Connecting psychophysical performance to neuronal response properties II: Contrast decoding and detection

Keith A. May

Centre for Applied Vision Research, City University London, London, UK

Joshua A. Solomon

Centre for Applied Vision Research, City University London, London, UK

The purpose of this paper is to provide mathematical insights into the results of some Monte Carlo simulations published by Tolhurst and colleagues. In these simulations, the contrast of a visual stimulus was encoded by a model spiking neuron, or a set of such neurons. The mean spike count of each neuron was given by a sigmoidal function of contrast, the Naka-Rushton function. The actual number of spikes generated on each trial was determined by a doubly stochastic Poisson process. The spike counts were decoded using a Bayesian decoder to give an estimate of the stimulus contrast. Tolhurst and colleagues used the estimated contrast values to assess the model's performance in a number of ways, and they uncovered several relationships between properties of the neurons and characteristics of performance. Although this work made a substantial contribution to our understanding of the links between physiology and perceptual performance, the Monte Carlo simulations provided little insight into why the obtained patterns of results arose, or how general they are. We overcame these problems by deriving equations that predict the model's performance. We derived an approximation of the model's decoding precision using Fisher information. We also analyzed the model's contrast detection performance, and discovered a previously unknown theoretical connection between the Naka-Rushton contrast-response function and the Weibull psychometric function. Our equations give many insights into the theoretical relationships between physiology and perceptual performance reported by Tolhurst and colleagues, explaining how they arise, and how they generalize across the neuronal parameter space.

Introduction

This work was inspired by a series of papers from Tolhurst and colleagues, which described Monte Carlo simulations of contrast coding in visual cortex. (Clatworthy, Chirimuuta, Lauritzen, & Tolhurst, 2003; Chirimuuta, Clatworthy, & Tolhurst, 2003; Chirimuuta & Tolhurst, 2005a, 2005b). In these simulations, the contrast of a visual stimulus was encoded by a model spiking neuron, or a set of such neurons, and the spike counts were decoded using a Bayesian decoder to give an estimate of the stimulus contrast. The contrast estimates were then used to assess the model's decoding precision or to generate decisions in simulations of psychophysical tasks.

While these studies made a very useful contribution to our understanding of the relationship between physiology and perceptual performance, the Monte Carlo simulations provided little insight into why the obtained patterns of results arose, or how general they are: It was not clear whether the findings applied just to the specific sets of model parameters that were used in the simulations or to any parameter values. The goal of the present study was to derive equations that predict the performance of this kind of model from its parameters, in order to better understand the relationships between physiology and perceptual performance, and to allow the model's performance to be calculated quickly and easily, removing the need for Monte Carlo simulations, which are difficult to program and time-consuming to run.

Our paper focuses on two key sets of results from the work of Tolhurst and colleagues. The first focus is the relationships between decoding precision and neuronal properties discovered by Clatworthy et al. (2003); we find that these relationships can be explained by approximating the decoding precision using Fisher information. The second focus is the model's psychometric function for detection, studied by Chirimuuta and Tolhurst (2005a); we derive an expression that approximates this function and

sheds light onto one of Chirimuuta and Tolhurst's (2005a) results.

Some of the derivations and technical details are included in appendices in the supplementary information. Each supplementary appendix is labelled with a letter. Equations and Figures in an appendix are labelled with the appendix's letter, followed by a dot, followed by the equation or figure number within that appendix. Supplementary Appendix A provides a list of the main symbols used in the text, and their meanings.

The contrast coding model

The contrast coding model used by Tolhurst and colleagues consisted of an encoding stage, in which the stimulus was encoded as neuronal spike counts, and a decoding stage, which used Bayesian inference to estimate the stimulus contrast from the spike counts. The encoding stage was based on standard physiological models of the neuronal response, while the decoding stage was simply a Bayesian decoder, with no proposed physiological implementation.

The model has three elements: (1) the form of the tuning function, which specifies the mean spike count of each neuron for a given stimulus; (2) the random process that generates spikes at the given mean rate; (3) the method of decoding the population response to give an estimate of the stimulus. We consider these three elements in turn.

We use the same basic terminology as May and Solomon (2014): We represent random variables using upper case letters, and their values on particular trials[1] using corresponding lower case letters. X and x represent the stimulus level; R and $r(x)$ represent the mean spike count; N and n represent the actual spike count; **N** and **n** are vectors holding the spike counts of all the neurons in the population. These variables are explained in more detail in our companion paper [the first paragraph of May and Solomon's (2014) section headed "The sensory coding model"].

