

# Optimal decisions for contrast discrimination

Department of Psychology, University of Warwick, Coventry, UK, &  
Gatsby Computational Neuroscience Unit, University College  
London, London, UK

Adam N. Sanborn

Peter Dayan

University College London, London, UK



Contrast discrimination functions for simple gratings famously look like a dipper. Discrimination thresholds are lower than detection thresholds for moderate pedestal contrasts, and the rate of growth of thresholds as the pedestal contrast gets larger typically lies between the values implied by two popular treatments of noise. Here, we suggest a new normative treatment of the dipper, showing how it emerges from Bayesian inference based on the responses of a population of orientation-tuned units. Our central assumption concerns the noise corrupting the outputs of these units as a function of the contrast: We suggest that it has the shape of a hinge. We show the match to the psychophysical data and discuss the neurobiological and statistical rationales for this form of noise. Finally, we relate our model to other major accounts of contrast discrimination.

Keywords: computational modeling, contrast sensitivity, detection/discrimination

Citation: Sanborn, A. N., & Dayan, P. (2011). Optimal decisions for contrast discrimination. *Journal of Vision*, 11(14):9, 1–13, <http://www.journalofvision.org/content/11/14/9>, doi:10.1167/11.14.9.

## Introduction

Human behavior can often be understood as optimal inference about the state of the world given noisy data. Tasks as varied as motor control (Faisal & Wolpert, 2009; Körding & Wolpert, 2004), animal learning (Courville, Daw, & Touretzky, 2006; Daw, Courville, & Dayan, 2008), semantic memory (Shiffrin & Steyvers, 1997; Steyvers, Griffiths, & Dennis, 2006), and categorization (Anderson, 1991; Heller, Sanborn, & Chater, 2009; Kemp, Perfors, & Tenenbaum, 2007) have all been successfully modeled in these terms. Vision has been a particular poster child for the approach, with a venerable and influential history of ideal observer (de Vries, 1943; Green & Swets, 1966; Peterson, Birdsall, & Fox, 1954; Rose, 1942) and sequential ideal observer (Geisler, 1989) analyses. As with other applications of Bayesian methods, optimal inference lays assumptions bare and, in its failures, suggests approximations. However, optimal inference has been used only somewhat sparingly in contrast discrimination (Chirimuuta & Tolhurst, 2005); this is our focus.

In a standard two-alternative forced-choice (2AFC) contrast discrimination experiment, participants are presented with two successive stimuli (A and B) that differ only in contrast. They are then asked to declare whether the first or second stimulus had higher contrast. We will treat the lower contrast stimulus as defining a pedestal contrast; the experiments indicate how much contrast needs to be added to this pedestal for the resulting

stimulus to be detected with suitable reliability; this is the threshold contrast increment. The crosses in Figure 1A shows a sample graph relating pedestal contrast to threshold (Foley, 1994). This is non-monotonic, having the shape of a dipper or ladle (Solomon, 2009).

Performance can be separated into two distinct regimes: one each for discrimination about low- and high-contrast pedestals. In the former, the key phenomenon is that the increment thresholds needed for discrimination can be lower than the detection threshold (Campbell & Kulikowski, 1966; Nachmias & Sansbury, 1974); this puts the “dip” in the dipper. Conversely, for high-contrast discrimination, the threshold increments needed to reach threshold rise very regularly, as a power function of contrast with an exponent somewhere between 0.5 and 1 (Barlow, 1957; Campbell & Kulikowski, 1966; Legge, 1981; Nachmias & Sansbury, 1974). On a log–log plot of threshold increment against pedestal contrast, these “handles” of the dipper function form a straight line with slope equal to the exponent. In Figure 1A, for thresholds when the pedestal contrast was high, the slope of the data points lies between 0.5 and 1.

The mechanisms that have been proposed to explain human contrast discrimination performance have largely been aimed at either the low- or the high-contrast regimes. To explain the dips in the contrast discrimination function at low contrast, mechanisms including threshold models (Crozier, 1950; Foley & Legge, 1981; Green & Swets, 1966), an expansive transducer followed by additive noise (Legge & Foley, 1980; Nachmias & Sansbury, 1974), and intrinsic uncertainty about stimuli at low contrast (Pelli,

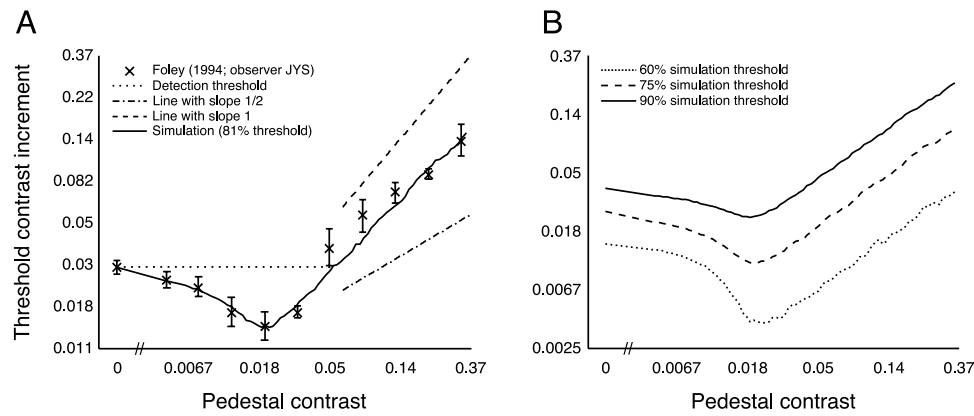


Figure 1. Contrast discrimination data and predictions. (A) Contrast discrimination data of Observer JYS in Experiment 2 of Foley (1994). Error bars are plus or minus one standard error of the mean. The dotted line extends the detection threshold for comparison. The dot-dashed line and dashed lines show handle slopes of 1/2 and 1, respectively. The solid line shows the prediction made by the hinge noise model. (B) Predictions of the hinge noise model for different percent correct thresholds.

1985; Tanner, 1961) have been proposed. These mechanisms can individually produce dips, but none also produces the increasing handle seen in human data. Instead, the handle has been explained as the result of noise that increases more quickly because of a compressive non-linearity followed by additive noise (e.g., Legge & Foley, 1980), a comparison between stimuli based on variance alone (Laming, 1986), or correlations in a particular kind of neural noise (Triesman, 1964).

In this paper, we propose an explanation (which results in the solid line in Figure 1A) for both the dip and the handle of the contrast discrimination function. It also reproduces other empirical contrast discrimination results, including deeper dips for lower percent correct thresholds and steeper slopes for detection compared to high-contrast discrimination (as shown in Figure 1B). Our account is based on a population of units that are tuned to the stimulus parameters, a linear transducer, increasing noise, and optimal decision making. First, we introduce the population model, which we term the hinge noise model, justify its components, and show its close match with the discrimination data. Next, we discuss the computational basis of the model, using the Cramér–Rao lower bound for intuition. Finally, we relate our account to mechanisms that have previously been proposed.

## A new approach to modeling contrast discrimination

The first key component is a population of units tuned to the oriented gratings that serve as input to the system. Populations of units have been used in many models of visual processing and contrast discrimination (Goris, Wichmann, & Henning, 2009; Itti, Koch, & Braun,

2000), and we make standard assumptions about how they are activated. We assume that each unit has a preferred orientation but responds to suboptimal target orientations according to a squared exponential function of the angular difference between preferred and target orientations. Its response also depends on the contrast of the target stimulus.

