

# Optimal methods for calculating classification images: Weighted sums

**Richard F. Murray**

Department of Psychology, University of Toronto,  
Toronto, Canada



**Patrick J. Bennett**

Department of Psychology, McMaster University,  
Hamilton, Canada



**Allison B. Sekuler**

Department of Psychology, McMaster University,  
Hamilton, Canada



In signal detection theory, an observer's responses are often modeled as being based on a decision variable obtained by cross-correlating the stimulus with a template, possibly after corruption by external and internal noise. The response classification method estimates an observer's template by measuring the influence of each pixel of external noise on the observer's responses. A map that shows the influence of each pixel is called a *classification image*. Other authors have shown how to calculate classification images from external noise fields, but the optimal calculation has never been determined, and the quality of the resulting classification images has never been evaluated. Here we derive the optimal weighted sum of noise fields for calculating classification images in several experimental designs, and we derive the signal-to-noise ratio (SNR) of the resulting classification images. Using the expressions for the SNR, we show how to choose experimental parameters, such as the observer's performance level and the external noise power, to obtain classification images with a high SNR. We discuss two-alternative identification experiments in which the stimulus is presented at one or more contrast levels, in which each stimulus is presented twice so that we can estimate the power of the internal noise from the consistency of the observer's responses, and in which the observer rates the confidence of his responses. We illustrate these methods in a series of contrast increment detection experiments.

**Keywords:** response classification, reverse correlation, signal detection theory

## Introduction

In signal detection theory, an observer's responses are often modeled as being based on a decision variable  $s$  obtained by cross-correlating a stimulus  $I$  with a template  $T$ , possibly after corruption by an external Gaussian white noise  $N$  and an internal Gaussian white noise  $Z$ . The decision variable for this noisy cross-correlator can be written as

$$s = (I + N + Z) \otimes T. \quad (1)$$

In a two-alternative identification task, the observer sets a criterion  $a$ , and gives one response if  $s \geq a$ , and the other response if  $s < a$  (Green & Swets, 1974; Peterson, Birdsall, & Fox, 1954). This model gives a good account of many aspects of observers' performance in many perceptual tasks. In the "General Discussion," we briefly review the evidence for the model.

The response classification method estimates an observer's template  $T$  by measuring the influence of each pixel of an external noise field on the observer's responses (Ahumada & Lovell, 1971; Beard & Ahumada, 1998). An image that shows the influence of each noise pixel is

called a *classification image*. Other authors have derived methods of calculating classification images whose expected value is proportional to the template  $T$  (Abbey, Eckstein, & Bochud, 1999; Richards & Zhu, 1994), but the optimal calculation has never been determined, and the quality of the resulting classification images has never been evaluated. Response classification experiments require large amounts of data, so it would be useful to know how to use the data most efficiently and to know how the quality of a classification image depends on experimental variables, such as the number of trials and the observer's level of performance. Here we show how to calculate a classification image by taking a weighted sum of external noise fields from individual trials, in such a way as to maximize the signal-to-noise ratio (SNR) of the classification image, and we give expressions for the SNR. We derive optimal methods for two-alternative identification experiments in which the stimulus is presented at one or more contrast levels, in which each stimulus is presented on two separate trials so that we can estimate the power of the internal noise from the consistency of the observer's responses, and in which an observer rates the confidence of his responses. From the

expressions for the SNR, we show how to choose experimental parameters, such as the observer's performance level and the power of the external noise, to obtain classification images with a high SNR. We illustrate these methods in a series of contrast increment detection experiments.

In this study, we derive the optimal weighted sum of noise fields for calculating classification images. It is an open question whether more elaborate calculations, e.g., multiple linear regression, can improve on the optimal weighted sum.

As a compromise between the need for a straightforward user's guide to response classification methods and the need to prove our theoretical results, we state the most useful results in the main text and give the proofs in appendices.

## Identification at a Single Contrast Level

On each trial of a typical two-alternative identification experiment, one of two signals, A or B, is presented in Gaussian white noise  $N$ , and the observer's task is to state which signal was presented. On some trials, the noise causes the observer to make mistakes if, by chance, the noise is distributed in such a way as to make A look more like B, or to make B look more like A. The response matrix of such an experiment has four cells, corresponding to the four stimulus-response pairs: AA, AB, BA, and BB (Figure 1). Other authors have shown that if the observer is a linear discriminator who responds 'A' if a decision variable  $s$  of form (1) is greater than some criterion  $a$  and responds 'B' otherwise, then the expected value of the external noise field  $N$  on trials where the observer responds 'A' is proportional to the template  $T$ , and the expected value on trials where the observer responds 'B' is proportional to the negated template  $-T$  (Abbey et al., 1999; Richards & Zhu, 1994). Hence we can estimate the template by finding the average of the external noise fields  $N$  over trials where the observer responds 'A', and subtracting the average of the noise fields over trials where the observer responds 'B'.

However, this difference of averages does not make efficient use of the data. We use the following definition of the SNR to measure the quality of a stochastic image  $M$  contaminated by white noise:

$$SNR[M] = \frac{\|E[M]\|^2}{\sigma_M^2} \quad (2)$$

Here  $\|U\| = \sqrt{\sum_i U_i^2}$  is the vector magnitude of  $U$ ,  $E[M]$

is the expected value of  $M$ , and  $\sigma_M^2$  is the variance of each pixel of  $M$ . [Some authors define SNR as the square root of the right-hand side of (2).] In "Appendix A" we show that in general the SNRs of the estimates of  $T$  given by individual noise fields in the four classes of trials are not equal. Specifically, noise fields have higher SNRs on trials where the observer gives the incorrect response than on trials where the observer gives the correct response, so it is inefficient to combine noise fields in a weighted average that does not distinguish between correct and incorrect trials.

		Response	
		A	B
Stimulus	A	AA	AB
	B	BA	BB

Figure 1. Response matrix of a two-alternative identification experiment.

In "Appendix A," we show that if the observer is unbiased, the weighted sum of noise fields that gives the highest SNR is

$$C = (\overline{N_{AA}} + \overline{N_{BA}}) - (\overline{N_{AB}} + \overline{N_{BB}}). \quad (3)$$

We use  $\overline{N_{SR}}$  to denote the average of the external noise fields in a stimulus-response class of trials, e.g.,  $\overline{N_{AB}}$  is the average of the external noise fields presented on trials where the stimulus was A and the observer responded 'B'. Expression (3) states that in a two-alternative identification experiment, the best classification image is obtained by calculating the average of the noise fields within each of the four classes of trials, then adding together the means of classes AA and BA and subtracting the means of classes AB and BB. This is the formula for calculating classification images that appears most often in the psychophysics literature (e.g., Beard & Ahumada, 1998), and in "Appendix A," we show that it is the optimal weighted sum of noise fields when the observer is unbiased. In "Appendix A," we also derive the optimal weighted sum (A10) for a biased observer. Although (3) is the most commonly used formula, other formulas have been proposed that are suboptimal because they weight correct and incorrect trials equally, and do not take account of observer bias (Abbey et al., 1999; Richards &

Zhu, 1994). The optimal expression (3) follows from the noisy cross-correlator model (1), but it is an empirical question whether it is actually optimal for human observers. In Experiment 2, we report data from a contrast increment detection experiment that indicate that (3) is the optimal expression.

In “Appendix A,” we also show that the SNR of the classification image calculated as in (3) is

$$SNR[C] = \frac{n\sigma_N^2}{(\sigma_N^2 + \sigma_Z^2)} \cdot \frac{g(d'/2)^2}{G(d'/2)G(-d'/2)}. \quad (4)$$

Here  $n$  is the total number of trials,  $\sigma_N^2$  is the variance of each pixel of external noise  $N$ ,  $\sigma_Z^2$  is the variance of each pixel of internal noise  $Z$ , and  $d'$  is the observer's performance level. The function  $g$  is the standard normal probability density function, and  $G$  is the standard normal cumulative distribution function.

Expression (4) reveals the influence of several variables on the SNR. The SNR is proportional to the number of trials  $n$ . The SNR depends nonlinearly on  $d'$ , varying as  $g(d'/2)^2 / G(d'/2)G(-d'/2)$ . As shown in Figure 2, this means that the quality of the classification image declines rapidly at high performance levels, but does not vary greatly below approximately 75% correct (e.g., the SNR at 60% correct is only 15% higher than the SNR at 75% correct). Finally, the SNR is proportional to the ratio of the external noise variance to the total noise variance,  $\sigma_N^2 / (\sigma_N^2 + \sigma_Z^2)$ , indicating that the more the observer's performance is limited by internal noise, the lower the quality of the classification image. Internal noise is typically the sum of a noise of fixed variance  $\sigma_{ZF}^2$  and a noise whose variance  $\sigma_{ZP}^2$  is proportional to a weighted sum of the signal energy  $E$  and the external noise variance  $\sigma_N^2$ :  $\sigma_Z^2 = \sigma_{ZF}^2 + q(kE + \sigma_N^2)$  (Burgess & Colborne, 1988; Lillywhite, 1981; Lu & Doshier, 1998). This implies that the noise ratio  $\sigma_N^2 / (\sigma_N^2 + \sigma_Z^2)$  is highest when the external noise power is high. In many foveal tasks, the power spectral density of the fixed internal noise is on the order of  $\sigma_{ZF}^2 = 10^{-6} \text{ deg}^2$  and the constant of proportionality of the internal proportional noise is approximately  $q \approx 1$ , so the internal noise variance is given by  $\sigma_Z^2 \approx 10^{-6} \text{ deg}^2 + (kE + \sigma_N^2)$  (e.g., Burgess & Colborne, 1988; Pelli & Farell, 1999). This suggests that to obtain a high noise ratio  $\sigma_N^2 / (\sigma_N^2 + \sigma_Z^2)$ , the power of the external noise should be several times the power of the internal fixed noise, e.g., at least  $10^5 \text{ deg}^2$ , which at a

typical pixel width of 0.02 degrees corresponds to a root mean square (RMS) noise contrast of 16%.

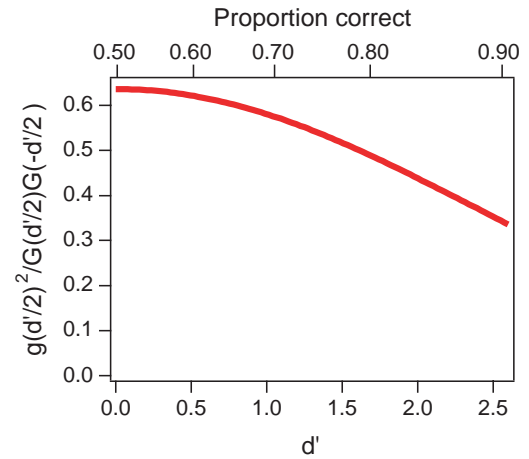


Figure 2. Effect of an observer's performance level on the signal-to-noise ratio of a classification image.

## Experiment 1: Identification at Several Contrast Levels

We often interleave several signal contrast levels in an identification experiment, e.g., to measure a threshold or a psychometric function. When calculating a classification image, how should we combine noise fields across trials with different signal contrast levels? Intuitively, we expect the informativeness of a noise field from a given trial to depend on both the signal contrast level and the correctness of the observer's response on that trial. For instance, we expect noise fields on incorrect trials to be more informative at high contrast levels than at low contrast levels, because at high contrast levels the noise must provide a larger amount of misleading evidence to induce an incorrect response. Conversely, we expect noise fields on correct trials to be more informative at low-contrast levels than at high contrast levels. The expression for the SNR of a single noise field confirms these intuitions [Equation (A4), derived in “Appendix A”], although this expression shows that the SNR is actually a function of the observer's performance level, and not of the signal contrast *per se*. It is less obvious whether the increase in information from the incorrect trials exceeds the decrease in information from the correct trials at higher performance levels, but our discussion of the SNR (4) of a classification image showed that the SNR is lower at higher performance levels, declining as  $g(d'/2)^2 / G(d'/2)G(-d'/2)$ . In “Appendix B,” we show that the optimal method of summing noise fields across trials with different performance levels is first to calculate separate classification images  $C_i$  from the trials at each

contrast level  $i$ , using Equation (3), and then to take the following weighted sum of the separate classification images  $C_i$ , in which each classification image is weighted according to the number of trials  $n_i$  and observer's performance level  $d'_i$  at the corresponding contrast level:

$$C = \sum_i n_i g(d'_i/2) C_i \quad (5)$$

This is the optimal method of summing the noise fields across contrast levels, and it follows (see "Appendix E") that the SNR of the classification image  $C$  is the sum of the SNRs of the classification images  $C_i$  at each contrast level:

$$SNR[C] = \frac{\sigma_N^2}{(\sigma_N^2 + \sigma_Z^2)} \sum_i \frac{n_i g(d'_i/2)^2}{G(d'_i/2)G(-d'_i/2)} \quad (6)$$

This SNR depends on the noise variances  $\sigma_N^2$  and  $\sigma_Z^2$  in the same way as the single-contrast classification image (3) discussed in the previous section. Furthermore, this SNR is highest when most trials are collected at low performance levels  $d'_i$ .

Taking account of the performance levels at which noise fields are presented can appreciably improve the quality of a classification image. When we measure a contrast threshold or a psychometric function in a two-alternative identification task, we typically use contrast levels that cover a performance range of 60% to 90% correct. Figure 2 shows that the SNR varies by approximately a factor of two across this range. It would be very inefficient to calculate classification images using expression (3) that we derived for the case of a single contrast level, as this would weight noise fields from high-performance trials as heavily as noise fields from low-performance trials. In the following experiment, we show that the weighting in (5) does improve the quality of classification images obtained in a contrast increment detection task.

In this experiment, and in the ones that follow, we compare new methods for calculating classification images that we have derived for particular experimental paradigms, to method (3), which we refer to as the standard method. The standard method was originally proposed for the case of an unbiased observer making binary responses to stimuli presented at a single contrast level, and as we have shown, in this case it is the optimal method. In the following experiments, we use the standard method as a benchmark against which to compare methods derived for different paradigms, without meaning to imply that it was ever intended for these paradigms. We use the standard method merely as a

plausible alternative for calculating classification images, to see whether other methods can improve on it.

## Methods

### Participants

Three undergraduate students at the University of Toronto, Toronto, Canada, participated. All had normal or corrected-to-normal Snellen acuity and were naïve as to the purpose of the experiment.

### Stimuli

The signal was a contrast increment in one of two disks shown in Gaussian white noise (Figure 3). The radius of each disk was 0.11 degrees of visual angle, and the center of each disk was 0.50 degrees to the left or right of a small fixation point. The base contrast of each disk was 10% Weber contrast, and the contrast increment varied from trial to trial, as explained in "Procedure." The noise formed a rectangle 1.0 degrees high and 2.0 degrees wide, centered on the fixation point, and its root mean square (RMS) Weber contrast was 20%. The noise was Gaussian, except that pixels more than two standard deviations from the mean were rejected and resampled, to keep contrast levels within the range displayable on the monitor. The stimulus duration was 200 ms.

Stimuli were displayed on an AppleVision monitor (640 × 480 resolution, pixel size 0.467 mm, refresh rate 67 Hz). Observers viewed the stimuli binocularly from a distance of 1 m, and head position was stabilized using a chin-and-forehead rest.

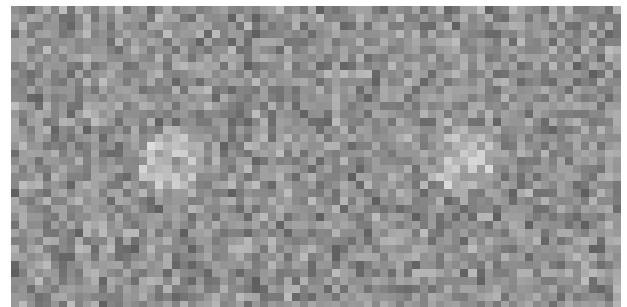


Figure 3. Stimulus in Experiment 1. The left disk has a contrast increment of 7%. [Movie of the stimulus.](#)

### Procedure

Each observer participated in five one-hour sessions of 2,000 trials. Each trial began with a 500-ms fixation interval, followed by the 200-ms stimulus, followed by a response interval in which the observer pressed one of two keys to indicate whether the contrast increment occurred in the left or right disk. Auditory feedback indicated whether the observer's response was correct.

