JOURNAL OF MATHEMATICAL PSYCHOLOGY 31, 366–409 (1987)

Signal-Detection Models for Multidimensional Stimuli: Probability Distributions and Combination Rules

NORMA GRAHAM

Columbia University

PATRICIA KRAMER

University of Connecticut

AND

DEAN YAGER

State University of New York

Probabilistic independence among multiple random variables (e.g., among the outputs of multiple spatial-frequency channels) has been invoked to explain two effects found with many kinds of stimuli: increments in detection performance due to "probability summation" and decrements in detection and identification performance due to "extrinsic uncertainty." Quantitative predictions of such effects, however, depend on the precise assumptions. Here we calculate predictions from multidimensional signal-detection theory assuming any of several different probability distributions characterizing the random variables (including two-state, Gaussian, exponential, and double-exponential distributions) and either of two rules for combining the multiple random variables into a single decision variable (taking the maximum or summing them). In general, the probability distributions predicting shallower ROC curves predict *greater* increments due to summation but *smaller* decrements due to extrinsic uncertainty. Some probability distributions yield steep-enough ROC curves to actually predict *decrements* due to summation in blocked-summation experiments. Probability distribution matters much less for intermixed-summation than for blocked-summation predictions. Of the two combination rules, the sum-of-outputs rule usually predicts both greater increments due to summation and greater decrements due to extrinsic uncertainty. Put another way, of the two combination rules, the sum-of-outputs rule usually predicts better performance on the compound stimulus under any condition but worse performance on simple stimuli under intermixed conditions.

Reprint requests should be addressed to Norma Graham, Department of Psychology, Columbia University, New York, NY 10027.

0022-2496/87 $3.00

1. Introduction

Probabilistic independence among multiple channels (e.g., those selective for the spatial frequency, orientation, and spatial position of visual patterns) has several potential effects. The first of two of interest here is *probability summation*: The greater the number of simple components in a compound (even though no one component excites more than a single channel), the greater the detectability of the compound. The second is *extrinsic-uncertainty effect*: The larger the number of possible alternative simple stimuli that might appear on a single trial, the lower the detectability or identifiability of a particular one. (For some recent reviews of and references to empirical work see Ball & Sekuler, 1980; Green & Weber, 1980; Green & Birdsall, 1978; Graham, 1981, 1985; Davis, Kramer, & Graham 1983; Shaw, 1980).

Increments due to probability summation are small and decrements due to extrinsic-uncertainty are often even smaller. Our understanding of spatial vision (and perhaps a number of other fields less well known to us) has progressed to a point where neither of these effects is negligible, however. A high-threshold model supplemented by a convenient analytic form for the psychometric function—called the Quick pooling model here—has proved invaluable in taking into account probability summation (e.g., Bergen, Wilson, & Cowan, 1979; Graham, 1977; Graham, Robson, & Nachmias, 1978; Mostafavi & Sakrison, 1976; Quick, Mullins, & Reichert, 1978; Robson & Graham, 1981; Watson, 1983). The interpretations of pattern-vision experiments now depend, however, on differences that are as small as extrinsic-uncertainty effects. (See, for example, Thomas, Gille, & Barker, 1982 discussion of the calculation of spatial-frequency and orientation bandwidths from two-by-two simultaneous detection and identification experiments). Unfortunately, high-threshold models cannot explain the decrements due to extrinsic uncertainty (Davis & Graham, 1981; Davis *et al.*, 1983; Yager, Kramer, Graham, & Shaw, 1984) nor several aspects of the detection of simple sinusoidal gratings (Nachmias, 1981).

A model adequate to deal with all these effects would certainly be useful in pattern vision and might be useful in other domains as well. We thought it might be found in extensions of traditional signal-detection models to the multidimensional case appropriate for multiple-channel models. These models share many assumptions with the Quick pooling model but differ from high-threshold models in the kind of probability distributions characterizing channels' outputs. In this paper we present numerical predictions from five other probability distributions as well as the high-threshold distribution. These are sufficient to demonstrate a systematic

relationship between the slope of the ROC curve and the predicted amounts of probability summation and uncertainty effect. On the basis of this general relationship, one would know a good deal about the predictions from any distribution (once its ROC curve is known).

Although some of these predictions are already available in the literature (Creelman, 1960; Davis, *et al.*, 1983; Green & Weber, 1980; Green & Birdsall, 1978; Green & Swets, 1966; Klein, 1985; Kramer, Graham, & Yager, 1985; Johnson, 1980; Nachmias, 1972; Nolte & Jaarsma, 1967; Pelli, 1985; Shaw, 1982; Yager *et al.*, 1984), no systematic comparison of the predictions from different probability distributions and combination rules has been made. In general, in fact, multidimensional signal-detection theory has been studied less well than unidimensional signal-detection theory perhaps because the optimal decision rule is somewhat intractable in the multidimensional case. Here we avoid that problem by studying two tractable decision variables (the sum and the maximum). Their relationship to the optimal decision rule is discussed where known.

2. Description of Models and Experiments

2.1. *Description of Stimulus Set*

A set of stimuli, as referred to in these assumptions, contains (1) a neutral stimulus which will be called "noise" (e.g., in pattern vision, a blank field of the same mean luminance as the sinusoidal grating), (2) a number N of simple stimuli S_j $(j = 1, N)$ that are far apart on some dimension (e.g., grating patches, each of a different spatial frequency or each at a different position) with intensities (e.g., contrasts of grating patches) chosen so that all are equally detectable, and (3) a compound stimulus C that contains all the N simple stimuli (e.g., a grating patch containing several spatial frequencies or a larger grating made up from the grating patches at different positions). Alternately, the compound stimulus can be thought of as multiple observations of one particular simple stimulus as will be discussed further in Section 2.6.

2.2. *Assumptions Relating Stimuli to Channels' Outputs*

Assumption 1. Multiple Random Variables R_i $(i = 1, N')$. There are N' channels (where $N' > N$ the number of simple stimuli). The output of the ith channel to a stimulus "stim" is represented as a random variable R_i(stim), where the distrubution of this random variable does not change across trials. For convenience the particular stimulus will be omitted when it is not necessary for clarity.

Let $f_i(z/\text{stim})$ and $F_i(z/\text{stim})$ be the probability density and probability distribution functions, respectively, of R_i(stim).

Assumption 2. Probabilistic Independence. For a given stimulus, the outputs of different channels are probabilistically independent. That is, the random

variables R_i(stim) and R_j(stim) are independent for all i, j and "stim." The outputs of the same channel on different trials are also probabilistically independent.

ASSUMPTION 3. COMPLETELY NON-OVERLAPPING AND NON-INTERACTING CHANNELS FOR FAR-APART STIMULI. For simple stimuli that are far enough apart (e.g., of widely separated spatial frequencies or orientations or spatial positions) no channel is sensitive to more than one of them. (A channel "is sensitive to a stimulus" if and only if the channel output's probability distribution to that stimulus is different from its probability distribution to the noise stimulus.) Further, if a channel is not sensitive to any component in a compound stimulus, it is not sensitive to the compound. If a channel is sensitive to one component in a compound, its probability distribution in response to the compound is the same as that to that one component. A channel can never be sensitive to more than one component for the stimulus set considered here.

Ashby and Townsend (1986) discuss the relationships between Assumptions 2 (closely related to what they call "perceptual independence") and 3 (closely related to what they call "perceptual separability") and several uses of the term "orthogonality."

ASSUMPTION 4. ONE RANDOM VARIABLE PER SIMPLE STIMULUS. We will further assume that for each simple stimulus in the stimulus set under consideration here (all of which are far apart and equally detectable), there is one and only one channel that is sensitive to it. This assumption is made without loss of generality if the proper interpretations are made (see Section 2.5).

ASSUMPTION 5. IDENTICAL SIGNAL AND NOISE DISTRIBUTIONS IN DIFFERENT CHANNELS. All channels' noise distributions—their probability distributions in response to noise—are assumed to be identical. Consider those channels that are sensitive to any of the equally detectable simple stimuli in the limited stimulus set being considered here. Their signal distributions—their probability distributions in response to a simple stimulus they are sensitive to—are assumed to be identical.

This assumption does reduce generality. It is plausible in the case of equally detectable far-apart simple stimuli, however. For spatial frequency and spatial position of visual stimuli, at least, this assumption is made more attractive by the strong similarities among the families of ROC functions and of psychometric functions measured for sinusoidal gratings of any of a wide range of spatial frequencies or spatial positions.

ASSUMPTION 6. THE EXACT FORM OF THE PROBABILITY DISTRIBUTIONS. To compute predictions, the signal and noise density functions must be specified. Predictions from six one-parameter families of probability density functions are considered in this paper. Changes in the value of the one parameter characterizing each of these families correspond to changes in stimulus intensity and, hence, detec-

tability. No assumption will be made, however, about the relationship between physical stimulus intensity and the value of this parameter.

For each of these six families, Fig. 1 illustrates the noise density function $f_j(z/\text{noise})$ and one signal density function $f_j(z/S_j)$, where the stimulus-strength parameter for the signal density function has been chosen for a moderate level of detectability as described below.

ASSUMPTION 6A. HIGH-THRESHOLD FAMILY. The random variables are discrete with only two possible values: 0 (the not-detect state) and 1 (the detect state). The parameter p_i is the probability of getting into the detect state and is taken to be the signal-strength parameter. For the noise stimulus, p_i equals zero; that is, the noise density is entirely concentrated on 0 (the not-detect state). In symbols,

$$f_i(z) = (1 - p_i) \cdot \delta(z) + p_i \cdot \delta(z - 1),$$

where $\delta(z)$ is the delta function or impulse (see, e.g., Bracewell, 1965). In the top row of Fig. 1, the signal and noise density functions have been offset horizontally for clarity.

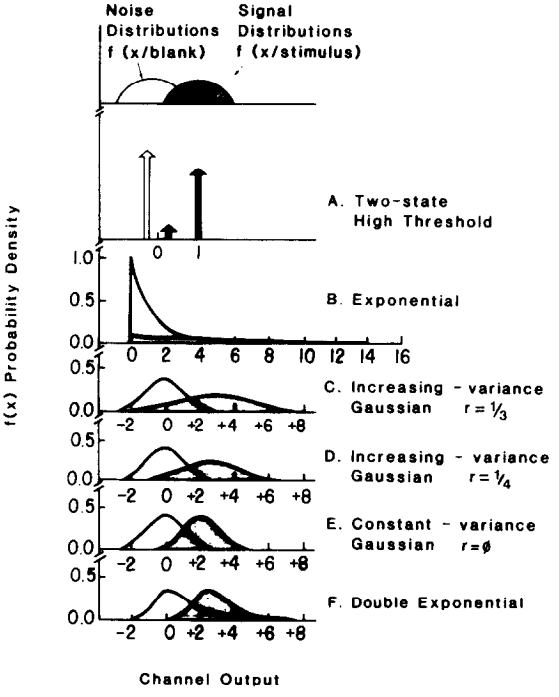

FIG. 1. The noise density function $f_i(z/\text{noise})$ and one signal density function $f_i(z/S_i)$ are illustrated for each of six probability distribution families.

ASSUMPTION 6B. SPREAD-EXPONENTIAL FAMILY. The exponential densities used here (second row Fig. 1) start at the same point but have different variances:

$$f_i(z) = 0 \qquad \text{for } z < 0$$
$$= k_i \exp(-k_i \cdot z) \qquad \text{for } z \geqslant 0.$$

Here $k_i = 1$ for the noise density function and gets smaller with increasing detectability. Rather than using k as a signal-strength parameter, we prefer

$$k' = 2 \log_{10}(1/k).$$

The value of k' increases from zero (for noise) with detectability. Further, the ROC curve generated with a particular value of k' has a negative-diagonal d' (defined below) almost identical to the value of k' (see Table 2).

The next three are one-parameter Gaussian families, for which the signal-strength parameter is taken to be

$$d' = \frac{\mu_S - \mu_N}{\sigma_N},$$

where μ_S and μ_N are the means of the signal and noise density functions respectively, and σ_S and σ_N are the standard deviations. The value of d' is zero for the noise stimulus.

The value of r distinguishing the three Gaussian families from one another describes the rate at which the variance of the signal distribution increases as signal strength increases:

$$\frac{\sigma_S}{\sigma_N} = 1 + r \cdot d'.$$

ASSUMPTION 6C. INCREASING-VARIANCE GAUSSIANS ($r = \frac{1}{3}$). Third row Fig. 1.

ASSUMPTION 6D. INCREASING-VARIANCE GAUSSIANS ($r = \frac{1}{4}$). Fourth row Fig. 1.

ASSUMPTION 6E. CONSTANT-VARIANCE GAUSSIANS ($r = 0$). Fifth row Fig. 1.

ASSUMPTION 6F. SHIFTED DOUBLE-EXPONENTIAL FAMILY. In general, the double-exponential density function (bottom row of Fig. 1) is

$$f_i(z) = \exp(-a\langle z - U_i \rangle) \exp(-\exp(-a\langle z + U_i \rangle)).$$

The members of the one-parameter family used here all have the same variance (the constant a will be set equal to one) but are horizontally translated with respect to one another. U_i is a convenient signal-strength parameter to use, letting it be equal to zero for the noise distribution.