In total, we model the response  $r_i$  of neuron  $i$ , given an input grating at angle  $\theta$  and contrast  $c$  as being Gaussian distributed (e.g., Goris et al., 2009; Itti et al., 2000; Kontsevich, Chen, & Tyler, 2002):

$$r_i|c, \theta \sim N\left(\bar{r}_i, \tau^2(\bar{r}_i)\right), \text{ where } \bar{r}_i = cf_i(\theta)$$

$$\text{with } f_i(\theta) = \exp\left(-\frac{-(\theta - \theta_i)^2}{2\phi^2}\right). \quad (1)$$

Here,  $\theta_i$  is the preferred orientation of the unit.  $\tau^2$  is the signal-dependent variance of the response, which we describe in detail below.  $\phi = 15^\circ$  is the tuning width, chosen based on the mean empirical values in cat visual cortex for both simple and complex cells (Li, Peterson, & Freeman, 2003), and also close to values used before in fitting human contrast discrimination performance (Itti et al., 2000). In all the simulations we report, we used a population of 36 equally spaced filters across the possible 180 degrees of orientation (for symmetric stimuli), with one unit directly centered on the stimulus orientation. We also make the approximation of allowing the activities  $r_i$  to be negative; this is a simplification away from a non-zero baseline.

The fact that the mean of  $r_i$  scales linearly with the contrast is equivalent to using a linear (in fact, an identity) transducer. Figure 2A cartoons the mean of responses  $r_i$  for various contrasts. Our model abstracts away some of

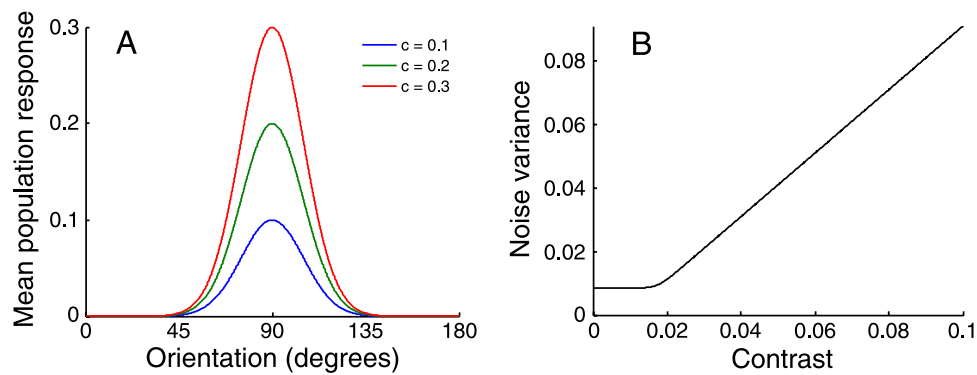


Figure 2. Population elements. (A) Mean response of units to a 90° grating with separate lines for each of a selection of contrasts. The population response scales linearly with contrast and has squared exponential orientation tuning. (B) Hinge noise. This shows the variance of the response as a function of mean response. To make it suitably smooth, we use the functional form  $\tau^2(\bar{r}_i) = \alpha + \gamma \log(\beta + e^{\bar{r}_i/\gamma})$  with  $\gamma = 0.0015$  being very small. The other parameters, used for all simulations, are  $\alpha = -0.009$  and  $\beta = 341.5$ .

the detail of the neural architecture, and so we cannot specify the precise mapping to individual cell types and regions. However, we note that cells in the visual cortex are linear in the middle of their range (Albrecht & Hamilton, 1982) and the responses of neurons very early in the visual pathway are often rather linear functions of contrast, especially for lower contrasts (Derrington & Lennie, 1984; Kaplan & Shapley, 1986), though these cells are not tuned to orientation.

The variance of the response depends on the mean. As shown in the solid lines in Figure 2B, we assume that this takes the form of a (soft) hinge function. For high-contrast values, the variance is nearly equal to the mean and so is consistent with the typical finding that the variability in neural firing rates is roughly a constant multiple of the mean. However, for zero contrast, we assume there to be an irreducible level of noise (allowing, for convenience, the activities to be negative). These dependencies collectively define the hinge function shown in the figure. Note that our model does not make the additional assumption that the irreducible and contrast-dependent noise are separate additive components (cf. Lu & Doshier, 1998).

As in sequential ideal observer analysis (Geisler, 1989), we consider the consequence of making optimal 2AFC decisions based on these noisy activations. This procedure makes the implicit assumption, which is credible for contrast discrimination experiments, that all stimulus and unit tuning properties other than the contrast are known, including the presented orientation, the preferred orientation and tuning width of each unit, the nature and slope of the linear transducer, and the function that relates the variance to the mean.

In this ideal observer context, the form of the noise has two critical effects. First, even if there was only a single unit, it would not necessarily be optimal to declare the 2AFC interval with the greater response as having the greater contrast. Although this decision rule is frequently a default assumption (Foley & Legge, 1981; Legge & Foley, 1980), it is well known that two thresholds are

typically required to separate signal from noise when the distributions concerned have different variances (Green & Swets, 1966, p. 63). Second, though perhaps less intuitively, the noise itself conveys information about the contrast, from which the ideal observer can benefit (Dayan & Abbott, 2001; Shamir & Sompolinsky, 2004; Snippe & Koenderink, 1992; Yoon & Sompolinsky, 1999).

The decision made by the model is a combination of the likelihoods of different contrasts,  $p_A[\mathbf{r}_A|c]$  and  $p_B[\mathbf{r}_B|c]$ , which are determined by the identity transducer and variance functions we assumed above, and the prior information about which contrasts are likely to appear. Different methods of collecting thresholds imply different prior distributions over the target and pedestal contrasts; we consider three canonical cases.

First, a task in which the pedestal and target contrasts are fixed over a block would, in the limit, imply a prior that only these two contrasts would be displayed (Both Known). Here, the best guess is to evaluate the sampled responses as coming from the target contrast,  $c_T$ , and pedestal contrast,  $c_P$ , choosing A as the target if  $p_A[\mathbf{r}_A|c_T]p_B[\mathbf{r}_B|c_P] > p_A[\mathbf{r}_A|c_P]p_B[\mathbf{r}_B|c_T]$ .<sup>1</sup> Second, a threshold obtained by a staircase on the contrast increment, such as used in Foley's (1994) data that we fit in Figure 1, implies that the pedestal contrast would be well known, but the target contrast could be a range of values (Pedestal Known). We captured this case by using a prior distribution in which the pedestal was known and the target contrast was equally likely to be any larger contrast for this task. In this case, interval A is chosen if  $p_A[\mathbf{r}_A|c_T] \int_{c_T=c_P}^1 dc_P p_B[\mathbf{r}_B|c_P] > p_B[\mathbf{r}_B|c_T] \int_{c_T=c_P}^1 dc_P p_A[\mathbf{r}_A|c_P]$ . Finally, if target and pedestal are chosen randomly on each trial, this might imply a uniform prior distribution over both pedestal and target contrasts with the assumption that the target contrast is higher (Neither Known). In this case, interval A is chosen if  $\int_{c_A=0}^1 dc_A p_A[\mathbf{r}_A|c_A] \int_{c_B=0}^{c_A} dc_B p_B[\mathbf{r}_B|c_B] > 0.5$ , where  $c_A$  and  $c_B$  are the possible contrasts of interval A and interval B, respectively. In various limits, the full Bayesian

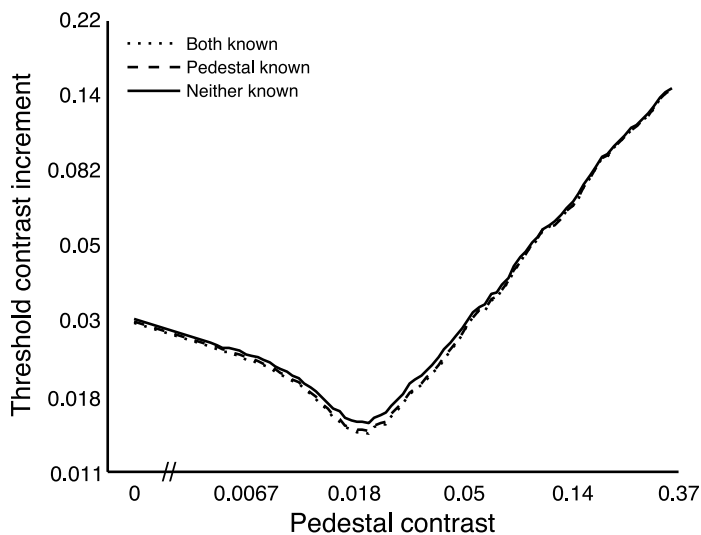


Figure 3. Contrast discrimination threshold predictions using different priors for the hinge noise model. The Both Known prior has exact knowledge of both the pedestal and target contrasts. The Pedestal Known prior has exact knowledge of the pedestal contrast and assumes that the target contrast could be any contrast value above that. The Neither Known prior puts a uniform distribution over the pedestal contrast and assumes that the target contrast is equally likely to be any greater value up to the maximum contrast.

treatment with this prior distribution will give the same answer as maximum likelihood inference (Gelman, Carlin, Stern, & Rubin, 2004; Jaynes, 2003).