Both the pedestal disks and the fixation point were shown throughout the entire trial. The method of constant stimuli was used to vary the magnitude of the contrast increment across trials. The contrast increments were chosen to span each observer's psychometric function, based on a pilot session. For observers A.N.C. and A.N.S., they were 2.0%, 2.7%, 3.8%, 5.4%, 7.5%, 11%, 15%, 22%, and 30% Weber contrast, and for observer O.R.W., they were 1.0%, 1.4%, 2.0%, 2.7%, 3.8%, 5.4%, 7.5%, 11%, 15%, and 21% Weber contrast. These values indicate the amount of contrast that was added to the pedestal contrast, not the proportion by which the pedestal contrast was increased; when we refer to a 5% contrast increment, we mean that a 10% contrast pedestal was increased to 15% contrast, not that it was increased to 10.5% contrast.

## Results and Discussion

Figure 4 shows each observer's psychometric function, plotting  $d'$  versus the contrast increment. The signal levels covered a wide performance range, from  $d'$  near zero to approximately 4.0, which in terms of proportion correct covers a range of 0.50 to 0.98. When measured in terms of  $d'$ , performance was an approximately linear function of signal contrast, at least up to high performance levels (around  $d' = 3.5$ , which corresponds to 96% percent correct) at which point even infrequent keypress errors and lapses of attention can cause performance to level off. This linearity is consistent with the noisy cross-correlator model (1), and with the findings of earlier studies of contrast increment detection (Legge, Kersten, & Burgess, 1987).

We calculated classification images for each observer using both the optimal weighted sum (5) that weights noise fields according to the observer's performance at

each contrast level and the standard method (3) that does not take account of the varying contrast level (Figure 5).

How can we compare the quality of the optimal and suboptimal classification images? A classification image is a random variable that can be written as  $kT + N_C$ , i.e., as the sum of a signal  $kT$  that is proportional to the observer's template  $T$ , and a sampling noise  $N_C$ . If we scale the template  $T$  to have unit energy, the signal energy in the classification image is  $k^2$ , the noise variance is the pixelwise variance of the sampling noise  $\sigma_C^2$ , and the SNR of the classification image is  $k^2 / \sigma_C^2$ . If we knew the observer's template  $T$  exactly, we could estimate the SNRs of the optimal and suboptimal classification images. First, the classification image is a weighted sum of noise fields,  $C = \sum_i w_i N_i$ , so if the weights and noise fields are

independent, then the pixelwise variance of the

classification image is  $\sigma_C^2 = \sum_i w_i^2 \sigma_N^2$ . The weights and

noise fields are not independent (e.g., the weight assigned to a noise field depends on the observer's response to the noise field, which in turn depends on how similar the noise field is to the observer's template), but in "Appendix A" [Equation (A5)], we show that this approximation to  $\sigma_C^2$  is very accurate and gives a simple and effective way of calculating the variance of the classification image. Second, if we knew the observer's template  $T$ , we could calculate the signal energy  $k^2$  from the cross-correlation  $T \otimes (kT + N_C)$ , which has an expected value of  $k$ . Of course, we do not know the observer's template  $T$  exactly, so we cannot directly compare the SNRs of the optimal and suboptimal classification images this way.

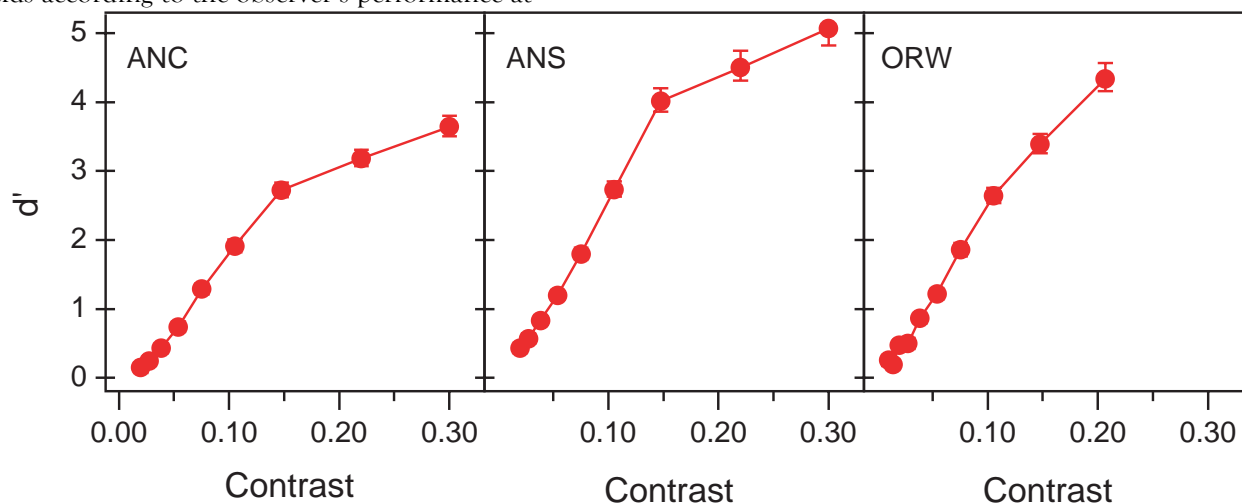


Figure 4. Results of Experiment 1. Psychometric functions plotting performance against contrast increment. The error bars show standard errors, and in most cases are smaller than the data points.



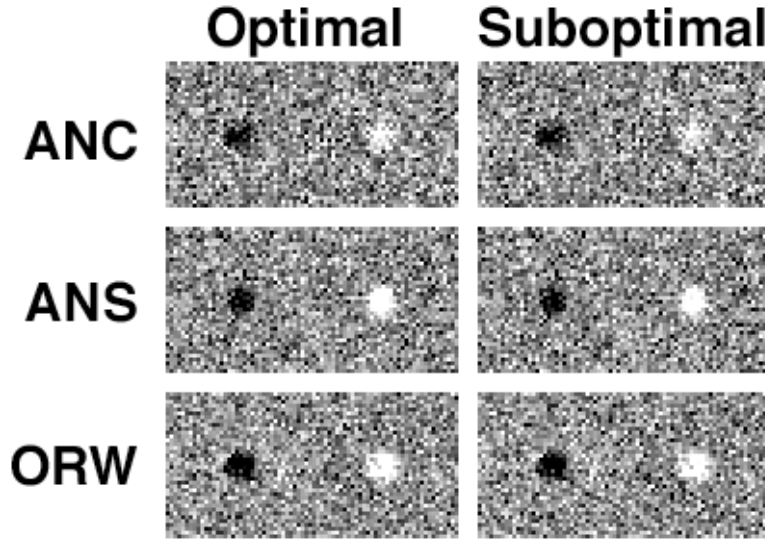


Figure 5. Results of [Experiment 1](#). Optimal and suboptimal classification images. Although the optimal and suboptimal classification images look quite similar, the optimal images are measurably better estimates of the observers' templates than the suboptimal images. This can be shown by calculating the SNR of the optimal and suboptimal images, as explained in the text.

However, using an approximation  $T$  to the observer's template with unit amplitude ( $\|T\| = 1$ ), we can calculate the SNR to within a scale factor. We define the relative SNR (rSNR) of a stochastic image  $M$  as

$$rSNR[M] = \frac{(T' \otimes M)^2}{\sigma_C^2} - 1. \quad (7)$$

For classification images, this amounts to

$$rSNR[kT + N_C] = \frac{(T' \otimes (kT + N_C))^2}{\sigma_C^2} - 1 \quad (8)$$

which a straightforward evaluation shows to have an expected value of  $k^2(T' \otimes T)^2 / \sigma_C^2$ . (The  $-1$  term corrects for a bias introduced by squaring  $T' \otimes N_C$ .) That is, the rSNR is proportional to the SNR,  $k^2 / \sigma_C^2$ , and can be used to estimate the SNRs of the optimal and suboptimal classification images up to a common scale factor. In principle, the choice of the approximation  $T$  is arbitrary, but if we make a poor approximation, the cross-correlation  $T' \otimes T$  is small compared to the noise term  $T' \otimes N_C$ , and our calculation of the rSNR will be noisy. The closer our approximation  $T$  is to the true template  $T$ , the better.

To compare the optimal and suboptimal classification images, we calculated their rSNRs. We used the ideal observer's template as the approximation  $T$  to the human observers' templates. The ideal template is the signal-right stimulus minus the signal-left stimulus, so it consists of a positive-contrast dot to the right of fixation and a

negative-contrast dot to the left. For the numerator of (8), we cross-correlated the ideal observer's template with the classification images, and for the denominator, we calculated the pixelwise variance from the variance of individual noise fields and the weights in the weighted sum that produced the classification image.

For observer A.N.C., the rSNR of the optimal classification image was  $247 \pm 32$  and the rSNR of the suboptimal classification image was  $227 \pm 30$ ; for observer A.N.S., the rSNRs were  $528 \pm 46$  and  $464 \pm 43$ ; and for observer O.R.W., they were  $766 \pm 55$  and  $665 \pm 52$ . The error values are standard errors, obtained by calculating the standard deviation of the cross-correlation of a unit-amplitude ideal template with a classification image of pixelwise variance  $\sigma_C^2$ , with  $\sigma_C^2$  calculated individually for each observer as described earlier. The optimal rSNRs were consistently higher than the suboptimal rSNRs, and although the differences were not statistically significant for individual observers, taken together they did reject the null hypothesis that the optimal rSNRs were the same as the corresponding suboptimal rSNRs: under the null hypothesis, differences this large for all three observers are improbable ( $p < .05$ ). On average, the rSNRs of the optimal classification images were 13% higher than the rSNRs of the suboptimal images. We conclude that the optimal method (5) improves on the standard method (3), although the standard method is reasonably efficient considering the wide range of performance levels covered in the experiment.

The derivation of the optimal method (5), in “[Appendix B](#),” makes it clear that the weight assigned to

each subordinate classification image  $C_i$  is determined by the SNR of the classification image  $C_i$ , as predicted by model (1): classification images that are predicted to have a high SNR are weighted heavily, and classification images that are predicted to have a low SNR are weighted lightly. Consequently, expression (5) is the optimal weighted sum for calculating the pooled classification image  $C$  only if our predictions of the SNRs of the subordinate classification images are correct. To see whether the SNR actually varied across performance levels as predicted by (4), we calculated the rSNR of the classification images at each performance level. Figure 6 plots the rSNRs of the classification images  $C_i$  at each performance level, divided by the number of trials  $n_i$  collected at each performance level, along with the predictions of expression (4) scaled to minimize the sum-of-squares error between the predictions and the data. The rSNRs roughly followed the predicted pattern of declining as a function of  $d'$ , indicating that expression (5) assigns approximately correct weights to the subordinate classification images  $C_i$  when calculating the pooled classification image  $C$ .

Two technical points may help clarify the meaning of Figure 6. First, we have scaled the predicted SNRs to fit the measured rSNRs, because (4) predicts absolute SNRs, whereas the rSNRs estimate the SNRs only up to a common scale factor. However, as we show in “Appendix E,” this scaling does not pose a problem because we only need to predict the SNRs correctly up to a common scale factor to compute the optimal weights. Second, in Figure 6 and similar figures that follow, we plot the rSNR divided by the number of trials, i.e., the average rSNR of a single trial. We do this because we are mostly interested in how accurately the noisy cross-correlator model

predicts the SNRs of single noise fields in each cell of the response matrix, and the number of trials in each cell is only of secondary interest.

Although the predictions shown in Figure 6 are roughly correct, there are also consistent deviations: for all observers, the rSNR was lower than predicted above a performance level of approximately  $d'=1$ , and for two of the observers (A.N.C. and O.R.W.), the rSNR may have been lower than predicted below this level as well. The deviations are small compared to the overall accuracy of the predictions, and the measurement errors are too large for us to say with certainty where the rSNR peaks. Nevertheless, this discrepancy calls for further investigation, because if it is genuine, then it indicates a failure of the noisy cross-correlator model, i.e., a failure of linearity. For instance, the keypress errors and lapses of attention that may have caused the psychometric functions to saturate at high performance levels (Figure 4) might have to be included in the model as a form of internal noise that grows with the signal level, lowering the SNR at high performance levels. Furthermore, this discrepancy implies that expression (5) is a slightly suboptimal method of calculating classification images, as it assigns too large a weight to noise fields collected at performance levels far from  $d' = 1$ . Most of the deviations are small, but a revised model that predicted the SNRs correctly would be useful because it would allow us to derive the truly optimal method of calculating classification images when combining trials across performance levels. Another possibility that we will not investigate here is empirical estimation of the rSNRs of trials at different performance levels, as in Figure 6, and the use of these estimates to weight the classification

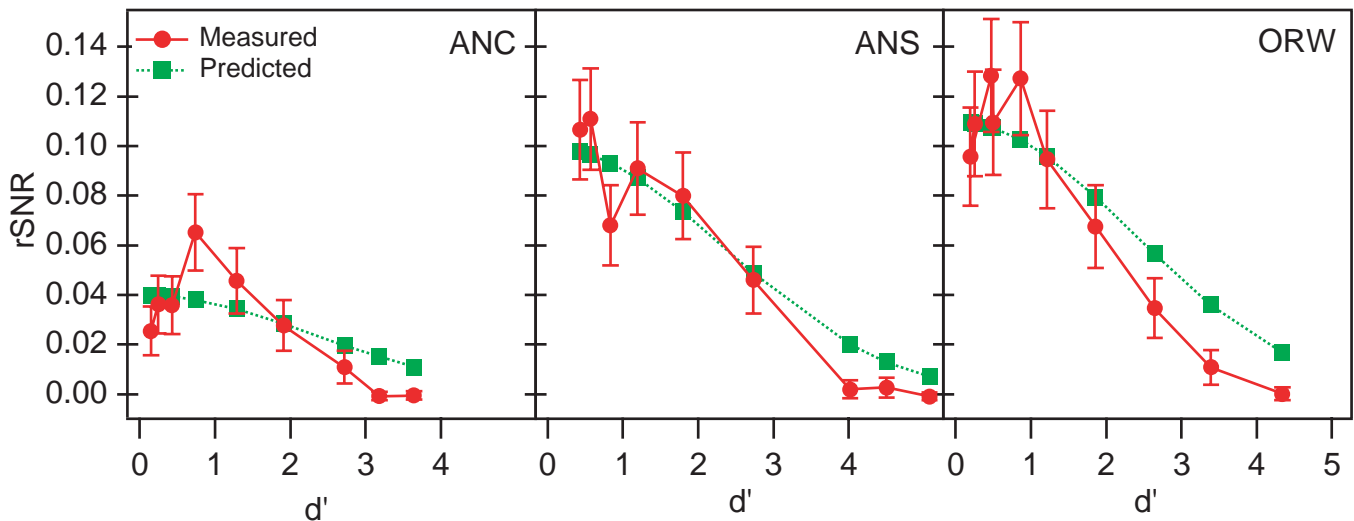


Figure 6. Results of Experiment 1. Mean rSNRs of individual noise fields as a function of performance level. Error bars show standard errors.

images at different performance levels optimally when combining them into a single classification image. In any case, one practical consequence of this finding is that classification images should be collected at a performance level of approximately  $d' = 1$ . Above this point, the SNR drops even more rapidly than predicted, and the SNR may drop below this point as well. Furthermore, there are other reasons for not collecting classification images at very low performance levels, such as the changes in observers' strategies that can occur due to spatial uncertainty (Ahumada & Beard, 1999; Pelli, 1985) or the frustration that observers experience when a task is too difficult.