The tuning function

The tuning function, $r(x)$, specifies the mean spike count for stimulus x. For visual stimulus contrast, the tuning function is called the *contrast-response function*. Tolhurst and colleagues modelled this function using the Naka-Rushton function (Naka & Rushton, 1966; Albrecht & Hamilton, 1982). For contrast, c, in linear (e.g. Michelson) units, the Naka-Rushton function has the following form:

$$\text{N-R}\, r(c) = \frac{r_{\max} c^q}{{c_{1/2}}^q + c^q} + r_0 . \quad (1)$$

If we measure contrast in log units, $x = \log_b c$, then the Naka-Rushton is given by

$$\text{N-R}\, r(x) = \frac{r_{\max} b^{qx}}{b^{qz} + b^{qx}} + r_0 , \quad (2)$$

where

$$z = \log_b c_{1/2} . \quad (3)$$

See our companion paper (May and Solomon, 2014) for a description of all the parameters, and plots of the Naka-Rushton function. On the log contrast scale, the gradient of the Naka-Rushton function peaks at a log contrast of $x = z$. The term $c_{1/2}$, called the *semi-saturation contrast*, is the contrast for which the mean response exceeds r_0 by $r_{\max}/2$. Many authors use c_{50} to represent this contrast, but we use the less common term $c_{1/2}$ (except when quoting other authors) because this form is easier to extend to other fractions of $r_{\max}$, e.g. $c_{1/3}$ (the contrast for which the mean response exceeds r_0 by $r_{\max}/3$, which we will show to be the contrast which is coded most accurately by a model neuron with a Naka-Rushton contrast-response function with $r_0 = 0$).

In what follows, whenever we use the term "contrast" without specifying the units, we mean "log Michelson contrast". To be compatible with Clatworthy et al. (2003) and Chirimuuta et al. (2003), we always used log to base 10 in our modelling, i.e. $x = \log_{10} c$, and $z = \log_{10} c_{1/2}$; however, our equations are derived for the general case of any base, b.

Note that the units of the Naka-Rushton function's output are often taken to be spikes per second, but we find it more convenient to use units of spikes, without making assumptions about the time interval over which the neuron's output is integrated.

The random process for spike generation

The tuning function gives the mean spike count, and we now turn to the stochastic process that generates spikes at the specified mean rate.

The Poisson process is sometimes used as a stochastic spiking model because of its considerable mathematical convenience (e.g., see Dayan & Abbott, 2001, Chapter 1). The Poisson distribution is defined as

$$P_{\text{Poisson}}(N = n \mid R = r(x)) = \frac{r(x)^n}{n!}\exp(-r(x)). \qquad (4)$$

In Equation (4), we use a right subscript on *P* to indicate the type of stochastic process being used.

For the Poisson process, the variance in the number of spikes for repeated trials is equal to the mean, $r(x)$, always giving a Fano factor (ratio of variance to mean) of 1. In visual cortex, the Fano factor is usually higher than 1 (Tolhurst, Movshon, & Thompson, 1981; Dean, 1981b; Tolhurst, Movshon, & Dean, 1983; Bradley, Skottun, Ohzawa, Sclar, & Freeman, 1987; Skottun, Bradley, Sclar, Ohzawa, & Freeman, 1987; Scobey & Gabor, 1989; Swindale & Mitchell, 1994; Geisler & Albrecht, 1997; Vogels, Spileers, & Orban, 1989; Snowden, Treue, & Andersen, 1992; Britten, Shadlen, Newsome, & Movshon, 1993; Softky & Koch, 1993; Bair & O'Keefe, 1998; McAdams & Maunsell, 1999; Durant, Clifford, Crowder, Price, & Ibbotson, 2007; Buracas, Zador, DeWeese, & Albright, 1998; Oram, Weiner, Lestienne, & Richmond, 1999). To get a Fano factor greater than 1, Tolhurst and colleagues (Clatworthy et al., 2003; Chirimuuta et al., 2003; Chirimuuta & Tolhurst, 2005a, 2005b) used a doubly stochastic Poisson process, which we will refer to as the "Tolhurst process". This process is a Poisson process whose mean is itself a random variable sampled from a Poisson process:

$$P_{\text{Tolhurst}}(N = n \mid R = r(x)) = \sum_{\mu=0}^{\infty} P_{\text{Poisson}}(N = n \mid R = \mu) P_{\text{Poisson}}(N = \mu \mid R = r(x)). \qquad (5)$$

For this process, the mean spike count is $r(x)$, and the variance is $2r(x)$, giving a Fano factor of 2. The infinite series in Equation (5) is difficult to handle, so, in Supplementary Appendix B, we derive a finite series expansion of the Tolhurst process that is more useful.