Why these six. The high-threshold model is discussed in the Introduction. The others are presented here because they turn out to span an interesting range of possible distributions. The original reasons for studying them were as follows: the constant-variance Gaussian case ($r = 0$) was used because it is so frequently used by others (often called the equal-variance case). The increasing-variance Gaussian cases ($r = \frac{1}{3}$, and $r = \frac{1}{4}$) as well as the exponential were used because they predict ROC curves like those found empirically for detection of simple sensory stimuli (e.g., Green & Swets, 1966; Nachmias & Kocher, 1970; unpublished results from Hirsch, Hylton, & Graham, 1982). The double-exponential distribution has some intriguing properties (Yellott, 1977) including the fact that its predictions correspond to those from choice theory in some situations, which have prompted investigators to consider it (e.g., Lappin & Staller, 1981). See further discussion in Section 3.1.

2.3. *Assumptions Relating Channels' Outputs to the Observer's responses*

ASSUMPTION 7. UNLIMITED SELECTIVE MONITORING. On each trial the observer monitors only the relevant random variables, i.e., only the outputs from channels sensitive to stimuli which might be presented on that trial. (The observer monitors all these relevant random variables.) A random variable's probability distribution is not affected by whether it is being monitored. The number of relevant random variables on a given trial will be called M'.

This assumption excludes the possibility of a limited capacity to attention or memory. It may work, however, quite well for near-threshold pattern vision (e.g., Davis *et al.*, 1983; Graham, Kramer, & Haber, 1985; Kramer, *et al.*, 1985; Yager *et al.*, 1984) and probably in many other domains tapping lower-level processes. In any case, before concluding that there are limits to attention or memory, one needs to know what models postulating no limits to attention or memory could predict (e.g., Egeth, 1977; Shaw, 1980; Sperling, 1984).

ASSUMPTION 8. DETECTION RULE. The decision variable R is assumed to be a combination of the outputs from the monitored channels. The relationship between the decision variable R and the observer's response is assumed to be given by standard unidimensional signal-detection theory:

Yes–no experiments. In a yes–no experiment, the observer is assumed to say "yes" on a trial if and only if R on that trial is greater than some criterion level λ. The predicted ROC curve for detection—which is Pr(yes/signal) plotted versus Pr(yes/noise)—is generated by plotting a point for a number of different values of λ. (See Note 1.) For the high-threshold family varying the criterion λ will produce an ROC curve that is just three discrete points. Since such ROC curves are unrealistic, it is conventional to generate a continuous ROC curve by connecting each pair of neighboring points by a straight line segment. Such connection is equivalent to assuming that the observer can adopt any of an infinite number of mixture

strategies. According to each mixture strategy, the observer uses one criterion on some proportion of the trials and on the rest of the trials he uses a criterion that leads to the neighboring point.

Forced-choice experiments. In two-interval forced-choice, the observer is assumed to say whichever interval yielded the larger value of the decision variable R. For the high-threshold family, ties can sometimes arise; we assume that the observer guesses without bias as to which alternative is correct.

Predictions described in this paper were calculated for two decision variables:

ASSUMPTION 8A. THE MAXIMUM-OUTPUT RULE.

$$R = \underset{i=1}{\overset{M'}{\text{MAX}}} R_i.$$

ASSUMPTION 8B. THE SUM-OF-OUTPUTS RULE.

$$R = \sum_{i=1}^{M'} R_i,$$

where the random variables have been re-arranged if necessary so that those being monitored are numbered from 1 to M'.

ASSUMPTION 9. THE MAXIMUM-OUTPUT IDENTIFICATION RULE. An observer identifies a stimulus as being the simple stimulus that corresponds to whichever one of the M' monitored channels gives the maximum output on that trial. (For the high-threshold model, there may be ties; then the observer guesses.) This rule is sensible only when the observer is required to identify the stimulus on each trial as being a simple stimulus. This will be true in the identification experiments discussed here (see Section 2.7).

Note that this identification rule uses *which* of the random variables R_i $(i = 1, N)$ produced the maximum output. Assuming that such information is available at higher stages is sometimes called an assumption of "labelled lines" or "specific nerve energies." (See discussion in, for example, Watson & Robson, 1981.) Knowing only the value of the maximum output or of the sum of outputs (used in the detection rules) without knowing which channel it came from would be of no use to the observer in identifying the stimulus.

2.4. *Assumptions in Combined and Abbreviated Form*

(1, 2). There are N' mutually probabilistically independent random variables.

(3, 4, 5). The whole set of N' random variables is numbered so that the ones sensitive to the N simple stimuli come first and are in the same order as the simple stimuli. (The compound stimulus C contains all N simple stimuli: S_j, $j = 1, N$.)

For the first N random variables $(j = 1, N)$,

$$f_j(z/C) = f_j(z/S_j) = f_1(z/S_1)$$

and, when i is not equal to j,

$$f_j(z/S_i) = f_j(z/\text{noise}) = f_1(z/\text{noise}).$$

For the last $N' - N$ random variables ($j = N + 1, N'$) and for any simple stimulus ($i = 1, N$),

$$f_j(z/C) = f_j(z/S_i)$$
$$= f_j(z/\text{noise}) = f_1(z/\text{noise}).$$

(6). The specific form of signal and noise densities must be specified—six alternatives listed above.

(7, 8, 9). The observer monitors (bases his decision on) all the (and only the) random variables sensitive to stimuli that might have been presented. In detection experiments, the observer's decision variable (used in the manner of standard signal detection theory) is either the maximum or the sum of the monitored random variables (to be specified). When the observer must identify which of several far-apart simple stimuli is presented, the observer responds stimulus S_j if and only if the maximum of the monitored random variables is R_j.

Ten distinct models. Of the 12 possible combinations of the six families of probability distributions (Assumptions 6A through 6F) paired with two different detection rules (Assumptions 8A and 8B) we present predictions from 11. The 12th—the adding-of-outputs double-exponential—required more work than was merited by what seemed likely to be learned from it.

Another combination—the sum-of-outputs rule with the high-threshold family—makes exactly the same predictions for the experiments considered here as the maximum-output rule with that same family. Therefore, only 10 of the 11 sets of predictions are distinct from one another.

2.5. *More Than One Channel per Simple Stimulus*

Sometimes more than one of the "channels" in a model have non-zero sensitivity for a given simple stimulus—in apparent contradiction to Assumption 4—although obeying the other assumptions. In many current models of pattern vision, for example, there are multiple mechanisms each of which is analogous to a single neuron. (See review in (Graham, 1985).) These mechanisms ("channels" in the original model) are often thought to obey Assumptions 1 through 3 and 5 through 9, and yet many of the mechanisms (those tuned to the same spatial frequency and orientation but having receptive fields at different spatial positions) will all be sensitive to a single sinusoidal grating in contradiction to Assumption 4.

If, however, a particular set of simple stimuli (e.g., gratings of different spatial frequencies) are very far apart, then by Assumption 3 no mechanism is sensitive to more than one of the simple stimuli in the set. Thus the outputs of all the mechanisms that are sensitive to a particular simple stimulus can be combined to form a new random variable, the output of a new "channel." The outputs of the

mechanisms must be combined in the manner consistent with whichever detection rule is used by the observer (that is, either their sum or their maximum must be taken appropriately). These new random variables—the outputs of new "channels"—will now satisfy Assumption 4 as well as Assumptions 1 through 3 and 7 through 9.

Given a particular original model, the question of whether or not the new random variables have identical distributions (per Assumption 5)—given that the original ones did—can be deduced.

Assumption 6—the specification of signal and noise densities for the random variables—will need to be modified since the new random variables will not have exactly the same probability distributions as did the original ones. Whether or not these new random variables have the same *kind* of probability distribution as did the original ones depends both on the combination rule and on the probability-distribution family. They will have the same kind for sum (or max) and high-threshold, sum and constant-variance Gaussians, and maximum and double exponentials. They will not have the same kind for the other models considered here. (When unequal-variance Gaussian families are used, the sum-of-outputs has a Gaussian distribution but the dependence of variance on signal strength will not, in general, be the same as that of the original random variables.)

2.6. *Stimulus Conditions*

Table 1 summarizes the application of the selective-monitoring assumption (7) and the sensitivity assumptions (3 and 4) to the stimulus conditions (first column) and experiments (second column) described below. The third column gives M', which is the number of random variables monitored on each trial under a given condition, and the fourth column gives P', which is the number of random variables characterized by a signal distribution on any trial of the non-blank stimulus indicated in the first column.

The simple-alone condition. All trials are either of the noise stimulus or of one particular simple stimulus. Under this condition the observer is assumed to monitor only the one random variable appropriate to the stimulus ($M' = 1$) and that one random variable will have a signal probability distribution on stimulus trials ($P' = 1$). On noise trials, of course, that random variable will have a noise probability distribution. Note that both the maximum-output and the sum-of-outputs rules produce the same decision variable for this condition; according to either, the one random variable being monitored is itself the decision variable.

This simple-alone condition will be used here in detection but not identification experiments (as described in the next section).

Simple-intermixed condition. A number (N) of far-apart simple stimuli—all equated for detectability—as well as a noise stimulus are randomly presented on different trials of the same block. Under this condition, the observer monitors N random variables ($M' = N$) but, on any trial of a simple stimulus, only one has a signal distribution ($P' = 1$).

TABLE 1

Application of Assumptions 3, 4, and 7 to
Different Stimulus Conditions and Kinds of Experiments

Row	Stimulus conditions	Kinds of experiments	Number of channels monitored M'	Number of channels with signal distributions P'
1	Compound-alone N components	Blocked-summation experiment	N	N
2	Simple-alone	Simple-uncertainty experiment	1	1
3	Simple-intermixed N simple stimuli		N	1
4	Compound-and-simple intermixed N simple stimuli Simple stimulus	Intermixed-summation experiment	N	1
5	Compound-stimulus N components		N	N

This simple-intermixed condition will be used in detection and identification experiments (as described in the next section).

Compound-alone condition. All trials are either of the noise stimulus or of the compound stimulus containing N equally detectable far-apart simple-stimulus components. The observer monitors N random variables ($M' = N$), therefore, all of which are characterized by a signal distribution on trials of the stimulus ($P' = N$).

This compound-alone condition will be used in detection but not in identification experiments.

Compound-and-simple intermixed condition. Trials of each of N equally detectable simple stimuli as well as of the compound made up from them are randomly intermixed along with trials of the noise stimulus. According to Assumption 7 the observer always monitors all N random variables ($M' = N$). On trials of simple stimuli, only one random variable has a signal distribution ($P' = 1$) but on trials of the compound stimulus, all N of them do ($P' = N$).

This condition will be used here only in detection experiments. It can also be used in some very interesting identification experiments but they are beyond the scope of this paper. (See, for example, Ashby & Townsend, 1986; Graham *et al.*, 1985; Klein, 1985).

The detection task. In detection experiments, the observer is to say whether a non-blank or a blank stimulus occurred on a trial (in the yes–no procedure) or which interval contained the non-blank stimulus (in a forced-choice procedure).

The identification task. In the only identification experiments considered here, the observer is to say on each trial which of several simple stimuli occurred. (If blank trials are included, the observer is still to say which simple stimulus was presented.)

Rephrased in terms of multiple observations. The above conditions were described in terms of presentations of various simple stimuli and one compound stimulus. For some readers it might be useful to redescribe the above experimental conditions in terms of an observer's having multiple opportunities to observe a single stimulus. This description leads to a formally identical situation. To be concrete, let us suppose an observer is listening for a signal during a number (M') of different listening intervals. (The intervals are so far apart it is reasonable to assume that events in different intervals are independent and non-interacting.) During some number (P') of these intervals a stimulus may occur (that is, the event will be from the signal distribution) while during all the other intervals (M'–P') the stimulus will not occur (the event will be from the noise distribution).

Under the simple-alone condition the observer is listening during only one interval ($M' = 1$) and only during that one interval may a stimulus occur ($P' = 1$). The observer doing the detection task is, in effect, testing between the hypothesis that one signal sample (with zero noise samples) occurred and the hypothesis that zero signal samples (with one noise sample) occurred.

Under the simple-intermixed condition, the observer is listening during many intervals ($M' = N$) during only one of which a stimulus may occur ($P' = 1$). The observer doing the detection task is testing between the hypothesis that one signal sample (and $M' - 1$ noise samples) occurred and the hypothesis that zero signal samples (with M' noise samples) occurred. (The observer doing the identification task is testing among many hypotheses, each of the form that the signal sample occurred during a specified interval and noise samples occurred during all the other intervals.)

Under the compound-alone condition, the observer is listening during many intervals ($M' = N$) during all or none of which a signal may occur ($P' = N' = N$). The observer doing the detection task is testing between the hypothesis that M' signal samples (with no noise samples) occurred and the hypothesis that zero signal samples (with M' noise samples) occurred.

Under the simple-and-compound intermixed condition, the observer is listening during many intervals ($M' = N$) during one of which ($P' = 1$) or during all of which

($P' = M' = N$) a signal may occur. The observer is testing between one composite and one simple hypothesis. The composite hypothesis is that either one signal sample (with $M' - 1$ noise samples) occurred or that M' signal samples (with zero noise samples) occured. The simple hypothesis against which it is tested is that zero signal samples (with M' noise samples) occurred.

2.7. *Kinds of Experiments*

Simple uncertainty experiments (*detection and identification*). The detectability of a simple stimulus is typically measured under a simple-alone condition ($M' = 1$ and $P' = 1$) and compared with its detectability measured under simple-intermixed conditions ($M' = N$ and $P' = 1$).

We will use the symbol D*n* to denote the detectability of the stimulus and condition in the *n*th row of Table 1.

The uncertainty effect (in detection) is D2-D3—that is, the difference in detectability between a simple stimulus measured under an alone condition and that stimulus measured under a simple-intermixed condition.