We should point out that because the predicted SNRs and observed rSNRs are scaled to minimize the sum-of-squares error in Figure 6, we cannot say for certain at what performance levels the model fails. For instance, we could say that given the rSNR at  $d' \approx 1$ , the rSNR at  $d' \approx 4$  is lower than expected, or we could equally well say that given the rSNR at  $d' \approx 4$ , the rSNR at  $d' \approx 1$  is higher than expected. This ambiguity does not matter for our test of whether expression (5) is the optimal method of calculating classification images, because as we mentioned earlier, we need only predict the SNRs correctly up to a common scale factor. This ambiguity does matter, though, when we attempt to explain why the model's predictions do not match the observed rSNRs.

## Experiment 2: Response Consistency

According to the noisy cross-correlator model (1), an observer's performance is limited both by the efficiency of the template  $T$ , and by the power of the internal noise  $Z$ . One way of measuring the power of the internal noise that limits an observer's performance is to present each stimulus twice, on separate trials, and to measure the proportion of repeated trials on which the observer gives the same response twice (Burgess & Colborne, 1988; Gold et al., 1999; Green, 1964). We emphasize that in this two-pass method, the repeated stimulus, including the external noise, is identical pixel-by-pixel on both presentations. The two presentations are separated by many trials, so the observer does not know when the stimulus is repeated, and treats the two trials as showing independent stimuli. If the observer's responses are based on a noiseless decision rule, then the observer will give the same response to a stimulus every time it is presented, whereas if the observer's performance is largely limited by internal noise, then the observer's responses to repeated presentations of the same stimulus will be less consistent.

Burgess and Colborne (1988) showed how to use the consistency of an observer's responses in a two-pass experiment to calculate the power of the internal noise.

		Response			
		AA	AB	BA	BB
Stimulus	A	AAA	AAB	ABA	ABB
	B	BAA	BAB	BBA	BBB

Figure 7. Response matrix of a two-pass experiment.

Figure 7 shows the response matrix of a two-pass experiment with two signals (A and B) and four possible pairs of responses on repeated presentations of a single stimulus (AA, AB, BA, and BB). In "Appendix C," we show that noise fields from trials in different cells of this matrix have different SNRs, and we show that for an unbiased observer, the optimal weighted sum for calculating a classification image is

$$\begin{aligned}
 C &= w(\overline{N_{AAA}} - \overline{N_{BBB}}) \\
 &+ (w - 0.5)(\overline{N_{BAB}} + \overline{N_{BBA}} - \overline{N_{AAB}} - \overline{N_{ABA}}) \\
 &+ (1 - w)(\overline{N_{BAA}} - \overline{N_{ABB}})
 \end{aligned} \quad (9)$$

$$\text{where } w = G \left( \frac{d'}{2\sqrt{1 + 2\sigma_N^2 / \sigma_Z^2}} \right).$$

To calculate the weighting parameter  $w$ , we need to know the observer's performance level ( $d'$ ), which is easily determined, and the internal-to-external noise ratio  $\sigma_Z^2 / \sigma_N^2$ , which can be calculated from the consistency of the observer's responses (Burgess & Colborne, 1988).

In "Appendix C," we also show that for an unbiased observer, the SNR of a classification image calculated as in (9) is

$$SNR[C] = \frac{4ng(d'/2)^2 \sigma_N^2}{(\sigma_N^2 + \sigma_Z^2)} \left( \frac{w^2}{p_{CC}} + \frac{4(w-0.5)^2}{p_{CI}} + \frac{(w-1)^2}{p_{II}} \right) \quad (10)$$

Here  $p_{CC}$  is the probability of the observer giving two correct responses on repeated trials,  $p_{CI}$  is the probability of one correct and one incorrect response, and  $p_{II}$  is the probability of two incorrect responses. For instance, when stimulus A is presented,  $p_{CC}$  is the probability of two A responses,  $p_{CI}$  is the probability of one A response and one B response, in either order, and  $p_{II}$  is the probability of two B responses.



A two-pass experiment gives information about an observer that a one-pass experiment does not, so it is natural to ask whether we can generate classification images more efficiently with two-pass experiments than with one-pass experiments. Expressions (4) and (10) for the SNRs obtained in one- and two-pass experiments, respectively, show that this is not possible. Figure 8 plots the ratio of the SNR (10) obtained in a two-pass experiment to the SNR (4) obtained in a one-pass experiment with the same number of trials. When  $\sigma_Z^2 / \sigma_N^2 = 0$ , it follows that  $p_{CC}=p_C$ ,  $p_{CI}=0$ ,  $p_{II}=p_I$ , and  $w = 0.5$ . With these values, the SNR (10) of the two-pass classification image is half the SNR (4) of the one-pass classification image. When  $\sigma_Z^2 / \sigma_N^2 \rightarrow \infty$ , it follows that  $p_{CC}=p_C^2$ ,  $p_{CI}=p_C p_I$ ,  $p_{II}=p_I^2$ , and  $w = p_C$ . With these values, the two-pass SNR (10) equals the one-pass SNR (4). That is, the quality of a classification image obtained in a two-pass experiment can approach but never exceed the quality of a classification image obtained in the corresponding one-pass experiment.

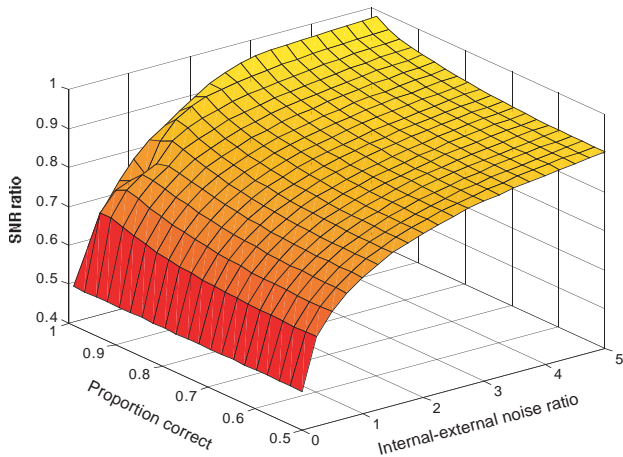


Figure 8. Ratio of the SNR of a two-pass classification image to the SNR of a one-pass classification image.

Does the optimal weighted sum (9) improve appreciably on the results we would obtain by using the standard method (3) in a two-pass experiment, incorrectly treating repeated trials as if they showed independent noise fields? With the standard method, each trial in the AAA cell of the two-pass response matrix would be counted twice in the AA cell of the one-pass matrix, each trial in the AAB cell of the two-pass matrix would be counted once in the AA cell and once in the AB cell of the one-pass matrix, and so on. Using this regrouping of trials and expressions (C1) through (C4) in “Appendix C” for the SNR of the noise fields in each cell of the two-pass response matrix, it is possible to show that over a wide

range of values of  $p_C$  and  $\sigma_Z^2 / \sigma_N^2$  (e.g., as  $p_C$  ranges from 0.50 to 0.95 and  $\sigma_Z^2 / \sigma_N^2$  ranges from 0 to 3), the SNR obtained using the standard method (3) is only a few percent lower than the SNR obtained using the optimal method (9), so the optimal method does not improve appreciably on the standard method. In the following experiment, we show that the standard method (3) does work almost as well as the optimal method (9) in a contrast increment detection task.

## Methods

One author (R.F.M.) and two observers from Experiment 1 (A.N.C. and O.R.W.) participated. The stimuli and procedure were the same as in Experiment 1, except in two respects. First, the magnitude of the contrast increment was fixed at the observer’s 70% threshold as calculated by fitting a normal cumulative distribution function to the psychometric function obtained in a pilot session (observer R.F.M.) or in Experiment 1 (observers A.N.C. and O.R.W.). For observer A.N.C., this threshold was 7%, for O.R.W., it was 5%, and for R.F.M., it was 3.5%. Second, each session was divided into ten 200-trial blocks, and the second 100 trials of each block were exact repetitions of the first 100 trials of the block.

## Results and Discussion

Observer A.N.C. gave  $74\% \pm 1\%$  correct responses and gave the same response on  $68\% \pm 1\%$  of repeated trials, corresponding to an internal-to-external noise ratio of  $2.47 \pm 0.20$ . Observer O.R.W. gave  $79\% \pm 1\%$  correct responses, and gave the same response on  $76\% \pm 1\%$  of repeated trials, corresponding to an internal-to-external noise ratio of  $1.13 \pm 0.05$ . Observer R.F.M. gave  $69\% \pm 1\%$  correct responses and gave the same response on  $69\% \pm 1\%$  of repeated trials, corresponding to an internal-to-external noise ratio of  $1.28 \pm 0.05$ . We calculated these internal-to-external noise ratios using the methods developed by Burgess and Colborne (1988).

We calculated a classification image for each observer using both the optimal weighted sum (9) that takes account of the consistency of an observer’s responses across repeated trials and the standard method (3), treating repeated trials as if they showed statistically independent noise fields. Calculating rSNRs as in Experiment 1, we found that for observer A.N.C., the rSNR of the optimal classification image was  $749 \pm 55$ , and the rSNR of the suboptimal image was  $759 \pm 55$ ; for

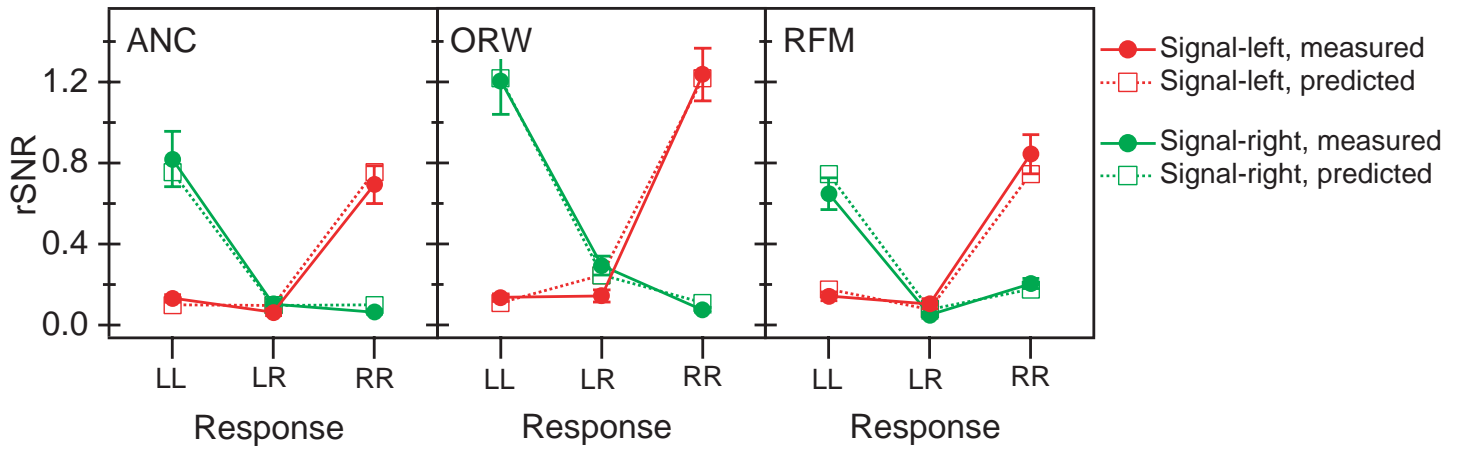


Figure 9. Results of Experiment 2. rSNR of individual noise fields in each cell of the two-pass response matrix. The red data points show the rSNR of noise fields on trials where the contrast increment was on the left, and the green data points show the rSNR on trials where the contrast increment was on the right. Response LL denotes two left responses, and the corresponding data points show the SNR in response matrix cells LLL and RLL. Response LR denotes one left and one right response, and these data points show the SNR in response matrix cells LLR, LRL, RLR, and RRL. Response RR denotes two right responses, and indicates the SNR in response matrix cells LRR and LRR. The error bars show standard errors, and are often smaller than the data points.

observer O.R.W., the rSNRs were  $1132 \pm 67$  and  $1146 \pm 68$ ; and for observer R.F.M., they were  $1164 \pm 68$  and  $1162 \pm 68$ . The rSNRs of the optimal and suboptimal classification images were practically identical, and we cannot reject the null hypothesis that the optimal and suboptimal rSNRs were the same. As predicted, the suboptimal method of calculating classification images was as good as the optimal method, to within experimental error.

As explained in Experiment 1, the optimal methods that we have derived rely on theoretical predictions of the SNRs of the noise fields in each cell of a response matrix. To see whether the SNR varied from cell to cell of the two-pass response matrix in the manner expected, we calculated the rSNR per trial of the noise fields in each cell, and compared them to the predicted SNRs (see Equations (C1) through (C4) in “Appendix C”), scaled to fit the rSNRs as in Experiment 1. The predictions were excellent (Figure 9), supporting the explanation that

model (1) gives of an observer’s response consistency in terms of internal and external noise, and demonstrating that (9) is the optimal weighted sum for calculating classification images in a two-pass experiment.

Finally, returning briefly to an earlier section of this work (Identification at a Single Contrast Level), we can use this experiment’s data to test whether expression (3) is actually the optimal weighted sum for calculating classification images in a two-alternative identification experiment. If we discard the repeated trials from this experiment (i.e., the second 100 trials of each 200-trial block), we are left with a simple two-alternative identification experiment. Figure 10 shows the rSNR of the noise fields in each of the four cells of the two-alternative response matrix, along with the predicted SNR [see expression (A6) in Appendix A]. The predictions are excellent, indicating that method (3) is the optimal weighted sum.

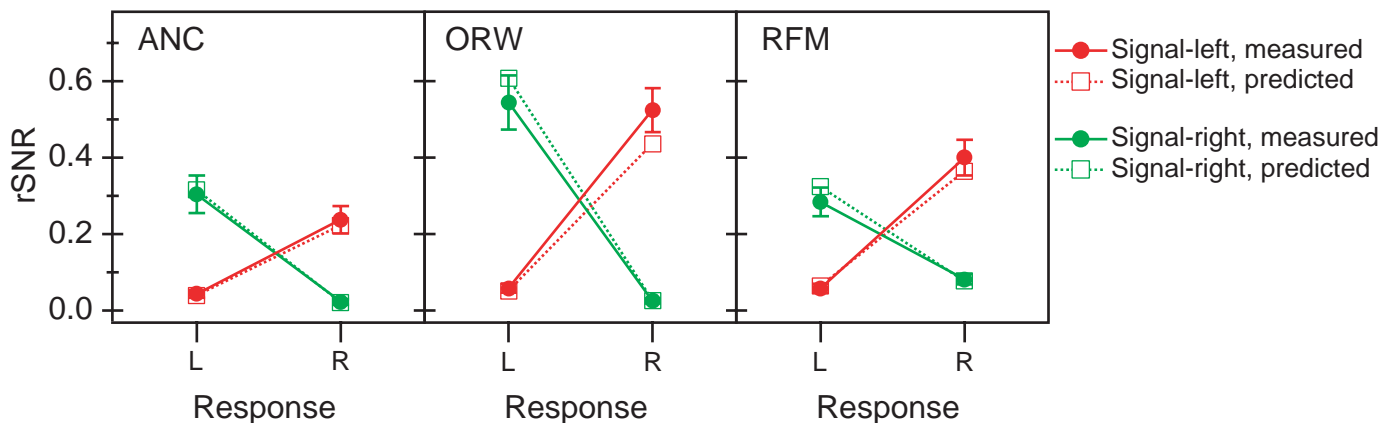


Figure 10. Results of Experiment 2. rSNR of individual noise fields in each cell of the one-pass response matrix.

## Experiment 3: Rating Scales

When observers make perceptual judgements, they can rate the confidence of their responses. In a typical rating scale experiment, the observer uses an  $r$  point rating scale to indicate his confidence that stimulus A or B was presented. We will take response 1 to mean that the observer is confident that the stimulus was B, and response  $r$  to mean that he is confident that it was A. Figure 11 shows the response matrix for an experiment with a six-point rating scale.

		Response					
		1	2	3	4	5	6
Stimulus	A	A1	A2	A3	A4	A5	A6
	B	B1	B2	B3	B4	B5	B6

Figure 11. Response matrix of a six-point rating scale experiment.

