373) stated: "Perhaps the single most pervasive characteristic of psychophysical data is the inconsistency of subjects when answering most questions we ask them about simple stimuli." One way of reducing the effects of inconsistency is to repeat an experiment multiple times with reproducible stimuli and then to use group operating characteristic (GOC) analysis (Drga, 1999; Metz & Shen, 1992; Taylor et al., 1991; Watson, 1962).

Observer inconsistency has been characterized by using the concepts of internal and external noise, where the emphasis is on the sources or causes of noise, or error, in experimental tasks (e.g., Green, 1964a; Pfafflin & Mathews, 1962; Raab & Goldberg, 1975). An alternative approach is to use the concepts of *unique* and *common* noise, which emphasize the statistical components or effects of error rather than the sources of error (Drga, 1999; Boven, 1976; Taylor, 1984; Taylor et al., 1991).

Unique noise is due to random perturbations in the observer, and in its environment, on each presentation of the same stimulus. Common noise is the result of instances of the events that are the same across repeated presentations of the stimuli. For any single experiment, the effects of common and unique noise are inseparable, but when more than one replication of an experiment is run ratings can be averaged across replications for the same stimulus, for each event. Averaging removes some of the unique noise while retaining common noise. This process is the basis of GOC analysis.

GOC analysis produces a GOC curve, which is an ROC curve based on group data (either within or between observers or both). The mean rating for each stimulus, averaged across replications, is calculated for all stimuli. The mean rating per stimulus forms the basis of a GOC curve in the same way that the rating made for each stimulus in a single replication forms the basis of an ROC curve. Measures

of detectability may be obtained from the GOC curve in the same way they are from an ROC curve. A useful comparison to the GOC curve is the mean ROC (mROC) curve, which gives the average, noise-degraded, ROC curve (Drga, 1999; Lapsley Miller, 1999; MacMillan & Kaplan, 1985). The mROC curve summarizes the multiple ROC curves from a multiple replication experiment. It is calculated by averaging the hit rates and false-alarm rates for each possible criterion, across replications, where the averaging may be of any form (such as arithmetic, z-transform, or arcsine-transform averaging).

If only a subset of the possible stimuli is used in an experiment (which is the usual case), then the common noise is affected by sampling variability, and the observer's performance will not necessarily reflect the underlying population statistics, even if unique noise is removed. One way of estimating the nature and extent of sampling variability is to estimate the variability of a measure of detectability associated with the ROC curve based on a sample of stimuli. This can be done statistically (Bamber, 1975) and by computer simulation (Lapsley Miller, 1999; Pollack & Hsieh, 1969). Sampling variability is only slightly dependent on the form of the evidence distributions but is strongly dependent on the number of samples.

The effects of both unique noise and sampling variability make it difficult to interpret psychophysical results, therefore both forms of error should be reduced as much as possible. The problem of minimizing both unique noise and sampling variability becomes a tradeoff between the number of stimuli per replication needed to minimize the sampling variability and the number of replications needed in order to minimize unique noise variability. Given a fixed number of trials, one form of error cannot be minimized without increasing the other.

3.1. Experiment I: Discrete $X_{\mathscr{S}}$ and $X_{\mathscr{N}}$

This experiment consisted of three conditions: (a) an SIFC task with a continuous rating scale, (b) a 2IFC task with a continuous rating scale (2IFC$_c$), and (c) a 2IFC task with a binary rating scale (2IFC$_b$), which was included for comparison with the 2IFC$_c$ condition, because there are only a few instances of 2IFC rating scales in the literature (Friedman & Carterette, 1964; Leshowitz, 1969; Markowitz & Swets, 1967; Schulman & Mitchell, 1966).

The SIFC task was to detect if the $\mathscr{S}$ event occurred in the observation interval. The evidence for each event had a discrete uniform distribution (Fig. 3). These distributions were realized as sets of sinusoidal transients, with center frequencies ranging in steps of 5 Hz from 595 to 640 Hz for the $\mathscr{N}$ event and from 620 to 665 Hz for the $\mathscr{S}$ event. The overlap of five frequencies meant the events were confusable, even for an ideal observer. The frequency separation of the transients, however, meant that they could easily be discriminated from one another. The 2IFC task was to detect in which observation interval the $\mathscr{S}$ event occurred. The evidence for the 2IFC event orderings $\langle \mathscr{S}, \mathscr{N} \rangle$ and $\langle \mathscr{N}, \mathscr{S} \rangle$ was modeled as the difference of the SIFC evidence distributions, $X_{\mathscr{S}} - X_{\mathscr{N}}$ and $X_{\mathscr{N}} - X_{\mathscr{S}}$ (Fig. 3). These finitely discrete distributions were designed so the entire stimulus set could be presented in both the SIFC and 2IFC conditions, thus there was no sampling variability.
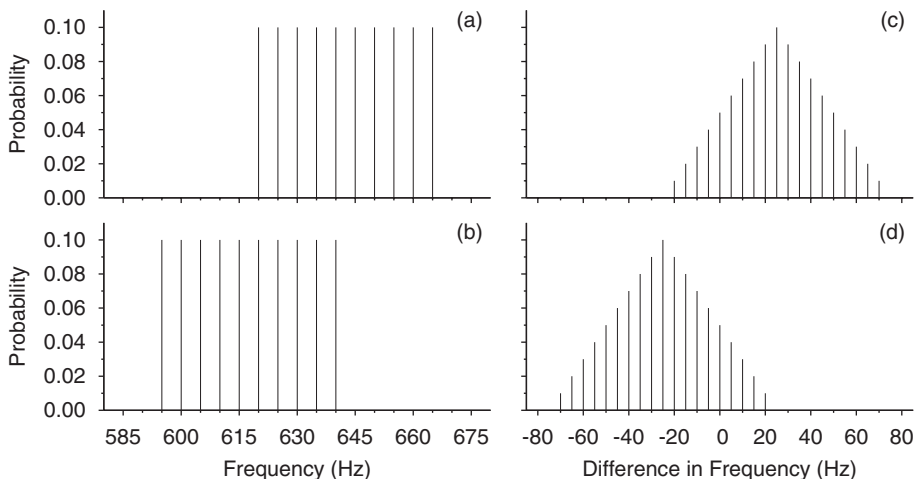
FIG. 3. The theoretical evidence distributions in Experiment I. The SIFC events (a) $\mathscr{S}$ and (b) $\mathscr{N}$ and the 2IFC event orderings (c) $\langle\mathscr{S},\mathscr{N}\rangle$ and (d) $\langle\mathscr{N},\mathscr{S}\rangle$, assuming a difference decision rule, are plotted.

Method

Observers. Four of the authors were observers. All had previous experience in SIFC and 2IFC psychoacoustic experiments.

Stimuli and equipment. The 15 sinusoidal transients were digitally generated, then shaped with a Kaiser data window (shaping parameter $\beta = 11$) over the first and last 15 ms (Kaiser, 1966; Rabiner & Gold, 1975). This produced an absolute duration of 100 ms and an equivalent rectangular duration (ERD) of 81.2 ms.