When identification performance is measured by having the observer say which of the N simple stimuli was presented on each trial, it makes little sense to have simple-alone blocks. Usually simple-intermixed blocks of several values of N are used instead. (The trials of the noise stimulus are sometimes omitted in these simple uncertainty identification experiments although there are advantages in retaining them.) The uncertainty effect in identification is the difference in proportion correct identification between two simple-intermixed conditions using different values of N.

Summation experiments (*detection*). In summation experiments, the detectability of a compound made up of several components is compared to the detectability of each component alone.

In *blocked-summation* experiments, the detectability of each component is measured under a simple-alone condition ($M' = 1$ and $P' = 1$) and the detectability of the compound is measured under a compound-alone condition ($M' = N$ and $P' = N$).

The blocked-summation effect is D1-D2—the difference in detectability between the compound under a compound-alone condition and a component under a simple-alone condition.

In *intermixed-summation* experiments, the detectability of each component and of the compound is measured simultaneously under a simple-and-compound intermixed condition ($M' = N$ on all trials, $P' = 1$ on trials of the simple stimulus, and $P' = N$ on trials of the compound stimulus).

The intermixed-summation effect is D5-D4—that is, the difference in detectability between the compound and a component both measured under a simple-and-compound intermixed condition.

Interrelationships among experiments. If the same stimuli were used throughout all the experiments in Table 1, two equivalences are implied by the assumptions here and indicated by the dashed brackets in Table 1. First, D4 = D3; that is, the

detectability of a simple stimulus when measured under the simple-and-compound intermixed condition is the same as that when measured under the simple-intermixed condition. Second, $D5 = D1$; that is, the detectability of the compound when measured under the simple-and-compound intermixed condition is the same as that when measured under the compound-alone condition.

The intermixed-summation effect (D5-D4) will, therefore, equal (D1-D3). Thus $D5 - D4 = D1 - D3 = (D1 - D2) + (D2 - D3)$, or, in words:

$$\begin{array}{ccccc} \text{the} & & \text{the} & & \text{the} \\ \text{intermixed-summation} & = & \text{blocked-summation} & + & \text{simple-uncertainty} \\ \text{effect} & & \text{effect} & & \text{effect} \end{array}$$

An immediate consequence of this relationship is that the intermixed-summation effect should be, in general, larger than the blocked-summation effect.

3. Methods for Calculating Predictions

The reader can skip to Section 4 without loss of continuity. (See Note 2.)

Let $f(z/\text{stim})$ be the probability density function for the decision random variable R (see Assumption 8) and let the distribution function F for R be $F(z/\text{stim})$. Similarly, let f_i and F_i be the density and distribution functions, respectively, for the random variable R_i (see Assumption 1). Let y be the hit rate and x be the false-alarm rate. Then the ROC curve for "stim" can be generated by varying λ in

$$\begin{aligned} x &= 1 - F(\lambda/\text{stim}) \\ y &= 1 - F(\lambda/\text{noise}). \end{aligned} \tag{1}$$

Simple-alone condition. It will be convenient to let x_1 and y_1 stand for the false-alarm and hit rates in the case where M' (the number of monitored random variables) equals P' (the number of random variables sensitive to the stimulus) $= 1$. In this case, according to either decision rule, the decision variable R will have the same distribution as the single relevant random variable R_1. Therefore,

$$\begin{aligned} 1 - y_1 &= F_1(\lambda/\text{stim}) \\ 1 - x_1 &= F_1(\lambda/\text{noise}). \end{aligned} \tag{2}$$

Note that, for the simple-alone condition, the models here reduce to ordinary unidimensional signal-detection theory where combination rule is irrelevant.

3.1. *Maximum-Output Detection Rule*

When the maximum-output detection rule is used (Assumption 8a), the decision variable R will be below a criterion λ if and only if every one of the monitored random variables is. For convenience, rearrange the R_i's (if necessary) so that the ones

being monitored are numbered from 1 to M' and those having a signal distribution are numbered from 1 to P' (where P' is less than or equal to M'). Then,

$$R \leqslant \lambda \quad \text{iff} \quad R_i \leqslant \lambda \text{ for all } i \text{ from 1 to } M'.$$

Because the random variables R_i are all independent (Assumption 2) and have identical distributions (Assumption 5),

$$\begin{aligned} F(z/\text{stim}) &= \prod_{i=1}^{M'} F_i(z/\text{stim}) \\ &= |F_1(z/\text{stim})|^{P'} \, |F_1(z/\text{noise})|^{M'-P'}. \end{aligned} \tag{3}$$

Substituting x_1 and y_1 (Eq. (2)) into this equation and then substituting the result into Eq. (1) for x and y gives

$$\begin{aligned} 1 - y &= (1 - y_1)^{P'} (1 - x_1)^{M'-P'} \\ 1 - x &= (1 - x_1)^{M'}. \end{aligned} \tag{4}$$

Thus, for the maximum-output rule with any probability distribution and for any condition of the summation and uncertainty experiments (i.e., any values of M' and P') the ROC curve that would be generated by varying λ can be generated from the ROC curve for the simple-stimulus alone case using Eq. (4). (For discrete probability distributions, the straight line segments between the discrete points generated by varying λ will not necessarily obey Eq. (4) although they do for high-threshold models.)

The predictions from all the maximum-output models discussed here were numerically generated using a program that embodied Eq. (4) above. In three cases, some analytic results were used to check the numerical calculations. For the exponential case, this amounted to using the above equation with symbols substituted for the numbers. For the high-threshold case, the maximum has again a high-threshold distribution, and one can solve for the new signal-strength parameter describing the maximum. Similarly, the maximum of doubly exponentially distributed random variables is again doubly exponentially distributed (and with the same variance) and one can solve for the new signal-strength parameters.

About the maximum-output double-exponential model and choice theory. The predictions from the maximum-output double-exponential model and from standard choice theory are identical for situations where the observer's response is based on which of several distributions produced the maximal value as in the forced-choice detection and the identification tasks used here (Yellott, 1977). The yes-no ROC curves predicted by the maximum-output double-exponential model here, however, are not the same as those predicted by the standard application of choice theory (Luce, 1963). The choice-theory yes-no ROC curves are symmetric around the negative diagonal quite unlike the maximum-output double-exponential model

here as is shown in Section 4.1. (It is the logistic distribution which predicts ROC curves like those predicted by choice theory, e.g., McNichol (1972)).

Yellot discusses a property he calls "invariance under uniform expansions of the choice set," a property which is slightly weaker than Luce's choice axiom. (The prediction in Section 4.2 by the maximum-output double-exponential model of no increment in the blocked-summation experiment is an example of it.) The double exponential is the only distribution which has this property when inserted into a class of Thurstonian models (essentially equivalent to the set of assumptions used here) and thus the only distribution which can predict choice probabilities satisfying Luce's choice axiom (Yellott, 1977).

3.2. *Sum-of-Outputs Detection Rule*

When the decision variable is the sum of M' random variables (Assumption 8B), there is no solution for the general ROC curve when $M' > 1$ and/or $P' > 1$ in terms of the ROC curve for the simple-alone condition.

To calculate the general ROC, one needs to compute the distribution function of the decision variable R, that is, of the sum of the R_i from 1 to M'. Of these M' random variables, P' will have signal distributions and the other $M'-P'$ will have noise distributions. In general, the probability density and distribution functions for the sum of independent random variables can be calculated as the appropriately defined convolution of the individual random variables' density or distribution functions. (See, for example, Feller 1966, Vol. 2, Chap. I, Sec. 2, p. 6, and Chap. V, Sec. 4.4, p. 142.) We did this convolution analytically for the high-threshold and exponential case. The sum of Gaussianly distributed random variables is well known to be Gaussian. Numerical methods could have been used for the double-exponential case, but the amount likely to be learned did not justify the cost.

3.3. *Forced Choice*

The proportion correct in a two-interval forced-choice procedure is predicted to equal the area under the ROC curve (Green & Swets, 1966) given continuous densities and the assumptions here. (Importantly, note that we are assuming that there is no bias toward saying one interval rather than the other; that is, the observer always says whichever interval yielded the highest value of the decision variable. And there is no correlation between the outputs on the two intervals). For the discrete high-threshold distribution, the forced-choice proportion correct can be easily derived from the distribution of the decision variable and the assumption about guessing. As it turns out, it also equals the area under the yes-no ROC curve.

3.4. *Identification*

The observer's task was to say which of the several possible simple stimuli had been presented on each trial under a simple-intermixed condition. In calculating proportion correct identification, trials of the blank stimulus (if there were any) are ignored and the proportion of all trials of simple stimuli on which the observer correctly identified the simple stimulus is counted.

Compared to detection under the simple-alone condition. Given a probability distribution such that the complete ROC curve is generated by varying λ and Assumptions 1 through 5 and 7, 8, and 9, then proportion correct identification under a simple-intermixed condition can be predicted from the ROC curve for detection under the simple-alone condition ($M' = P' = 1$) by the well-known area theorem (e.g., Green & Birdsall, 1978; Green & Swets, 1974, pp. 45–51). In the case of the high-threshold discrete distributions, the probabilities of correct identification are easy to derive from the distribution and the assumption of guessing in case of ties. These probabilities are, in fact, equal to the integral given by the area theorem.

For the constant-variance Gaussian case, the numerical values of integrals can be obtained from Elliot's table in Swets (1964) which was recently improved by Hacker and Ratcliff (1979) and an algorithm for extending them suggested by Smith (1982).

For the double exponential, the equivalence to choice theory allows quicker computations (although we did, in fact, do the area-theorem integration numerically as well).

Compared to detection under simple-intermixed. The ROC curve for detection under the simple-intermixed condition (assuming a maximum-output detection rule) can be directly related to identification performance under that same intermixed condition (Starr, Metz, Lusted, & Goodenough, 1975; presented and discussed in, for example, Green & Birdsall, 1978; Swensson & Judy, 1981). The known theorem makes a prediction for the joint probability of saying "yes" and of correctly identifying conditional on the presence of the signal and a particular criterion λ. To prove that theorem, the hit and false-alarm probabilities are assumed to be differentiable functions of λ (as they are for families *B* through *F* here—see Note 3) and assumptions 1 through 5 above as well as the maximum-output detection rule (Assumption 8A) and the maximum-output identification rule (Assumption 9) are assumed to hold (Green & Birdsall, 1978). The theorem can apparently not be extended to the case of discrete distributions. The theorem was not used in computing calculations for the maximum-output models here, but its use would presumably have led to the same predictions as those done directly from the simple-alone ROC curves.

4. Predictions

4.1. *Predicted ROC Curve Slopes*

Simple-alone condition. Figure 2 (top panel) shows the ROC curve for the simple-alone condition from each of the six pairs of density functions shown in Fig. 1. The plot is conventional (but see Note 4 for an argument against this convention). The probability of a yes given a stimulus (the hit rate) is plotted on the vertical axis and the probability of a yes given noise (the false-alarm rate) is plotted on the horizontal axis. In this figure, the axes give the normal deviate (the *z*-score)

corresponding to the probabilities. On z-axes like these, many empirical and theoretical ROC curves are approximately linear.

The signal-strength parameter for each family in Fig. 1 was chosen to make the full ROC curve intersect the negative diagonal at z-scores of -1 and $+1$ as can be seen in the top panel of Fig. 2. For the constant-variance Gaussian family (curve E), twice the ordinate of this intersection with the negative diagonal (2.0 in this example) always equals the value of that family's signal-strength parameter d' as

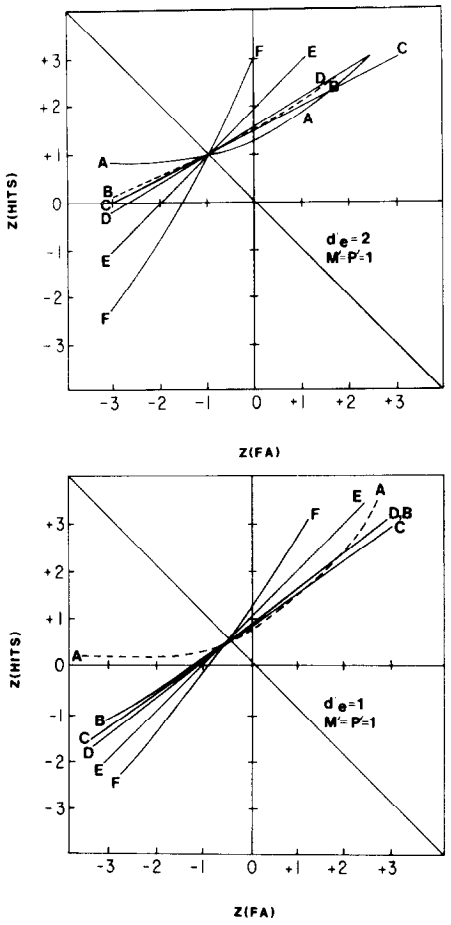

FIG. 2. ROC curves for the simple-alone condition from each of the six pairs of density functions shown in Fig. 1 (top panel) and for pairs from the same six families but yielding lower detectability values (bottom panel). The probability of a yes given signal (hit rate) is plotted against the probability of a yes given noise (false-alarm rate) on axes that are linear with the corresponding z-score (the standard normal deviate). Curve A is for the high threshold, B for the exponential, C for the increasing-variance Gaussian ($r=\frac{1}{3}$), D for the increasing-variance Gaussian ($r=\frac{1}{4}$), E for the constant-variance Gaussian, and F for the double-exponential probability distribution families. Curve A in the top panel and curve B in the bottom panel are dashed for clarity.

defined above. Since values of d' now have intuitive meaning for many people, we find it useful to use the following quantity in discussing predictions from any distributions:

DEFINITION. The negative-diagonal d' value for any pair of signal and noise densities is equal to twice the ordinate at the intersection of the negative diagonal with the ROC curve generated by that pair of densities. We will call that value d'_e.