In signal detection theory, one typically assumes that the observer makes responses by setting  $r+1$  criteria  $a_i$ , and giving response  $i$  if the decision variable  $s$  falls between  $a_i$  and  $a_{i+1}$  (Egan, Schulman, & Greenberg, 1959). (This formulation requires that  $a_1 \rightarrow -\infty$  and  $a_{r+1} \rightarrow +\infty$ .) In “Appendix D” we show that noise fields from different cells of the response matrix of a rating scale experiment have different SNRs, and we show that the optimal weighted sum for calculating a classification image is

$$C = \sum_{i=1}^r (g(z_i) - g(z_{i+1})) \overline{N_{Ai}} + (g(z_i - d') - g(z_{i+1} - d')) \overline{N_{Bi}} \quad (11)$$

$$\text{where } z_i = \frac{G^{-1}(p_{Ai-}) + G^{-1}(p_{Bi-}) + d'}{2}.$$

Here  $p_{Ai-}$  is the probability that the observer gives a rating less than  $i$  when stimulus A is presented, and  $p_{Bi-}$  is the probability that the observer gives a rating less than  $i$  when stimulus B is presented. The function  $G^{-1}$  is the inverse of the normal cumulative distribution function (i.e., it is the z-transform function used in signal detection theory). Expression (11) states that the optimal weighted sum adds the average noise fields in each cell of the response matrix, with the average of each cell weighted by a quantity that is a function of the normal deviates  $z_i$  and  $z_{i+1}$  of the criteria  $a_i$  and  $a_{i+1}$  that bound the decision variable in that cell.

In “Appendix D,” we also show that the SNR of the classification image calculated as in (11) is

$$SNR[C] = \frac{(n/2)\sigma_N^2}{\sigma_N^2 + \sigma_Z^2} \times \sum_{i=1}^r \left( \frac{(g(z_i) - g(z_{i+1}))^2}{G(z_{i+1}) - G(z_i)} + \frac{(g(z_i - d') - g(z_{i+1} - d'))^2}{G(z_{i+1} - d') - G(z_i - d')} \right). \quad (12)$$

How much do we gain by recording rating responses instead of binary identification responses? If the observer uses a six-point rating scale and places his criteria at  $-\infty$ ,  $-0.39d'$ ,  $0.10d'$ ,  $0.50d'$ ,  $0.90d'$ ,  $1.39d'$ , and  $+\infty$  so that he gives each response equally often, expression (12) predicts that the SNR will be 1.67 times the SNR obtained in the corresponding binary-response identification experiment. Evidently, the advantage of using a rating scale can be substantial. In the following experiment, we show that recording rating scale responses can improve the quality of classification images obtained in a contrast increment detection task.

## Methods

The same three observers participated as in Experiment 1. The stimuli and procedure were also the same as in Experiment 1, except in two respects. First, the contrast increment was fixed at the observer's 70% threshold as determined in Experiment 1 (observer A.N.C., 7%; observers A.N.S. and O.R.W., 5%). Second, observers gave keypress responses on a six-point rating scale, giving response 1 to indicate confidently that the contrast increment occurred in the left disk, and response 6 to indicate confidently that it occurred in the right disk. We instructed observers to adjust their criteria so that they gave each response equally often, and after every 200 trials, they were given feedback on the computer monitor, indicating how many times they had given each response.

## Results and Discussion

Figure 12 shows each observer's receiver operating characteristic (ROC) curve on z-scaled axes. Clearly, the observers succeeded in maintaining several widely spaced criteria, and we found that the observers gave each response approximately equally often, as instructed. The ROC curves were approximately linear, indicating that the decision variable had a roughly Gaussian distribution, and the slope of the best-fitting line was approximately 1, indicating that the decision variable had the same variance on signal-left and signal-right trials (Green & Swets, 1974). However, the curves were slightly but consistently bowed, indicating that either the noise limiting the observers' performance was non-Gaussian or the observers were less consistent in their use of the more extreme responses.

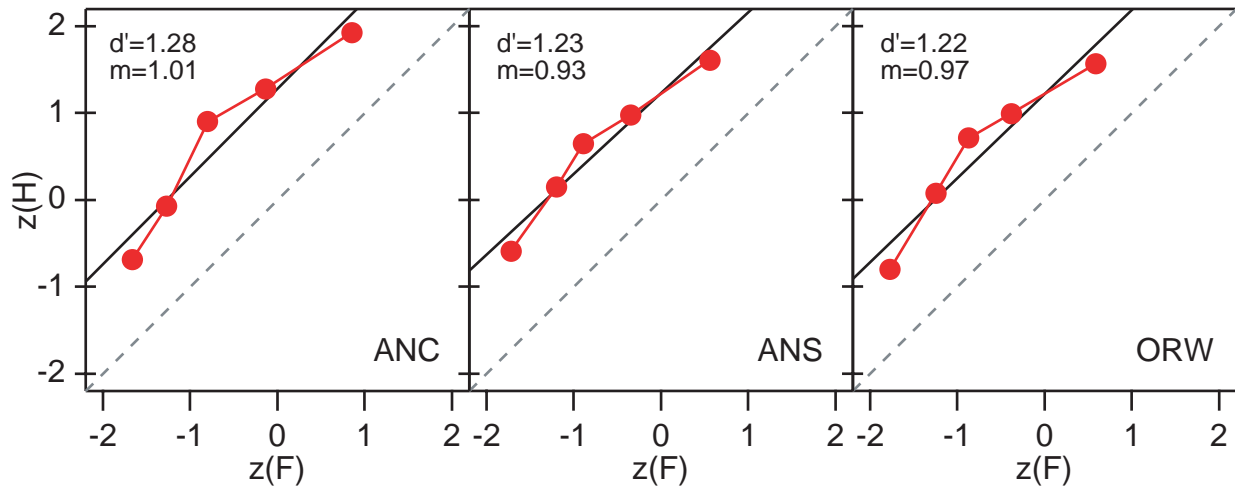


Figure 12. Results of Experiment 3. Receiver operating characteristic (ROC) curves. These plots show each observer's  $z$ -transformed hit rates  $z(H_i)$  plotted against the corresponding  $z$ -transformed false alarm rates  $z(F_i)$  for each rating response. Each hit rate  $H_i$  is the probability of the observer giving a rating of  $i$  or lower (i.e., responding left with at least a certain amount of confidence) on a trial where the left disk had a contrast increment, and each false alarm rate  $F_i$  is the probability of the observer responding  $i$  or lower on a trial where the right disk had a contrast increment. We have omitted the uninformative point  $(z(H_6), z(F_6))$  from the graphs: observers used six rating responses, so all ratings are six or less, and  $H_6 = F_6 = 1$  in every case. The best-fitting lines of the form  $z(H) = d' + mz(F)$  are shown in solid black, and the chance-performance lines are shown in dashed grey. The error bars are smaller than the data points.

We calculated classification images using the optimal weighted sum (11) that takes account of the observers' confidence ratings, and using the standard binary-response method (3), with responses 1, 2, and 3 grouped together as a left response, and responses 4, 5, and 6 grouped together as a right response. We calculated the rSNRs of the optimal and suboptimal classification images using the same method as in Experiments 1 and 2. For observer A.N.C., the rSNR of the optimal classification image was  $682 \pm 52$ , and the rSNR of the suboptimal classification image was  $870 \pm 59$ ; for observer A.N.S., the rSNRs were  $1026 \pm 64$  and  $1122 \pm 67$ ; and for observer O.R.W., they were  $1460 \pm 76$  and  $1580 \pm 80$ . Surprisingly, the rSNRs of the suboptimal classification images were significantly higher than the rSNRs of the optimal classification images ( $p < .01$ ), and on average were 15% higher. Equation (11) gives the optimal method of calculating classification images for an observer who performs the rating scale task by comparing a Gaussian-distributed decision variable of form (1) to a number of fixed criteria, as in the standard signal detection account that we outlined above. Clearly, our observers did not follow this strategy.

Method (11) of calculating classification images depends on the noisy cross-correlator model's predictions of the SNRs of noise fields in each cell of the response matrix. The failure of our allegedly optimal method in this rating scale experiment indicates that the model's predictions were incorrect. To see how observers departed

from the model, we calculated the rSNR of noise fields in each cell of the response matrix and compared these to the predicted SNRs [see Equations (D9) and (D10) in "Appendix D"] (Figure 13). As in Experiments 1 and 2, we scaled the predicted SNRs to minimize the sum-of-squares error in their fit to the rSNRs. A consistent pattern in these graphs is that the rSNR plots are not as sharply concave upwards as the predicted plots, i.e., the rSNRs corresponding to conservative responses (2, 3, 4, and 5) are consistently higher than the predictions, and the rSNRs corresponding to extreme responses (1 and 6) tend to be lower than the predictions. That is, the model predicts that extreme responses should be much more informative than conservative responses, but Figure 13 shows that they were only slightly more informative. Method (11) weights noise fields in each cell according to their predicted SNR and hence assigns a large weight to noise fields that produced extreme responses, which turn out to be much less informative than expected.

We instructed observers to use each rating response equally often because expression (12) for the SNR indicated that this strategy would produce a classification image with an SNR 67% higher than a classification image from a binary identification experiment, whereas at the other extreme, if an observer concentrated his responses in the most conservative response categories, the rating scale experiment would reduce to a binary identification experiment. However, our results

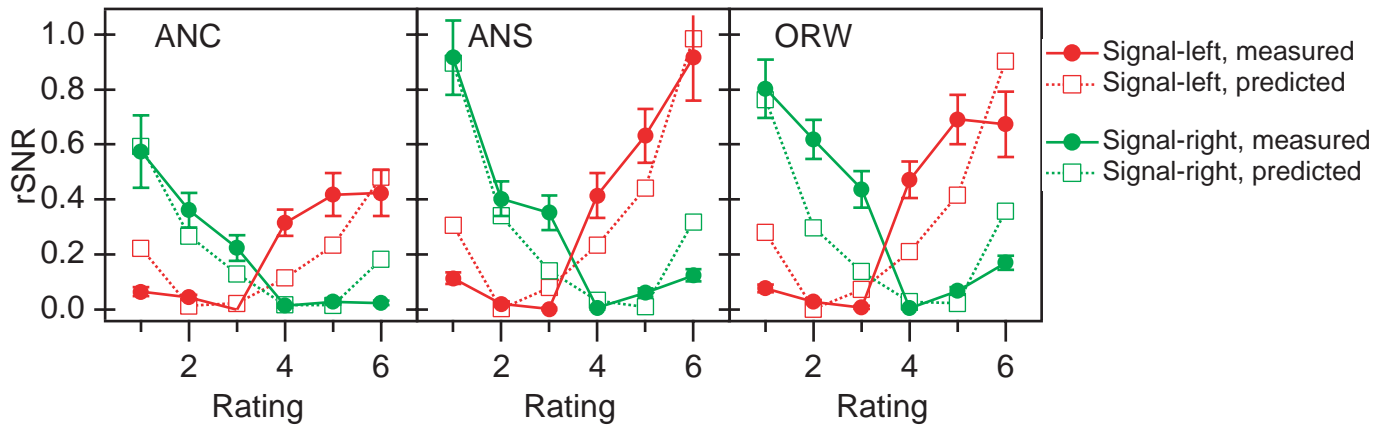


Figure 13. Results of Experiment 3. rSNRs of noise fields in each cell of the rating scale response matrix. The error bars show standard errors.

suggest that observers had difficulty following our instructions. In particular, the bowed ROC curves in Figure 12 suggest that observers were unable to use the extreme criteria consistently: if an observer varies his criterion from trial to trial, this variability appears as a form of internal noise that reduces sensitivity (Wickelgren, 1968), and the bowed ROC curves show that observers did perform more poorly when they gave extreme responses. Furthermore, expression (12) indicates that the quality of a classification image declines as the internal-to-external noise ratio grows, and Figure 13 shows that the rSNRs of noise fields on extreme-response trials were lower than expected.

To see whether the instructions to use each rating response equally often caused this marked departure from the model, we re-ran the experiment with three new observers. This time we gave no instructions about how often each rating response should be used, and we did not give feedback about how often each rating had been used in each block of 200 trials. All three observers were naïve, and none had participated in any of the preceding

experiments. The contrast increment for each observer was set to the observer's 70% threshold, based on a pilot session (observers D.I.H. and T.F.S., 6% Weber contrast; observer L.C.S., 5% Weber contrast).

Figure 14 shows the new observers' ROC curves. The curves are much less bowed than the previous ones, indicating that observers used the extreme criteria more consistently when they were free to set the criteria where they wished. We found large individual differences in how often the observers used each response, and no observer used each response equally often. All observers used the conservative responses 3 and 4 most often. Observer D.I.H. used the extreme responses 1 and 6 second most often, and the middle responses 2 and 5 least often. Observer L.C.S. used the middle responses 2 and 5 second most often, and rarely used the extreme responses 1 and 6, which is reflected in the wide placement of the endpoints of this observer's ROC curve. (Note that this observer's plot axes are scaled differently from the other two observers'.) Observer T.F.S. used

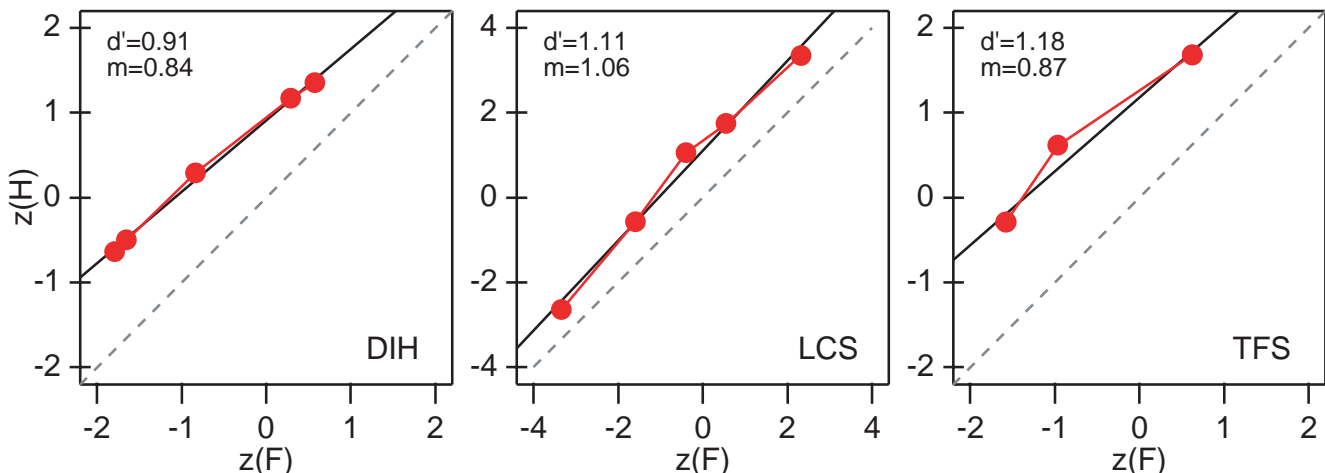


Figure 14. Results of Experiment 3. Receiver operating characteristic (ROC) curves for a second set of observers. See caption of Figure 12 for details.