The diotic transients were output from a digital-to-analog converter (DAC) at 7.2 kHz, low-pass filtered (1.25 kHz), and mixed with a continuous, analog, Gaussian-noise masker (4 kHz, low-pass). During the observation interval, the transients were presented to the observer through a purpose-built headset amplifier to TDH-39 100-Ω earphones mounted in Rudmose Tracor RA125 Otocups and MX41/AR cushions. The timing was controlled by a Hewlett Packard 9825A real-time clock and the experimental system was controlled by Hewlett Packard 9825 and 9826 computers.

The continuous background masker spectrum level was 40 dB SPL. The signal level was 53.2 dB SPL (13.2 dB SNR) for Observer 1 and Observer 2 and 55 dB SPL (15.0 dB SNR) for Observer 3 and Observer 4. These settings made the unique-noise degraded performance for all observers approximately the same. The same theoretical models were still appropriate for all observers, because the task was not dependent on overall SPL.

Procedure. All observers completed 16 replications of each condition. Each replication was run in two, approximately 10 minute, sessions. In the SIFC condition, the 10 transients in each event were presented 20 times giving 400 trials per replication. In the two 2IFC conditions, all possible pairings of the $X_{\mathscr{S}}$

transients and $X_{\mathcal{N}}$ transients were presented twice, also giving 400 trials per replication. Each presentation of the same transient within a replication was treated as a unique stimulus to increase resolution in the ROC space. Replications in the two 2IFC conditions were alternated two at a time, then the SIFC replications were completed. The trial order was randomized for each replication, with event run limits randomly constrained to a maximum of four, five, or six. The durations of the warning (50 ms), observation (100 ms), decision (1000 ms), and reset (minimum of 750 ms) intervals were indicated by LED lights. There was a 500-ms interstimulus interval in the 2IFC task.

The observers ran sessions in a sound-attenuating chamber. They indicated their decision on each trial by either moving a slider, for continuous rating tasks, or pressing one of two buttons, for binary rating tasks. The slider was moved from the left-hand end of a 12-cm slot. At the end of the trial the observer was required to reset the slider to the left before the next trial would begin. In the SIFC condition, the observers indicated a decision by moving the slider to the right-hand side of the scale if they were confident that the $\mathcal{S}$ event occurred, and to the left-hand side of the scale if they were confident that the $\mathcal{N}$ event occurred. In the 2IFC conditions, the observers used the right-hand side of the slider (or right button) to indicate confidence that the $\langle \mathcal{S}, \mathcal{N} \rangle$ event ordering occurred and the left-hand side of the scale (or left button) to indicate that the $\langle \mathcal{N}, \mathcal{S} \rangle$ event ordering occurred. If an observer failed to respond during the decision interval in the binary rating condition an incorrect decision was recorded.
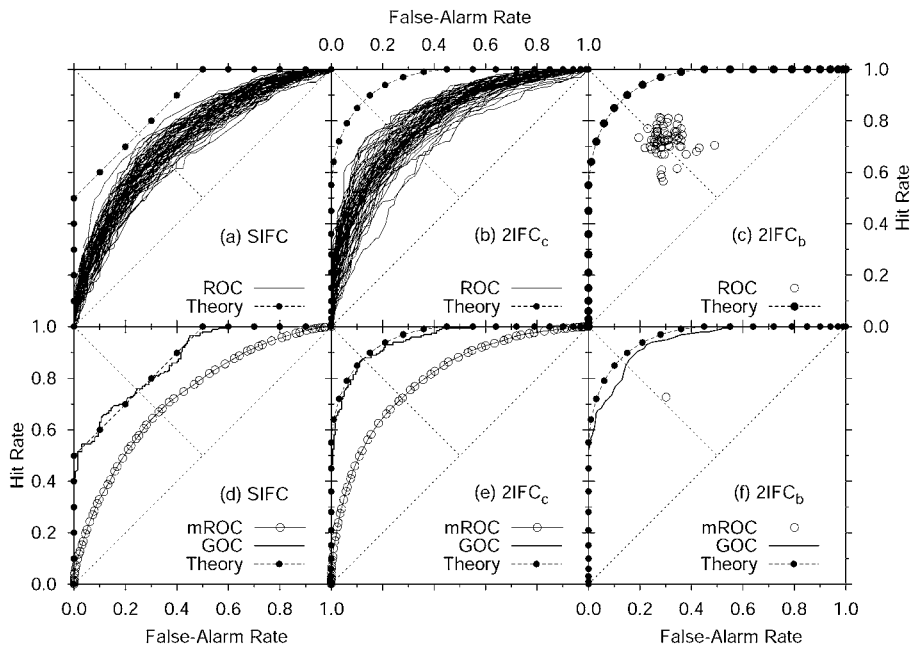

FIG. 4. ROC, mROC, and GOC curves for all observers in Experiment I, compared with the known theoretical ROC curves. The 64 ROC curves for the (a) SIFC, (b) 2IFC$_c$, and (c) 2IFC$_b$ conditions are shown in the top three panels. The GOC curves and the mROC curves for the (d) SIFC, (e) 2IFC$_c$, and (f) 2IFC$_b$ conditions are shown in the bottom three panels.

Results and discussion. The continuous rating scale readings were converted into 64 equally spaced categories and ROC, mROC (using arcsine-transform averaging), and GOC analyses were performed for each observer and over all observers, for each condition.

Figure 4(a–c) shows the 64 ROC curves (or the 64 ROC points for the binary condition) for all observers in each condition. The amount of variability and the fact that all the ROC curves are below the theoretical ROC curves indicate that observer inconsistency has degraded performance. It is obvious that with this degree of variability single-replication data could not be used to assess Green's relationship. Figure 4(d–f) shows the 64-replication GOC and mROC curves taken over all observers. Both SIFC and 2IFC mROC curves are consistently below the theoretical ROC curves and seem typical of psychophysical data—it looks like they would be fitted well by a normal model. The GOC curves, on the other hand, tend toward the theoretical ROC curves. In comparing the two 2IFC conditions, both the $2IFC_c$ mROC curve and $2IFC_b$ mROC point are consistently below the theoretical ROC curve. Both of the 2IFC GOC curves approach the theoretical ROC curve, but only the $2IFC_c$ GOC curve (Fig. 4b) is a good fit to the theoretical ROC curve after 64 replications.

Figure 5 is of particular interest for it shows the GOC, mROC, and theoretical ROC curves in z-coordinates. Consistent with Hanley and McNeil's (1982) observation that most psychophysical data would be well fitted by a normal model (regardless of the appropriateness of this assumption), the mROC curves look linear with slope near unity. The GOC curves are, however, curvilinear and are good approximations to the theoretical ROC curves.