The values of signal-strength parameter necessary in each of the six distribution families to give negative-diagonal d' values of 1, 2, and 3 under the simple-alone condition are given in the left half of Table 2.

Figure 2 (bottom panel) shows another set of ROC curves for the six probability distribution families, this time at a lower detectability level (negative-diagonal d' values of 1.0). For families A through D (high-threshold, exponential, increasing-variance Gaussians), the ROC curves become steeper as detectability decreases. In the constant-variance Gaussian family (E), the slope stays constant. In the double-exponential family (F), the curves become shallower as detectability decreases. For these six families of probability distribution, however, the relative ordering of the slopes stayed constant throughout the range of detectabilities studied (up to negative-diagonal d' values somewhat above 3.0 under the simple-alone condition) except for some minor reversals among those having very similar slopes (the exponential and the increasing-variance Gaussians—see Note 5).

Other experimental conditions. Full ROC curves were plotted from the 10 models at many levels of detectability for two or four ($N = 2$ and 4) simple stimuli intermixed under simple-intermixed conditions ($M' = N$; $P' = 1$) and for a compound stimulus containing two or four components ($N = 2$ or 4) under the compound-alone condition or, equivalently, under the simple-and-compound intermixed condition ($M' = P' = N$ on trials of the compound stimulus under either condition). In general the slopes predicted by any one model with a given value of signal-strength parameter do change with condition (but see Note 6), but the relative ordering of slopes stays the same as that in Fig. 2 with minor reversals among the ones with similar slopes.

4.2. *Predicted Uncertainty and Summation Effects*

The values of signal-strength parameters in Table 2 were used for the predictions shown in Figs. 3 through 5. Values of M' and P' are given on the horizontal axes. For uncertainty experiments, predicted proportion correct identification (Fig. 3) and detectability measured as negative-diagonal d' (Fig. 4) are plotted against the number of intermixed simple stimuli ($N = M'$). For blocked-summation experiments, detectability measured as negative-diagonal d' (Fig. 5) is plotted against the number of components in the compound ($N = P' = M'$). For Figs. 4 and 5, the sum-of-output predictions are shown in the left panel and maximum-output predictions in the right.

Figure 6 shows predicted detectability in intermixed-summation experiments as a

80/31/4-6

TABLE 2

Signal-Strength Parameter Values for Which the Negative-Diagonal d' Value Is 1, 2, or 3

Model	Parameter	Simple-alone blocks $M' = P' = 1$			Simple intermixed blocks $M' = 2,\ P' = 1$			Simple intermixed blocks $M' = 4,\ P' = 1$		
		$d'_e = 1.0$	2.0	3.0	1.0	2.0	3.0	1.0	2.0	3.0
a. High-threshold	p_i	0.555	0.811	0.928	0.555	0.811	0.928	0.555	0.811	0.928
b. Exponential	k'	1.01	2.05	3.18	1.17	2.24	3.36	1.36	2.40	3.54
c. Increasing-variance Gaussian	d'	1.20	3.00	6.00	1.4	3.4	6.2	1.7	3.75	7.
d. Increasing-variance Gaussian	d'	1.14	2.66	4.80	1.35	3.0	5.2	1.7	3.4	5.6
e. Constant-variance Gaussian	d'	1.00	2.00	3.00	1.3	2.3	3.3	1.55	2.6	3.6
f. Double Exponential	U_i	1.17	2.37	3.67	1.69	3.00	4.35	2.28	3.70	5.05

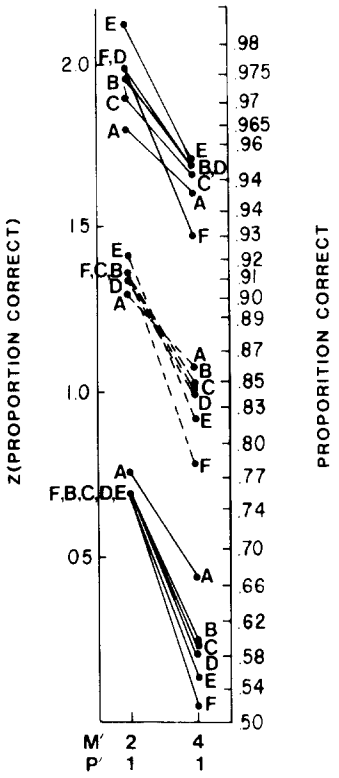

FIG. 3. For extrinsic-uncertainty experiments, predicted proportion correct identification is plotted on the vertical axis. The horizontal axis shows M' (the number of channels monitored) and P' (the number of channels characterized by the signal probability distribution). The number of intermixed simple stimuli equals M'. Curve A is for the high threshold, B for the exponential, C for the increasing-variance Gaussian ($r = \frac{1}{3}$), D for the increasing-variance Gaussian ($r = \frac{1}{4}$), E for the constant-variance Gaussian, and F for the double-exponential probability distribution families. Some curves are dashed for clarity.

function of number of components ($N = P'$). In this figure, the negative-diagonal d' of each of the components is equal to 1.0, 2.0, or 3.0 (as in Fig. 5) but this detectability is measured under an intermixed condition. Hence the values of signal-strength parameter necessary to obtain these negative-diagonal d' values must be computed separately depending on whether the number of components is 2 or 4; these values are given in the right hand sections of Table 2. (The predicted negative-diagonal d' levels for a compound containing four components ($N = P' = 4$) when the negative-diagonal d' for each component alone was 3.0 all exceeded 6.0, and so are not plotted in Fig. 6. Some of the points in Fig. 6 were obtained by interpolation between plotted ROC curves spaced about $0.1d'$ units apart rather than by direct calculation at the desired signal-strength levels.)

Conclusions about the predictions in Figs. 3–6 are discussed at length in Sections 4.3 and 4.4. For the moment, however, note that the curves in each group tend to

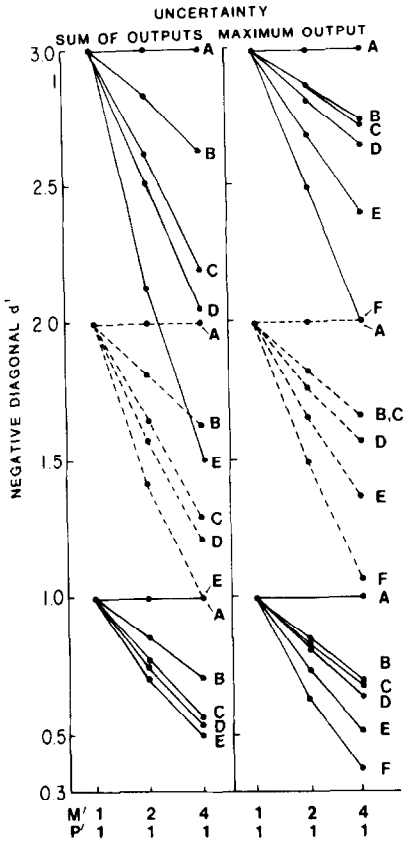

FIG. 4. For extrinsic-uncertainty experiments, predicted detectability measured as negative-diagonal d' is plotted on the vertical axis. The predictions using the sum-of-outputs rule are shown in the left panel and those using the maximum-output rule are shown in the right panel. Curve A is for the high threshold, B for the exponential, C for the increasing-variance Gaussian ($r=\frac{1}{3}$), D for the increasing-variance Gaussian ($r=\frac{1}{4}$), E for the constant-variance Gaussian, and F for the double-exponential probability distribution families. Other conventions as in Fig. 3.

be ordered from top to bottom much the way the distributions themselves were ordered in Fig. 1 which is also the ordering of their ROC slopes in Fig. 2.

Forced-choice detection predictions. When figures analogous to Figs. 4, 5, and 6 are drawn with percentage correct in two-interval forced-choice plotted on the vertical axis (rather than the negative-diagonal d' from the yes–no ROC curve), the figures look remarkably similar to Figs. 4–6 and so are not shown. Instead, plots giving predicted forced-choice proportion correct for a wide range of signal-strength levels are shown in Fig. 7 for each of the 10 distinct models and five experimental conditions individually.

Still another form of plot, where for each type of experiment predicted

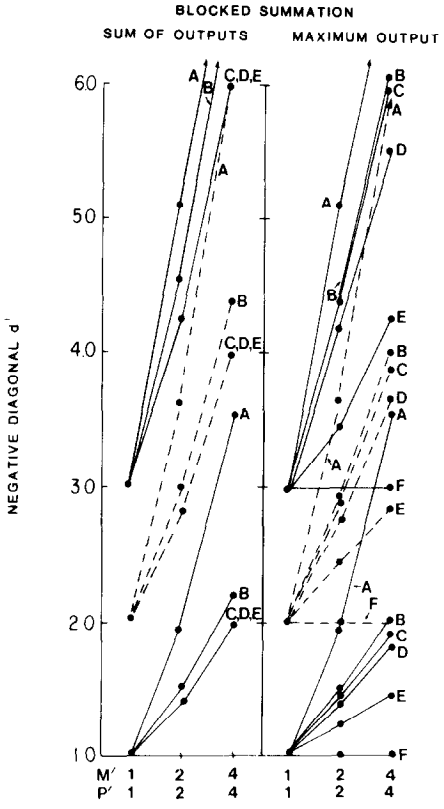

FIG. 5. For blocked-summation experiments, predicted detectability is plotted on the vertical axis. The number of components in the compound equals P'. Curve A is for the high threshold, B for the exponential, C for the increasing-variance Gaussian ($r=\frac{1}{3}$), D for the increasing-variance Gaussian ($r=\frac{1}{4}$), E for the constant-variance Gaussian, and F for the double-exponential probability distribution families. Other conventions as in Fig. 4.

probability correct at one value of N is shown as a function of predicted probability correct at another value of N, can be found in Kramer (1984).

4.3. *Conclusions about Predictions*

Four conclusions about the prediction of Figs. 2 through 7 are described in this subsection. An attempt to give some insight into the first three conclusions is given in subsequent subsections. These conclusions as stated apply only to the particular probability distributions studied here. The extent to which they may or may not generalize is discussed in Section 5.

CONCLUSION 1. The shallower the ROC curve, the smaller the decrement in extrinsic-uncertainty experiments, the greater the increment in summation experiments.

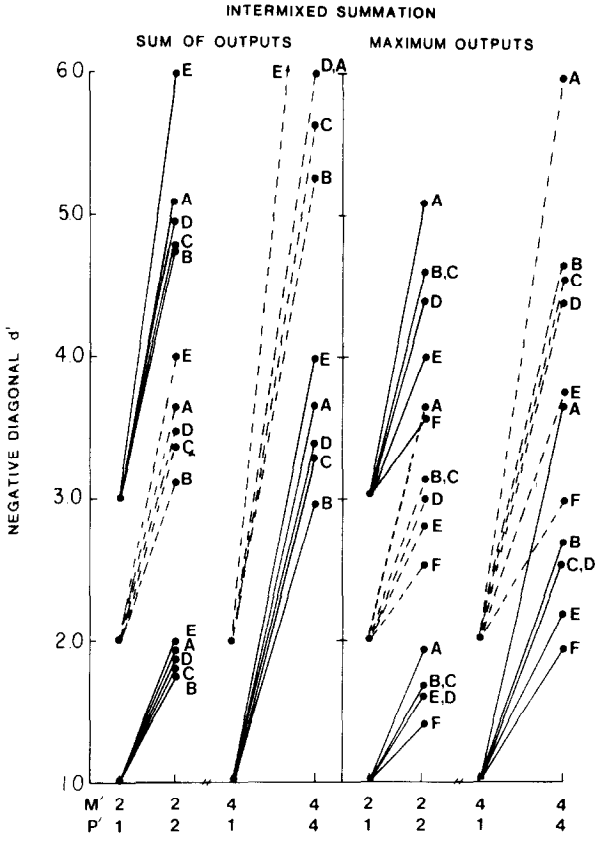

FIG. 6. For intermixed-summation experiments, predicted detectability is plotted on the vertical axis. For each combination rule, there are separate curves for experiments with different numbers of components in the compound. Those for two components are on the left of each panel and those for four components are on the right. Curve A is for the high threshold, B for the exponential, C for the increasing-variance Gaussian ($r = \frac{1}{3}$), D for the increasing-variance Gaussian ($r = \frac{1}{4}$), E for the constant-variance Gaussian, and F for the double-exponential probability distribution families. Other conventions as in Fig. 4.

In general, the curves in Figs. 2 through 6 are in the same order—top to bottom—as are the probability distribution families in Fig. 1. The predictions go from the extreme of no decrement due to extrinsic uncertainty but a large increment due to summation (the high-threshold model—A) to that of no increment due to summation in a blocked-summation experiment but a large decrement due to extrinsic uncertainty (the maximum-output double-exponential model—F). The trade-off between increments in summation experiments and decrements in uncertainty experiments is particularly orderly for blocked-summation experiments and the maximum-output detection rule.

The trade-off is less orderly with the sum-of-outputs rule than with the

maximum-output rule even when just considering blocked-summation experiments, and it breaks down rather completely in intermixed-summation experiments (reversing, in fact, for forced-choice).