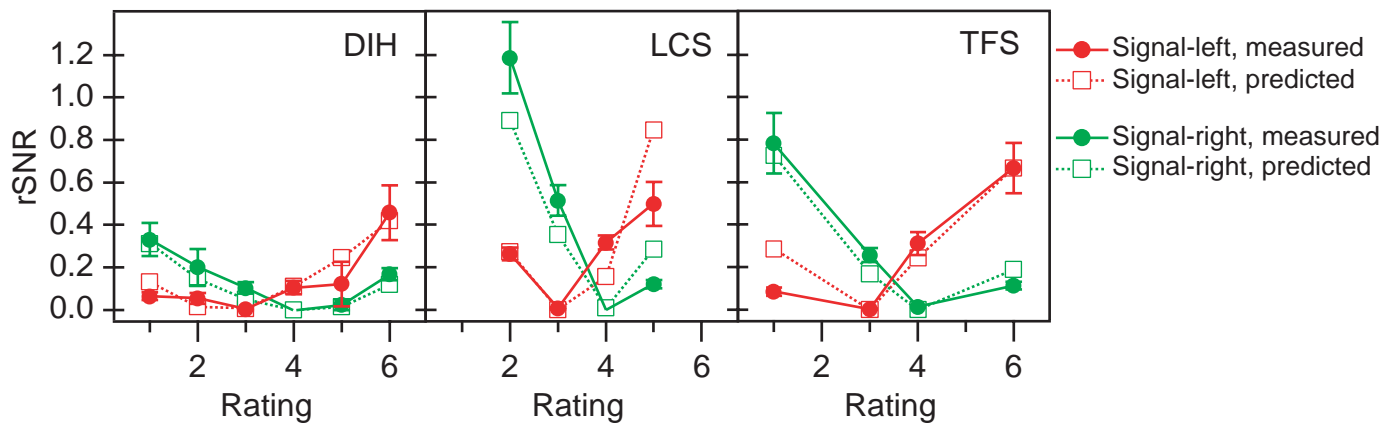


Figure 15. Results of Experiment 3. rSNRs of individual noise fields in each cell of the rating scale response matrix, for a second set of observers. Data points are not shown for observer L.C.S.'s responses 1 and 6 or for observer T.F.S.'s responses 2 and 4, because these observers used these responses so rarely that we cannot estimate the rSNR with precision.

the extreme responses second most often, and almost never used the middle responses 2 and 5, so each endpoint of this observer's ROC curve is actually two points superimposed.

For observer D.I.H., the rSNR of the optimal classification image was  $458 \pm 43$ , and the rSNR of the suboptimal classification image was  $337 \pm 37$ ; for observer L.C.S., the rSNRs were  $1242 \pm 71$  and  $1013 \pm 64$ ; and for observer T.F.S., they were  $774 \pm 56$  and  $739 \pm 54$ . The rSNRs of the optimal classification images were significantly higher than the rSNRs of the suboptimal classification images ( $p < .001$ ), and on average were 21% higher. Figure 15 shows the average rSNRs of noise fields in each cell of the rating scale response matrix. The agreement between the predicted SNRs and actual rSNRs is much better for these observers than for the first three, which explains why the optimal method (11) gave better results for this set of observers.

We conclude that using a rating scale in a response classification experiment can improve the quality of classification images. However, we found that observers were unable to reliably maintain the criteria specified in our instructions, which we chose to give an especially large improvement in SNR. When observers chose their own criteria, the average improvement in rSNR was 21%.

A final caveat is that observers' reaction times are typically longer when giving rating responses than when giving binary responses, and this must be traded off against the increase in SNR (Burgess, 1995). In Experiments 1 and 2, which recorded binary responses, the mean reaction time across all observers was 270 ms, and in Experiment 3, which recorded six-point rating responses, it was 460 ms. Taking into account the 500-ms fixation interval and the 200-ms stimulus interval, this means that rating scale trials took about 20% longer than

binary response trials. The SNR of a classification image is proportional to the number of trials, so given a fixed amount of time for an experiment, the 21% increase in SNR gained by recording rating scale responses is almost exactly undone by the reduction in the number of trials. Perhaps by using a four-point rating scale instead of a six-point scale, hence simplifying the observer's task, and by imposing a response deadline, we could combine the advantages of the rating scale and binary response methods.

## General Discussion

### The Noisy Cross-Correlator Model

The noisy cross-correlator model (1) describes performance in many visual tasks reasonably well. It accounts for the linear relationship between discrimination threshold energy and external noise power (Pelli, 1990), and the fact that performance measured in  $d'$  is often a linear function of signal contrast (e.g., Legge et al., 1987). It leads to the concepts of sampling efficiency and internal-to-external noise ratio, which are useful ways of describing many factors that limit observers' performances (Burgess & Colborne, 1988; Burgess, Wagner, Jennings, & Barlow, 1981). Nevertheless, it does not account for all aspects of observers' performances, and the model has been elaborated in various ways by many authors. Most of these elaborations are unimportant for our purposes, because we require only that the model described by (1) is *locally* valid, in the sense that it describes observers' performance in a single discrimination task. In the following paragraphs, we consider a few examples of how models that differ from the noisy cross-correlator described by

Equation (1) may be locally equivalent to it.

Linear models differ in where they place internal noise sources in the calculation that leads to the decision variable. Some variants of the noisy cross-correlator model place a noise before the cross-correlation (e.g., Pelli, 1990), and some place one after (e.g., Lu & Doshier, 1998). In a single discrimination task, these differences are mostly irrelevant, and we can model the effects of all internal noise sources as a single noise  $Z$  added to the stimulus at the input (Ahumada, 1987; Ahumada & Watson, 1985). The internal noise  $Z$  affects the observer's decisions only via the  $Z \otimes T$  term that is added to the decision variable, so any late noise  $Z_L$  added after the cross-correlation is equivalent to an early noise  $Z$  added before the cross-correlation that satisfies  $\text{var}[Z \otimes T] = \text{var}[Z_L]$ . Hence the difference between early- and late-noise models is not important for our purposes.

Similar comments apply to nonwhite internal noise (e.g., Burgess et al., 1981). Any nonwhite noise term  $Z_N$  that affects the observer's decisions only after it has passed through the template is equivalent to a white noise  $Z$  that produces a term of the same variance after the template, i.e.,  $Z_N$  and  $Z$  are equivalent so long as  $\text{var}[Z_N \otimes T] = \text{var}[Z \otimes T]$ .

Many models include a proportional noise (sometimes called multiplicative noise) whose power grows with the stimulus energy (Burgess & Colborne, 1988; Lillywhite, 1981; Lu & Doshier, 1998). If a signal is shown in strong external noise, much of the observer's proportional noise will be induced by the external noise, and small differences in signal power between signal-A and signal-B trials in a threshold discrimination task will produce little difference in the observer's proportional noise. For this reason, we can consider proportional noise as just another form of internal noise that can be incorporated in the early noise ( $Z$ ). This said, we should also point out that it is easy to modify the methods we have presented to handle tasks where the internal noise power is very different on signal-A and signal-B trials. The derivation in "Appendix A" considers signal-A and signal-B trials separately, and if we need to obtain a more general expression for calculating classification images, we can simply drop the assumption that the internal noise power is the same on the two types of trials.

Models with transduction nonlinearities and stimulus-dependent noise are often equivalent to linear models with stimulus-independent noise, if the range of relevant stimuli is small compared to the range over which the transduction nonlinearities and stimulus-dependent noise amplitudes change appreciably

(Ahumada, 1987). To take just one example, in Foley and Legge's (1981) model of grating detection and discrimination, observers use a decision variable with mean  $(c/c_0)^f$  and fixed variance, where  $c$  is the signal grating contrast and  $c_0$  is an arbitrary reference contrast. This is clearly a nonlinear model, but in a task where the observer discriminates between two gratings of fixed contrast  $c_A$  and  $c_B$ , the nonlinearity can be accommodated within the noisy cross-correlator model. Let  $I$  be a unit-contrast grating, so that  $cI$  is a grating of contrast  $c$ . We can incorporate Foley and Legge's (1981) power-law transduction nonlinearity by writing the decision variable in response to a stimulus  $cI+N$  as

$$s = \left( \frac{(cI+N) \otimes T}{c_0 I \otimes T} \right)^f + Z_L. \quad (13)$$

With no external noise, this decision variable has mean  $(c/c_0)^f$  and fixed variance, as in Foley and Legge's (1981) model. If the external noise  $N$  causes the term  $(cI+N) \otimes T$  to vary over only a small range, as in an experiment where observers discriminate between gratings of similar contrasts  $c_A$  and  $c_B$ , we can use a Taylor series approximation that is linear in the external noise term  $N$ :

$$s \approx \left( \frac{cI \otimes T}{c_0 I \otimes T} \right)^f + f \left( \frac{cI \otimes T}{c_0 I \otimes T} \right)^{f-1} \left( \frac{N \otimes T}{c_0 I \otimes T} \right) + Z_L \quad (14)$$

$$= (c/c_0)^f + \frac{f c^{f-1}}{c_0^f I \otimes T} N \otimes T + Z_L. \quad (15)$$

If we rescale the decision variable, multiplying by  $c_0^f I \otimes T$ , we can rewrite it as

$$s = c^f I \otimes T + f c^{f-1} N \otimes T + (c_0^f I \otimes T) Z_L. \quad (16)$$

As we pointed out when we discussed early and late noise, we can choose  $Z$  so that  $\text{var}[Z \otimes T] = \text{var}[Z_L]$ , and rewrite the decision variable as

$$s = \left( c^f I + f c^{f-1} N + (c_0^f I \otimes T) Z \right) \otimes T. \quad (17)$$

Hence over a small contrast range, the observer behaves like a noisy cross-correlator, except that the internal-to-external noise ratio depends on the signal contrast  $c$ . When an observer discriminates between gratings of two similar contrasts  $c_A$  and  $c_B$ , the internal noise power will be approximately the same on signal-A and signal-B trials, and the methods we have derived will be approximately optimal. When the grating contrast ratio  $c_A/c_B$  is very different from 1, as when  $c_A$  or  $c_B$  is zero in a detection experiment, the internal-to-external noise ratio may be

very different on the two types of trials. As we pointed out in our discussion of proportional noise, the methods we have derived can be easily modified to handle this case as well.

One type of nonlinearity that does pose a problem for the noisy cross-correlator model is stimulus uncertainty. Even when observers are told the exact shape and location of the signals that they are to discriminate between, they sometimes behave as if they are uncertain as to exactly where the stimulus will appear or what shape it will take (e.g., [Manjeshwar & Wilson, 2001](#); [Pelli, 1985](#)). We can model spatial uncertainty by assuming that the observer has many identical templates that he applies over a range of spatial locations in the stimulus, but the effects of this operation are complex, and it is not obvious precisely how a classification image is related to the template of such an observer, or how the SNR of the classification image is related to quantities such as the observer's performance level or internal-to-external noise ratio. If an observer is very uncertain about some stimulus properties, such as the phase of a grating signal, a response classification experiment may produce no classification image at all ([Ahumada & Beard, 1999](#)).

Early pointwise nonlinearities also pose a problem. These nonlinearities transform the contrast of each pixel of the stimulus by a static function, converting contrast  $c$  to  $f(c)$ . [Chubb and Nam \(2000\)](#) reported an extreme example of such a nonlinearity: they found that observers used a half- or full-wave rectifying nonlinearity to judge the contrast variance of a texture patch. Clearly, an observer who used full-wave rectification would not produce a classification image because the contrast of each pixel of the stimulus would be uncorrelated with the observer's response. The precise effect of less extreme nonlinearities, such as a logarithmic transform, is unclear. On the other hand, [Nam and Chubb \(2000\)](#) found that early pointwise nonlinearities were negligible when observers judged the luminance of a texture patch, and we have found similar results in complex shape discrimination tasks ([Murray, Bennett, & Sekuler, 2001](#)), suggesting that such nonlinearities might be unimportant in first-order tasks.

A final point is that the methods we have derived rely only on the noisy cross-correlator model to predict the SNRs of individual noise fields so that we may know how to combine the noise fields optimally in a weighted sum. As long as the model succeeds in this respect, any other failures are irrelevant for the purpose of calculating classification images efficiently. As we have shown by comparing measured rSNRs to predicted SNRs, the model's predictions are approximately correct in several experimental paradigms. We have shown this only for the

contrast increment detection task, but using the method of measuring rSNRs that we have outlined, it is straightforward to validate the model in any other task.

## Summary

For several experimental designs, we have derived the optimal weighted sum of noise fields for calculating classification images. In a series of contrast increment detection experiments, we confirmed our theoretical predictions that the standard formula (3) is the optimal weighted sum in a two-alternative identification experiment, that expression (5) improves on the standard formula in an experiment where the signal is presented at several different contrast levels, and that expression (11) improves on the standard formula in a rating scale experiment. Our experiments also confirmed that the optimal weighted sum (9) does not improve appreciably on the standard formula in a two-pass experiment.

For the same set of experimental designs, we derived expressions for the SNRs of the classification images calculated using the optimal weighted sum of noise fields. These expressions show how to choose experimental parameters to maximize the SNR of a classification image. First, of course, one should collect as many trials as possible. Second, the external noise should have much more power than the observer's fixed equivalent input noise, and we suggested that this condition is usually met if the external noise power is  $10^{-5} \text{ deg}^2$ , which at a typical pixel width of 0.02 degrees corresponds to an RMS noise contrast of 16%. Third, we found that the SNR of our classification images peaked at a performance level of approximately  $d' = 1$ . Finally, we found that classification images had a higher SNR when we recorded responses on a six-point rating scale than when we recorded binary identification responses.

## Appendix A: Single-Contrast level

In this appendix, we derive the optimal weighted sum of noise fields for calculating classification images in a two-alternative identification experiment having only one signal contrast level, and we derive the SNR of the resulting classification image.

### A vector space description of the noisy cross-correlator

We assume that the observer identifies a noisy stimulus  $I+N$  as one of two alternatives, A or B, by corrupting it with an additive internal noise  $Z$ , cross-

correlating the corrupted stimulus with a template  $T$  to obtain a decision variable  $s$ , and responding  $A$  if and only if  $s$  exceeds a criterion  $a$ :

$$s = (I + N + Z) \otimes T \quad (\text{A1})$$

$$R = \text{sgn}(s - a) = \begin{cases} +1 & \text{if } s \geq a \\ -1 & \text{if } s < a \end{cases} \quad (\text{A2})$$

We will call  $I^* = I + N + Z$  the corrupted stimulus.

We can consider the signal  $I$ , the noises  $N$  and  $Z$ , and the template  $T$  as vectors in an  $m$ -dimensional vector space, where  $m$  is the number of pixels in the stimulus. The cross-correlation  $(I + N + Z) \otimes T$  then becomes the vector dot product  $(I + N + Z) \bullet T$ . An observer who follows strategy (A2) divides the  $m$ -space in two with a hyperplane  $\Pi_T$  perpendicular to  $T$ , and responds 'A' if and only if the corrupted stimulus  $I^*$  falls on one side of  $\Pi_T$ . Without loss of generality, we can assume that  $\|T\| = 1$ .

## The SNR of each cell of the response matrix

Consider the trials on which the signal is  $A$ . We will adopt an orthonormal coordinate frame  $F'$  with the origin at  $A$ , with the first coordinate vector  $\bar{x}_1$  parallel to  $T$  and with the remaining coordinate vectors  $\bar{x}_2, \dots, \bar{x}_m$  parallel to  $\Pi_T$ . We can represent the transformation from our original coordinate frame  $F$  to the new frame  $F'$  by  $\bar{v}_{F'} = R(\bar{v}_F - A)$ , where  $R$  is a rotation matrix  $R = [\bar{x}_1 \ \dots \ \bar{x}_m]^T$ . In  $F'$ , a signal- $A$  stimulus  $I^* = A + N + Z$  is represented as  $I^{*'} = R(A + N + Z - A) = RN + RZ$ . We will define  $N' = RN$  and  $Z' = RZ$ , and write  $I^{*'} = N' + Z'$ . The coordinates of  $N$  are independent, equal-variance Gaussian random variables, so the coordinates of  $N'$  are also independent, equal-variance Gaussian random variables. Similarly, the coordinates of  $Z$  and  $Z'$  are independent, equal-variance Gaussian random variables.

We have defined the decision variable as  $s = (I + N + Z) \bullet T$ , which on signal- $A$  trials amounts to  $s = (A + N + Z) \bullet T$ , and we have assumed that the observer's responses depend on whether  $s \geq a$ . Equivalently, we can define the decision variable as  $s' = (N + Z) \bullet T$ , and assume that the observer uses a criterion  $a' = a - A \bullet T$ . The vector dot product is invariant under rotation, so we can rewrite this new decision variable as  $s' = R(N + Z) \bullet RT = (N' + Z') \bullet RT$ .

We have defined the rotation  $R$  such that

$T' = RT = (1, 0, 0, \dots, 0)$ , so the decision variable  $s'$  takes the particularly simple form  $s' = N'_1 + Z'_1 = I^{*'}_1$ . That is, the observer's responses depend only on whether the first coordinate  $I^{*'}_1$  of the 0 stimulus exceeds a criterion  $a'$ , and the observer's responses are statistically independent of coordinates 2 through  $m$  of the noises  $N'$  and  $Z'$ .