As shown in Figure 3, we found that using any of the three prior distributions over contrast made remarkably little difference in the thresholds predicted from the hinge noise model.<sup>2</sup> The reason that the prior only has a slight influence on the threshold is analogous to being told only the final score in a match between two teams: It is often not much help in guessing which team was the winner. This prediction corresponds to empirical results in contrast detection and discrimination, in which prior knowledge of contrast has surprisingly little effect. In forced-choice detection, Davis, Kramer, and Graham (1983) found no difference between blocks with a single contrast (Both Known) and blocks with multiple target contrasts (Pedestal Known). Huang and Dobkins (2005) used a staircase procedure to determine contrast discrimination thresholds and found little difference between blocks in which a single pedestal contrast was used (Pedestal Known) and blocks in which pedestal contrasts were mixed (Neither Known).

The solid line in Figure 1A shows the result of fitting this hinge noise model to the data and constitutes the main result of this paper. It is apparent that the hinge noise model fits both the dip and the handle of the dipper. As for other population models motivated by the characteristics of the first stages of orientation processing in the cortex (Goris et al., 2009; Itti et al., 2000), this fit is based on

more parameters than some more abstract models. However, as we see below, the constraints of optimal inference render our model quite inflexible, making this a significant finding.

In the following sections, we use simplifying analyses to provide some intuition as to how this model produces the dip and the handle and explain why this model is not equivalent to previous, mechanistic, suggestions.

## The Cramér–Rao lower bound

A key tool for analyzing population codes is the Cramér–Rao (CR) lower bound (Oram, Földiák, Perrett, & Sengpiel, 1998; Seung & Sompolinsky, 1993; van den Bos, 2007). This starts from the density function  $p[\mathbf{r}|c]$  that relates contrast to population response  $\mathbf{r}$  (we omit parameters other than contrast, since these are known) and considers estimators  $\hat{c}(\mathbf{r})$  of the contrast given a sample of  $\mathbf{r}$ . The bound limits the minimum variance  $\sigma_{\hat{c}}^2(c)$  of any unbiased such estimators (i.e., having mean  $c$ ) to

$$\sigma_{\hat{c}}^2(c) \geq \frac{1}{\text{FI}(c)}, \quad (2)$$

where FI is the Fisher information, which is a characteristic of the population response  $\mathbf{r}$  and not the estimator, and is given in Appendix A. The variance of the estimator is important because it helps to determine discrimination performance. For instance, for an unbiased estimator with Gaussian statistics with nearly equal variances, the probability of getting the contrast discrimination correct when  $c_p$  is the pedestal contrast and  $c_t$  is the target contrast would be approximately

$$P_C = \Phi^{-1} \left( \frac{c_t - c_p}{\sqrt{\sigma_{\hat{c}}^2(c_t) + \sigma_{\hat{c}}^2(c_p)}} \right), \quad (3)$$

where  $\Phi^{-1}$  is the inverse of the standard cumulative Normal distribution. This would then determine the threshold contrast increment shown in Figure 1.

The discrimination bound in Equation 3 is, in some cases, a good quantitative predictor of model performance but, in other cases, can only be used as a heuristic guide. The bound is only a heuristic guide when the prior on the contrasts is flat over the region of the likelihoods, but given the results of Figure 3 it can still be an accurate guide. For the low-contrast case that we will explore first, the bound is suggestive but less accurate. There are two reasons for this: the bias in the estimator and the non-Gaussian nature of the distribution of maximum likelihood estimates. The first problem concerns how the derivative of the bias of the estimator  $b'_c(c)$  affects the CR lower bound, since bias acts in the same way as non-linear



transduction. Seriés, Stocker, and Simoncelli (2009) show that the resulting change in the variability of the estimator is exactly matched by a change in its systematic output, thus leaving thresholds unaffected. This only holds when the slope of the bias is relatively constant; for the case with low contrast, this slope changes significantly.

In the high-contrast case, the bound does provide a good quantitative prediction of our simulations. Here,  $\sigma_c^2(c)$  either accurately reflects the variance of the Bayesian estimator or can be used to calculate the thresholds. In those cases, with a flat prior on  $c$  (ignoring limits on the contrast), the posterior distribution is approximately Gaussian, with mean and median being the maximum likelihood value of  $c$ :

$$c_{\text{ML}}(\mathbf{r}) = \operatorname{argmax}_c \{p[\mathbf{r}|c]\}, \quad (4)$$

which is asymptotically unbiased and saturates the CR lower bound. The regime in which this is true broadly requires there to be sufficient activation in the population. This is typically true in our case again for all but the lowest contrasts. In all cases, we show the results of substantial, realized, stochastic simulations, as in Figure 1, to back up claims made on the basis of the CR bound.

## Contrast discrimination at low contrasts

The presence of the “dip” in the contrast discrimination function shows that the contrast needed to discriminate between a blank interval and a target can be more than the contrast needed to discriminate between two stimuli both of which have non-zero contrast (Campbell & Kulikowski,

1966; Nachmias & Sansbury, 1974). This may be a surprising result in the face of growing noise: Indeed in standard models, the combination of a linear transducer, increasing noise, and perfect knowledge of all stimulus parameters aside from contrast has not been shown to produce dips. However, optimal decision making with hinge noise does produce a dip. To motivate this result, we examine the form of the Cramér–Rao bound for Gaussian noise in a single unit. We can solve for the bound on the standard deviation of the maximum likelihood estimator when the variance is a function of contrast:

$$\sigma_{\text{ML}}(c) \geq \frac{\tau(\bar{r}_i)}{\sqrt{1 + 2 \left[ \frac{d\tau(\bar{r}_i)}{dc} \right]^2}}, \quad (5)$$

where  $\tau(\bar{r}_i)$  is the hinge noise at a particular mean response and  $\frac{d\tau(\bar{r}_i)}{dc}$  is the derivative of the hinge noise at that mean response with respect to contrast. The denominator in this expression can be shown to arise from the information about  $c$  that is present in signal-dependent noise (Abbott & Dayan, 1999).

For purely additive noise,  $\frac{d\tau(\bar{r}_i)}{dc} = 0$ . Thus, the CR lower bound would depend only on the absolute level of the noise. Hence, from Equation 3, so would the threshold. However, the growth in  $\tau(\bar{r}_i)$  with  $\bar{r}_i$  specified by hinge noise has two opposite effects: The rising standard deviation will increase the bound and push the threshold upward, but an increasing slope will decrease the bound and push the threshold downward. For the case of hinge noise, the change in regime, from nugatory dependence on  $c$  to substantial dependence, makes for the dip.