Although we have not tried to specify a set of assumptions that would necessarily entail this relationship between ROC slopes and changes due to extrinsic uncertainty and summation, this relationship is much more than a peculiar coincidence. Some insight into the reasons behind this relationship—why it usually occurs but might break down—is presented in Section 4.4.

A Consequence of Conclusion 1. The effect of probability distribution is less in intermixed- than in blocked-summation.

Overall, there is less difference due to probability distribution in the intermixed-summation than in the blocked-summation. As indicated in Table I and discussed earlier, the effect in an intermixed-summation experiment equals the effect in a blocked-summation experiment plus the absolute value of the decrement in the

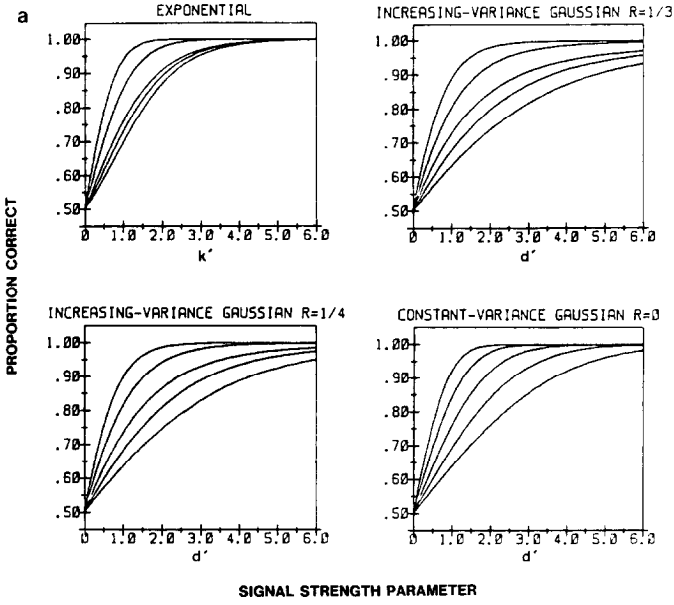

FIG. 7. Proportion correct detection in forced-choice predicted by each of the 10 distinct models described here (10 panels) plotted as a function of signal-strength parameter. The sum-of-output models are in (a) and the maximum-output models are in (b). Each panel shows predictions for five experimental conditions. For the eight panels which have five distinct curves, these conditions are two compounds (uppermost curve assumes $M' = P' = 4$ and next uppermost assumes $M' = P' = 2$), simple-alone (middle-curve, assumes $M' = P' = 1$), and two simple-intermixed (the second from the bottom curve assumes $M' = 2$ and $P' = 1$, the lowest curve assumes $M' = 4$ and $P' = 1$). For the high-threshold model, the compounds are as above but the curves for the simple-alone and the two simple-intermixed conditions coincide. For the maximum-output double-exponential model, the two simple-intermixed are the two bottom curves (as in the first eight models) but the curves for the simple-alone and the two compounds coincide.

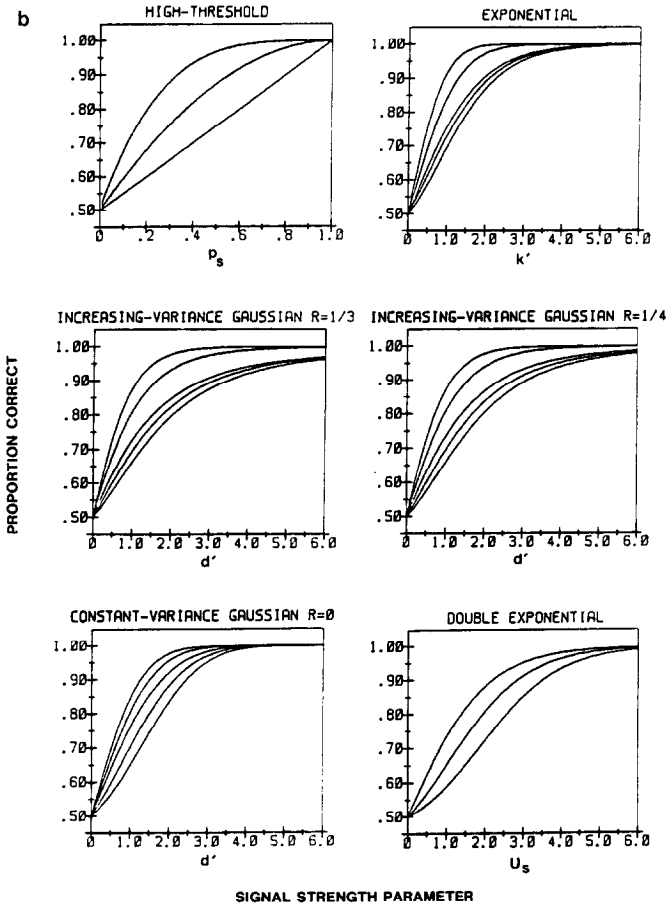

FIG. 7—*Continued.*

uncertainty experiment (using the same stimuli). Changing the probability distribution has opposite effects on these two terms which, therefore, tend to cancel one another out.

This cancellation may explain why the Quick pooling model (see references in Introduction and Note 9 for more description), which is a high-threshold model as usually derived, has done so well predicting experimental results in pattern vision. Most of the experimental summation results to which the Quick pooling model have been applied have been collected under intermixed conditions.

CONCLUSION 2. The sum-of-outputs rule tends to predict both greater increments due to summation and greater decrements due to extrinsic-uncertainty than does the maximum-output rule.

In Figs. 4 through 6, the curves in the left panels (for the sum-of-outputs rule) are usually steeper than the corresponding curves in the right panels (for the maximum-output rule). Loosely speaking, summing the outputs makes an observer more susceptible to either the effects of added noise distributions (in uncertainty experiments) or added signal distributions (in summation experiments).

Optimal strategies. In other words, of the two combination rules the sum-of-outputs rule predicts better detection performance on the compound stimulus under any condition and the maximum-output rule predicts better performance on simple stimuli under the intermixed conditions. Thus, if an observer has control over his combination rule, he ought to exercise it. He ought to use the sum-of-outputs rule under the compound-alone condition and the maximum-output rule under simple-intermixed conditions. Which he should use under the compound-and-simple intermixed conditions will depend on presentation probabilities, etc. Of course, under all conditions, he ought only to monitor the relevant channels.

This finding is quite consistent with previous results about optimality, although they have been derived only for constant-variance Gaussians. For constant-variance Gaussians, the optimal combination rule for detection of the compound is already known to be the sum-of-outputs rule (Green & Swets, 1974, Appendix 9A) and the optimal rule for detection under the simple-intermixed condition is known to be very well approximated by the maximum-output rule (Nolte & Jaarsma, 1967; see also Klein, 1985, and Note 7 here). The optimal rule for detection (of all stimuli) under the simple-and-compound intermixed condition is neither of these rules but lies inbetween as one would expect (calculated by James P. Thomas 1979, personal communication).

CONCLUSION 3. Combination rule and probability distribution interact.

The relative effects of the two different combination rules are not the same for all probability distributions. Consider, for example, the exponential family (B) and the increasing-variance Gaussian families (C and D) of probability distributions. They always make very similar predictions when combined with the maximum-output rule. (They must since their ROC curves under the simple-alone condition are so similar. See Section 3.1.) The effect of changing to the sum-of-outputs rule, however, is different for the two types of distributions. For uncertainty predictions (Fig. 4), there is almost no effect of changing rules for the exponentials (curves B) although there is a large one for the increasing-variance Gaussians (curves C and D). The opposite is true for blocked-summation experiments (Fig. 5), where the effect of changing rules is bigger for the exponential than for the increasing-variance Gaussians.

CONCLUSION 4. Identification performance predicted from simple-alone detectability depends little on distribution, particularly when two simple stimuli are intermixed ($N = 2$).

Consider first the $N = 2$ case. When detection performance measured as negative-diagonal d' under the simple-alone condition is equated for, all six probability distributions predict rather similar identification performances for two far-apart equally detectable stimuli (the $N = 2$ predictions in Fig. 3) particularly at low detectabilities. (Remember that any trials of blank stimuli are ignored in calculating proportion correct identification.) This similarity is not surprising, for—when forced-choice detection performance under the simple-alone condition is equated for (not shown)—all six distributions predict exactly the same identification performance at any level of detectability as is easy to prove from our assumptions. Both two-interval forced-choice detection under the simple-alone condition ($N = 1$) and identification under a simple-intermixed condition with two far-apart simple stimuli ($N = 2$) have the same structure according to our assumptions: Each is a choice between one random variable described by the signal distribution and one random variable described by the noise distribution. In two-interval forced-choice detection ($N = 1$), these two random variables correspond to the two outputs of a single channel in two intervals. In identification of two simple stimuli ($N = 2$—remember that trials of the blank stimulus are ignored), the two random variables correspond to the outputs from two channels. In both the detection and the identification cases, the observer is assumed to pick the maximum of the two random variables. And, in both cases, the observer's probability correct equals the probability that the maximum of the two comes from the signal rather than the noise distribution. Both probabilities correct are, in fact, predicted to equal the area under the yes–no ROC curve.

When four far-apart stimuli (equated for detectability under the simple-alone condition) are intermixed, identification performance does depend on probability distribution to some extent; such a dependence must exist to be consistent with the earlier conclusion that the size of the identification uncertainty effect depends on distribution.

Predicting identification performance from detectability under intermixed conditions. Identification performance after equating for detection performance in intermixed blocks (not shown here) does show a clear pattern of dependence on distribution. The best performance is predicted by the double exponential and worst by high threshold because the larger detection uncertainty effects dominate the smaller identification uncertainty effects.

4.4. *Some Insight into the Effect of Probability Distribution* (*Conclusion* 1)

When an observer in a detection task monitors additional random variables while keeping the criterion λ constant, there are always two effects—extra hits (due to above-criterion values of the decision variable on stimulus trials) and extra false alarms (due to above-criterion values of the decision variable on noise trials). Figure 8 shows an example. Suppose that under a simple-alone condition the observer's false-alarm and hit rates for a particular simple stimulus were 0.3085 and 0.6915 as plotted by the closed circle. The three lines through this point show three

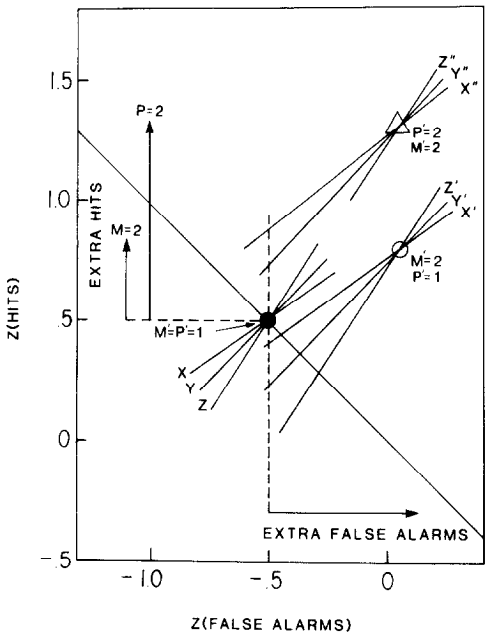

FIG. 8. An illustration of why the sizes of uncertainty and summation effects depend on ROC slope. An observer's predicted performance for a constant level of the criterion λ assuming the maximum-output detection rule is shown in three situations: simple-alone (closed circle), that same simple stimulus when trials of two simple stimuli are intermixed (open circle), a compound stimulus containing that simple stimulus and another equally detectable one (open triangle). These calculated points do not depend on probability distribution. The extra hits and extra false alarms resulting from monitoring more than one channel are indicated by the horizontal and vertical arrows. The vertical arrow marked "$M=2$" is for the comparison between the simple-intermixed (open circle) and the simple-alone (closed circle) conditions; the vertical arrow marked "$P=2$" is for the comparison between the compound stimulus (open triangle) and simple-alone condition (closed circle). The three lines through each point show three possible ROC curves (with slopes of $\frac{1}{2}$, 1, and 2) coming from three different hypothetical families of probability distributions.

possible ROC curves (with slopes of $\frac{1}{2}$, 1, and 2) coming from three different hypothetical families of probability distributions.

Uncertainty experiments. The observer's performance on this same simple stimulus under a simple-intermixed condition with $N=2$, assuming the observer continues to use the same criterion λ, is plotted by the open circle in Fig. 8. The observer is now monitoring two random variables on each trial ($M'=2$): the original random variable plus one that is sensitive to another simple stimulus but not to the original one ($P'=1$).

On noise trials, both these random variables have a probability of 0.3085 of exceeding criterion (or 0.6915 of *not* exceeding it), and the probability that at least one (and therefore the maximum) exceed the criterion is $(1-0.6915^2)=0.5218$ which is plotted as the horizontal coordinate of the open circle. This increase from

0.3085 to 0.5218 in false-alarm rate is indicated by the horizontal arrow at the bottom of the drawing labelled "extra false alarms."

On trials of the original simple stimulus, the original random variable (the one sensitive to the stimulus) exceeds the criterion with a probability of 0.6915 but the other random variable exceeds the criterion with a probability of only 0.3085. The probability that at least one (the maximum) exceeds the criterion has a value, therefore, of $(1 - 0.3085 \times 0.6915) = 0.7867$ which is plotted as the vertical coordinate of the open circle. (On trials of the second simple stimulus, the random variables reverse roles but the probability is the same.) According to the maximum-output rule, in short, the hit rate has increased from 0.6915 to 0.7867 as indicated by the short vertical arrow at the left hand edge of the drawing labelled ($M = 2$) where M indicates the number of simple stimuli intermixed.