To find the SNR of a single noise field in each cell of the response matrix, we need to know the expected value and variance of each class of noise field, which we will now derive.

What is the expected value of the external noise field  $N'$  on trials where the signal is  $A$  and the observer responds 'A'? We will denote this expected value by  $E[N' | AA]$ . The observer's response is independent of components  $N'_2$  through  $N'_m$ , so the mean of these components, conditional on the observer having responded  $A$ , is equal to their unconditional mean, which is zero. The conditional mean of the first component is  $E[N'_1 | N'_1 + Z'_1 \geq a']$ . In "Appendix F," we derive an expression (F1) for conditional means of this form, and this expression shows that the expected value of  $N'_1$  is

$$E[N'_1 | AA] = \frac{\sigma_N^2}{\sqrt{\sigma_N^2 + \sigma_Z^2}} \cdot \frac{g(G^{-1}(p_{AA}))}{p_{AA}} \quad (\text{A3})$$

Here  $p_{AA}$  is the probability that the observer gives response  $A$  on a trial where stimulus  $A$  is presented,  $\sigma_N^2$  is the variance of each pixel of the external noise field  $N'$ , and  $\sigma_Z^2$  is the variance of each pixel of the internal noise field  $Z'$ . The function  $g$  is the standard normal probability density function, and  $G^{-1}$  is the inverse of the standard normal cumulative distribution function.  $N'_1$  is the only nonzero component of the expected value of the entire noise field  $E[N' | AA]$ , so expression (A3) also gives the magnitude of this expected value. Furthermore, because  $N'_1$  is the only nonzero component, the expected value  $E[N' | AA]$  is proportional to the coordinate vector  $\bar{x}_1$ , and hence proportional the observer's template  $T$ ; this is why the response classification method gives an estimate of the observer's template.

What is the mean value of the external noise field  $N'$  on trials where the signal is  $A$  and the observer responds  $B$ , i.e.,  $E[N' | AB]$ ? Again, the conditional mean of components  $N'_2$  through  $N'_m$  is zero, and now the conditional mean of the first component is  $E[N'_1 | N'_1 + Z'_1 < a']$ . This mean can be rewritten as



$-E[-N'_1 | -N'_1 - Z'_1 > -a']$ , and we can use (F1) again to evaluate this expression:

$$E[N'_1 | AB] = -\frac{\sigma_N^2}{\sqrt{\sigma_N^2 + \sigma_Z^2}} \cdot \frac{g(G^{-1}(P_{AB}))}{P_{AB}}. \quad (\text{A4})$$

What is the variance of each pixel of the external noise field  $N$  on trials of type AA and AB? The observer's response is independent of components  $N'_2$  through  $N'_m$  of the transformed noise field  $N'$ , so the conditional variance of these components is equal to their unconditional variance  $\sigma_N^2$ . In "Appendix F," we give expressions (F1, F2) from which the conditional variance of  $N'_1$  can be computed, and these expressions show that under typical experimental conditions (e.g., 75% correct and an internal-to-external noise ratio of 1.0) the variance of  $N'_1$  is slightly less than  $\sigma_N^2$ .  $N$  is a rotation of  $N'$ , so each pixel of  $N$  can be expressed as a weighted sum of the components of  $N'$ , i.e.,  $N = R^{-1}N'$ . When the stimulus contains many pixels (e.g., 10,000 pixels in a  $100 \times 100$  stimulus) the variance of the single component  $N'_1$  makes a negligible contribution to the variance of the pixels of  $N$ . Furthermore, the expression for the conditional variance of  $N'_1$  is cumbersome, and requires us to know the observer's internal-to-external noise ratio  $\sigma_Z^2 / \sigma_N^2$ . Consequently, we will use the approximation that the variance of  $N'_1$  is  $\sigma_N^2$  on trials of type AA and AB, and that the variance of each pixel in  $N$  is

$$\sigma_{N|AA}^2 \approx \sigma_{N|AB}^2 \approx \sigma_N^2. \quad (\text{A5})$$

How accurate is this approximation? At a performance level of 75% correct and an internal-to-external noise ratio of 1.0, (F1) and (F2) show that on AA trials the true variance of  $N'_1$  is  $0.77\sigma_N^2$ , whereas we approximate it as  $\sigma_N^2$ . Hence in an  $m$ -pixel noise field, we approximate a total noise field variance of  $((m-1) + 0.77)\sigma_N^2$  as  $m\sigma_N^2$ . Even in a small  $16 \times 16$  pixel noise field, this is an error of less than 0.1%, which is negligible compared to other sources of error, such as our estimate of the observer's response bias that as we show below must also be known to calculate the optimal weighted sum. In any case, in situations where (A5) is an inadequate approximation, we can use the correct expression for the conditional variance of  $N'$  which is immediately derivable from (F1) and (F2).

The same analysis applies to trials on which signal  $B$  is presented, with analogous results. We need only

introduce a coordinate frame  $F''$  with its origin at  $B$  and its first coordinate vector parallel to  $T$ , and a decision variable  $s'' = (N + Z) \cdot T$  with criterion  $a'' = a - B \cdot T$ . We then obtain expressions like (A3), (A4), and (A5), with stimulus  $A$  replaced by stimulus  $B$ . Hence the SNR of an external noise field  $N$  on a single trial of type SR (where we use SR to stand for trial types AA, AB, BA, and BB) is the ratio of an expression of the form (A3) or (A4) squared, and an expression of the form (A5):

$$\text{SNR}[N' | \text{SR}] = \frac{\sigma_N^2}{(\sigma_N^2 + \sigma_Z^2)} \cdot \frac{g(G^{-1}(P_{\text{SR}}))^2}{P_{\text{SR}}^2}. \quad (\text{A6})$$

In general, the four trial types have different probabilities  $p_{\text{SR}}$ , so noise fields from the four cells of the response matrix have different SNRs.

## Combining noise fields

In the preceding section, we showed that the expected value of a noise field in each class of trials is proportional to the template  $T$ , but also that, in general, noise fields from different classes of trials have different SNRs, given by (A6). In "Appendix E," we show that the SNR of a weighted sum  $C = \sum w_i C_i$  of several noisy images  $C_i$  with proportional means is maximized when  $w_i = \|E[C_i]\| / \text{VAR}[C_i]$ . According to the expressions for the expected value (A3, A4), and variance (A5) of the noise fields, this means that we should calculate the classification image by taking a weighted sum of noise fields across trials, with the weight  $w_{\text{SR}}$  assigned to the noise fields in each cell SR of the response matrix equal to

$$w_{\text{SR}} = \pm \frac{1}{\sqrt{\sigma_N^2 + \sigma_Z^2}} \cdot \frac{g(G^{-1}(P_{\text{SR}}))}{P_{\text{SR}}}. \quad (\text{A7})$$

The sign is positive for trials in classes AA and BA, and negative for trials in classes AB and BB, as the expected values of noise fields on response-A and response-B trials are anti-parallel.

Any factors that appear in the weights  $w_{\text{SR}}$  corresponding to all noise fields can be dropped, as these factors simply scale the classification image without affecting its SNR. The external noise variance  $\sigma_N^2$  is the same on all trials, and if the contrast energy of stimuli  $A$  and  $B$  is approximately the same then the internal noise variance  $\sigma_Z^2$  will also be approximately the same on all trials (see our discussion of fixed and proportional



internal noise in the general discussion), and we can drop the factor  $1/\sqrt{\sigma_N^2 + \sigma_Z^2}$ . Hence the weights reduce to

$$w_{SR} = \pm \frac{g(G^{-1}(p_{SR}))}{p_{SR}}. \quad (\text{A8})$$

We will denote the number of trials in response cell SR in a particular experiment as  $\hat{n}_{SR}$  and estimate the response probabilities by  $\hat{p}_{SR} = \hat{n}_{SR}/(n/2)$ . If we substitute  $\hat{p}_{SR}$  for  $p_{SR}$  (we discuss this substitution later), and scale each weight by dividing it by the number of trials  $n/2$  on which stimulus A or stimulus B was presented, the weights evaluate to

$$w_{SR} = \pm \frac{g(G^{-1}(\hat{p}_{SR}))}{\hat{n}_{SR}}. \quad (\text{A9})$$

Hence, the optimal weighted sum for calculating the classification image of a possibly biased observer is

$$C = \left( g(G^{-1}(\hat{p}_{AA}))\overline{N_{AA}} + g(G^{-1}(\hat{p}_{BA}))\overline{N_{BA}} \right) - \left( g(G^{-1}(\hat{p}_{AB}))\overline{N_{AB}} + g(G^{-1}(\hat{p}_{BB}))\overline{N_{BB}} \right). \quad (\text{A10})$$

We use the notation  $\overline{N_{SR}}$  to denote the average of the external noise fields in a stimulus-response class of trials, e.g.,  $\overline{N_{AB}}$  is the average of the external noise fields presented on trials where the stimulus was A and the observer responded B.

If the observer is unbiased, a further simplification is possible. In any response matrix, the probabilities in each row sum to one:  $\hat{p}_{AA} + \hat{p}_{AB} = 1$  and  $\hat{p}_{BA} + \hat{p}_{BB} = 1$ . Furthermore, it is a property of the normal distribution that  $g(G^{-1}(p)) = g(G^{-1}(1-p))$ . If the observer is unbiased, then  $p_{AA} = p_{BB}$ , and from these equalities it follows (again using the substitution  $p_{SR} = \hat{p}_{SR}$ ) that the numerator of (A9) is the same for all four classes of trials, and this factor can be dropped, leaving

$$w_{SR} = \pm \frac{1}{\hat{n}_{SR}}. \quad (\text{A11})$$

That is, the optimal weighted sum for calculating a classification image for an unbiased observer is derived by averaging the noise fields in each cell of the response matrix, adding the averages from response-A cells, and subtracting the averages from response-B cells:

$$C = (\overline{N_{AA}} + \overline{N_{BA}}) - (\overline{N_{AB}} + \overline{N_{BB}}). \quad (\text{A12})$$

To calculate the optimal weights as in (A8), we need to know the true response probabilities  $p_{SR}$ . In practice, we can only count the number of trials  $\hat{n}_{SR}$  in each cell and make estimates of the probabilities,  $\hat{p}_{SR} = \hat{n}_{SR}/(n/2)$ , from a limited total number of trials. This may introduce biases into our formulas for classification images. For example, although  $\hat{n}_{SR}$  is an unbiased estimator of  $n_{SR}$ ,  $1/\hat{n}_{SR}$  is not an unbiased estimator of  $1/n_{SR}$ , so the weights calculated according to (A11) will be biased. However, the biases are very small: when we evaluate the expected value of (A11) numerically, we find that at a response probability of  $p_{SR} = 0.75$ , even with only 100 trials, the bias is less than 1%. The number of trials in a response classification experiment is typically very large (e.g., 10,000 trials per condition in Gold, Murray, Bennett, & Sekuler [2000]), so our estimates of  $\hat{n}_{SR}$  and  $\hat{p}_{SR}$  are very accurate, and the biases in (A10) and (A15) are negligible.

## SNR of the optimal classification image

Because (A12) is the optimally weighted sum of individual noise fields, the SNR of the classification image C is the sum of the SNRs of the individual noise fields (see “Appendix E”). The SNR of a single noise field on trial type SR is given by (A6). We can find the SNR obtained from all trials of type SR by multiplying (A6) by the expected number of trials of type SR,  $(n/2)p_{SR}$ , and we can find the SNR of the classification image C by summing these SNRs across all four types of trials. For a unbiased observer with proportion correct  $p_C = p_{AA} = p_{BB}$ ,

$$\begin{aligned} \text{SNR}[C] = & \frac{np_{AA}}{2} \cdot \frac{\sigma_N^2}{(\sigma_N^2 + \sigma_Z^2)} \cdot \frac{g(G^{-1}(p_{AA}))^2}{p_{AA}^2} \\ & + \frac{np_{BA}}{2} \cdot \frac{\sigma_N^2}{(\sigma_N^2 + \sigma_Z^2)} \cdot \frac{g(G^{-1}(p_{BA}))^2}{p_{BA}^2} \\ & + \frac{np_{AB}}{2} \cdot \frac{\sigma_N^2}{(\sigma_N^2 + \sigma_Z^2)} \cdot \frac{g(G^{-1}(p_{AB}))^2}{p_{AB}^2} \\ & + \frac{np_{BB}}{2} \cdot \frac{\sigma_N^2}{(\sigma_N^2 + \sigma_Z^2)} \cdot \frac{g(G^{-1}(p_{BB}))^2}{p_{BB}^2} \end{aligned} \quad (\text{A13})$$

$$= \frac{n\sigma_N^2}{(\sigma_N^2 + \sigma_Z^2)} \cdot \frac{g(G^{-1}(p_C))^2}{p_C(1-p_C)} \quad (\text{A14})$$

$$= \frac{n\sigma_N^2}{(\sigma_N^2 + \sigma_Z^2)} \cdot \frac{g(d'/2)^2}{G(d'/2)G(-d'/2)}. \quad (\text{A15})$$

Expression (A15) assumes that we use the true optimal weights  $w_{SR}$ , whereas we can only estimate the optimal weights from a limited amount of data. Hence (A15) overestimates the quality of the classification image calculated according to (A12). However, as we have pointed out, the number of trials in a response classification experiment is typically very large, and in this case, our estimates of the weights will be quite accurate, and the error in (A15) will be small.

## Appendix B: Several Contrast Levels

In this appendix, we derive the optimal weighted sum for calculating a classification image in a two-alternative identification experiment in which the signal is presented at several different contrast levels.

Expression (A6) shows that the SNR of a noise field depends on the observer's performance at the signal contrast level of the trial on which the noise field was presented, so when calculating a classification image, it is inefficient to simply average together noise fields from trials with different signal contrast levels. In an experiment with several signal contrast levels and an unbiased observer, we can use (A12) to find the optimal weighted sum of noise fields at each contrast level, to obtain a number of classification images  $C_i$ . In "Appendix E," we show that the weighted sum  $C = \sum_i w_i C_i$  has the

highest SNR when  $w_i = \|E[C_i]\| / \sigma_{C_i}^2$ . From expressions (A3), (A4), and (A5) for the mean and variance of individual noise fields, and the method (A12) of combining noise fields to calculate a classification image, it follows that the magnitude and variance of the classification images  $C_i$  obtained at each stimulus level are

$$\|E[C_i]\| = \frac{2\sigma_N^2}{\sqrt{\sigma_N^2 + \sigma_Z^2}} \cdot \frac{g(d'_i/2)}{G(d'_i/2)G(-d'_i/2)} \quad (\text{B1})$$

$$\sigma_{C_i}^2 = \frac{4\sigma_N^2}{n_i G(d'_i/2)G(-d'_i/2)}. \quad (\text{B2})$$

The optimal weights  $w_i$  are the ratio of (B1) and (B2):

$$w_i = \frac{(n_i/2)g(d'_i/2)}{\sqrt{\sigma_N^2 + \sigma_Z^2}}. \quad (\text{B3})$$

Eliminating scale factors common to the weights at all stimulus levels, we are left with

$$w_i = n_i g(d'_i/2). \quad (\text{B4})$$

Hence the optimal weighted sum for calculating a classification image when several signal contrast levels are interleaved and the observer is unbiased is to calculate classification images  $C_i$  at each stimulus level using (A12), and to take the weighted sum  $C = \sum_i n_i g(d'_i/2) C_i$ .

In replacing (B3) with (B4), we have assumed that the internal noise level  $\sigma_Z^2$  is the same at all signal contrasts. As we mentioned in the "General Discussion" section, internal noise power actually grows with the signal contrast as well as with the external noise power, so this assumption is false. However, if a large part of the stimulus energy comes from external noise, as is often the case in a response classification experiments because of the high power of the noise added to the signal, then the proportional noise will vary little with the signal contrast. (See our discussion of fixed and proportional noise in the general discussion.) In any case, internal proportional noise is not yet understood well enough for us to anticipate how it will vary with the signal contrast in general (e.g., how the constant of proportionality depends on the shape of the signal), so we will use the rough approximation that the internal noise power does not vary with signal contrast when a signal is shown in strong external noise. In an experiment where one is in a position to measure the internal noise power at each signal contrast, the weights in (B3) can be used instead of those in (B4).