Since the aim of this paper is to explain the modelling results of Tolhurst and colleagues, all the Monte Carlo simulations in this paper use the Tolhurst process to generate spikes. However, to make our theoretical results more general, we consider two other Poisson-based spiking processes.

The first additional spiking process is the Generalized Poisson Distribution, devised by Consul and Jain (1973). Sakata and Harris (2009) used this process to model neuronal spiking distributions. This process is more flexible than the Tolhurst Process, because the Fano factor, *F*, is set as a parameter, and can take any value ≥ 1. The Consul-Jain distribution takes the following form:

$$P_{\text{C-J}}(N = n \mid R = r(x)) = \frac{r(x)}{n!\sqrt{F}} A^{n-1} \exp(-A), \qquad (6)$$

where

$$A = \frac{r(x) + n\left(\sqrt{F} - 1\right)}{\sqrt{F}}. \qquad (7)$$

Like the Poisson and Tolhurst processes, the Consul-Jain process generates only non-negative numbers of spikes, and the spike count variance is proportional to the mean. When $F = 1$, the Consul-Jain process reduces to the ordinary Poisson process, given in Equation (4). However, the Consul-Jain process with $F = 2$ is *not* identical to the Tolhurst process, even though they have the same Fano factor.

The second additional spiking process that we consider is the doubly stochastic Poisson process recently proposed by Goris, Movshon and Simoncelli (2014). We will refer to this as the "Goris process". It is a Poisson process whose mean is modulated by a multiplicative gain mechanism. The gain value on each stimulus presentation is a gamma-distributed random variable. The Goris process is described in detail in our companion paper (May & Solomon, 2014). The gain fluctuations result in a Fano factor that increases with the firing rate, which is more physiologically plausible than the fixed Fano factor produced by the Tolhurst and Consul-Jain processes. For mathematical simplicity, we assume that each neuron in the population has the same gain signal, and each neuron's Poisson spiking process is independent (see May & Solomon, 2014, for justification of these restrictions). The shared gain signal causes the neuronal responses to be correlated, with a realistic correlation structure. However, because each neuron's Poisson spiking process is independent, a decoder that knows the gain signal on each stimulus presentation can express the spiking distributions as independent Poisson distributions, so the neurons can be decoded as if they were statistically independent (see May and Solomon, 2014, for a more detailed explanation of this).

Bayesian population decoding

All the results that we report from Tolhurst and colleagues were obtained using maximum-likelihood decoding. The estimated stimulus level is the one that had the highest probability of giving rise to the obtained set of spike counts, i.e. it is the value of *x* that maximizes the likelihood, $P(\mathbf{N} = \mathbf{n} \mid X = x)$. In Tolhurst and colleagues' model, the neurons were statistically independent; in addi-

tion, if we use May and Solomon's (2014) parameterization of the Goris process, the neurons are implicitly decorrelated if the decoder knows the gain signal. For statistically independent neurons, the population likelihood is then given by the product of the likelihoods of the individual neurons:

$$P(\mathbf{N}=\mathbf{n} \mid X=x)=\prod_{j=1}^{K} P(N_j=n_j \mid X=x) \quad (8)$$

$$=\prod_{j=1}^{K} P(N_j=n_j \mid R_j=r_j(x)). \quad (9)$$

K is the number of neurons. N_j is a random variable representing the spike count of neuron j, and n_j is its value. R_j is a random variable representing the mean spike count for neuron j, and $r_j(x)$ is its value. The second equality (Equation (9)) arises because each $r_j(x)$ is a deterministic function of the stimulus value, x. To evaluate the probabilities in Equation (9), we use the appropriate expression, depending on which spiking process we are using. For the Consul-Jain process, we use Equation (6). For the Goris process, assuming the decoder knows the gain value, we use the gain-modulated Poisson distribution (May & Solomon, 2014, Equation (10)). All the modelling in this paper used the Tolhurst process, which is defined in Equation (5). However, the infinite sum in Equation (5) makes this expression cumbersome to evaluate, and it is usually better to use the finite series expansion that we derived in Supplementary Appendix B (Equation (B.23)).