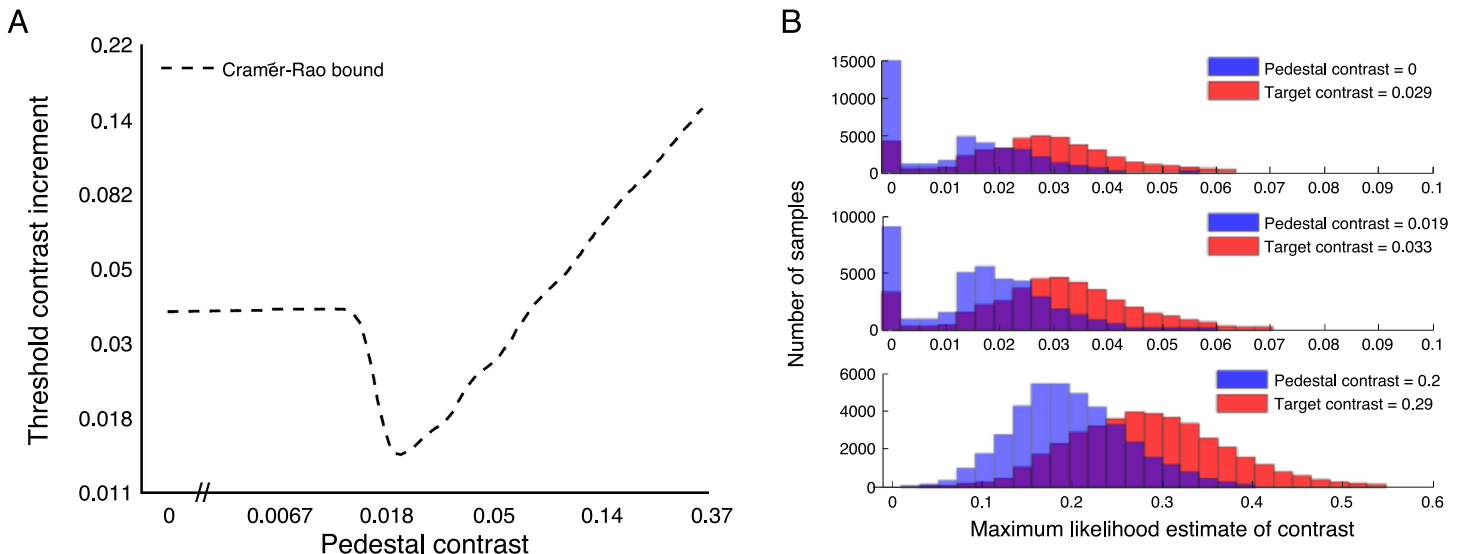


Figure 4. Lower bound on maximum likelihood estimator (MLE) variance and distributions of MLE values at various contrasts. (A) The CR lower bound for the case of hinge noise. (B) Distributions of MLEs for the threshold target and pedestal contrasts for detection (top), discrimination at the maximum dip (middle), and discrimination at high contrast (bottom). The purple region is where the pedestal and target bars overlap. The range of axes is set in each plot to provide the best view of the overlap between target and pedestal distributions.

Figure 4A shows the threshold increments implied by the CR lower bound for our form of hinge noise. However, as mentioned, the CR lower bound can only provide an approximation to the actual performance of the estimator; we show full distributions of the maximum likelihood estimator (constrained to lie between 0 and 1) for various pedestal and threshold target contrasts in Figure 4B. These distributions provide a more accurate approximation of the performance of the hinge noise model, but the predictions shown in Figure 1 make use of the more informative posterior distributions over contrast rather than just the maximum likelihood estimates.

In addition to the dip, our model predicts other empirical effects found at low contrast. The first is that the depth of the dip is larger for lower percentage correct thresholds than for higher (Bird, Henning, & Wichmann, 2002), as shown in Figure 1B. The second is that the slope of the psychometric function for contrast discrimination is steeper for detection than for discrimination on a log plot (Bird et al., 2002; Nachmias & Sansbury, 1974). The smaller distances between the various percentage thresholds for detection relative to discrimination in Figure 1B show that the hinge noise model matches this empirical finding.

## Contrast discrimination at high contrast

The central puzzle for contrast discrimination at high contrast is that the noise corrupting neural responses depends on the mean response in a way that is inconsistent with overall behavior. Both the photon counts that arise from stimuli in most psychophysical experiments and spike counts of neurons have a variance that is proportional to the mean rate (as, for instance, for the case of Poisson noise). However, were contrast discrimination thresholds determined by choosing the largest sample with a linear transducer and noise variance that scaled in this way, then the handle of the contrast discrimination dipper function should approximately be a function of the square root of the contrast (Barlow, 1957; de Vries, 1943; Rose, 1942) and so have a slope of 0.5 on a log–log plot.

Indeed, consider the effect, contrary to Equation 1, of using the canonical model of neural noise, namely, the Poisson distribution, about the same mean  $\bar{r}_i = cf_i(\theta)$ . In Appendix A, we derive the Cramér–Rao lower bound as having the expected square-root dependence:

$$\sigma_{\text{ML}}(c) \geq \frac{\sqrt{c}}{\sqrt{\sum_i f_i(\theta)}}. \quad (6)$$

We also show in Appendix A that even for large differences between target and pedestal contrasts, the log–log slope of the contrast discrimination function should not exceed 0.5. However, contrast discrimination thresholds for human observers appear to have a larger

exponent, between 0.5 and 1 (Barlow, 1957; Campbell & Kulikowski, 1966; Legge, 1981; Nachmias & Sansbury, 1974).

The fact that these slopes are generally less than 1 is also significant, since Weber’s law would suggest that contrast increment thresholds should be proportional to the contrast itself, i.e., to have a slope of 1. In sum, the empirical thresholds lie between the values associated with the two classic models.

For the actual, mean-dependent, Gaussian noise of Equation 1, Appendix A shows that the CR lower bound enjoys a subtly different dependence on  $c$ :

$$\sigma_{\text{ML}}(c) \geq \frac{c}{\sqrt{n/2 + cf_i(\theta)}}. \quad (7)$$

Here, the scaling resembles Weber’s law for low values of contrast and square-root scaling for very high values of contrast as shown in Figure 5. In effect, a constant has been added to the denominator for each unit in the population—this constant reflects a fixed amount of information that each unit contributes, regardless of stimulus contrast.<sup>3</sup>

The hinge that caused the dip in the contrast discrimination function at low contrasts has a secondary effect: It forces the slope of the contrast discrimination function to be below 1 at high contrasts. Including also the hinge in the noise, we found that the slope of the simulation thresholds in Figure 1 is 0.84 for pedestal contrasts

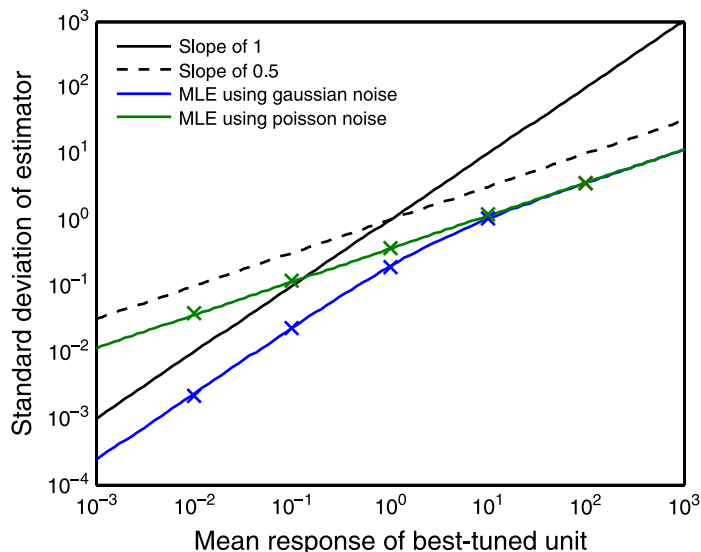


Figure 5. Standard deviation of the maximum likelihood estimator for different mean responses. The mean response is the contrast multiplied by the tuning curve of a unit perfectly tuned to the stimulus, as in Equation 1. Lines are Cramér–Rao bounds and the markers are individual simulation results. The Gaussian distribution has been constrained, so like the Poisson distribution, the variance is equal to the mean.

between  $c = 0.1$  and  $c = 0.36$ . This result matches the qualitative finding of a slope between 0.5 and 1, though quantitatively misses the slope of the data at high contrasts.<sup>4</sup> The intermediate slope of the function is due to hinge noise being partly constant and partly growing. A high-contrast stimulus produces high activations in units that are well tuned to the stimulus but low activations in units that are poorly tuned to the stimulus, as shown in Figure 2. For very low activations, the noise is essentially flat, which would produce a contrast discrimination function with a slope of 0. For higher activations, the variance of the response is equal to the mean, which produces a slope of 1 as shown above. The result is a compromise between the two possible slopes—a result of the interaction between the population of variously tuned units, hinge noise, and optimal decisions.