To assess Green's relationship, the difference between $\mathscr{A}_{SI}$ and $P(C)_{2I}$ was considered. $\mathscr{A}_{SI}$ was calculated using the trapezoidal rule. $P(C)_{2I}$ was estimated from the intersection of the ROC, mROC, or GOC curve with the negative diagonal. In all cases, $\mathscr{A}_{SI}$ and $P(C)_{2I}$ are larger for GOC curves than for the mROC curves (see Table 1). The variability of these measures of detectability for the ROC curves (shown in standard deviations in Table 1 and graphically in Fig. 4) indicates the degree of observer inconsistency. GOC analysis, however, removes most of this
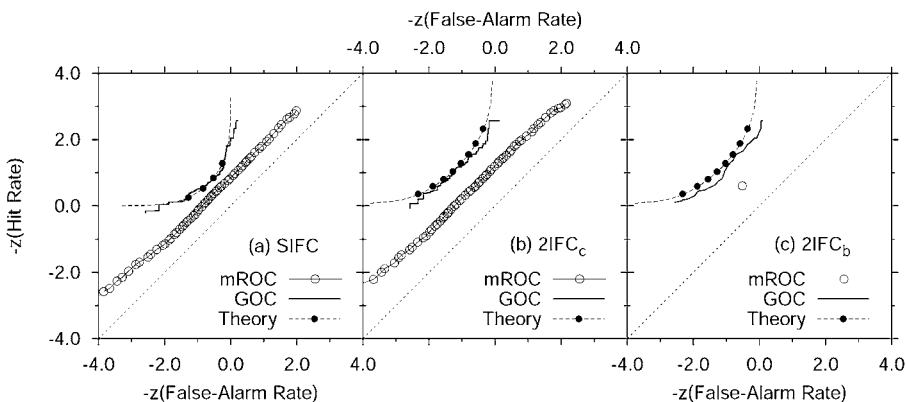

FIG. 5. The GOC curve and the mROC curve, in z-coordinates, for all observers, and the theoretical ROC curves in Experiment I are given for: (a) the SIFC condition, (b) the $2IFC_c$ condition, and (c) the $2IFC_b$ condition.

TABLE 1

The Difference between $\mathscr{A}_{SI}$ and $P(C)_{2I}$ (from the 2IFC$_c$ Condition), after GOC, mROC, and ROC Analysis (Mean and Standard Deviations of $\mathscr{A}_{SI}$ and $P(C)_{2I}$), over All Replications, Compared with the Theoretical Model in Experiment I

Observer	Analysis	$\mathscr{A}_{SI}$	$P(C)_{2I}$	Difference
1	GOC	0.8547	0.8650	−0.0103
	mROC	0.7096	0.7076	0.0020
	ROC M (SD)	0.7087 (0.0263)	0.7088 (0.0356)	−0.0001
2	GOC	0.8636	0.8550	0.0086
	mROC	0.7414	0.7474	−0.0060
	ROC M (SD)	0.7401 (0.0203)	0.7443 (0.0255)	−0.0042
3	GOC	0.8560	0.8600	−0.0040
	mROC	0.7176	0.7093	0.0083
	ROC M (SD)	0.7159 (0.0253)	0.7046 (0.0150)	0.0113
4	GOC	0.8597	0.8100	0.0497
	mROC	0.7517	0.7337	0.0180
	ROC M (SD)	0.7504 (0.0242)	0.7364 (0.0202)	0.0140
All	GOC	0.8720	0.8750	−0.0030
	mROC	0.7283	0.7233	0.0050
	ROC M (SD)	0.7288 (0.0295)	0.7235 (0.0305)	0.0053
Theory		0.8750	0.8750	0.0000

inconsistency, producing measures of detectability that are good estimates of the theoretical measures of detectability.

The difference between $\mathscr{A}_{SI}$ and $P(C)_{2I}$ is small for all observers except Observer 4 (see Table 1). Observer 4's performance is depressed in the 2IFC task, possibly indicating that more replications need to be run to remove unique noise. Even if the mROC measures indicate that $\mathscr{A}_{SI} = P(C)_{2I}$, this should not be taken as evidence that the relationship holds empirically, for it is confounded by unique noise. The results of the GOC analysis tend to the theoretical prediction that $\mathscr{A}_{SI} = P(C)_{2I}$. This

TABLE 2

$\mathscr{D}_{SI}$, $\mathscr{C}_{2I}$, and Their Difference (bits), after GOC and mROC Analysis, Compared with the Theoretical Model in Experiment I

	GOC			mROC		
Observer	$\mathscr{D}_{SI}$	$\mathscr{C}_{2I}$	Difference	$\mathscr{D}_{SI}$	$\mathscr{C}_{2I}$	Difference
1	0.4020	0.4290	−0.0270	0.1308	0.1282	0.0026
2	0.4253	0.4028	0.0225	0.1754	0.1846	−0.0093
3	0.4054	0.4158	−0.0104	0.1413	0.1304	0.0109
4	0.4150	0.2985	0.1164	0.1914	0.1639	0.0275
All	0.4481	0.4564	−0.0084	0.1561	0.1491	0.0070
Theory	0.4564	0.4564	0.0000	0.4564	0.4564	0.0000

TABLE 3

Relationships $\mathscr{A}_{SI} \leqslant \mathscr{A}_{2I} \leqslant \mathscr{A}_T$ and $\mathscr{D}_{SI} \leqslant \mathscr{D}_{2I} \leqslant \mathscr{D}_T$ after GOC Analysis in Experiment I

Observer	$\mathscr{A}_{SI}$	$\leqslant$	$\mathscr{A}_{2I}$	$\leqslant$	$\mathscr{A}_T$	$\mathscr{D}_{SI}$	$\leqslant$	$\mathscr{D}_{2I}$	$\leqslant$	$\mathscr{D}_T$
1	0.8547	$\leqslant$	0.9316	$\leqslant$	0.9578	0.4020	$\leqslant$	0.6401	$\leqslant$	0.7476
2	0.8636	$\leqslant$	0.9393	$\leqslant$	0.9628	0.4253	$\leqslant$	0.6698	$\leqslant$	0.7706
3	0.8560	$\leqslant$	0.9450	$\leqslant$	0.9585	0.4054	$\leqslant$	0.6927	$\leqslant$	0.7510
4	0.8597	$\leqslant$	0.9203	$\leqslant$	0.9606	0.4150	$\leqslant$	0.5989	$\leqslant$	0.7606
All	0.8720	$\leqslant$	0.9483	$\leqslant$	0.9672	0.4481	$\leqslant$	0.7064	$\leqslant$	0.7919
Theory	0.8750	$\leqslant$	0.9588	$\leqslant$	0.9688	0.4564	$\leqslant$	0.7520	$\leqslant$	0.7994

gives strong support to the theoretical relationship of $\mathscr{A}_{SI} = P(C)_{2I}$ for this finitely discrete random variable. The fact that mROC analysis also gives empirical support may indicate that the unique-noise processes are similar in each task.