Note that the coordinates just computed for the open circle do not depend on probability distribution at all.

The slope of the ROC curve through the open circle will depend on the probability distribution, however. The three straight lines again illustrate three possible ROC curves and have the same slopes as those through the solid point. They are labelled by the same letters (although primed) as the curves through the simple-alone condition point to represent the fact observed in our calculations that the relative ordering across probability-distribution families of ROC slopes remained the same under all conditions.

Suppose now that the observer changes his criterion λ under the simple-intermixed condition ($M' = 2$ and $P' = 1$) to reduce the false-alarm rate to the rate under the simple-alone condition ($M' = P' = 1$). The performance predicted by a particular probability distribution is found by travelling leftwards in Fig. 8 from the open circle along the appropriate ROC curve to its intersection with the dashed vertical line through the closed circle. Note that that intersection is generally below the closed circle—the hit rate under the simple-intermixed condition is less than the hit rate under the simple-alone condition when false-alarm rates have been equated; that is, a decrement due to uncertainty is predicted. Note also that the shallower the ROC curve, the higher the intersection and the *smaller* the amount of predicted decrement due to uncertainty.

Inspection of Fig. 8 shows that the conclusion about the relationship between ROC slope and size of the decrement due to uncertainty applies also to the relative positions of intersections with the negative-diagonal (which give the value of d'_e). Further, given curves that are close to linear on these axes and maintain the same relative slopes under all conditions (as the hypothetical curves in Fig. 8 do), the conclusion can also be extended to the relative positions of the full ROC curves (curve X' is entirely below curve X, for example, but by less than curve Z' is below Z), and hence to forced-choice performance as well.

Blocked-summation experiments. An entirely analogous argument holds for blocked-summation experiments. Briefly, the open triangle shows the predicted false-alarm and hit rates under a compound-alone condition where the compound

contained two equally detectable components ($N = M' = P' = 2$) assuming the same criterion λ as that for the solid and open circles and assuming the maximum-output rule. The false-alarm rate is $0.5218 = (1 - 0.6915^2)$. The hit rate is 0.9048 (one minus the probability that neither random variable exceed λ in response to the compound to which both are sensitive—that is $1 - 0.3015^2$). The extra hits are indicated by the long vertical arrow labelled $P = 2$ (where P is used to indicate number of components in a compound).

When travelling leftwards from the open triangle along some ROC curve, the intersection is generally above the closed circle showing that the hit rate under the compound-alone condition will be greater than that under the simple-alone condition when false-alarm rates are equated. Similarly, performance under the compound-alone condition measured as d'_e is better than that under the simple-alone condition. Further if ROC curves are reasonably linear and maintain the same relative slopes under all conditions, the full ROC curve for the compound-alone condition will be higher than that for the simple-alone (curve X'' is higher than curve X). In all these comparisons, the shallower the ROC curve, the *greater* the predicted effect of summation (curve X'' is higher than curve X, for example, by more than curve Z'' is higher than curve Z).

Intermixed-summation experiments. As illustrated previously in Table 1, the comparison between a compound-alone condition (e.g., the open triangle in Fig. 8) and a simple-intermixed condition (e.g., open circle in Fig. 8) is equivalent to the comparison in an intermixed-summation experiment. Thus the difference between the hit rates of those two points is the predicted size of the intermixed-summation effect at a constant false-alarm rate and is given directly by the kind of calculations done for Fig. 8, calculations which made no assumption about probability distribution or ROC slope (although they do assume the maximum-output rule). Thus, the size of the intermixed-summation effect at a constant false-alarm rate is the same for all probability distributions!

Other measures of relative performance (e.g., hit rate at the intersection with the negative-diagonal, forced-choice) will produce some effect of distribution (ROC slope) in intermixed-summation experiments although not as great as that in blocked-summation experiments.

An abbreviated argument. Loosely speaking, on a shallow ROC curve, each hit is worth several false alarms. (You can add several extra false alarms for every extra hit and still stay on the same ROC curve.) The summation effect is the effect of extra hits on compound-stimulus trials "minus" the effect of extra false alarms on noise trials. So, the more effective a hit is relative to a false alarm (the shallower the slope of the ROC), the bigger the increment due to summation.

Conversely, on a steep ROC curve, each false alarm is worth several hits. The extrinsic-uncertainty effect is the effect of extra false alarms on noise trials "minus" the effect of extra hits on simple-stimulus trials (the hits of which are spurious hits by insensitive channels). So, the more effective a false alarm is relative to a hit (the steeper the slope of the ROC curve) the bigger the extrinsic-uncertainty effect.

Generalizability of argument. At best the argument illustrated in Fig. 8 is far from complete or rigorous. For example, if performance under the simple-alone condition (the solid point) were considered in a very different part of ROC space, a figure corresponding to Fig. 8 would show somewhat different relative positions of the three points. The exact ROC slope necessary to get no uncertainty effect, for example, would change. (It changes, in fact, in exactly the way that ROC slopes from the high-threshold model change.)

Further, the argument assumes that ROC curves from different probability distributions tend to have the same relative slopes under different experimental conditions. They do (see Section 4.1 and Note 6), but this argument provides no insight into why.

In spite of the argument's lack of rigor, the argument makes it clear that slope will matter everywhere in ROC space. As long as the slopes of different probability distributions maintain the same relative order in different parts of ROC space and under different conditions, their predicted uncertainty and summation effects will retain the same order in different parts of ROC space and thus in forced-choice.

It is also clear from this argument that probability distributions producing ROC curves that have vastly different slopes in different parts of ROC space and/or under different conditions will have no invariant position in an ordering of probability distributions like that in Figs. 1 through 6. Whether they predict more or less of an effect than some other distribution will depend on what part of the ROC space is being considered and exactly which effect is being considered. See Section 5.3 for examples.

Note that, although the argument as shown in Fig. 8 assumes linearity on normal–normal probability axes (z-axes), the argument can clearly be used on any kind of axes where there was approximate linearity.

The argument in Fig. 8 is for the maximum-output decision rule but can be applied in a watered-down version to the sum-of-outputs rule. Using the sum-of-outputs rule, the ROC curves from different distributions for the simple-intermixed or for the compound-alone condition will not all intersect at the same point as they do in Fig. 8, and thus the argument will involve more approximation and depend on particular probability distributions.

4.5. *Some Insight into the Effect of Combination Rules*

With the sum-of-outputs rule, the additional random variables monitored as N gets larger affect the decision variable (the sum) on every trial. With the maximum-output rule, however, the additional random variables affect the decision variable (the maximum) only on those trials on which one of them happens to produce the maximum output. Looked at this way, it seems not unreasonable that both decrements due to uncertainty effects and increments due to summation—both of which are effects of monitoring additional random variables—should tend to be greater with the sum-of-outputs than with the maximum-output rule.

That the effects of changing from the maximum-output to the sum-of-output rule are different for different probability distributions is not surprising either. The

maximum-output predictions depend only on the ROC curve for the simple-alone condition or, to put it another way, only on the ratios of the noise and signal probability densities at various criteria. If one takes a given pair of signal and noise density functions and then non-uniformly stretches the z-axis underneath both, the maximum-output predictions stay the same. The sum-of-output predictions do not stay the same after non-uniform stretching, however, for the distribution of the sum depends on the actual magnitudes of all the numbers that went into it.

5. Discussion

The argument in Fig. 8 suggests that Conclusion 1 should generalize to all models predicting ROC curves that are approximately linear on z-axes and of the same slope under all experimental conditions. Section 5.1 below presents some further examples which are of interest also on their own.

We wondered if models could be found with such shallow or such steep ROC curves that they would predict the opposite of the usual extrinsic-uncertainty or probability-summation effects. We know of no cases with ROC curves shallower than the high-threshold model's, that is, no cases where increasing uncertainty (increasing the number of intermixed simple stimuli) actually improved performance. But we did find two cases with ROC curves steeper than the double-exponential maximum-output model. These two cases do indeed predict a *decrement* in the detectability of a compound stimulus as the number of components is increased (in a blocked-summation experiment); that is, a compound stimulus is predicted to be less detectable than any of its components (although the components stimulate independent and non-overlapping channels). These two cases are described in Section 5.2 below.

Conclusion 1 cannot generalize as originally stated to models where the slope changes dramatically in different regions of the ROC space or across experimental conditions. We asked two questions therefore: (i) Can models be found that predict radically different effects in different regions of the ROC space? (ii) For such models, will behavior in local regions of ROC space show the relationship described in Conclusion 1? Yes is the answer to both questions as described in Section 5.3 below.

As our insight into the effect of combination rule (Conclusions 2 and 3 above) is even less rigorous than that into the effect of probability distribution, we wondered whether cases could be found where the sum-of-outputs rule or the maximum-output rule was always the better. Reading Green and Birdsall (1978) and further computation of our own convinced us the answer was yes as described below in Section 5.4.

5.1. *Other Distribution Families That Fit into the Ordering of the Original Six*

Other Gaussian families. Gaussians with other ratios of signal/noise variance certainly would fall into place in the orderings shown in Figs. 1 through 6. In fact, for a sum-of-outputs rule, it is easy to prove analytically that they do.

Poisson families. Poisson distributions are particularly interesting to visual scientists as they describe the probabilistic nature of light. Nachmias (1972; also see Nachmias and Kocher, 1970) computed the predictions for uncertainty detection experiments from two families of Poisson distributions combined with the maximum-output rule ($M = 1$, 4, and 16). In one family, the mean of the noise density function was 1.3 and in the other it was 40. The slopes of the ROC curves under the simple-alone condition are approximately equal to 1, and the predicted uncertainty effect is very similar to that of constant-variance Gaussians.

The mean and variance of a Poisson are always equal to each other, and in the limit a Poisson is well approximated by a Gaussian. Thus if the noise distribution is Poisson with a high mean and the signal distribution is Poisson with slightly higher mean, the pair will be well approximated by Gaussian distributions of constant variance and will act accordingly (e.g., see Note 7).

Intrinsic-uncertainty distribution. Some inability of the observer to completely ignore non-informative sources of information can be represented by building some "intrinsic uncertainty" into the random variable R_i corresponding to a single simple stimulus (e.g., Tanner, 1961; Nachmias and Kochar, 1970; Nachmias, 1972; Pelli, 1981, 1985). The random variable R_i is itself taken to be the maximum of many (M'') probabilistically independent random variables R_{ij}, most of which represent the outputs of "micro-channels" that are not sensitive to the stimulus:

$$R_i = \underset{j=1}{\overset{M''}{\text{MAX}}} R_{ij}.$$

In Pelli's (1981, 1985) extensive calculations, one of these random variables (let it be R_{i1}) has a unit-variance Gaussian distribution with a mean that increases linearly with the intensity of the stimulus. All the other micro-channels' outputs (R_{ij}, $j = 2$ to M'') have unit-variance Gaussian distributions with means always equal to zero. We will call the distribution of R_i under these assumptions the "intrinsic-uncertainty distribution."

Pelli's calculations showed that this intrinsic-uncertainty distribution predicts quite well the shapes of psychometric functions for simple visual patterns. (More precisely, it predicts that the psychometric functions are well approximated by the formula suggested by Quick with parameters changing with false-alarm rate in the manner found empirically— see Nachmias, 1981.) It must also predict, therefore, ROC curves like those for simple visual patterns, and these ROC curves get shallower with increased detectability in the way those from the exponential (B) or increasing-variance Gaussian (C and D) families do. Hence, multiple random variables R_i that are each described by the intrinsic-uncertainty probability distribution—particularly when the observer's performance obeys the maximum-output detection rule (Assumption 8A)—must predict increments due to summation and decrements due to extrinsic uncertainty much like those predicted by the exponential or increasing-variance Gaussian family. For intermixed-summation

experiments the intrinsic-uncertainty distribution family should predict much the same increments due to summation as do high-threshold models, in particular, the Quick pooling model. Some explicit calculations of Pelli (1985) show that they do. (See Note 8).

Other extreme-value distributions. The double exponential is one of three distributions that can arise as the asymptotic distribution for the maximum of a large number of independent and identically distributed random variables (Gumbel, 1958; Galambos, 1978; a short and readable introduction appears in Wandell & Luce, 1978). The double exponential is the asymptotic distribution in the case of distributions with exponential upper tails (normal, gamma, exponential, and the lognormal) and is sometimes referred to as *the* extreme-value distribution. We will call the two other asymptotic distributions the "high-tailed limit distribution" and the "bounded-from-above limit distribution" reflecting the kinds of distributions for which they are the asymptotic function. (Equations can be found in the references above.) We formed two families of distributions from each of these (see Note 2). In the "shifted" family, the member distributions all had the same variance but different central tendencies (like the double-exponential distributions of Assumption 6E); in the "spread" family, the member distributions all had the same anchor point but different variances (like the exponential family of Assumption 6B). Predictions were computed from the maximum-output rule for detection under three conditions: simple-alone ($M' = P' = 1$), simple-intermixed ($M' = 4$, $P' = 1$), and compound-alone ($M' = P' = 4$). As expected, the relationships between ROC slope and sizes of uncertainty and summation effects (Conclusion 1) held for all four families. The shifted bounded-from-above family fell into the ordering of probability distributions at approximately the same place as the constant-variance Gaussians. (See Note 9 for a comparison with the Quick pooling model.) The two spread families fell into place with the shifted double-exponential family. (In fact these two families turn out to be exactly equivalent to that family. See Note 2.) The fourth case was by far the most interesting, however, as described in the next subsection.