## Comparison to other models of contrast discrimination

Many different models of contrast discrimination have already been developed (Solomon, 2009). Most have been

designed to apply to either the dip or the handle but not both. In this section, we discuss the different models of high- and low-contrast discrimination and how our approach differs from each.

### Low contrast

Cutoff models provide an intuitive explanation for the dip evident for low-contrast discrimination. Constant noise is added to each stimulus, and any net sample that fails to exceed the cutoff is set to zero, while any sample that does exceed the cutoff is unaffected. This is shown schematically in Figure 6A. If the sample from the pedestal and the sample from the target are both below the cutoff, then the choice is a pure guess. However, as an increasing pedestal causes both types of samples to exceed the cutoff more frequently, discrimination becomes more accurate. Several varieties of cutoff accounts have been proposed, including high cutoff models, which assume that samples from a blank interval never exceed the cutoff (Blackwell, 1963), and low cutoff models, which allow the cutoff to be exceeded by noise (Foley & Legge, 1981).

The hinge noise we proposed above looks a bit like a cutoff, but instead of being applied to the mean, it affects

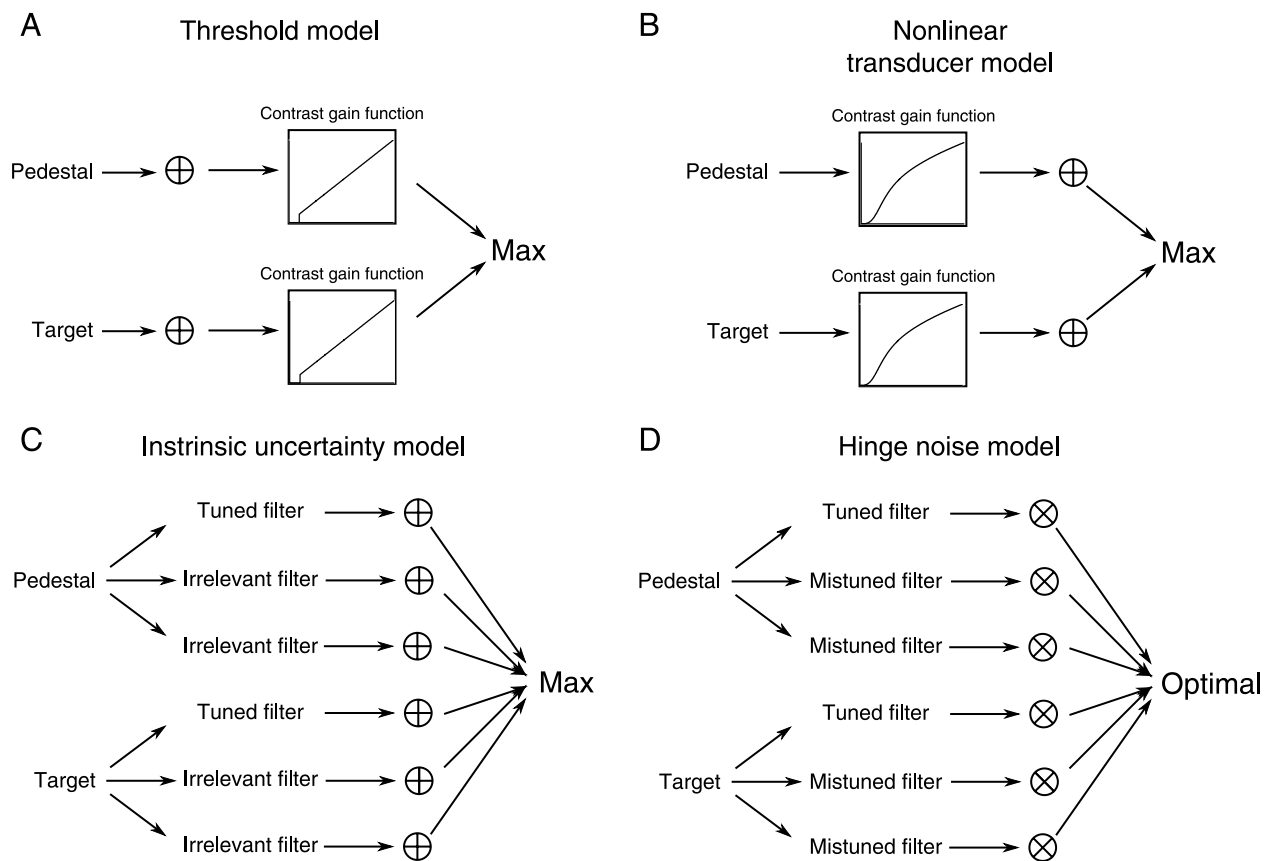


Figure 6. Schematic diagrams of various models of dipper functions. Circled pluses indicated additive noise and circled xs indicate signal-dependent noise.

the variance of the response. Thus, it does not squash the activations, which is the way cutoff models generate the dip.

A second approach to producing dips is an expansive non-linearity followed by noise (Legge & Foley, 1980; Nachmias & Sansbury, 1974). In this account, the mean response of a unit is a non-linear function of contrast, schematically shown in Figure 6B. A dip can be produced by a transducer with a slope that is initially flat, then is steep. Late additive noise is crucial to this account as well, because Birdsall's theorem implies that transduction alone will not have any impact on accuracy for single units. It is the noise that causes transduced contrasts that are closer together to be less discriminable. As a result, the greater slopes of the transducer at higher contrasts allow for more accurate responses, producing the dip.

It is well known that a non-linear transducer with additive noise is, in many circumstances, equivalent to a linear transducer with multiplicative noise (Legge, Kersten, & Burgess, 1987). The hinge noise we propose is not just the non-linear transducer written in a different form. To see this, we note that an expansive transducer followed by constant noise is equivalent to a linear transducer with decreasing noise,<sup>5</sup> while our proposed model uses a linear transducer and noise that never decreases.

Threshold models and expansive non-linearities can both produce dips in the contrast discrimination function based on the responses of a single unit. A very different way of producing a dip is through intrinsic uncertainty (Pelli, 1985; Tanner, 1961), which is shown schematically in Figure 6C. This approach assumes that many filters are active in each interval, of which only a small number are tuned to the target. The remaining filters only contribute noise to the decision process. The optimal decision is approximated by choosing the interval that has produced the largest overall sample (Pelli, 1985). For higher pedestal contrasts, the largest samples almost exclusively result from the filters tuned to the features of the stimuli, meaning less noise in the decision process and lower thresholds.

Our approach, shown schematically in Figure 6D, works differently from intrinsic uncertainty because the decoder is not uncertain about any parameter of the stimulus other than its actual contrast (as even its potential contrast is known). If we were to apply our form of optimal decisions to the noisy population of units used in intrinsic uncertainty, all the irrelevant units tuned solely to noise would be ignored by the estimator, and so it would not matter how many of these there were. By contrast, each member of our population of orientation-tuned units had some (possibly extremely weak) tuning to the signal, and the decoder, being aware of the orientation, could take advantage of them according to their own signal/noise ratios. Instead of uncertainty about the stimulus properties, it is the sudden increase in noise at the hinge that drives the dip.

## High contrast

Explanations of the contrast discrimination function at high pedestal contrasts have tended to focus on processes that transform the mean of the signal distribution, including a compressive non-linearity (Legge & Foley, 1980) and a transformation that removes any differences in the mean signal entirely (Laming, 1986). The predominant explanation is to appeal to a transducer that suffers from a compressive non-linearity at high contrasts. If the output of this transducer is followed by additive noise, then the difference between equally spaced contrasts will be less the higher that these contrasts are—meaning that the additive noise that follows will have a larger effect. Unlike the case for the expansive non-linearity in the low-contrast regime, such a compressive non-linearity would be compatible with a linear transducer and growing noise. However, in order to produce a handle with a slope that is greater than 0.5, the noise would have to grow faster than implied by neural noise. Our results are able to produce a handle with a slope greater than 0.5 using noise that would imply a slope of 0.5 when used with the standard decision model.