The relationship $\mathscr{D}_{SI} = \mathscr{C}_{2I}$ is also empirically supported, except for Observer 4 (Table 2). These measures were calculated from the empirical $\mathscr{A}_{SI}$ and $P(C)_{2I}$, using Eqs. (34) and (36) respectively. The discrepancy in Observer 4's results is highlighted using $\mathscr{D}_{SI}$, because the difference between $\mathscr{D}_{SI}$ and $\mathscr{C}_{2I}$ is an order of magnitude larger than for the other three observers.

Theorem 2.5 and Corollary 2.4 provide bounds for $\mathscr{A}_{2I}$ in terms of $\mathscr{A}_{SI}$ and for $\mathscr{D}_{2I}$ in terms of $\mathscr{D}_{SI}$, respectively. Table 3 shows that in all cases, after GOC analysis, the experimental results are consistent with these theoretical predictions.

A predicted 2IFC ROC (pROC) curve was estimated by tallying the SIFC sum-of-ratings for each event, then taking the discrete cross correlation of the sum-of-ratings distributions for the two events, for each 2IFC event ordering,
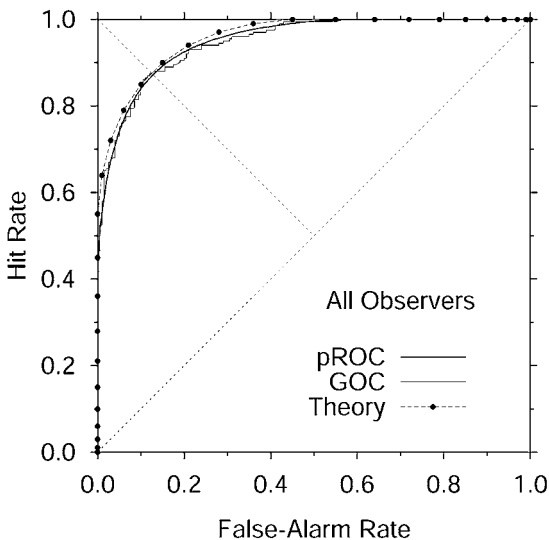

FIG. 6. A comparison of the predicted 2IFC ROC (pROC) curve (derived from the SIFC GOC sum-of-ratings), the 2IFC_c condition GOC curve, and the theoretical 2IFC ROC curve, for all observers in Experiment I.

using Eqs. (11) and (12). This resulted in a probability mass value for each value of δ for each 2IFC event ordering, where Δ is the decision axis for the difference of the sum-of-ratings. The mass functions were then cumulated in the usual way to obtain the 2IFC hit rates and false-alarm rates. Figure 6 shows the 2IFC pROC curve, the obtained 2IFC GOC curve, and the theoretical 2IFC ROC curve. The similarities among the three curves are striking, once observer inconsistency is removed.

3.2. Experiment II: Continuous $X_{\mathscr{S}}$ and $X_{\mathscr{N}}$

Experiment I indicated good support for the theoretical relationships developed here for discrete random variables. It is important to also establish whether the theoretical relationships hold for a task such as amplitude discrimination where the evidence is assumed to be distributed continuously. In Experiment II, observers were asked to detect narrow-band short-duration Gaussian noise with a bandwidth-duration product of unity, masked by wide-band Gaussian noise. Three ideal observers have been suggested for this signal-known-statistically observer: the energy detector (Green & McGill, 1970), the envelope detector (Drga, 1988; Whitmore *et al*., 1968), and the full-linear detector (Lapsley Miller, 1999). The theoretical evidence distributions are unknown but are suggested to be χ^2 (energy) or Rayleigh (envelope) for the SIFC task, although Lapsley Miller (1999) found evidence that neither distribution was appropriate. Both a difference and a ratio decision rule have been used to model the 2IFC case (Drga, 1988; Green, 1960a; Green & McGill, 1970). The inability to draw firmer conclusions about the nature of amplitude discrimination is partly due to the similar predictions of competing models and partly due to the problem of observer inconsistency in experimental data.

A computer simulation was used to estimate the minimum number of stimuli required, by considering the sampling variability of $\mathscr{A}_{SI}$ for ROC curves from Rayleigh distributions. Based on this simulation, 250 $\mathscr{S}$ and 250 $\mathscr{N}$ stimuli were chosen. Ideally, more stimuli would have been used, but there was a tradeoff between the number of stimuli and the number of replications it was feasible to run.

Each observer was allocated a unique set of stimuli for use in both the SIFC and the 2IFC conditions. In the SIFC condition, observers listened to each stimulus once in each replication. In the 2IFC conditions, it was practically impossible to present every pairing of $\mathscr{S}$ and $\mathscr{N}$ stimuli because there were 250^2 combinations of 250 stimulus pairs per event ordering. Here, the $\langle \mathscr{S}, \mathscr{N} \rangle$ stimulus pairs were assigned by randomly pairing one $\mathscr{S}$ transient with one $\mathscr{N}$ transient, without replacement. The $\langle \mathscr{N}, \mathscr{S} \rangle$ stimulus pairs were assigned by swapping the order of the $\langle \mathscr{S}, \mathscr{N} \rangle$ stimulus pairs, thus each transient was presented twice in each replication. Using this sampling method, only one possible set of 250 stimulus pairs was sampled per event ordering. To assess the effect of stimulus-pair sampling, the 2IFC condition was repeated three times; each block using the same $\mathscr{S}$ and $\mathscr{N}$ stimuli but different random stimulus pairings. The three 2IFC blocks were analyzed separately (as 250 stimulus pairs per event), then pooled (as 750 stimulus pairs per event) to decrease variability due to stimulus-pair sampling.

Method

Observers.   The four observers from Experiment I participated in Experiment II.

Stimuli.   The digitally generated, narrow-band, Gaussian transients were generated using an inverse fast Fourier transform (IFFT) method (Drga, 1999). This method is better than a digital filter approach because it produces reproducible noise transients with steep roll-offs in the frequency domain and minimal leakage in the time domain.

The signal-alone transients were generated using 250 2^{15}-point IFFTs with an input bandwidth of 35 Hz centered on 250 Hz. Four 276-point transients (one for each observer) were randomly selected (without overlap) from each IFFT. 500 maskers (noise-alone transients) were generated similarly, except that the input bandwidth was 1500 Hz (low-pass). The distributions of these unwindowed transients were consistent with a Gaussian distribution out to 2.5 SD for the signal-alone transients and to 3 SD for the maskers.

The 20-ms unmixed signal-alone transients and maskers were windowed over their entire duration with a Kaiser data window ($\beta = 9$), giving an ERD of 8.2 ms (Kaiser, 1966; Rabiner & Gold, 1975). The ERB of the windowed signals was 92 Hz and the ERB of the windowed maskers was 1500 Hz. Thus the equivalent bandwidth–duration product for the signal-alone transients was approximately 0.75. This corresponds to an essential bandwidth–duration product of approximately unity (Lapsley Miller, 1999, 2001).