Measuring decoding precision

Definition of decoding precision

Decoding precision is usually taken to be the reciprocal of the variance of the estimated stimulus value; this is how it is defined in our companion paper (May & Solomon, 2014). However, Tolhurst and colleagues (Clatworthy et al., 2003; Chirimuuta et al., 2003) used a slightly different measure of decoding performance in their Monte Carlo simulations, which they called "accuracy", defined as the reciprocal of the mean squared difference between the individual estimate and the true stimulus value:

$$\text{accuracy}=\frac{T}{\sum(\hat{x}-x)^2}, \quad (10)$$

where T is the number of trials (= 10,000), $\hat{x}$ is the estimated log contrast estimate (note we use an upper case $\hat{X}$ to represent the random variable, and lower case $\hat{x}$ to represent its value on a particular trial), and the denominator is the sum over all trials. For the models that we consider in this study, the log contrast estimate is largely unbiased (except at very low performance levels), so we have $\text{mean}[\hat{X}] \approx x$. In this case, the accuracy score in Equation (10) is essentially the same as the precision. For consistency with Tolhurst and colleagues, we used their measure of decoding accuracy (Equation (10)) when analyzing our Monte Carlo simulations. However, we refer to it as "precision" (except when directly quoting Tolhurst and colleagues), because our analytical approximations of it are formally measures of precision, and the term "accuracy" is often used to mean the inverse of bias (e.g., Smith, 1999, Chapter 2). We found that, except in degenerate conditions where the model's performance level was very low, it made a negligible difference whether we plotted Tolhursts's accuracy score or true precision. Note, in the Monte Carlo simulations in our companion paper (May & Solomon, 2014), we calculated true decoding precision (reciprocal of the variance of the estimated stimulus value), not the accuracy score defined in Equation (10).

Approximating decoding precision using Fisher Information

For an unbiased maximum-likelihood decoder, the precision cannot exceed a quantity called the *Fisher information*. This limit is known as the Cramér-Rao bound (Cramér, 1946; Rao, 1945; see Dayan & Abbott, 2001, pp. 120-121). For populations that give sufficiently large numbers of informative spikes, the precision of a maximum-likelihood decoder is very close to the Fisher information; when the tuning function is a sigmoid function like the Naka-Rushton function, this applies even if the "population" consists of a single neuron. This means that we can use the Fisher information as an analytical approximation of the decoding precision, thereby giving us a real insight into why Tolhurst and colleagues' results occurred, and how general they are.

The Fisher information depends on the tuning function and the spiking process. Unfortunately, for the Tolhurst and Consul-Jain spiking processes, the derivation of an exact formula for the Fisher information turned out to be intractable. In Supplementary Appendix E, we derive approximations of the Fisher information for each of these spiking processes: $\tau_{\text{Tolhurst}}(x)$ and $\tau_{\text{C-J}}(x)$. As long as the mean spike count of the most informative neurons is not too low, both $\tau_{\text{Tolhurst}}(x)$ and $\tau_{\text{C-J}}(x)$ are very close

to the following general approximation of the decoding precision, $\tilde{\tau}(x)$:

$$\tilde{\tau}(x) = \frac{1}{v}\sum_{j=1}^{K}\frac{r_j'(x)^2}{r_j(x)}, \qquad (11)$$

where $r_j'(x)$ is the first derivative of neuron j's tuning function with respect to x. To parameterize $\tilde{\tau}(x)$ so that it approximates $\tau_{\text{Tolhurst}}(x)$ or $\tau_{\text{C-J}}(x)$, v should be equal to the Fano factor; the Fano factor is fixed at 2 for the Tolhurst process, and can take any value ≥ 1 for the Consul-Jain process. The tilde (~) above τ is to indicate that this estimate of decoding precision is based on an approximation of the Fisher information that is not always accurate. With a Fano factor of 1, the Consul-Jain process is the ordinary Poisson, in which case $\tilde{\tau}(x)$ with $v = 1$ is exactly equal to the Fisher information (see Dayan & Abbott, 2001, Chapter 3). For the Tolhurst process and the Consul-Jain process with $F \neq 1$, $\tilde{\tau}(x)$ with v equal to the Fano factor is an approximation of the Fisher information. For the Goris process, May and Solomon (2014) showed that an appropriate estimate of the decoding precision is given exactly by Equation (11) with $v = 1/(1-\sigma_G^2)$, where σ_G^2 is the variance of the gain signal (see May & Solomon, 2014, Equations (26) and (31)). In this case, v is not the Fano factor – the Fano factor for Goris et al.'s spiking process is variable, and depends on the mean spike count. When the gain is known by the decoder, the Fisher information of the Goris process varies from trial to trial due to the fluctuating gain, and $\tilde{\tau}(x)$ with $v = 1/(1-\sigma_G^2)$ gives the modal value of the Fisher information exactly.