Laming (1986) developed a mechanistic model in which noise whose variance is proportional to its mean is nevertheless consistent with Weber's law for discrimination. Using the motivation that the visual system is only sensitive to gradients, the stimulus is divided into positive and negative aspects, each of which is represented by Gaussian noise with variance proportional to the mean. These two aspects are subtracted from one another, producing a zero-mean Gaussian distribution, but with the variance still proportional to the mean. Discriminability between two zero-mean Gaussian distributions with different variances is proportional to the ratio of the variances, producing thresholds that would be consistent with Weber's law. Essentially, the approach works by discarding the means. Our results differ from this approach in that all the information is available, including mean contrasts.

A final, perhaps less well-known explanation for the discrepancy between neural and behavioral noise is based on a different statistical structure. Triesman (1964) proposed a compound Poisson process, in which each quantum resulted in a new sample from a Gaussian distribution being drawn and added to a decision variable. This decision variable, which is the sum of a Poisson number of Gaussian samples, would have a variance that is equal to the mean, unless the samples are correlated. By assuming that all the samples were perfectly correlated, Triesman showed that for high contrasts the sum will have a standard deviation equal to the mean and thus follow Weber's law. Unlike this explanation, we assume that there are no correlations between the units. This is in keeping with many treatments of cortical activity. Indeed, empirical assessments of correlations in neural noise vary



from tiny to modest (Cohen & Kohn, 2011; Ecker et al., 2010).

## Comparison to models that use non-linear transduction

The hinge noise model provides a unitary normative account of both low and high pedestal contrast regimes. Perhaps the most influential model that has attempted to accommodate both regimes involves non-linear transduction (Legge & Foley, 1980), which has been adopted by models of contrast discrimination that use populations of units (Chirimuuta & Tolhurst, 2005; Goris et al., 2009; Itti et al., 2000) and has also served as the basis for models that use optimal decoding (Chirimuuta & Tolhurst, 2005; Itti et al., 2000). This is perhaps best seen as two explanations glued into one: an expansive non-linearity at low contrasts and a compressive non-linearity at high contrasts. The hinge noise model is a unitary explanation rather than the combination of two explanations: Both regimes of noise are necessary to produce the dip, and both are necessary to produce a handle with a slope between 0.5 and 1.

## Discussion

We developed a statistical computational explanation for empirical observations of contrast discrimination thresholds. The central claim of our account is that both the main features of the contrast discrimination function arise from optimal inference subject to a plausible form of noise. This contrasts with the rather numerous suggestions that combine separate explanations for the origins of the dips and handles.

According to our account, it is the structure of the noise, together with appropriate inference, that substantially controls the shape of the dipper. The dip in the threshold occurs around the input contrast at which the noise provides information about the signal; the shape of the handle in the threshold is partly determined by doing full inference rather than making the approximation of considering the interval with the larger input as automatically being the interval with the greater contrast.

Of course, unlike the mechanistic suggestions described above, we have offered a computational account without an algorithmic or network implementation. An estimator that is purely linear in the input activities would likely not be able to extract information from the covariance; however, mild non-linearities are known to suffice (Shamir & Sompolinsky, 2004), opening up a range of possibilities.

Our use of a population code, as in various others such as Goris et al. (2009), Itti et al. (2000) and Pelli (1985), leads us to look at population phenomena such as the tilt

illusion or the tilt aftereffect (Solomon & Morgan, 2006) as extensions. We would hope, for instance, to combine this present account of contrast analysis with the Gaussian scale mixture model approach of Schwartz, Sejnowski, and Dayan (2009), which eliminates the bulk of contrast effects through a statistically normative form of divisive normalization. Its account of orientation processing should, therefore, be well placed to survive the contrast processing implied by the inferences discussed here.

It would also be interesting to investigate further aspects of prior information. We explored some simple models of prior structures motivated by standard psychophysical paradigms, but there is broad scope for generalization. As an example, attentional tasks have explored how cues influence performance: Findings show that people use these cues to exclude the irrelevant noisy aspects of the stimuli (Doshier & Lu, 2000; Lu, Lesmes, & Doshier, 2002). In addition, perceptual learning tasks have shown similar findings where later trials show more efficient processing than earlier trials (Gold, Bennett, & Sekuler, 1999; Gold, Sekuler, & Bennett, 2004). Both of these results could be interpreted as the use of prior knowledge: outside information or experience that allow people to exclude irrelevant hypotheses. Of interest is whether we could reproduce the form and rate of learning.

## Conclusions

A general conclusion from our study is that surprisingly subtle features of noise can have a qualitative effect on the output of an ideal observer. Other studies have shown that optimal inference with a non-linear transducer can produce dipper functions (Chirimuuta & Tolhurst, 2005; Itti et al., 2000), but here we have fixed the transducer to be linear and shown that changes in the higher order moments result in the same predictions. People have shown a sensitivity to the variance and higher order moments of stimuli in their behavior (Körding & Wolpert, 2004; Morgan, Chubb, & Solomon, 2008; Symmonds, Wright, Bach, & Dolan, *in press*), and here we propose that the sensitivity extends to how the noise variance changes with signal intensity.

Along with their application to higher level aspects of visual processing such as multimodal integration (Ernst & Banks, 2002) and attention (Chikkerur, Serre, Tan, & Poggio, 2010; Yu, Dayan, & Cohen, 2009), the current work is part of a movement bringing Bayesian principles into relatively early visual processing (Chirimuuta & Tolhurst, 2005; Dayan & Solomon, 2010; Whiteley & Sahani, 2008). Of course, there is a need to develop stronger data on the uncertainties faced by organisms and how inference is implemented in the brain. However, the normativity of these approaches gives us the hope of encompassing more of the complexity evident in behavior

in accounts that refer to the nature of the environment, resulting in a set of simple principles that can predict human behavior across a broad array of situations.

## Appendix A

In this appendix, we derive the standard deviations of the maximum likelihood estimators for Poisson noise, Gaussian noise when the variance is constrained to be equal to the mean response, and the time to the first event for a Poisson process. In addition, we derive an upper bound on the slope of the contrast discrimination function when the noise is Gaussian with the variance equal to the mean response. We write  $f_i(\theta)$  as  $f_i$  in this appendix for convenience.

### Poisson noise

For each unit  $i$  of the total of  $n$  units, let the mean response equal  $\bar{r}_i = cf_i$ , where  $f_i$  is the orientation-dependent response of unit  $i$  based on the discrepancy between its preferred orientation and the stimulus orientation at a contrast of 1. Let  $\mathbf{r}$  be the vector of outputs drawn as independent Poisson samples with these mean values.

For distributions that satisfy certain regularity conditions, we can substitute in the negative second derivative for the square of the first derivative in the expected Fisher information calculation:

$$\text{FI}(c) = E \left[ \frac{d^2}{dc^2} (-\log p[\mathbf{r}|c]) \right], \quad (\text{A1})$$

$$= E \left[ \frac{d^2}{dc^2} \left( -\log \prod_i e^{-cf_i} \frac{(cf_i)^{r_i}}{r_i!} \right) \right], \quad (\text{A2})$$

$$= E \left[ \sum_i \frac{r_i}{c^2} \right], \quad (\text{A3})$$

$$= \frac{1}{c} \sum_i f_i \quad \text{and so,} \quad (\text{A4})$$

$$\sigma_{\text{ML}}(c) \geq \frac{\sqrt{c}}{\sqrt{\sum_i f_i}}. \quad (\text{A5})$$

### Time to first Poisson event

For a Poisson process with mean count  $cf_i$  on a particular interval, the interarrival times  $\ell_i$  follow an

exponential distribution with mean  $1/cf_i$ . Thus, writing  $r_i$  as the time to the first event of unit  $i$ , we have

$$\text{FI}(c) = E \left[ \frac{d^2}{dc^2} (-\log p[\mathbf{r}|c]) \right], \quad (\text{A6})$$

$$= E \left[ \frac{d^2}{dc^2} \left( -\log \prod_i cf_i e^{-cf_i \ell_i} \right) \right], \quad (\text{A7})$$

$$= E \left[ \sum_i \frac{1}{c^2} \right], \quad (\text{A8})$$

$$= \frac{n}{c^2}, \quad \text{and so,} \quad (\text{A9})$$

$$\sigma_{\text{ML}}(c) \geq \frac{c}{\sqrt{n}}. \quad (\text{A10})$$

Simulations showed that the variance of the maximum likelihood estimator is nearly equal to this bound for a wide range of contrasts using the same population parameters as the simulations in Figure 5. Unlike the bound on Gaussian noise, the bound on the time to the first Poisson event produces scaling consistent with Weber's law for any value of the mean response.