To make the signal-plus-noise transients, the signal-alone and masker transients were mixed to give 7.5 dB SNR. The masker spectrum level was 69 dB SPL. During an observation interval, a randomly selected digital transient was output to the 12-bit DAC at 13.8 kHz, low-pass filtered (3 kHz), then mixed with a continuous background Gaussian-noise masker (4 kHz, low-pass, 33 dB SPL spectrum level).

Procedure.   Each observer ran 32 SIFC replications and 16 2IFC replications (for each 2IFC block). Each replication was run in two approximately 20-min sessions. The SIFC replications were completed first, followed by the three, counterbalanced, 2IFC blocks. Within a replication, trial sequences were randomized and constrained so that sequences were limited to four, five, or six events in a row. A new run limit was randomly chosen each trial. A trial consisted of a 100-ms warning interval, a 20-ms observation interval, a 500-ms interstimulus interval and a 20-ms observation interval (2IFC condition only), a 1500-ms decision interval, and a 900-ms (minimum) reset interval.

In the SIFC condition, the observer's task was to detect whether the $\mathcal{S}$ event occurred. In the 2IFC condition, the observer's task was to detect whether the $\langle \mathcal{S}, \mathcal{N} \rangle$ event ordering occurred. Observers indicated their decisions on the continuous rating scale like in Experiment I.

Results and discussion.   The continuous ratings were binned into 64 categories, and ROC, mROC, and GOC analyses were performed for each observer, in each condition, like in Experiment I. Data were not combined across observers, because each observer used a different stimulus set. All of the comparisons between tasks were based on the pooled 2IFC results unless otherwise stated.
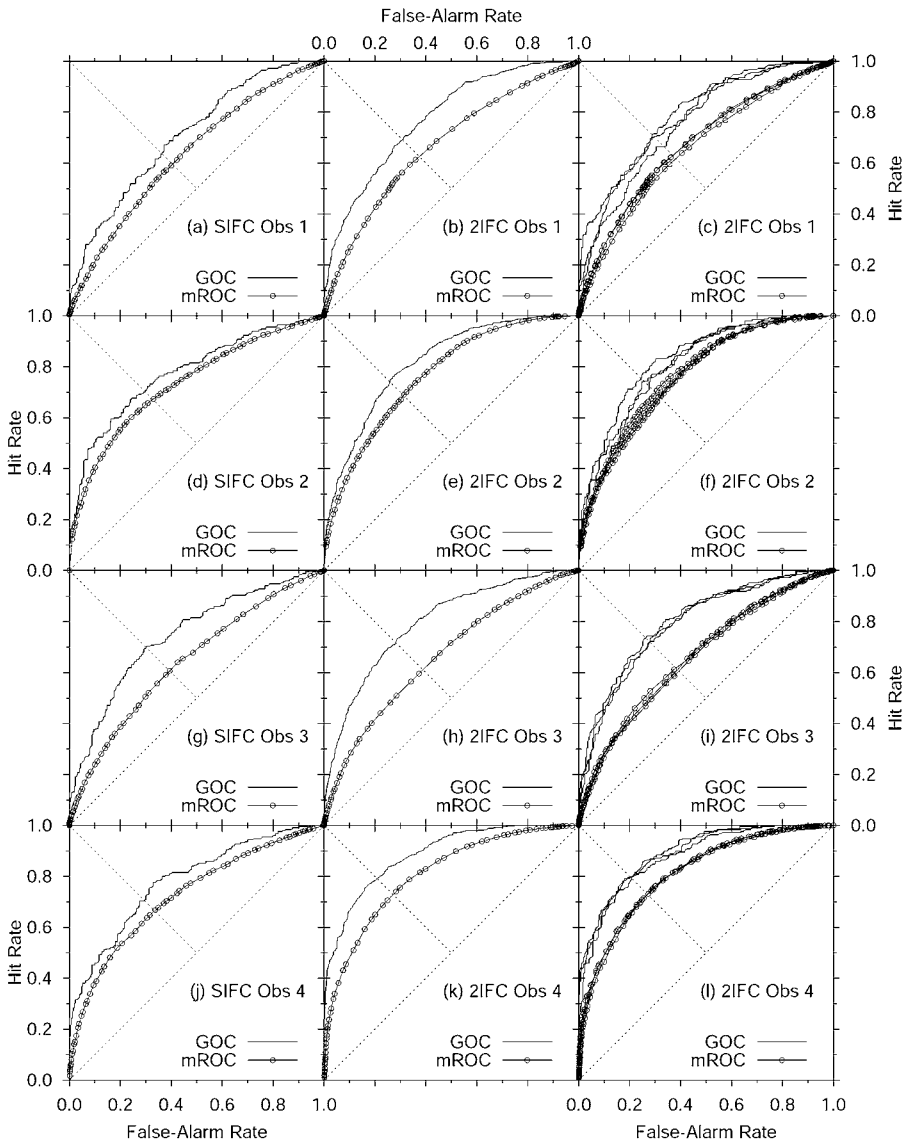
FIG. 7. GOC and mROC curves for each observer in the SIFC condition (left column), in the pooled 2IFC condition (middle column), and in the three unpooled 2IFC conditions (right column) in Experiment II are shown.

Figure 7 shows that performance improved from mROC analysis to GOC analysis for all observers in each condition. Figure 7 (right column) shows the mROC and GOC curves for the three 2IFC blocks. The pooled 2IFC curves in Fig. 7 (middle column) are smoother compared to the three unpooled curves. Assuming most of the unique noise is removed, the variability among the three 2IFC blocks (right column in Fig. 7) is possibly due to sampling variability of the three 250-stimuli-per-event sample sets.

There is an improvement in all measures of detectability after GOC analysis (see Table 4). Table 4 shows that Observer 1 and Observer 4 have the smallest

TABLE 4

The Difference between $\mathscr{A}_{SI}$ and $P(C)_{2I}$, after GOC, mROC, and ROC Analysis (Mean and Standard Deviations of $\mathscr{A}_{SI}$ and $P(C)_{2I}$) in Experiment II

Observer	Analysis	$\mathscr{A}_{SI}$	$P(C)_{2I}$	Difference
1	GOC	0.7077	0.6920	0.0157
	mROC	0.6346	0.6254	0.0092
	ROC M (SD)	0.6353 (0.0237)	0.6236 (0.0139)	0.0117
2	GOC	0.7767	0.7411	0.0356
	mROC	0.7333	0.6877	0.0457
	ROC M (SD)	0.7334 (0.0143)	0.6885 (0.0164)	0.0449
3	GOC	0.7466	0.7183	0.0283
	mROC	0.6438	0.6114	0.0324
	ROC M (SD)	0.6452 (0.0223)	0.6115 (0.0185)	0.0337
4	GOC	0.7822	0.7920	−0.0098
	mROC	0.7242	0.7288	−0.0046
	ROC M (SD)	0.7238 (0.0221)	0.7263 (0.0231)	−0.0025

differences between $\mathscr{A}_{SI}$ and $P(C)_{2I}$, after both mROC and GOC analysis. The relationship $\mathscr{D}_{SI} = \mathscr{C}_{2I}$ is also somewhat supported (Table 5). Whether or not one should take these results as evidence that $\mathscr{A}_{SI} = P(C)_{2I}$ or $\mathscr{D}_{SI} = \mathscr{C}_{2I}$ is debatable. The differences could be due to (a) sampling variability, (b) unique noise, or (c) inappropriate assumptions about the observers' decision processes, especially in the 2IFC task. The relationship $\mathscr{A}_{SI} \leqslant \mathscr{A}_{2I} \leqslant \mathscr{A}_{T}$ and its corollary $\mathscr{D}_{SI} \leqslant \mathscr{D}_{2I} \leqslant \mathscr{D}_{T}$, however, are supported in all cases (see Table 6).