In summary, Equation (11) gives a general approximation of the decoding precision for a variety of different Poisson-based spiking distributions. For two processes (Poisson and Goris) it is based on an exact expression for the Fisher information; for the other two processes (Tolhurst and Consul-Jain) it is based on an approximation of the Fisher information that is accurate as long as the mean spike count of the most informative neurons is not too low.

If we expand $r_j(x)$ and $r_j'(x)$ in Equation (11) using the Naka-Rushton function (Equation (2)), then we have

$$\tilde{\tau}(x) = \frac{1}{v}\sum_{j=1}^{K}\frac{r_{\max}\left(q\ln b\right)^2 b^{q(2z+x)}}{\left(b^{qz}+b^{qx}\right)^3} \quad \text{if } r_0 = 0. \qquad (12)$$

Tolhurst and colleagues always used $r_0 = 0$ in their modelling; since we are focusing on their modelling results, we will only consider the case of $r_0 = 0$ in this paper. In Equation (12), each neuronal parameter can vary from neuron to neuron, so, strictly speaking, each parameter should be indexed by the neuron number, j, but we omit these indices to reduce notational clutter.

In the next section, many of the "populations" of neurons that we analyze consist of either just a single neuron or a set of identical neurons. In this case, Equation (12) reduces to

$$\tilde{\tau}(x) = \frac{K r_{\max}\left(q\ln b\right)^2 b^{q(2z+x)}}{v\left(b^{qz}+b^{qx}\right)^3} \quad \text{if } r_0 = 0, \qquad (13)$$

with $K = 1$ in the case of a single neuron.

Explaining Clatworthy et al.'s contrast decoding results

Clatworthy et al. (2003) applied their Bayesian contrast decoding algorithm to single Tolhurst-spiking neurons, or populations of them, and discovered several relationships between the neuronal parameters and the peak of decoding precision along the contrast axis:

1. "Increasing $R_{\max}$ increases the contrast identification accuracy of single neurons at all contrasts, most obviously the peak accuracy, without changing the contrast at which accuracy is a maximum" (Clatworthy et al., 2003, p. 1991; note, they use $R_{\max}$ where we use $r_{\max}$); this can be summarized by saying that there is an approximately multiplicative effect of $r_{\max}$ on decoding precision;
2. "the position of the accuracy peak along the contrast axis is consistently close to but, interestingly, slightly below the neuron's c_{50}" (Clatworthy et al., 2003, p. 1989);
3. "the peak value of accuracy is independent of c_{50}" (Clatworthy et al., 2003, p. 1989);
4. "the relationship between the maximum accuracy and q is a steep straight line on log-log coordinates" (Clatworthy et al., p. 1989);
5. "To change the maximum accuracy ... requires only a change in the product of $R_{\max}$ and number of neurons, i.e. the total number of action potentials generated on average ...; for a given accuracy, there is a simple trade-off between the number of neurons and the response amplitude of individual neurons." (Clatworthy et al., 2003, p. 1990).

The numerical nature of Clatworthy et al.'s method means that it gave little insight into why these relationships arose, or how generally they apply. In this section we explain Clatworthy et al.'s findings by deriving equations that explain corresponding findings for the analytical approximation of decoding precision, $\tilde{\tau}(x)$, given in Equation (13).

First, we examined whether $\tilde{\tau}(x)$ was sufficiently close to the true decoding precision to make this approach valid. This is important, because the Fisher information (on which $\tilde{\tau}(x)$ is based) is only an upper bound on the decoding precision, and can far exceed the true decoding precision for small population sizes (Xie, 2002). In one of their investigations (shown in their Figure 5A), Clatworthy et al. examined the effect of $r_{\max}$ on decoding precision for a single Tolhurst-spiking neuron with Naka-Rushton exponent (q) equal to 2, $c_{1/2} = 0.1$, and $r_0 = 0$. $r_{\max}$ took values of 5, 20, 50, 100, or 180 spikes. We replicated their methods (see Supplementary Appendix G for details) and obtained essentially identical contrast decoding precision scores (calculated using Equation (10)) to those that we read off from their Figure 5A, the small differences almost certainly being due to the stochastic nature of the simulations and small inaccuracies in our transcription of the data from Clatworthy et al.'s figure. The symbols in Figure 1 show these precision scores. The thin, coloured curves show the true Fisher information, calculated numerically (for methods, see Supplementary Appendix H). The thick, black curves show $\tilde{\tau}(x)$ with $v = 2$. The two panels in Figure 1 are identical except that Figure 1A has a linear vertical axis, and Figure 1B has a logarithmic one. The linear vertical axis facilitates comparison with Clatworthy et al.'s Figure 5A, which used linear vertical axes, and the logarithmic vertical axis gives a clearer picture of the data for low $r_{\max}$ values, which are flattened out on the linear axis. The logarithmic vertical axis tends to exaggerate the deviations of the decoding precision from Fisher information because the precision scale is expanded for the worst-matching conditions (those with low precision) relative to the best-matching conditions.