### Gaussian noise

In Equation 1, we employ a Gaussian rendition of the Poisson case. According to this, each unit is an independent Gaussian random variable with mean  $\bar{r}_i = cf_i$  and  $\tau^2(c) = cf_i$ . In this case, the Fisher information is

$$\text{FI}(c) = E \left[ \frac{d^2}{dc^2} (-\log p[\mathbf{r}|c]) \right], \quad (\text{A11})$$

$$= E \left[ \frac{d^2}{dc^2} \left( -\log \prod_i \frac{1}{\sqrt{2\pi cf_i}} \exp \left( -\frac{(r_i - cf_i)^2}{2cf_i} \right) \right) \right], \quad (\text{A12})$$

$$= \sum_i -\frac{1}{2c^2} + \frac{2E[r_i] - cf_i}{c^2} + \frac{E[(r_i - cf_i)^2]}{c^3 f_i}, \quad (\text{A13})$$

$$= \frac{1}{c} \sum_i f_i + \frac{n}{2c^2}, \quad \text{and so,} \quad (\text{A14})$$

$$\sigma_{\text{ML}}(c) \geq \frac{c}{\sqrt{n/2 + c \sum_i f_i}}. \quad (\text{A15})$$

## Bound on slope of threshold for Gaussian variance equal to its mean

Consider the case in which we have a single output from a stimulus, either the output of a model with a single channel or the maximum likelihood estimate from a population of responses. Assume that the output is Gaussian distributed and that both the mean of the output and the variance of the output are equal to the contrast. In a contrast discrimination experiment in which two outputs are compared, we show that the slope of the contrast discrimination function on a log–log plot is never greater than 1/2.

The probability correct  $P_C$  simplifies from Equation 3 to

$$P_C = \Phi^{-1}\left(\frac{\Delta c}{\sqrt{2c_p} + \Delta c}\right), \quad (\text{A16})$$

where  $\Delta c$  is the threshold difference in contrast between the pedestal  $c_p$  and the target  $c_t$ . We assume that  $P_C$  is equal to the threshold probability correct. Solving for this threshold increment using the quadratic formula:

$$\Delta c = \frac{A + \sqrt{A^2 + 8Ac_p}}{2}, \quad (\text{A17})$$

where assuming  $\Delta c$  is positive eliminates the negative root and where  $A = [\Phi(P_C)]^2 \geq 0$ . To find the slope of the threshold in log–log coordinates, we take the derivative of  $\log \Delta c$  with respect to  $\log c_p$ :

$$\frac{d \log \Delta c}{d \log c_p} = \frac{4c_p}{8c_p + A + \sqrt{A^2 + 8Ac_p}} < \frac{1}{2}, \quad (\text{A18})$$

with the inequality resulting from the positivity of all the terms in the denominator of the middle expression. For large values of  $c_p$ , this slope will converge to 1/2.

## Acknowledgments

The authors thank Josh Solomon, Jeff Beck, and David Schulz for helpful discussions. This work was funded by the Gatsby Charitable Foundation and the Cognitive Systems Foresight Project (Grant BB/E000444/1).

Commercial relationships: none.

Corresponding author: Adam Sanborn.

Email: a.n.sanborn@warwick.ac.uk.

Address: Department of Psychology, University of Warwick, Coventry CV4 7AL, UK.

## Footnotes

<sup>1</sup>Making the evident assumption that there is no bias in the order of presentation.

<sup>2</sup>As a result of the near equivalence between Both Known and Pedestal Known priors, we used the computationally simpler Both Known prior for the predictions made in Figure 1.

<sup>3</sup>Interestingly, even in the Poisson case, pure Weber law scaling arises if the relevant statistic consists of the times to the first spikes of each neuron in the population rather than the summed activities over a fixed period (shown in Appendix A).

<sup>4</sup>A more complex form for the noise could, of course, produce a better quantitative fit.

<sup>5</sup>Looking at Equation 3, in order for a smaller difference between  $c_t$  and  $c_p$  to produce an equivalent  $P_C$ , the sum of the variances in the denominator must decrease.

## References

- Abbott, L. F., & Dayan, P. (1999). The effect of correlated variability on the accuracy of a population code. *Neural Computation*, *11*, 91–101.
- Albrecht, D. G., & Hamilton, D. B. (1982). Striate cortex of monkey and cat: Contrast response function. *Journal of Neurophysiology*, *48*, 217–237.
- Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychological Review*, *98*, 409–429.
- Barlow, H. B. (1957). Increment thresholds at low intensities considered as signal/noise discriminations. *The Journal of Physiology*, *136*, 469–488.
- Bird, C. M., Henning, G. B., & Wichmann, F. A. (2002). Contrast discrimination with sinusoidal gratings of different spatial frequency. *Journal of the Optical Society of America A*, *19*, 1267–1273.
- Blackwell, H. R. (1963). Neural theories of simple visual discriminations. *Journal of the Optical Society of America*, *53*, 129–160.
- Campbell, F. W., & Kulikowski, J. J. (1966). Orientational selectivity of the human visual system. *The Journal of Physiology*, *187*, 437–445.
- Chikkerur, S., Serre, T., Tan, C., & Poggio, T. (2010). What and where: A Bayesian inference theory of attention. *Vision Research*, *50*, 2233–2247.
- Chirimuuta, M., & Tolhurst, D. J. (2005). Does a Bayesian model of V1 contrast coding offer a neurophysiological account of human contrast discrimination? *Vision Research*, *45*, 2943–2959.