5.2. *Distributions That Are Even More Extreme*

Shifted high-tailed limit family. The shifted high-tailed limit family fell into place below the shifted double-exponential family. That is, it predicted very steep ROC curves (slopes of about 2.5, 5, and 10 at negative-diagonal d' values of 1, 2, and 3, respectively). It predicted an uncertainty effect even larger than the shifted double-exponential family does. And it predicted the opposite of summation in a blocked-summation experiment—that is, it predicted lowered detectability as the number of components became larger. The extra false alarms due to monitoring the extra channels more than cancelled out the extra hits!

We found a second distribution that predicts decrements due to summation in a blocked-summation experiment—a low-threshold two-state discrete distribution—

but that was a case that predicted different behaviors in different regions of ROC space and is, therefore, discussed in the next section.

5.3. *Different Behaviors in Different Regions of ROC Space*

Low-threshold two-state discrete distributions. The low-threshold two-state discrete distributions are like the high-threshold distributions (Assumption 6A) except that p_i(noise) is greater than 0. This distribution has been studied frequently (e.g., Luce, 1963a, b; Atkinson, Bower, & Crothers, 1965, pp. 199–200; Green & Birdsall, 1978; Green & Swets, 1974, pp. 140–145; Krantz, 1969). When combined with a sum-of-outputs rule, Green & Birdsall (1978) called this model the discrete-sum model. When combined with a maximum-output rule, they called it the discrete disjunctive model. Since the decision variables are discrete, mixture strategies are invoked to complete the ROC curves.

We did calculations for a particular one-parameter family of low-threshold distributions, sometimes called the symmetric two-state model, where the probability that noise *not* exceed the threshold is assumed to equal the probability that the stimulus *exceed* the threshold. (See Note 2.) This model predicts that the limb of the ROC curve to the left of the negative diagonal under the simple-alone condition (which is what determines behavior at the negative diagonal for other conditions) is very steep, steeper than that of the double exponential in fact. And it turns out that with the maximum-output rule the symmetric two-state distribution does predict very large uncertainty effects and a *decrement* due to summation (at the negative diagonal and throughout the range of low false-alarm rates). In other words, at low false-alarm rates the maximum-output symmetric two-state model falls into place below the maximum-output double-exponential model in the ordering of Figs. 2–6. With a sum-of-outputs rule, comparison with the double-exponential sum-of-outputs is not possible (since calculations were not done for it), but the symmetric two-state distribution does fall into place below the next distribution (the constant-variance Gaussian) for low false-alarm rates.

At very high false-alarm rates, the ROC curve for both the sum- and the maximum-output rules with the symmetric-two-state distribution is the same and is very shallow. Indeed, both the sum and the maximum symmetric-two-state models act like the high-threshold model at high false-alarm rates, predicting no uncertainty effect and a good deal of summation.

For forced-choice behavior, that depends on the full ROC curves, one might expect some intermediate behavior from this distribution. Green and Weber (1980) showed that the predicted uncertainty effect in forced-choice detection is very close to that predicted by the constant-variance Gaussian. We calculated identification uncertainty effects (they also depend on the full ROC curve) and found that they too were like those from constant-variance Gaussians.

Piecewise-continuous high-threshold distributions. In general, piecewise-continuous (rectangular) two-state density functions have the form

$$
\begin{aligned}
f_i(z) &= 0 && \text{for} \quad z \leqslant 0 \\
&= (1 - p_i)/w && \text{for} \quad 0 < z \leqslant w \\
&= 0 && \text{for} \quad w < z \leqslant 1 \\
&= p_i/w && \text{for} \quad 1 < z \leqslant 1 + w \\
&= 0 && \text{for} \quad z > 1 + w.
\end{aligned}
$$

We initially considered this kind of distribution because (Green & Swets, 1966) if p_i(noise) equals zero, it is equivalent to standard high-threshold theory under the simple-alone condition and yet will produce a full ROC curve when criterion λ is varied, avoiding the necessity of postulating mixture strategies. (In general, however, when more than one channel is involved these piecewise-continuous density functions turn out not always to be equivalent to the corresponding discrete ones. See Note 2. For the maximum-output rule and high-threshold case, the equivalence continues, however.)

Of interest in this section is the fact that with the sum-of-outputs rule, the above equation with p_i(noise) $= 0$ and $w = 1$ predicts different behavior in different parts of ROC space. At a false-alarm rate of zero, the ROC slopes for N equal to 2 are very steep, and this model acts like the maximum-output double-exponential model (predicting no increment at all in a blocked-summation experiment and a very large decrement in an intrinsic uncertainty experiment). For high false-alarm rates, on the other hand, the ROC slope is shallow under all conditions, and this model behaves like the discrete high-threshold model (predicting a large increment due to summation and no decrement due to uncertainty).

5.4. *Limitations on Conclusion 2 about Combination Rules*

Both models providing examples above of local variations in ROC space also provide examples limiting our previous conclusion about combination rules (Conclusion 2).

Low-threshold discrete models: Sum better than maximum even under a simple-intermixed condition. With discrete low-threshold distributions, no matter what the condition, the ROC curve for the sum-of-outputs rule is always further from the positive-diagonal (chance performance) than the maximum-output ROC curve except at high false-alarm rates where they are identical. In fact, not only is the sum-of-outputs rule better than the maximum-output rule, Green and Birdsall (1978, p. 203) proved that the sum-of-outputs is monotonic with the likelihood ratio and hence is better than any rule (is an optimal detection rule) for this distribution.

Piecewise-continuous high-threshold: Maximum better than sum even for the compound stimulus. The rectangular high-threshold distribution limits our earlier generalization about combination rules in the opposite way: for this distribution the maximum-output rule always predicts detectability that is at least as good as that

predicted by the sum-of-outputs rule even on trials of the compound stimulus! This lower performance with the sum-of-outputs rule occurs because a sum greater than 1 sometimes occur (as it should) when one or more channel outputs are in the detect state (between 1 and 2) but it also sometimes occurs even though all channel outputs are in the not-detect state (between 0 and 1). Thus combining the outputs into a sum loses diagnostic information.

Although limiting the generality of Conclusion 2 about the effect of combination rule (the blanket statement that the maximum was better for simple-intermixed conditions and the sum for the compound stimulus), the distributions in this subsection are further examples buttressing Conclusion 3—that the effect of combination rule depends on distribution (and vice versa). It is, regrettably, the usual condition of models that the effect of one assumption depends on another.

5.5. *Concluding Remarks*

As shown above, the decrements due to uncertainty (sometimes called set-size effects) and the increments due to summation (sometimes called probability summation effects) that are predicted to arise from multiple, independently variable channels are remarkably dependent on the assumed form of the probability distributions characterizing the channels' outputs. Or, to put it another way, they are remarkably correlated with the slope of the predicted ROC curves. Indeed, the dependence on assumed probability distribution is as large as the effects themselves. Predictions range from absolutely no uncertainty effect and a good deal of summation (in blocked-summation experiments) to a good deal of uncertainty effect and no summation or even anti-summation (in blocked-summation experiments).

On consideration (illustrated in Fig. 8) it became obvious why these particular effects of assumed probability distribution occur. They result from the trade-off between the helpful effect of extra hits and the harmful effect of extra false alarms that occur when the observer monitors additional channels.

These large effects of assumed probability distribution make it dangerous to use any particular assumed distribution (e.g., constant-variance Gaussian) in one's models without at least some evidence as to the form of the probability distribution or the slope of the ROC curve. To put it another way, the assumed probability distribution can affect one's substantive conclusions dramatically. On the high-threshold model, any decrement at all due to uncertainty suggests that there is a limit to attention capacity; but, on the double-exponential model, very large decrements due to uncertainty reflect nothing but noisy channels. Similarly, on the double-exponential model, any increment due to summation (in blocked-summation experiments) suggests that individual channels are overlapping (sensitive to two or more components in the compound) while on the high-threshold model, considerable increment is consistent with non-overlapping channels. As these are examples of substantive conclusions that investigators have often wanted to make, their dependence on probability distribution must be taken very seriously.

The predicted decrements and increments also depend on the combination rule, both tending to be larger when the outputs of the channels are summed than when

the maximum output is taken. Or, to put it the other way, the sum-of-outputs combination rule is better for the compound stimulus and the maximum-output rule is better for intermixed simple stimuli. (Exceptions to this rule can be found, however, when two-state distributions are considered.)

The systematic nature of the effects of probability distribution and combination rules leads one to wonder what other effects might depend in a systematic but dramatic way on the assumed probability distribution or ROC slope.

These results also make us wonder what features of the shape of probability distributions correlate with the predicted slope of the ROC curve (on z-coordinates). It is well known, of course, that for ROC curves entirely generated by varying a criterion λ the slope of the ROC curve on linear coordinates at any point equals the likelihood ratio $F(\lambda/\text{stim})/F(\lambda/\text{noise})$ at the criterion λ which led to that point. This, however, does not seem to be a complete understanding of why various shapes of distributions (e.g., those in Fig. 1) lead to different slopes and hence to different uncertainty and summation effects.

Situations that are less susceptible to the assumed probability distribution and justification for use of the Quick pooling model. Interestingly, the effect of assumed probability distribution is a good deal smaller in intermixed-summation experiments (when the observer is monitoring the same set of channels on every trial) than in blocked-summation experiments (when the observer is monitoring different sets of channels on trials of different stimuli).

This insensitivity to distributional assumptions under intermixed conditions may explain the success of the Quick pooling model—which in its most straightforward form is a high-threshold model—and is thus demonstrably wrong for pattern vision. (See Introduction and Note 9 here for more description, but the other sources given in the Introduction will have to be consulted for a fuller description.) The thresholds to which this model has been quantitatively compared have generally been collected under what are essentially intermixed-summation conditions (the observer is monitoring all the channels sensitive to any of the stimuli being compared). The insensitivity of intermixed-summation predictions to assumed probability distribution provides some justification for the continued use of the Quick pooling model (as an approximation to more reasonable models that predict both uncertainty effects and ROC curves of the appropriate slope). Since the Quick pooling model is an exceedingly convenient way to do calculations when multiple channels are involved, thus justification for its continued use is desirable.

APPENDIX: NOTES

1. *Criterion variance.* In comparing experimental ROC curves to theoretical predictions (which we do not do here) the value of the criterion λ is almost always assumed to be constant throughout the trials from which results are pooled to get an empirical ROC curve. Implications of criterion variance across trials (random

and systematic) have been considered by Nachmias and Kocher (1970), Treisman and Williams (1984), and Wickelgren (1968).

2. *Supplemental manuscript.* Many details omitted from this manuscript are given in an unpublished manuscript available from the authors.

3. *Differentiability of hit and false-alarm rates.* The hit and false-alarm rates are just one minus the appropriate probability distribution (see Eq. (1) in Section 3) and hence will be differentiable functions of the criterion λ whenever the distribution function is. If $f(x)$ is continuous at a value, the distribution function $F(x)$ is differentiable at that value. For the Gaussian and double-exponential families (C through F) considered here, the density functions $f(x)$ are continuous everywhere. For the exponential family (B) the densities are continuous everywhere except at zero, a discontinuity of no consequence here as a criterion of zero produces hits and false alarms equal to 1.0. The high-threshold family is discussed in the text.

4. *Objection to conventional ROC plots.* Sperling and Dosher (1986, Sec. 3.3) argue against this convention on the grounds that it counter-intuitively plots good performance on signal trials against bad performance on noise trials. They prefer mirror ROC curves in which hit rates are plotted against correct rejections (1-false-alarm rates) so good performance is plotted up and to the right. This also is the convention used for other types of "performance operating characteristics."

5. *Comparison of the exponential and unequal-variance Gaussian families.* The exponential (B in Figs. 1–6) and the two unequal-variance Gaussian families considered here (C and D) lead to ROC curves under the simple-alone condition (and thus under any condition using the maximum-output rule) that are very similar. In particular, the relationship between the slope of the ROC curve on z-axes and detectability is very much the same for all three families when both slope and detectability are measured in the mid-range of the ROC plot—in the range from the negative horizontal axis to the positive vertical axis. Furthermore, throughout the mid-range the exponential curves are very straight on z-axes although departures from straightness will occur across the whole range. (More examples of exponential ROC curves can be found in Green & Swets, 1974, p. 80)). The unequal-variance Gaussian ROC curves are completely straight on z-axes, of course.

In the exponential family, likelihood ratio is actually monotonic with the channel output but it is not monotonic for the unequal-variance Gaussian families. Thus, the ROC curve plotted on linear coordinates for the exponential distribution will have a slope that is continually decreasing from left to right (so the curve is always bowed downwards on these axes) but the ROC curves plotted on linear coordinates for the unequal-variance Gaussians will have a slope that decreases for most of the range (so the curve is bowed downwards) but increases again in the upper right (so the curve is bowed upwards at very high hit and false-alarm rates). See illustrations on p. 63 of Green & Swets (1974).

6. *Invariance of signal-strength families of ROC curves.* For any one probability distribution and any one experimental condition, consider the family of ROC

curves generated by plotting an ROC curve for every value of the signal-strength parameter. We will call this infinite family of curves a signal-strength family. One can ask whether two signal-strength families from the same probability distribution family for two different conditions are identical even if, in general, different values of signal-strength parameter are required under the two different conditions to generate the same member of the family. In other words, one can ask whether the ROC curve going through any arbitrary point (any particular combination of false-alarm rate and hit rate) is the same under both experimental conditions although it may require a larger signal-strength parameter to generate that curve under one condition than under the other.