In analyzing amplitude-discrimination data, the theoretical evidence distributions are unknown. Compounding this with observer inconsistency and sampling variability, it is difficult to come to any firm conclusions about the relationship between the SIFC and 2IFC tasks. It is therefore useful to generate a 2IFC ROC curve from the SIFC ratings, because 2IFC performance can then be predicted based on empirical SIFC performance, without needing to know the underlying distributions —providing that the same stimuli are used in each task.

TABLE 5

$\mathscr{D}_{SI}$, $\mathscr{C}_{2I}$, and Their Difference (bits) after GOC and mROC Analysis in Experiment II

Observer	GOC			mROC		
	$\mathscr{D}_{SI}$	$\mathscr{C}_{2I}$	Difference	$\mathscr{D}_{SI}$	$\mathscr{C}_{2I}$	Difference
1	0.1283	0.1091	0.0192	0.0529	0.0459	0.0071
2	0.2339	0.1749	0.0589	0.1635	0.1042	0.0593
3	0.1834	0.1422	0.0411	0.0605	0.0361	0.0244
4	0.2439	0.2624	−0.0185	0.1503	0.1568	−0.0065

<div align="center">

TABLE 6

Relationships $\mathcal{A}_{SI} \leqslant \mathcal{A}_{2I} \leqslant \mathcal{A}_T$ and $\mathcal{D}_{SI} \leqslant \mathcal{D}_{2I} \leqslant \mathcal{D}_T$ after GOC Analysis in Experiment II

</div>

Observer	$\mathcal{A}_{SI}$	$\leqslant$	$\mathcal{A}_{2I}$	$\leqslant$	$\mathcal{A}_T$	$\mathcal{D}_{SI}$	$\leqslant$	$\mathcal{D}_{2I}$	$\leqslant$	$\mathcal{D}_T$
1	0.7077	$\leqslant$	0.7755	$\leqslant$	0.8291	0.1283	$\leqslant$	0.2317	$\leqslant$	0.3403
2	0.7767	$\leqslant$	0.8174	$\leqslant$	0.9003	0.2339	$\leqslant$	0.3143	$\leqslant$	0.5319
3	0.7466	$\leqslant$	0.7962	$\leqslant$	0.8716	0.1834	$\leqslant$	0.2705	$\leqslant$	0.4469
4	0.7822	$\leqslant$	0.8866	$\leqslant$	0.9051	0.2439	$\leqslant$	0.4899	$\leqslant$	0.5475

It would be expected that Observer 1 and Observer 4 would have the best fit between predicted and obtained 2IFC ROC curves, because the difference between $\mathcal{A}_{SI}$ and $P(C)_{2I}$ was small for them. Figure 8 indicates that this is indeed the case. The difference between the predicted 2IFC ROC curve and the obtained 2IFC ROC curve is pronounced for Observer 2 and Observer 3, with the pROC curve overestimating the actual 2IFC ROC curve. This discrepancy may be due to observer inconsistency or sampling variability. Only 16 replications were run in the 2IFC conditions compared with the 32 replications for the SIFC condition; thus there may have been more residual unique noise in the 2IFC conditions. The 2IFC sample-ROC curve may also have been different from the predicted sample-ROC curve because only a subset of the possible 2IFC pairings was presented or because of the monotonic transform between the observer's evidence and the observer's rating (Drga, 1999). Finally, the assumptions underlying the tasks may not have
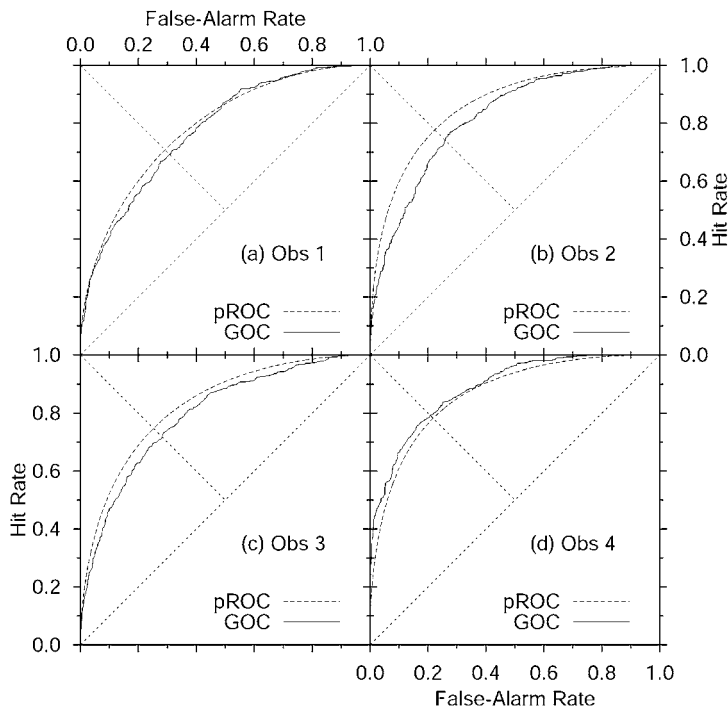

FIG. 8. The predicted 2IFC ROC (pROC) curve, derived from the SIFC GOC ratings, and the 2IFC GOC curve for each observer in Experiment II are shown.

been fulfilled (e.g., the observer may not have used a decision rule monotonic with the difference decision rule, or there may have been interactions between the intervals in the 2IFC task).

4. DISCUSSION

The theoretical relationships between the SIFC and 2IFC tasks for discrete random variables, developed in this paper, were strongly supported by the results of Experiment I. Once observer inconsistency was reduced using GOC analysis, human observers were in most cases able to perform as well as an ideal observer. It was also shown by comparison to the theoretical ROC curves that mROC curves do not necessarily reflect the underlying evidence distributions because of the effects of observer inconsistency. Experiment II provided some evidence that the nonparametric relationships hold for a continuous random variable, but the results were not as strong. These results are more robust than the results of previous experimental research (Emmerich, 1968a; Green & Moses, 1966; Whitmore *et al.*, 1968), however, because error due to observer inconsistency was reduced using GOC analysis.