- Cohen, M. R., & Kohn, A. (2011). Measuring and interpreting neuronal correlations. *Nature Neuroscience*, *14*, 811–819.
- Courville, A. C., Daw, N. D., & Touretzky, D. S. (2006). Bayesian theories of conditioning in a changing world. *Trends in Cognitive Sciences*, *10*, 294–300.
- Crozier, W. J. (1950). On the visibility of radiation at the human fovea. *Journal of General Physiology*, *34*, 87–136.
- Davis, E. T., Kramer, P., & Graham, N. (1983). Uncertainty about spatial frequency, spatial position, or contrast of visual patterns. *Perception & Psychophysics*, *33*, 2028.
- Daw, N. D., Courville, A. C., & Dayan, P. (2008). Semi-rational models of conditioning: The case of trial order. In N. Chater & M. Oaksford (Eds.), *The probabilistic mind* (pp. 431–452). Oxford, UK: Oxford University Press.
- Dayan, P., & Abbott, L. (2001). *Theoretical neuroscience*. Cambridge, MA: MIT Press.
- Dayan, P., & Solomon, J. A. (2010). Selective Bayes: Attentional load and crowding. *Vision Research*, *50*, 2248–2260.
- Derrington, A. M., & Lennie, P. (1984). Spatial and temporal contrast sensitivities of neurones in lateral geniculate nucleus of macaque. *The Journal of Physiology*, *357*, 219–240.
- de Vries, H. L. (1943). The quantum character of light and its bearing upon threshold of vision, the differential sensitivity and visual acuity of the eye. *Physica*, *10*, 553–564.
- Dosher, B. A., & Lu, Z.-L. (2000). Noise exclusion in spatial attention. *Psychological Science*, *11*, 139–146.
- Ecker, A. S., Berens, P., Keliris, G. A., Bethge, M., Logothetis, N. K., & Tolias, A. S. (2010). Decorrelated neuronal firing in cortical microcircuits. *Science*, *327*, 584–587.
- Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*, 429–433.
- Faisal, A. A., & Wolpert, D. M. (2009). Near optimal combination of sensory and motor uncertainty in time during a naturalistic perception-action task. *Journal of Neurophysiology*, *101*, 1901–1912.
- Foley, J. M. (1994). Human luminance pattern mechanisms: Masking experiments require a new model. *Journal of the Optical Society of America A*, *11*, 1710–1719.
- Foley, J. M., & Legge, G. E. (1981). Contrast detection and near-threshold discrimination in human vision. *Vision Research*, *21*, 1041–1053.
- Geisler, W. S. (1989). Sequential ideal-observer analysis of visual discriminations. *Psychological Review*, *96*, 267–314.
- Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2004). *Bayesian data analysis*. Boca Raton, FL: Chapman & Hall/CRC.
- Gold, J. M., Bennett, P. J., & Sekuler, A. B. (1999). Signal but not noise changes with perceptual learning. *Nature*, *402*, 176–178.
- Gold, J. M., Sekuler, A. B., & Bennett, P. J. (2004). Characterizing perceptual learning with external noise. *Cognitive Science*, *28*, 167–207.
- Goris, R. L. T., Wichmann, F. A., & Henning, G. B. (2009). A neurophysiologically plausible population code model for human contrast discrimination. *Journal of Vision*, *9*(7):15, 1–22, <http://www.journalofvision.org/content/9/7/15>, doi:10.1167/9.7.15. [PubMed] [Article]
- Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.
- Heller, K., Sanborn, A. N., & Chater, N. (2009). Hierarchical learning of dimensional biases in human categorization. In J. Lafferty & C. Williams (Eds.), *Advances in neural information processing systems* (pp. 727–735, vol. 22). Cambridge, MA: MIT Press.
- Huang, L., & Dobkins, K. R. (2005). Attentional effects on contrast discrimination in humans: Evidence for both contrast gain and response gain. *Vision Research*, *45*, 1201–1212.
- Itti, L., Koch, C., & Braun, J. (2000). Revisiting spatial vision: Toward a unifying model. *Journal of the Optical Society of America A*, *17*, 1899–1917.
- Jaynes, E. T. (2003). *Probability theory: The logic of science*. Cambridge, UK: Cambridge University Press.
- Kaplan, E., & Shapley, R. M. (1986). The primate retina contains two types of ganglion cells, with high and low contrast sensitivity. *Proceedings of the National Academy of Sciences of the United States of America*, *83*, 2755–2757.
- Kemp, C., Perfors, A., & Tenenbaum, J. B. (2007). Learning overhypotheses with hierarchical Bayesian models. *Developmental Science*, *10*, 307–321.
- Kontsevich, L. L., Chen, C.-C., & Tyler, C. W. (2002). Separating the effects of response nonlinearity and internal noise psychophysically. *Vision Research*, *42*, 1771–1784.
- Körding, K., & Wolpert, D. M. (2004). Bayesian integration in sensorimotor learning. *Nature*, *427*, 244–247.
- Laming, D. (1986). *Sensory analysis*. London/San Diego: Academic Press.



- Legge, G. E. (1981). A power law for contrast discrimination. *Vision Research*, *21*, 457–467.
- Legge, G. E., & Foley, J. M. (1980). Contrast making in human vision. *Journal of the Optical Society of America*, *70*, 1458–1471.
- Legge, G. E., Kersten, D., & Burgess, A. E. (1987). Contrast discrimination in noise. *Journal of the Optical Society of America A*, *4*, 391–404.
- Li, B., Peterson, M. R., & Freeman, R. D. (2003). Oblique effect: A neural basis in the visual cortex. *Journal of Neurophysiology*, *90*, 204–217.
- Lu, Z.-L., & Doshier, B. A. (1998). External noise distinguishes attentional mechanisms. *Vision Research*, *38*, 1183–1198.
- Lu, Z.-L., Lesmes, L. A., & Doshier, B. A. (2002). Spatial attention excludes external noise at the target location. *Journal of Vision*, *2*(4):4, 312–323, <http://www.journalofvision.org/content/2/4/4>, doi:10.1167/2.4.4. [PubMed] [Article]
- Morgan, M., Chubb, C., & Solomon, J. A. (2008). A “dipper” function for texture discrimination based on orientation variance. *Journal of Vision*, *8*(11):9, 1–8, <http://www.journalofvision.org/content/8/11/9>, doi:10.1167/8.11.9. [PubMed] [Article]
- Nachmias, J., & Sansbury, R. V. (1974). Grating contrast: Discrimination may be better than detection. *Vision Research*, *14*, 1039–1042.
- Oram, M. W., Földiák, P., Perrett, D. I., & Sengpiel, F. (1998). The ‘ideal homunculus’: Decoding neural population signals. *Trends in Neuroscience*, *21*, 259–265.
- 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.
- 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.
- Rose, A. (1942). The relative sensitivities of television pickup tubes, photographic film, and the human eye. *Proceedings of the Institute of Radio Engineers*, *30*, 293–300.
- Schwartz, O., Sejnowski, T. J., & Dayan, P. (2009). Perceptual organization in the tilt illusion. *Journal of Vision*, *9*(4):19, 1–20, <http://www.journalofvision.org/content/9/4/19>, doi:10.1167/9.4.19. [PubMed] [Article]
- Seriés, P., Stocker, A. A., & Simoncelli, E. P. (2009). Is the homunculus ‘aware’ of sensory adaptation? *Neural Computation*, *21*, 3271–3304.
- Seung, H. S., & Sompolinsky, H. (1993). Simple models for reading neuronal population codes. *Proceedings of the National Academy of Sciences of the United States of America*, *90*, 10749–10753.
- Shamir, M., & Sompolinsky, H. (2004). Nonlinear population codes. *Neural Computation*, *16*, 1105–1136.
- Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM: Retrieving Effectively from Memory. *Psychonomic Bulletin & Review*, *4*, 145–166.
- Snippe, H. P., & Koenderink, J. J. (1992). Discrimination thresholds for channel-coded systems. *Biological Cybernetics*, *66*, 543–551.
- Solomon, J. A. (2009). The history of dipper functions. *Attention, Perception & Psychophysics*, *71*, 435.
- Solomon, J. A., & Morgan, M. J. (2006). Stochastic re-calibration: Contextual effects on perceived tilt. *Proceedings of the Royal Society B*, *273*, 2681–2686.
- Steyvers, M., Griffiths, T. L., & Dennis, S. (2006). Probabilistic inference in human semantic memory. *Trends in Cognitive Sciences*, *10*, 327–334.
- Symmonds, M., Wright, N. D., Bach, D. R., & Dolan, R. J. (2011). Deconstructing risk: Separable encoding of variance and skewness in the brain. *Neuroimage*, *58*, 1139–1149.
- Tanner, W. P. J. (1961). Physiological implications of psychophysical data. *Annals of the New York Academy of Sciences*, *89*, 752–765.
- Triesman, M. (1964). Noise and Weber’s law: The discrimination of brightness and other dimensions. *Psychological Review*, *71*, 314–330.
- van den Bos, A. (2007). *Parameter estimation for scientists and engineers*. New Jersey: Wiley.
- Whiteley, L., & Sahani, M. (2008). Implicit knowledge of visual uncertainty guides decisions with asymmetric outcomes. *Journal of Vision*, *8*(3):2, 1–15, <http://www.journalofvision.org/content/8/3/2>, doi:10.1167/8.3.2. [PubMed] [Article]
- Yoon, H., & Sompolinsky, H. (1999). The effect of correlations on the Fisher information of population codes. In M. J. Kearns, S. A. Solla, & D. A. Cohn (Eds.), *Advances in neural information processing systems* (pp. 167–173, vol. 11). Cambridge, MA: MIT Press.
- Yu, A. J., Dayan, P., & Cohen, J. D. (2009). Dynamics of attentional selection under conflict: Toward a rational Bayesian account. *Journal of Experimental Psychology: Human Perception and Performance*, *35*, 700–717.