Finally, because (B4) is the optimal weighted sum of the images  $C_i$ , the SNR of the combined image  $C$  is the sum of the SNRs of the individual images. The SNR of the individual images  $C_i$  is given by (A15), and it follows directly that the SNR of the pooled image  $C$  is

$$\text{SNR}[C] = \frac{\sigma_N^2}{(\sigma_N^2 + \sigma_Z^2)} \sum_i n_i \frac{g(d'_i/2)^2}{G(d'_i/2)G(-d'_i/2)}. \quad (\text{B5})$$

## Appendix C: Response Consistency

In this appendix, we derive the optimal weighted sum for calculating a classification image in a two-pass experiment. We assume that the reader is familiar with the derivation in “Appendix A”; we follow “Appendix A” closely, and we simply allude to arguments that are developed there more fully.

Consider the trials on which signal A is presented. We will adopt the coordinate frame  $F'$  introduced in “Appendix A.” On a repeated pair of trials, the external noise  $N'$  is the same on both trials, but the internal noise samples  $Z^1$  (first trial) and  $Z^2$  (second trial) are independent. As shown in “Appendix A,” the observer's response is independent of noise components  $N'_2$  through  $N'_m$ , so the conditional mean of these components is zero. The conditional mean of  $N'_1$  is

$E[N'_1 | N'_1 + Z^1_1 > a' \wedge N'_1 + Z^2_1 > a']$ . In “Appendix F,” we give an expression (F3) for conditional means of this form, and this expression shows that

$$E[N'_1 | AAA] = \frac{2}{p_{AAA}} \frac{\sigma_N^2}{\sqrt{\sigma_N^2 + \sigma_Z^2}} g(z) G\left(z \sqrt{\frac{1}{1 + 2\sigma_N^2 / \sigma_Z^2}}\right) \quad (C1)$$

$$\text{where } z = a' / \sqrt{\sigma_N^2 + \sigma_Z^2}.$$

Here  $p_{AAA}$  is the probability that the observer responds A on both trials of a trial pair on which the signal is A.

A similar argument shows that on trial pairs where the observer responds A on the first trial and B on the second trial, the mean of components  $N'_2$  through  $N'_m$  is zero, and the mean of  $N'_1$  is

$E[N'_1 | N'_1 + Z^1_1 > a' \wedge N'_1 + Z^2_1 < a']$ , which expression (F4) shows to be equal to

$$E[N'_1 | AAB] = \frac{2}{p_{AAB}} \frac{\sigma_N^2}{\sqrt{\sigma_N^2 + \sigma_Z^2}} g(z) \times \left(0.5 - G\left(z \sqrt{\frac{1}{1 + 2\sigma_N^2 / \sigma_Z^2}}\right)\right). \quad (C2)$$

The same expression gives the mean on trial pairs where the observer responds B on the first trial and A on the second trial.

Finally, on trial pairs where the observer incorrectly responds B twice, the mean of components  $N'_2$  through  $N'_m$  is zero, and the mean of component  $N'_1$  is

$E[N'_1 | N'_1 + Z^1_1 < a' \wedge N'_1 + Z^2_1 < a']$ . This can be

rewritten as  $-E[-N'_1 | -N'_1 - Z^1_1 > -a' \wedge -N'_1 - Z^2_1 > -a']$ , and we can use (F3) again to evaluate this expression:

$$E[N'_1 | ABB] = \frac{2}{p_{ABB}} \frac{\sigma_N^2}{\sqrt{\sigma_N^2 + \sigma_Z^2}} g(z) \times \left(1 - G\left(z \sqrt{\frac{1}{1 + 2\sigma_N^2 / \sigma_Z^2}}\right)\right). \quad (C3)$$

For the variance of single noise fields, we will use the same approximation as in “Appendix A”:

$$\sigma_{N|AAA}^2 \approx \sigma_{N|AAB}^2 \approx \sigma_{N|ABA}^2 \approx \sigma_{N|ABB}^2 \approx \sigma_N^2 \quad (C4)$$

The same analysis can be carried out for trials on which the signal is B, and we obtain expressions of the form (C1) through (C4), with signal A replaced by B.

Taking the ratios of (C1), (C2), and (C3) with (C4), dropping scale factors common to all weights, and replacing the probability  $p_{AAA}$  with  $n_{AAA}$ ,  $p_{AAB}$  with  $n_{AAB}$ , etc., and using the fact that for an unbiased observer  $z = d'/2$ , we find that for an unbiased observer, the optimal weighted sum for calculating a classification image is

$$\begin{aligned} C &= w(\overline{N_{AAA}} - \overline{N_{BBB}}) \\ &+ (w - 0.5)(\overline{N_{BAB}} + \overline{N_{BBA}} - \overline{N_{AAB}} - \overline{N_{ABA}}) \\ &+ (1 - w)(\overline{N_{BAA}} - \overline{N_{ABB}}) \end{aligned} \quad (C5)$$

$$\text{where } w = G\left(\frac{d'}{2\sqrt{1 + 2\sigma_N^2 / \sigma_Z^2}}\right).$$

The SNR of C is the sum of the SNRs of the individual noise fields. The SNRs of the individual noise fields are the ratios of (C1), (C2), and (C3) squared with (C4), and summing these ratios times the expected number of trials in each category, we obtain

$$\begin{aligned} SNR[C] &= \frac{4n\sigma_N^2}{(\sigma_N^2 + \sigma_Z^2)} g(d'/2)^2 \times \\ &\left(\frac{w^2}{p_{CC}} + \frac{4(w - 0.5)^2}{p_{CI}} + \frac{(w - 1)^2}{p_{II}}\right). \end{aligned} \quad (C6)$$

In this expression,  $p_{CC}$  is the probability of two correct responses ( $p_{CC} = p_{AAA} = p_{BBB}$ ),  $p_{CI}$  is the probability of one correct response and one incorrect response (this probability covers two cells of the response matrix, AAB and ABA, as the order of the correct and incorrect

responses are normally of no interest; hence  $p_{CI}=2p_{AAB}=2p_{ABA}=2p_{BAB}=2p_{BBA}$ , and  $p_{II}$  is the probability of two incorrect responses ( $p_{II}=p_{ABB}=p_{BAA}$ ).

## Appendix D: Rating Scales

In this appendix, we derive the optimal weighted sum for calculating a classification image in a rating experiment. Again, we assume that the reader is familiar with “Appendix A.”

First, consider the trials on which signal A is presented. We will adopt the coordinate frame  $F'$  introduced in “Appendix A.” On trials where the observer’s response is rating  $i$ , the decision variable  $s'$  falls between criteria  $a_i'$  and  $a_{i+1}'$ . The observer’s response is independent of noise components  $N'_2$  through  $N'_m$ , so the mean of these components on trials where the observer responds  $i$  is zero. The mean of the first component,  $E[N'_1 | N'_1 + Z'_1 \geq a_i' \wedge N'_1 + Z'_1 < a_{i+1}']$ , can be evaluated using (F5) from “Appendix F,” and we find

$$E[N'_1 | A_i] = \frac{\sigma_N^2}{\sqrt{\sigma_N^2 + \sigma_Z^2}} \cdot \frac{g(z_{Ai}') - g(z_{Ai+1}')} {G(z_{Ai+1}') - G(z_{Ai}')} \quad (D1)$$

$$\text{where } z_{Ai}' = a_i' / \sqrt{\sigma_N^2 + \sigma_Z^2}.$$

The quantities  $z_{Ai}'$  are the normal deviates of the criteria  $a_i'$  with respect to the distribution of the decision variable  $s'$  on trials on which the stimulus is A. These normal deviates can be estimated from the observer’s response probabilities,  $z_i' = G^{-1}(P_{Ai-})$ , where  $p_{Ai-}$  is the probability that the observer gives a rating less than  $i$  on a trial where stimulus A is presented.

The same analysis applies to trials where signal B is presented. We adopt the coordinate frame  $F''$  introduced in “Appendix A.” Again, the observer’s response is independent of external noise components  $N''_2$  through  $N''_m$ , so the mean of these components on trials where the observer responds  $i$  is zero. Expression (F5) shows that the mean of the first component

$$E[N''_1 | N''_1 + Z''_1 \geq a_i'' \wedge N''_1 + Z''_1 < a_{i+1}''] \text{ is}$$

$$E[N''_1 | B_i] = \frac{\sigma_N^2}{\sqrt{\sigma_N^2 + \sigma_Z^2}} \cdot \frac{g(z_{Bi}'') - g(z_{Bi+1}'')} {G(z_{Bi+1}'') - G(z_{Bi}'')} \quad (D2)$$

$$\text{where } z_{Bi}'' = a_i'' / \sqrt{\sigma_N^2 + \sigma_Z^2}.$$

Note that the quantities  $z_{Ai}'$  are the normal deviates of the criteria  $a_i'$  with respect to the distribution of the decision variable on trials where the signal is A, and the quantities  $z_{Bi}''$  are the normal deviates of the criteria  $a_i''$

with respect to the distribution of the decision variable on trials where the signal is B. The signal detection analysis of this task holds that the criteria are the same on signal-A and signal-B trials. The normal deviate of the mean on signal-A trials with respect to the distribution on signal-B trials is  $d'$ , and it follows that  $z_{Ai}' = z_{Bi}'' - d'$ . We can form more accurate estimates of the observer’s criteria by averaging the estimates from signal-A and signal-B trials:

$$z_i = \frac{z_{Ai}' + z_{Bi}'' - d'}{2} \quad (D3)$$

$$= \frac{G^{-1}(p_{Ai-}) + G^{-1}(p_{Bi-}) - d'}{2}. \quad (D4)$$

These values  $z_i$  are estimates of the normal deviates of the observer’s criteria with respect to the distribution of the decision variable  $s$  on signal-A trials; the normal deviates with respect to the distribution on signal-B trials are  $z_i + d'$ . We could form better estimates  $z_i$  by combining the estimates  $z_{Ai}'$  and  $z_{Bi}'' - d'$  in a weighted average that takes account of the standard errors associated with these quantities (Taylor, 1982), but we will not develop these details.

For the variance of single noise fields, we will use the same approximation as in “Appendix A”:

$$\sigma_{N|Ai}^2 \approx \sigma_{N|Bi}^2 \approx \sigma_N^2. \quad (D5)$$

Taking the ratio of (D1, D2) and (D5), we find that the optimal weights for noise fields from each response cell  $A_i$  and  $B_i$  are

$$w_{Ai} = \frac{1}{\sqrt{\sigma_N^2 + \sigma_Z^2}} \cdot \frac{g(z_i) - g(z_{i+1})}{G(z_{i+1}) - G(z_i)} \quad (D6)$$

$$w_{Bi} = \frac{1}{\sqrt{\sigma_N^2 + \sigma_Z^2}} \cdot \frac{g(z_i + d') - g(z_{i+1} + d')}{G(z_{i+1} + d') - G(z_i + d')}. \quad (D7)$$

Each denominator term of the form  $G(z_i) - G(z_{i+1})$  is the probability that the decision variable  $s$  falls between criteria  $a_i$  and  $a_{i+1}$ , i.e., the probability that a trial receives rating  $i$ , which is proportional to the number of trials that receive rating  $i$ . Using this fact, and dropping scale factors common to all weights, we find that the optimal weighted sum is

$$C = \sum_{i=1}^r (g(z_i) - g(z_{i+1})) \overline{N_{Ai}} + (g(z_i + d') - g(z_{i+1} + d')) \overline{N_{Bi}} \quad (D8)$$

where the normal deviates  $z_i$  are calculated from both signal-A trials and signal-B trials as in (D4).

Note that the optimal weighted sum for calculating a classification image for a biased or unbiased observer in a two-alternative identification task (A10, A12) can be derived from this expression, by considering the case of a rating scale with only two responses.

Finally, because (D8) is the optimal weighted sum of the individual noise fields, the SNR of  $C$  is the sum of the SNRs of the individual images. The SNR of the individual noise fields is given by the ratio of (D1, D2) squared and (D5):

$$SNR[N | Ai] = \frac{\sigma_N^2}{\sigma_N^2 + \sigma_Z^2} \left( \frac{g(z_i) - g(z_{i+1})}{G(z_{i+1}) - G(z_i)} \right)^2 \quad (D9)$$

$$SNR[N | Bi] = \frac{\sigma_N^2}{\sigma_N^2 + \sigma_Z^2} \left( \frac{g(z_i + d') - g(z_{i+1} + d')}{G(z_{i+1} + d') - G(z_i + d')} \right)^2. \quad (D10)$$

The SNR of  $C$  is the sum of the SNRs of the noise fields in each cell of the response matrix, given by (D9) and (D10), times the expected number of noise fields in each cell:

$$SNR[C] = \frac{(n/2)\sigma_N^2}{\sigma_N^2 + \sigma_Z^2} \times \sum_{i=1}^r \left( \frac{(g(z_i) - g(z_{i+1}))^2}{G(z_{i+1}) - G(z_i)} + \frac{(g(z_i + d') - g(z_{i+1} + d'))^2}{G(z_{i+1} + d') - G(z_i + d')} \right). \quad (D11)$$

## Appendix E: Combining Images of Different SNRs

Suppose we have  $n$  stochastic images  $C_i$ , all of whose means are proportional to an image  $T$ , but that have different SNRs. If we take a weighted sum  $C = \sum_i w_i C_i$  to

estimate  $T$  more precisely, what weights  $w_i$  should we use to maximize the SNR of the weighted sum  $C$ ? (We are interested only in obtaining an image  $C$  that is proportional to  $T$ , and we do not care about estimating the magnitude of  $T$ .)

Let us define  $\mu_i = \|E[C_i]\|$  and  $\sigma_i^2$  as the variance of each pixel of  $C_i$ . Rescaling an image does not change its SNR, so we can rescale the images  $C_i$  to have unit magnitude,  $C_i' = C_i / \mu_i$ , and consider weighted sums of  $C_i'$ . The unit-magnitude images  $C_i'$  have variance  $\sigma_i'^2 = \sigma_i^2 / \mu_i^2$ . Similarly, we can require that the weights be chosen so that the magnitude of the sum  $C'$  is one. The sum then takes the form

$$C' = u_1 C_1' + u_2 C_2' + \dots + u_{n-1} C_{n-1}' + (1 - u_1 - \dots - u_{n-1}) C_n'. \quad (E1)$$

The SNR of this sum is

$$SNR[C'] = \frac{\|E[C']\|^2}{\sigma_{C'}^2} \quad (E2)$$

$$= \frac{1}{u_1^2 \sigma_1'^2 + \dots + u_{n-1}^2 \sigma_{n-1}'^2 + (1 - u_1 - \dots - u_{n-1})^2 \sigma_n'^2}. \quad (E3)$$

This SNR is maximized when the denominator is minimized. Taking the derivative of the denominator with respect to each weight  $u_i$ , and requiring that each derivative equal zero, we obtain a system of linear equations for the weights  $u_i$  whose solution is

$$u_i = \frac{1/\sigma_i'^2}{\sum_j 1/\sigma_j'^2}. \quad (E4)$$

It follows that the  $n$ th weight  $(1 - u_1 - \dots - u_{n-1})$  is

$$u_n = \frac{1/\sigma_n'^2}{\sum_j 1/\sigma_j'^2}. \text{ The SNR of the sum } C' \text{ is unaffected if}$$

we multiply each weight by  $\sum_j 1/\sigma_j'^2$ , so the SNR of the weighted sum of images  $C_i'$  is maximized with weights  $u_i = 1/\sigma_i'^2$ , and the SNR of the weighted sum of images  $C_i$  is maximized with weights  $w_i = u_i / \mu_i = \mu_i / \sigma_i^2$ :

$$C = \sum_i \frac{\mu_i}{\sigma_i^2} C_i. \quad (E5)$$

The SNR of the optimal weighted sum  $C$  is

$$SNR[C] = \frac{\left\| E \left[ \sum_i \frac{\mu_i}{\sigma_i^2} C_i \right] \right\|^2}{\text{var}_p \left[ \sum_i \frac{\mu_i}{\sigma_i^2} C_i \right]} = \frac{\left\| \sum_i \frac{\mu_i}{\sigma_i^2} E[C_i] \right\|^2}{\sum_i \left( \frac{\mu_i}{\sigma_i^2} \right)^2 \text{var}_p[C_i]}. \quad (E6)$$

The images  $C_i$  have proportional means,  $E[C_i] = \mu_i \frac{T}{\|T\|}$ , and it follows that

$$= \frac{\left( \sum_i \frac{\mu_i^2}{\sigma_i^2} \right)^2}{\sum_i \frac{\mu_i^2}{\sigma_i^2}} = \sum_i \frac{\mu_i^2}{\sigma_i^2} = \sum_i SNR[C_i]. \quad (E7)$$



Hence the sum of the optimally weighted sum  $C$  is the sum of the SNRs of the individual images  $C_i$ .