Observer inconsistency helps to explain psychophysicists' preoccupation with normal models. Hanley and McNeil (1982) comment that the majority of psychophysical data can be fitted very well on the assumption that the underlying random variables are normal, regardless of the appropriateness of this assumption. The results from Experiment I demonstrated this phenomenon by using known, finitely-discrete evidence distributions. The mROC curves appeared normal, but after GOC analysis this apparent normality disappeared and the GOC curve reflected the shape of the theoretical ROC curve. These results, and those of Drga (1999), Lapsley Miller (1999), and Taylor *et al.* (1991) using experiments and computer simulations, suggest that using data from any single replication of an experiment can not provide strong support for any theory.

Experiment I showed that rating scales may be used instead of binary responses in 2IFC experiments. Green and Swets (1966) suggested that the crucial issue is whether or not the two methods yield essentially the same curves. Although performance in the binary-decision 2IFC task was lower than in the rating-scale 2IFC task, both 2IFC GOC curves (and their related measures of detectability) tended toward the theoretical 2IFC ROC curve as replications were added (Drga, 1999). Decisions made on a rating scale supply more information than binary decisions (Watson *et al.*, 1964), and trials do not necessarily take longer. We recommend that rating scales be used in preference to binary decisions, for both SIFC and 2IFC tasks, because fewer replications are needed to achieve the same GOC performance and the entire ROC curve is approximated—not just one point.

Despite the equivalence between measures of detectability in the SIFC and 2IFC tasks, the 2IFC task is often considered preferable in practice because it is (incorrectly) assumed to be unbiased (e.g., Mackworth, 1970). There are, however, good reasons why researchers should consider using an SIFC task.

1.   Interpreting SIFC experimental results with respect to a theory may be easier because fewer assumptions are required.

2. 2IFC ROC curves (with appropriate assumptions about X) and their associated measures of detectability may be derived from SIFC ratings or from the hit rates and false-alarm rates, but SIFC ROC curves cannot be derived from the 2IFC ROC ratings or hit rates and false-alarm rates.

3. Empirical 2IFC results can be confounded if the evidence in one interval interacts with the other. This may occur if there is temporal asymmetry in the task (e.g., due to memory) or, separately, if there is an effect of the duration of the interstimulus interval. 2IFC theory assumes symmetry and does not specify the duration, or even the existence, of the interstimulus interval.

4. Additional sampling may be required to form 2IFC stimulus pairings (as described in Section 3.2).

5. An SIFC trial is faster than a 2IFC trial because the 2IFC trial has an extra observation interval and an interstimulus interval.

In summary, the relationship $\mathscr{A}_{\mathrm{SI}} = P(\mathrm{C})_{\mathrm{2I}}$ was derived from the cross-correlation function of the SIFC evidence distributions, thereby relating both $\mathscr{A}_{\mathrm{SI}}$ and $P(\mathrm{C})_{\mathrm{2I}}$ to $P(X_{\mathscr{N}} < X_{\mathscr{G}})$, something which previous theoretical developments have not explicitly done. $\mathscr{A}_{\mathrm{SI}} = P(\mathrm{C})_{\mathrm{2I}}$ was demonstrated for continuous and finitely and infinitely discrete unidimensional random variables, arbitrary prior probabilities, and with no requirement for using a likelihood ratio decision axis. Guessing probabilities in the discrete case were shown to be important, because $\mathscr{A}_{\mathrm{SI}} = P(\mathrm{C})_{\mathrm{2I}}$ holds if and only if the 2IFC guessing probabilities are equal, or the 2IFC prior probabilities are equal, or both. Connections between the two tasks were extended to the new measure of detectability, $\mathscr{D}_{\mathrm{SI}}$, and the channel capacity, $\mathscr{C}_{\mathrm{2I}}$, of the 2IFC observer. These measures have been shown elsewhere to have a number of properties desirable in a measure of detectability (Scurfield, 1996, 1998). Nonparametric bounds on $\mathscr{A}_{\mathrm{2I}}$ in terms of $\mathscr{A}_{\mathrm{SI}}$ were derived, and similar bounds for $\mathscr{D}_{\mathrm{2I}}$ in terms of $\mathscr{D}_{\mathrm{SI}}$. Furthermore, it was shown how a 2IFC ROC curve may be calculated from the SIFC ratings or the SIFC hit rates and false-alarm rates. Finally, all these results are based on the relationship $\mathscr{A}_{\mathrm{SI}} = \mathrm{HR}(0, \frac{1}{2})_{\mathrm{2I}}$.

Elsewhere, Scurfield (1996, 1998) has extended $\mathscr{A}_{\mathrm{SI}} = P(\mathrm{C})_{\mathrm{2I}}$ to three events and to unidimensional and multidimensional decision axes by specifying an equality between one of the three-event, single-interval, three-alternative ROC volumes and the corresponding proportion of correct decisions in the three-event, three-interval, six-alternative task (for equal prior probabilities). Furthermore, he showed that $\mathscr{D} = \mathscr{C}$ for the same tasks, respectively, and that results are generalizable to n-events.

REFERENCES

Abramson, N. (1963). *Information theory and coding*. New York: McGraw–Hill.

Bamber, D. (1975). The area above the ordinal dominance graph and the area below the receiver operating characteristic graph. *Journal of Mathematical Psychology, 12*, 387–415.

Boven, R. (1976). *The use of multiple observers in signal detection theory: A method to remove the effect of unique noise from experimental data* (Masters thesis). Victoria University of Wellington, New Zealand.

Drga, V. (1988). *2IFC ROC analysis of narrow–band noise detection* (Honours thesis). Victoria University of Wellington, New Zealand.

Drga, V. (1999). *Theory of group operating characteristic analysis in discrimination tasks* (Doctoral dissertation). Victoria University of Wellington, New Zealand [available electronically at www.psychophysics.org].

Egan, J. P. (1975). *Signal detection theory and ROC analysis.* New York: Academic Press.

Egan, J. P., & Clarke, F. R. (1966). Psychophysics and signal detection. In J. B. Sidowski (Ed.) *Experimental methods and instrumentation in psychology*. New York: McGraw–Hill.

Emmerich, D. S. (1968a). Receiver-operating characteristics determined under several interaural conditions of listening. *Journal of the Acoustical Society of America, 43*(2), 298–307.

Emmerich, D. S. (1968b). ROCs obtained with two signal intensities presented in random order, and a comparison between yes–no and rating ROCs. *Perception & Psychophysics, 3*(1A), 35–40.

Falmagne, J. C. (1985). *Elements of psychophysical theory.* Oxford, UK: Oxford University Press.

Friedman, M. P., & Carterette, E. C. (1964). Detection of Markovian sequences of signals. *Journal of the Acoustical Society of America, 36*(12), 2334–2339.

Green, D. M. (1960a). Auditory detection of a noise signal. *Journal of the Acoustical Society of America, 32*(1), 121–131.

Green, D. M. (1960b). Psychoacoustics and detection theory, *Journal of the Acoustical Society of America, 32*(10), 1189–1203.

Green, D. M. (1964a). Consistency of auditory detection judgments. *Psychological Review, 71*(5), 392–407.