Firstly, note that $\tilde{\tau}(x)$ is very close to the true Fisher information at the peaks (compare the thick, black lines with the thin, coloured lines). This is important, because the five observations that we analyze in this section are about the peaks, and our justification for using $\tilde{\tau}(x)$ to predict decoding precision is that it is approximately equal to the Fisher information.

For $r_{\max}$ of 50 spikes or more, the horizontal and vertical position of the peak of precision coincides closely with the peak of $\tilde{\tau}(x)$. This allows us to explain many of Clatworthy et al.'s observations about the precision peak by deriving equations that explain corresponding findings for $\tilde{\tau}(x)$. For $r_{\max}$ values substantially lower than 50 spikes, the decoding precision does not match $\tilde{\tau}(x)$ at all well, even at the peak, but Clatworthy et al.'s observations do not apply here either. In the following five subsections, we explain each of the five observations above using the approximation of the decoding precision, $\tilde{\tau}(x)$, given by Equation (13).

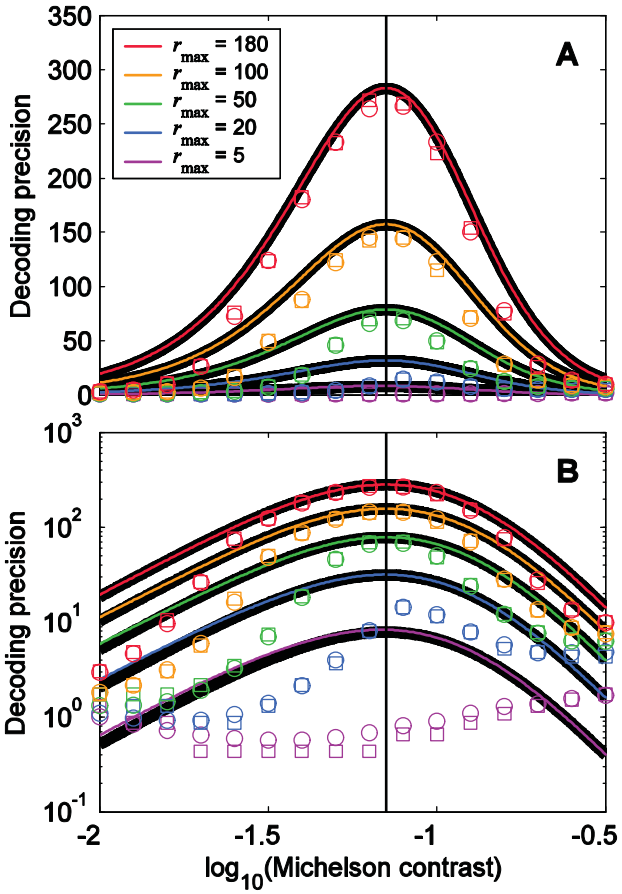

Figure 1. Comparison of contrast decoding precision against the analytical prediction for single Tolhurst-Spiking neurons. (A) Squares show Clatworthy et al.'s Bayesian decoding precision scores for a Tolhurst-spiking neuron with $q = 2$, $c_{1/2} = 0.1$ (i.e. $z = -1$), $r_0 = 0$, and five different $r_{\max}$ values, indicated by different colours; the data were read off from Clatworthy et al.'s Figure 5A. Circles show the results of our replication of their simulations. The thin, coloured curves plot each neuron's true Fisher information for decoding contrast, calculated numerically (see Supplementary Appendix H). The thick, black curves show the analytical approximation of the decoding precision, $\tilde{\tau}(x)$, calculated using Equation (13) with $K = 1$ and $v = 2$. The vertical black line passing through the peaks of the black curves is positioned along the log contrast axis at $\log_{10}(c_{1/3})$, i.e. the log of the Michelson contrast for which the mean response is $r_{\max}/3$. (B) The same as (A), except that the vertical axis has a log scale.