Finally, we will justify the claim we made in the results section of [Experiment 1](#), that we can calculate the optimal weights for calculating classification images so long as we know the SNRs of noise fields in each cell of the response matrix up to a common scale factor. Expression (E5) shows that, in general, we must know both the expected values  $\mu_i$  and the variances  $\sigma_i^2$  of a set of stochastic images, to find the optimal weights for a weighted sum of the images. In “[Appendices A through D](#),” in which we derive optimal expression for calculating classification images, only the expected values  $\mu_i$  of the noise fields depend on the noisy cross-correlator model (1). The variances  $\sigma_i^2$  of the noise fields follow directly from the statistics of sampling error. Hence when we compare the predicted SNRs of each class of noise fields to the measured rSNRs, we are testing the model’s prediction of the expected value of each class of noise fields. If our predictions about the expected value of each class of noise fields are mistaken by a common scale factor  $b$ , the weights we use to calculate a classification image will nevertheless be correct: rather than weighting each noise field by  $\mu_i / \sigma_i^2$ , as in (E5), we will weight each noise field by  $b\mu_i / \sigma_i^2$ , and multiplying each weight by a common scale factor does not change the SNR of the weighted sum. Hence the resulting classification image will still have the maximum possible SNR.

## Appendix F: Conditional Statistics of Normal Random Variables

In signal detection theory, an observer’s responses are often modeled as being based on a decision variable that is the sum of a random variable  $X$  determined by the stimulus and a random variable  $Y$  determined by noise internal to the observer. It is sometimes useful to know the statistics of  $X$  conditional on the sum  $X+Y$  falling in a particular range. In this appendix, we state some results of this type, and we give a full proof of one such result. The proofs are simply evaluations of integrals, and we hope it is clear from the example given how the other integrals can be completed.

*Theorem.* Let  $X \sim N(\mu_X, \sigma_X^2)$  and  $Y_1, Y_2 \sim N(\mu_Y, \sigma_Y^2)$  be independent Gaussian random variables, and let  $a_1$  and  $a_2$  be real numbers. Define

$$z_1 = (a_1 - (\mu_X + \mu_Y)) / \sqrt{\sigma_X^2 + \sigma_Y^2} \text{ and}$$

$z_2 = (a_2 - (\mu_X + \mu_Y)) / \sqrt{\sigma_X^2 + \sigma_Y^2}$ . That is,  $z_1$  and  $z_2$  are the normal deviates of  $a_1$  and  $a_2$ , respectively, with respect to the distribution of  $X + Y_1$  or  $X + Y_2$ . Then,

$$E[X | X + Y_1 > a_1] = \mu_X + \frac{\sigma_X^2}{\sqrt{\sigma_X^2 + \sigma_Y^2}} \cdot \frac{g(z_1)}{G(-z_1)} \quad (F1)$$

$$E[X^2 | X + Y_1 > a_1] = \sigma_X^2 \left( 1 + z_1 \frac{g(z_1)}{G(-z_1)} \cdot \frac{\sigma_X^2}{\sigma_X^2 + \sigma_Y^2} \right) + 2\mu_X E[X | X + Y_1 > a_1] + \mu_X^2 \quad (F2)$$

$$E[X | X + Y_1 > a_1 \wedge X + Y_2 > a_1] = \mu_X + \frac{2}{p_{YY}} \cdot \frac{\sigma_X^2}{\sqrt{\sigma_X^2 + \sigma_Y^2}} \cdot g(z_1) G\left(z_1 \sqrt{\frac{1}{1 + 2\sigma_X^2 / \sigma_Y^2}}\right) \quad (F3)$$

where  $p_{YY} = P(X + Y_1 > a_1 \wedge X + Y_2 > a_1)$

$$E[X | X + Y_1 > a_1 \wedge X + Y_2 < a_1] = \mu_X + \frac{2}{p_{YN}} \cdot \frac{\sigma_X^2}{\sqrt{\sigma_X^2 + \sigma_Y^2}} \cdot g(z_1) \left( 0.5 - G\left(z_1 \sqrt{\frac{1}{1 + 2\sigma_X^2 / \sigma_Y^2}}\right) \right) \quad (F4)$$

where  $p_{YN} = P(X + Y_1 > a_1 \wedge X + Y_2 < a_1)$

$$E[X | X + Y_1 > a_1 \wedge X + Y_1 < a_2] = \mu_X + \frac{\sigma_X^2}{\sqrt{\sigma_X^2 + \sigma_Y^2}} \cdot \frac{g(z_1) - g(z_2)}{G(z_2) - G(z_1)}. \quad (F5)$$

The function  $g(x, \mu, \sigma)$  is the normal probability density function, and  $G(x, \mu, \sigma)$  is the normal cumulative distribution function. When we omit the arguments  $\mu$  and  $\sigma$ , they default to 0 and 1, respectively.

Note that  $G(-z_1) = P(X + Y_1 > a_1)$  and  $G(-z_2) = P(X + Y_1 > a_2)$ . Sometimes we make this substitution when we use these results in other parts of the article.

*Proof of (F1).* First, consider the case where  $\mu_X = 0$ .

$$E[X | X + Y_1 > a_1] = \int_{-\infty}^{\infty} x P(X = x | X + Y_1 > a_1) dx$$

$$= \frac{1}{P(X + Y_1 > a_1)} \int_{-\infty}^{\infty} x P(X = x) P(Y_1 > -x + a_1) dx$$

$$\begin{aligned}
&= \frac{1}{1 - G(a_1, \mu_Y, \sqrt{\sigma_X^2 + \sigma_Y^2})} \\
&\times \int_{-\infty}^{\infty} xg(x, 0, \sigma_X)(1 - G(-x + a_1, \mu_Y, \sigma_Y))dx \\
&= \frac{1}{G(-a_1, -\mu_Y, \sqrt{\sigma_X^2 + \sigma_Y^2})} \\
&\times \int_{-\infty}^{\infty} xg(x, 0, \sigma_X)G(x, a_1 - \mu_Y, \sigma_Y)dx.
\end{aligned}$$

Integrating by parts, this becomes

$$\begin{aligned}
&= \frac{1}{G(-a_1, -\mu_Y, \sqrt{\sigma_X^2 + \sigma_Y^2})} \\
&\times \left( (-\sigma_X^2 g(x, 0, \sigma_X)G(x, a_1 - \mu_Y, \sigma_Y)) \Big|_{-\infty}^{\infty} \right. \\
&\quad \left. - \int_{-\infty}^{\infty} -\sigma_X^2 g(x, 0, \sigma_X)g(x, a_1 - \mu_Y, \sigma_Y)dx \right) \\
&= \frac{\sigma_X^2}{G(-a_1, -\mu_Y, \sqrt{\sigma_X^2 + \sigma_Y^2})} \\
&\times \int_{-\infty}^{\infty} g(\mu_Y - a_1, 0, \sqrt{\sigma_X^2 + \sigma_Y^2})g(x, \mu^*, \sigma^*)dx.
\end{aligned}$$

Here we have used the fact that the pointwise product of two normal density functions is a scaled normal density function, and we have set  $\mu^* = (a_1 - \mu_Y)\sigma_X^2 / \sqrt{\sigma_X^2 + \sigma_Y^2}$  and  $\sigma^{*2} = \sigma_X^2 \sigma_Y^2 / (\sigma_X^2 + \sigma_Y^2)$ . The density function with these parameters integrates to 1, and we are left with

$$\begin{aligned}
&= \frac{\sigma_X^2}{G(-a_1, -\mu_Y, \sqrt{\sigma_X^2 + \sigma_Y^2})} g(\mu_Y - a_1, 0, \sqrt{\sigma_X^2 + \sigma_Y^2}) \\
&= \frac{\sigma_X^2}{\sqrt{\sigma_X^2 + \sigma_Y^2}} \cdot \frac{g((\mu_Y - a_1) / \sqrt{\sigma_X^2 + \sigma_Y^2})}{G((\mu_Y - a_1) / \sqrt{\sigma_X^2 + \sigma_Y^2})}. \quad (F6)
\end{aligned}$$

This is the expression for the conditional mean when  $\mu_X = 0$ .

When  $\mu_X \neq 0$ , we can reduce the problem to the zero-mean case by defining  $X' = X - \mu_X$  and  $a_1' = a_1 - \mu_X$ . Then,

$$\begin{aligned}
E[X | X + Y_1 > a_1] &= E[(X' + \mu_X) | (X' + \mu_X) + Y_1 > (a_1' + \mu_X)] \\
&= E[X' | (X' + \mu_X) + Y_1 > (a_1' + \mu_X)] \\
&\quad + E[\mu_X | (X' + \mu_X) + Y_1 > (a_1' + \mu_X)] \\
&= E[X' | X' + Y_1 > a_1'] + \mu_X.
\end{aligned}$$

This conditional mean can be evaluated using (F6), and we find

$$\begin{aligned}
&= \mu_X + \frac{\sigma_X^2}{\sqrt{\sigma_X^2 + \sigma_Y^2}} \cdot \frac{g((\mu_Y - (a_1 - \mu_X)) / \sqrt{\sigma_X^2 + \sigma_Y^2})}{G((\mu_Y - (a_1 - \mu_X)) / \sqrt{\sigma_X^2 + \sigma_Y^2})} \\
&= \mu_X + \frac{\sigma_X^2}{\sqrt{\sigma_X^2 + \sigma_Y^2}} \cdot \frac{g(z_1)}{G(-z_1)}.
\end{aligned}$$

## Acknowledgments

We would like to thank Craig Abbey and Brent Beutter for helpful discussions. We would also like to thank two anonymous reviewers for their insightful comments. This research was supported by National Science and Engineering Research Council grants OGP0105494 and OGP0042133. Commercial relationships: None.

## References

- Abbey, C. K., Eckstein, M. P., & Bochud, F. O. (1999). Estimation of human-observer templates in two-alternative forced-choice experiments. *Proceedings of SPIE*, 3663, 284-295.
- Ahumada, A., Jr., & Lovell, J. (1971). Stimulus features in signal detection. *Journal of the Acoustical Society of America*, 49, 1751-1756.
- Ahumada, A. J., & Beard, B. L. (1999). Classification images for detection [Abstract]. *Investigative Ophthalmology and Visual Science*, 40, S572.
- Ahumada, A. J., Jr. (1987). Putting the visual system noise back in the picture. *Journal of the Optical Society of America A*, 4, 2372-2378. [PubMed]
- Ahumada, A. J., Jr., & Watson, A. B. (1985). Equivalent-noise model for contrast detection and discrimination. *Journal of the Optical Society of America A*, 2, 1133-1139. [PubMed]
- Beard, B. L., & Ahumada, A., Jr. (1998). A technique to extract relevant image features for visual tasks. *Proceedings of SPIE*, 3299, 79-85.

- Burgess, A. E. (1995). Comparison of receiver operating characteristic and forced choice observer performance measurement methods. *Medical Physics*, 22, 643-655. [\[PubMed\]](#)
- Burgess, A. E., & Colborne, B. (1988). Visual signal detection. IV. Observer inconsistency. *Journal of the Optical Society of America A*, 5, 617-627. [\[PubMed\]](#)
- Burgess, A. E., Wagner, R. F., Jennings, R. J., & Barlow, H. B. (1981). Efficiency of human visual signal discrimination. *Science*, 214, 93-94. [\[PubMed\]](#)
- Chubb, C., & Nam, J. H. (2000). Variance of high contrast textures is sensed using negative half-wave rectification. *Vision Research*, 40, 1677-1694. [\[PubMed\]](#)
- Egan, J. P., Schulman, A. I., & Greenberg, G. Z. (1959). Operating characteristics determined by binary decisions and by ratings. *Journal of the Acoustical Society of America*, 31, 768-773.
- Foley, J. M., & Legge, G. E. (1981). Contrast detection and near-threshold discrimination in human vision. *Vision Research*, 21, 1041-1053. [\[PubMed\]](#)
- Gold, J., Bennett, P. J., & Sekuler, A. B. (1999). Signal but not noise changes with perceptual learning. *Nature*, 402, 176-178. [\[PubMed\]](#)
- Gold, J. M., Murray, R. F., Bennett, P. J., & Sekuler, A. B. (2000). Deriving behavioural receptive fields for visually completed contours. *Current Biology*, 10, 663-666. [\[PubMed\]](#)
- Green, D. M. (1964). Consistency of auditory judgments. *Psychological Review*, 71, 392-407.
- Green, D. M., & Swets, J. A. (1974). *Signal detection theory and psychophysics*. Huntington, NY: R. E. Krieger.
- Legge, G. E., Kersten, D., & Burgess, A. E. (1987). Contrast discrimination in noise. *Journal of the Optical Society of America A*, 4, 391-404. [\[PubMed\]](#)
- Lillywhite, P. G. (1981). Multiplicative intrinsic noise and the limits to visual performance. *Vision Research*, 21, 291-296. [\[PubMed\]](#)
- Lu, Z. -L., & Doshier, B. A. (1998). External noise distinguishes attention mechanisms. *Vision Research*, 38, 1183-1198. [\[PubMed\]](#)
- Manjeshwar, R. M., & Wilson, D. L. (2001). Hyperefficient detection of targets in noisy images. *Journal of the Optical Society of America A*, 18, 507-513. [\[PubMed\]](#)
- Murray, R.F., Bennett, P.J., & Sekuler, A.B. (2001). No pointwise nonlinearity in shape discrimination [Abstract]. *Journal of Vision*, 1(3), 52a, <http://journalofvision.org/1/3/52/>, DOI 10.1167/1.3.52. [\[Link\]](#)
- Nam, J. H., & Chubb, C. (2000). Texture luminance judgments are approximately veridical. *Vision Research*, 40, 1695-1709. [\[PubMed\]](#)
- Pelli, D. G. (1985). Uncertainty explains many aspects of visual contrast detection and discrimination. *Journal of the Optical Society of America A*, 2, 1508-1532. [\[PubMed\]](#)
- Pelli, D. G. (1990). The quantum efficiency of vision. In C. Blakemore (Ed.), *Vision: Coding and efficiency* (pp. 3-24). Cambridge, MA: Cambridge University Press.
- Pelli, D. G., & Farell, B. (1999). Why use noise? *Journal of the Optical Society of America A*, 16, 647-653. [\[PubMed\]](#)
- Peterson, W. W., Birdsall, T. G., & Fox, W. C. (1954). The theory of signal detectability. *Transactions of the Institute of Radio Engineers, Professional Group on Information Theory*, 4, 171-212.
- Richards, V. M., & Zhu, S. (1994). Relative estimates of combination weights, decision criteria, and internal noise based on correlation coefficients. *Journal of the Acoustical Society of America*, 95, 423-434. [\[PubMed\]](#)
- Taylor, J. R. (1982). *An introduction to error analysis: The study of uncertainties in physical measurements*. Mill Valley, CA: University Science Books.
- Wickelgren, W. A. (1968). Unidimensional strength theory and component analysis of noise in absolute and comparative judgements. *Journal of Mathematical Psychology*, 5, 102-122.