Green, D. M. (1964b). General prediction relating yes–no and forced-choice results. *Journal of the Acoustical Society of America, 36*, 1042(A).

Green, D. M., & Luce, R. D. (1974). Counting and timing mechanisms in auditory discrimination and reaction time. In D. H. Krantz, R. O. Atkinson, R. D. Luce, & P. Suppes (Eds.), *Contemporary development in mathematical psychology,* (pp. 372–415). San Francisco: Freeman.

Green, D. M., & McGill, W. J. (1970). On the equivalence of detection probabilities and well-known statistical quantities. *Psychological Review, 77*(4), 294–301.

Green, D. M., & Moses, F. L. (1966). On the equivalence of two recognition measures of short-term memory. *Psychological Bulletin, 66*(3), 228–234.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics.* New York: Wiley [reprinted by Krieger, Huntingdon, NY, 1974].

Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Diagnostic Radiology, 143*(1), 29–36.

Hodos, W. (1970). Nonparametric index of response bias for use in detection and recognition experiments. *Psychological Bulletin, 74*(5), 351–354.

Iverson, G. J., & Sheu, C.-F. (1992). Characterizing random variables in the context of signal detection theory. *Mathematical Society Sciences, 23*, 151–174.

Kaiser, J. F. (1966). Digital filters. In F. F. Kuo & J. F. Kaiser (Eds.), *System analysis by digital computer* (pp. 218–285). New York: Wiley.

Lapsley Miller, J. A. (1999). *The rôle of the bandwidth-duration product $\mathcal{WT}$ in the detectability of diotic signals* (Doctoral dissertation). Victoria University of Wellington, New Zealand [available electronically at www.psychophysics.org].

Lapsley Miller, J. A. (2001). An algorithm for calculating the essential bandwidth of a discrete spectrum and the essential duration of a discrete time–series (Tech. Rep. 1221). Groton, CT: Naval Submarine Medical Research Laboratory [available electronically at www.psychophysics.org].

Leshowitz, B. (1969). Comparison of ROC curves from one- and two-interval rating-scale procedures. *Journal of the Acoustical Society of America, 46*(2(2)), 399–402.

Luce, R. D. (1997). Some unresolved conceptual problems in mathematical psychology. *Journal of Mathematical Psychology, 41*, 79–87.

Mackworth, J. F. (1970). *Vigilance and attention: A signal detection approach*. London: Penguin Books.

MacMillan, N. A., & Kaplan, H. L. (1985). Detection theory analysis of group data: Estimating sensitivity from average hit and false-alarm rates. *Psychological Bulletin, 98*(1), 185–199.

Marill, T. (1956). Detection theory and psychophysics (Tech. Rep. 319). MIT Research Laboratory of Electronics.

Markowitz, J., & Swets, J. A. (1967). Factors affecting the slope of empirical ROC curves: Comparison of binary and rating responses. *Perception & Psychophysics, 2*, 91–97.

McNicol, D. (1972). *A primer of signal detection theory*. London: George Allen & Unwin.

Metz, C. E., & Shen, J.-H. (1992). Gains in accuracy from replicated readings of diagnostic images: prediction and assessment in terms of ROC analysis. *Medical Decision Making, 12*, 60–75.

Peterson, W. W., Birdsall, T. G., & Fox, W. C. (1954). The theory of signal detectability. *Transactions of the IRE,* **PGIT-4**, 171–212.

Pfafflin, S. M., & Mathews, M. V. (1962). Energy-detection model for monaural auditory detection. *Jounal of the Acoustical Society of America, 34*(12), 1842–1853.

Podd, J. V. (1982). *The aural discrimination of simultaneous changes in amplitude and frequency* (Doctoral dissertation). Victoria University of Wellington, New Zealand.

Pollack, I., & Hsieh, R. (1969). Sampling variability of the area under the ROC-curve and of d'_e. *Psychological Bulletin, 71*(3), 161–173.

Raab, D. H., & Goldberg, I. A. (1975). Auditory intensity discrimination with bursts of reproducible noise. *Journal of the Acoustical Society of America, 57*(2), 437–447.

Rabiner, L. R., & Gold, B. (1975). *Theory and application of digital signal processing*. Englewood Cliffs, NJ: Prentice–Hall.

Schulman, A. I., & Mitchell, R. R. (1966). Operating characteristics from yes–no and forced-choice procedures. *Journal of the Acoustical Society of America, 40*(2), 473–477.

Scurfield, B. K. (1995). *Discrimination among events by neural networks* (Doctoral dissertation). Victoria University of Wellington, New Zealand.

Scurfield, B. K. (1996). Multiple-event forced-choice tasks in the theory of signal detectability. *Journal of Mathematical Psychology, 40*(3), 253–269.

Scurfield, B. K. (1998). Generalization of the theory of signal detectability to *m*-dimensional *n*-event forced-choice tasks. *Journal of Mathematical Psychology, 42*(1), 5–31.

Shannon, C. E. (1949). The mathematical theory of communication. In C. E. Shannon & W. Weaver (Eds.), *The mathematical theory of communication* (pp. 3–99). Urbana, IL: The University of Illinois Press.

Simpson, A. J., & Fitter, M. J. (1973). What is the best index of discriminability. *Psychological Bulletin, 80*(6), 481–488.

Swets, J. A. (1959). Indices of signal detectability obtained with various psychophysical procedures. *Journal of the Acoustical Society of America, 31*(4), 511–513.

Swets, J. A., & Green, D. M. (1961). Sequential observations by human observers of signals in noise. In J. A. Swets (Ed.), *Signal detection and recognition by human observers: Contemporary readings* (1964 ed.). New York: Wiley.

Taylor, A. (1984). *Auditory psychophysics in birds: The effects of unique noise on sensitivity* (Doctoral dissertation). Victoria University of Wellington, New Zealand.

Taylor, A., Boven, R., & Whitmore, J. (1991). Reduction of unique noise in the psychophysics of hearing by group operating characteristic analysis. *Psychological Bulletin, 109*(1), 133–146.

Watson, C. S. (1962). *Signal detection and certain physical characteristics of the stimulus during the observation interval* (Doctoral dissertation). Indiana University [reprinted in *Dissertation Abstracts International* (Vol. 24, p. 2995)].

Watson, C. S., Rilling, M. E., & Bourbon, W. T. (1964). Receiver-operating characteristics determined by a mechanical analog to the rating scale. *Journal of the Acoustical Society of America, 36*(2), 283–288.

Whitmore, J. K., Williams, P. I., & Ermey, H. L. (1968). Psychometric function from Rayleigh–Rayleigh ROC curves. *Journal of the Acoustical Society of America,* **44**(1), 370–371(A).

Wickelgren, W. A. (1968). Unidimensional strength theory and component analysis of noise in absolute and comparative judgments. *Journal of Mathematical Psychology,* **5**, 102–122.

Wilcox, G. W. (1968). *Inter-observer agreement and models of monaural auditory processing in detection tasks* (Doctoral dissertation). Ann Arbor, MI: University of Michigan.