This kind of invariance can be shown to hold across all values of N under the conditions considered here (Table 1) for 3 of the 10 distinct models being considered here: for the high-threshold model (with either combination rule—6A with 8A or 8B), for the maximum-output double-exponential model (6F with 8A), and for the sum-of-outputs constant-variance Gaussian model (6E). It does not hold for the others. See discussion of Assumption 6 in Section 2.5.

7. *Optimality at low intensities.* For asymptotically low intensities and for both constant-variance Gaussian and Poisson distribution families, the optimal rule for detection under the simple-intermixed condition has been reported to equal the sum-of-outputs rule (Cohn, 1978). Combining this result with that of Nolte and Jaarsma mentioned above, one would expect the sum-of-outputs rule and the maximum-output rule to make similar predictions for the constant-variance Gaussian family at low detectability levels under the simple-intermixed conditions of uncertainty experiments as they did in our calculations (curves E in bottom panels of Fig. 4).

8. *The pedestal effect.* Of less relevance here, but potentially quite important, the assumption about the changes of the intrinsic-uncertainty distribution with contrast leads it to also predict the "pedestal effect"—the fact that intensity discrimination near threshold is better than detection (e.g., Nachmias & Kocher, 1970; Pelli, 1985). If all the predictions just reviewed turn out to be obtainable with the same parameter M'' (when data are collected from the same observer under the same conditions) this intrinsic-uncertainty model will indeed be a marvelously good description of the detection process, at least for visual patterns.

9. *The Quick pooling model and extreme-value distributions.* There is an interesting and potentially confusing relationship between the maximum-output shifted bounded-from-above model and the Quick pooling model (which is one of the reasons we investigated these extreme-value distributions).

The Quick pooling model (e.g., Graham *et al.*, 1978) is the high-threshold model above with the additional assumption that the signal-strength parameter p_i (the probability that the channel's output is in the detect state) grows with signal strength c according to the function

$$p_i = 1 - e^{-c^k}. \tag{5}$$

The equation for the bounded-from-above limit density function (the asymptotic density function of the maximum of independent and identically distributed random variables that are bounded from above) is

$$f_i(z) = 0 \qquad \text{for } z \geqq c_i$$

$$= -k \frac{(z-c_i)^{k-1}}{V_i^k} \exp\left\{ -\left| \frac{z-c_i}{V_i} \right|^k \right\} \qquad \text{for } z < c_i,$$

where $V_i < 0$ and $k > 0$.

In the shifted family of these density functions, the signal-strength parameter is c_i and the spread parameter V_i stays constant. Suppose that the signal-strength parameter c_i is directly proportional to signal strength. Then the probability of a channel's output being above zero is a function of signal strength that has the same form as the right hand side of Eq. (5). Thus this model—the maximum-output bounded-from-above limit model with the added assumption that signal-strength parameter grows linearly with signal strength—can have the same "channel psychometric function" as the Quick pooling model. The rest of these two models' predictions are not the same, however; the full ROC curves for high-threshold models (e.g., the Quick pooling model) are much shallower than those for the maximum-output bounded-from-above limit model.

The function on the right side of Eq. (5) is itself a candidate for a probability distribution function since it goes from zero to one. As such, it is commonly called the Weibull distribution.

ACKNOWLEDGMENTS

We thank Jeanette Churche for carrying out many of the computations and Rob McAdow for his help as a research assistant. We are grateful to Jacob Nachmias, John Yellott, Elizabeth Davis, Denis Pelli, and three referees for many comments and discussions.

Some of this work was done in partial fulfillment of requirements for the Ph.D. degree by Patricia Kramer. This work was partially supported by a grant from the Foundation of the Optometric Center of New York and NSF Grant BNS-83-11350 to Dean Yager and NSF Grant BNS-76-18839 to Norma Graham.

A report of some of this work was presented at the May 1983 meetings of the Association for Research in Vision and Opthalmology held in Sarasota, Florida, and at the June 1984 Symposium on the Mechanisms of Spatial Vision held at the Center for Visual Science, University of Rochester, New York.

REFERENCES

ASHBY, F. G. & TOWNSEND, J. T. (1986). Varieties of perceptual independence. *Psychological Review*, **93**, 154–179.

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

BALL, K. & SEKULER, R. (1980). Models of stimulus uncertainty in motion perception. *Psychological Review*, **87**, 435–469.

BERGEN, J. R., WILSON, H. R. & COWAN, J. O. (1979). Further evidence for four mechanisms mediating vision at threshold: Sensitivities to complex gratings and aperiodic stimuli. *Journal of the Optical Society of America*, **69**, 1580–1586.

BRACEWELL, R. (1965). *The Fourier transform and its applications*. New York: McGraw-Hill.

COHN, T. E. (1978). Detection of 1-of-M orhtogonal signals: Asymptotic equivalence of likelihood ratio and multiband models. *Optics Letters*, **3**, 22–23.

COOMBS, C. H., DAWES, R. M. & TVERSKY, A. (1970). *Mathematical Psychology*. Englewood Cliffs, NJ: Prentice-Hall.

CREELMAN, C. D. (1960). Detection of signals of uncertain frequency. *Journal of the Acoustical Society of America*, **32**, 805–810.

DAVIS, E. T. & Graham, N. (1981). Spatial frequency uncertainty effects in the detection of sinusoidal gratings. *Vision Research*, **21**, 705–712.

DAVIS, E. T., KRAMER, P. & GRAHAM, N. (1983). Uncertainty about spatial frequency, spatial position, or contrast of visual patterns. *Perception and Psychophysics*, **33**(1), 20–28.

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

EGETH, H. (1977). Attention and preattention. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 11). New York: Academic Press.

FELLER, W. (1966). *An introduction to probability theory and its applications*. New York: Wiley.

GALAMBOS, J. (1978). *The asymptotic theory of extreme-order statistics*. New York: Wiley.

GRAHAM, N. (1977). Visual detection of aperiodic spatial stimuli by probability summation among narrow band channels. *Vision Research*, **17**, 637–652.

GRAHAM, N. (1981). Psychophysics of spatial-frequency channels. In M. Kubovy and J. Pomerantz (Eds.), *Perceptual organization* (pp. 1–25). Hillsdale, NJ: L. Erlbaum.

GRAHAM, N. (1985). Detection and identification of near-threshold visual patterns. *Journal of the Optical Society of America A*, **2**, 1468–1482.

GRAHAM, N., KRAMER, P. & HABER, N. (1985). Attending to the spatial frequency and spatial position of near-threshold visual patterns. In M. I. Posner and O. S. M. Marin (Eds.) *Attention and Performance* XI (pp. 269–284). Hillsdale, NJ: L. Erlbaum.

GRAHAM, N., ROBSON, J. G. & NACHMIAS, J. (1978). Grating summation in fovea and periphery. *Vision Research*, **18**(7), p. 815–826.

GREEN, D. M. & BIRDSALL, T. G. (1978). Detection and recognition. *Psychology Review*, **85**, 192–206.

GREEN, D. M. & SWETS, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley. (1974) Reprinted with corrections. Huntington, NY: Krieger.

GREEN, D. M. & WEBER, D. L. (1980). Detection of temporally uncertain signals. *Journal of the Acoustical Society of America*, **67**, 1304–1311.

GUMBEL, E. J. (1958). *Statistics of extremes*. New York: Columbia Univ. Press.

HACKER, M. J. & RATCLIFF, R. (1979). A revised table of d' for M-alternative forced choice. *Perception and Psychophysics*, **26**, 168–170.

HIRSCH, J., HYLTON, R. & GRAHAM, N. (1982). Simultaneous recognition of two spatial-frequency components. *Vision Research*, **22**, 365–375.

JOHNSON, K. O. (1980). Sensory discrimination: Decision process. *Journal of Neurophysiology*, **43**(6), 1771–1792.

KLEIN, S. (1985). Double-judgment psychophysics: Problems and solutions. *Journal of the Optical Society of America A*, **2**, 1560–1585.

KRAMER, P. (1984). Summation and uncertainty effects in the detection of spatial frequency. Ph.D. dissertation, Columbia University, New York.

KRAMER, P., GRAHAM, N. & YAGER, D. (1985). Simultaneous measurement of spatial-frequency summation and uncertainty effects. *Journal of the Optical Society of America A*, **2**, 1533–1542.

KRANTZ, D. H. (1969). Threshold theories of signal detection. *Psychological Review*, **76**, 308–324.

LAPPIN, J. S. & STALLER, J. D. (1981). Prior knowledge does not facilitate the perceptual organization of dynamic random-dot patterns. *Perception and Psychophysics*, **29**, 445–456.

LUCE, R. D. (1963a). A threshold theory for simple detection experiments. *Psychological Review*, **70**, 61–79.

LUCE, R. D. (1963b). Detection and recognition. In R. D. Luce, R. R. Bush, and G. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. I, pp. 103–190). New York: Wiley.

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

MOSTAFAVI, H. & SAKRISON, D. J. (1976). Structure and properties of a single channel in the human visual system. *Vision Research*, **16**, 957–968.

NACHMIAS, J. & KOCHER, E. C. (1970). Visual detection and discrimination of luminance increments. *Journal of the Optical Society of America*, **60**, 382–389.

NACHMIAS, J. (1972). Signal detection theory and its application to problems in vision. In *Handbook of sensory physiology* (Vol. VII/4). New York/Berlin: Springer-Verlag.

NACHMIAS, J. (1981). On the psychometric function for contrast detection. *Vision Research*, **21**, 215–223.

NOLTE, L. W. & JAARSMA, D. (1967). More on the detection of one of M orthogonal signals. *J. Acoustical Society of America*, **41**, 497–505.

OLZAK, L. A. & THOMAS, J. P. (1981). Gratings: Why frequency discrimination is sometimes better than detection. *Journal of the Optical Society of America*, **71**, 64–70.

PELLI, D. (1981). *Effects of visual noise.* Ph.D. dissertation, Cambridge University, England.

PELLI, D. (1985). Uncertainty explains many aspects of visual contrast detection and discrimination. *Journal of the Optical Society of America A*, **2**, 1508–1531.

QUICK, R. F. (1974). A vector-magnitude model of contrast detection. *Kybernetik*, **16**, 65–67.

QUICK, R. F., MULLINS, W. W. & Reichert, T. A. (1978). Spatial summation effects on two-component grating thresholds. *Journal of the Optical Society of America*, **68**, 116–121.

ROBSON, J. G. & GRAHAM, N. (1981). Probability summation and regional variation in contrast sensitivity across the visual field. *Vision Research*, **21**, 409–418.

SHAW, M. L. (1980). Identifying attentional and decision-making components in information processing. In Nickerson (Ed.), *Attention and performance VIII.* Hillsdale, NJ: L. Erlbaum.

SHAW, M. L. (1982). Attending to multiple sources of information. I. The integration of information in decision making. *Cognitive Psychology*, **14**, 353–409.

SMITH, J. E. K. (1982). Simple algorithms for M-alternative forced-choice calculations. *Perception and Psychophysics*, **31**, 95–96.

SPERLING, G. (1984). A unified theory of attention and signal detection. In R. Parasuraman, and D. R. Davies (Eds.), *Varieties of attention* (pp. 103–181). New York: Academic Press.

SPERLING, G. & DOSHER, B. A. (1986). Strategies and optimization in human information processing. In K. Boff, J. Thomas, and L. Kaufman (Eds.), *Handbook of perception and performance.* New York: Wiley.

STARR, S. J., METZ, C. E., LUSTED, L. B. & GOODENOUGH, D. J. (1975). Visual detection and localization of radiographic images. *Radiology*, **116**, 533–538.

SWENSSON, R. G. & JUDY, P. F. (1981). Detection of noisy, visual targets: Models for the effects of spatial uncertainty and signal-to-noise ratio. *Perception and Psychophysics*, **29**, 521–534.

SWETS, J. A. (1964). *Signal detection and recognition by human observers.* New York: Wiley.

TANNER, W. P. (1961). Physiological implications of psychophysical data. *Annals of the New York Academy of Sciences*, **89**, 752.

THOMAS, J. P., GILLE, J. & BARKER, R. A. (1982). Simultaneous visual detection and identification: Theory and data. *Journal of the Optical Society of America*, **72**(12), 1642–1651.

TREISMAN, M. & WILLIAMS, T. C. (1984). A theory of criterion setting with an application to sequential dependencies. *Psychological Review*, **91**(1), 68–111.

WANDELL, B. & LUCE, R. D. (1978). Pooling peripheral information: Averages versus extreme values. *Journal of Mathematical Psychology*, **17**, 220–235.

WATSON, A. B. (1983). Detection and recognition of simple spatial forms. In O. J. Braddick, and A. C. Sleigh (Eds.), *Physical and biological processing of images.* New York: Springer-Verlag. Also available as NASA Technical Memorandum 84353. Ames Research Center, Moffett Field, CA.

WATSON, A. B. & ROBSON, J. G. (1981). Discrimination at threshold: Labelled detectors in human vision. *Vision Research*, **21**, 1115–1122.

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

YAGER, D., KRAMER, P., SHAW, M. & GRAHAM, N. (1984). Detection and identification of spatial frequency: Models and data. *Vision Research*, **24**, 1021–1035.

YELLOT, J. I. JR. (1977). Relationship between Luce's choice axiom, Thurstone's theory of comparative judgement, and the double exponential distribution. *Journal of Mathematical Psychology*, **15**, 109–144.

RECEIVED: June 10, 1987