Observation 1: Multiplicative effect of increasing r_{max}

From Equation (13), we see that $\tilde{\tau}(x)$ is proportional to $r_{\max}$, so increasing $r_{\max}$ increases the Fisher information by the same multiplicative factor for each contrast level; the position of the peak across contrast is unchanged, and the largest absolute change is for the peak precision. This multiplicative change gives rise to the vertical shifts seen for the black curves on the logarithmic vertical axis in Figure 1B.

Observation 2: Precision peaks slightly below the semi-saturation contrast

The exact position of the peak of $\tilde{\tau}(x)$ along the contrast axis can be found by differentiating $\tilde{\tau}(x)$ with respect to *x*, setting the derivative to zero, and solving for *x*. From Equation (13), we have

$$\frac{d\tilde{\tau}(x)}{dx} = \frac{Kr_{\max}\left(q\ln b\right)^3 b^{q(2z+x)}}{v\left(b^{qz}+b^{qx}\right)^4}\left(b^{qz}-2b^{qx}\right). \quad (14)$$

Setting this to zero gives

$$\frac{d\tilde{\tau}(x)}{dx} = 0 \Rightarrow b^{qz} = 2b^{qx}$$

$$\Rightarrow x = z - \log_b\left(2^{1/q}\right) \quad (15)$$

$$\Rightarrow c = 2^{-1/q} c_{1/2}\,. \quad (16)$$

Using Equation (15) to substitute for *x* in Equation (2) gives (for $r_0 = 0$)

$$_{\text{N-R}}\, r(x) = r_{\max}/3 \text{ if } \tilde{\tau}(x) \text{ is at its peak.} \quad (17)$$

So, regardless of the values of *v*, $r_{\max}$, *q*, or $c_{1/2}$, $\tilde{\tau}(x)$ for a single neuron will always peak at the contrast for which the mean response is $r_{\max}/3$. We introduce the term $c_{1/3}$ for the Michelson contrast that gives rise to a mean response of $r_{\max}/3$, to be consistent with the term $c_{1/2}$ for the semi-saturation contrast. $\log_{10}(c_{1/3})$ is indicated by the black vertical line in each panel of Figure 1, and it passes through the peak of each thick, black curve.

Observation 3: The peak value of precision is independent of the semi-saturation contrast

The peak value of $\tilde{\tau}(x)$ occurs at a log contrast given by Equation (15). Using this equation to substitute for *x* in Equation (13), we find that the peak value of $\tilde{\tau}(x)$ is given by

$$\max\left[\tilde{\tau}(x)\right] = \frac{4Kr_{\max}\left(q\ln b\right)^2}{27v}, \quad (18)$$

which is independent of $c_{1/2}$.

Observation 4: The relationship between the maximum precision and *q* is a steep straight line on log-log coordinates

From Equation (18), we see that $\max\left[\tilde{\tau}(x)\right]$ is proportional to q^2, giving a straight line on log-log coordinates. Figure 2 plots $\max\left[\tilde{\tau}(x)\right]$ as a function of *q* for a single Tolhurst-spiking neuron with $r_{\max} = 50$ spikes, and shows that $\max\left[\tilde{\tau}(x)\right]$ is very close to the true peak of precision for this model neuron obtained using Clatworthy et al.'s Monte Carlo methods.

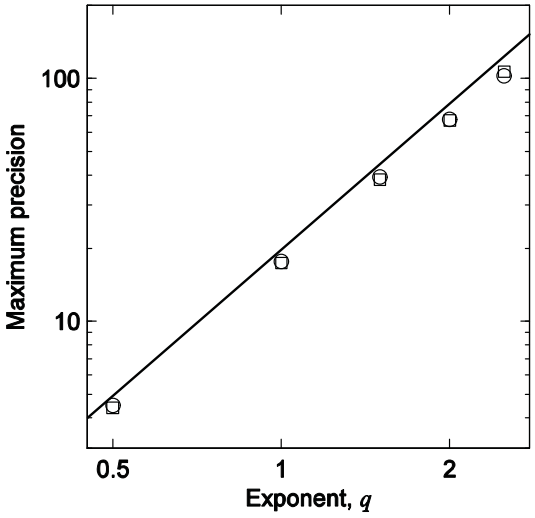

Figure 2. The relationship between the maximum precision and the Naka-Rushton exponent for $r_{\max} = 50$ spikes. Squares show maximum precision as a function of exponent, *q*, when $r_{\max} = 50$ spikes, read off from Clatworthy et al.'s Figure 4C. Circles show the results of our replication of their methods; the differences are negligible. The solid line plots $\max\left[\tilde{\tau}(x)\right]$, calculated using Equation (18) with $K = 1$ and $v = 2$.

Observation 5: Linear trade-off between $r_{\max}$ and number of neurons

Clatworthy et al. (2003) and Chirimuuta and Tolhurst (2005a) both noted that $r_{\max}$ and the number of neurons, K, trade off so that, regardless of the individual values of $r_{\max}$ and K, the decoding precision is a function of their product. In fact, unless $r_{\max} \times K$ is very low, the decoding accuracy is close to being proportional to $r_{\max} \times K$. This observation is easily explained by Equation (13), which shows that $\tilde{\tau}(x)$ is proportional to $r_{\max} \times K$. The crosses in Figure 3 re-plot data from Clatworthy et al.'s (2003) Figure 5C. These data show the maximum decoding precision of various populations of neurons, all with $q = 2$ and $c_{1/2} = 0.1$; the near-proportionality between decoding accuracy and $r_{\max} \times K$ is indicated by the fact that most of the points lie on a straight line of gradient 1 on these log-log axes. The straight line in our Figure 3 is a plot of $\max[\tilde{\tau}(x)]$, calculated according to Equation (18). It clearly provides a very good match to the decoding precision data for $r_{\max} \times K \geq 50$.

Note that, although $\tilde{\tau}(x)$ gives an exact linear trade-off between $r_{\max}$ and K, the true Fisher information for the Tolhurst process does not. As $r_{\max}$ decreases, $\tilde{\tau}(x)$ tends to underestimate the true Fisher information.

A more accurate estimate of the decoding precision is given by $\tau_{\text{Tolhurst}}(x)$, derived in Supplementary Appendix E:

$$\tau_{\text{Tolhurst}}(x) = \sum_{j=1}^{K} H_{\text{Tolhurst}}(x) \times \frac{r_j'(x)^2}{r_j(x)}, \qquad (19)$$

where

$$H_{\text{Tolhurst}}(x) = \left(\frac{1}{2} - \frac{1 + 0.06630 \times r(x)}{e}\right) \exp[r(x)(1/e - 1)] + \frac{1}{2}. \qquad (20)$$

$H_{\text{Tolhurst}}(x)$ is plotted in Figure 4. As the spike rate of each individual neuron increases, $H_{\text{Tolhurst}}(x)$ for that neuron approaches 1/2, and so $\tau_{\text{Tolhurst}}(x)$ approaches $\tilde{\tau}(x)$ (Equation (11)) with $v = 2$. However, as the spike rate decreases, $H_{\text{Tolhurst}}(x)$ increases, and so $\tau_{\text{Tolhurst}}(x)$ exceeds $\tilde{\tau}(x)$, and better reflects the true decoding precision.

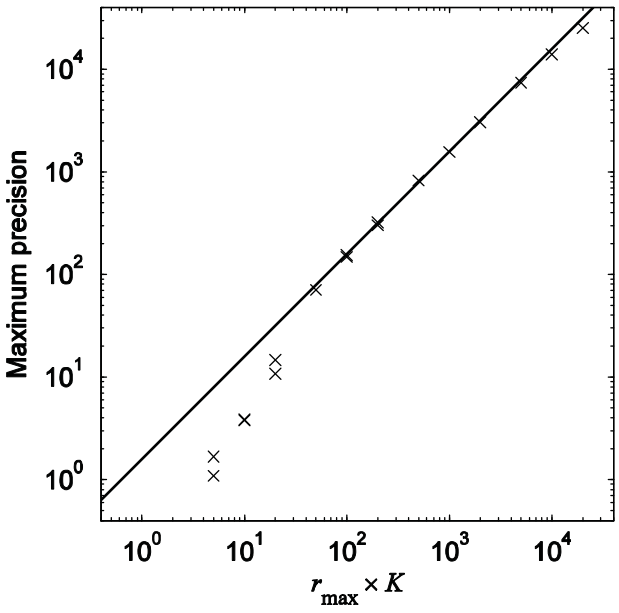

Figure 3. Maximum precision as a function of $r_{\max} \times K$. The crosses plot data from Clatworthy et al.'s (2003) Figure 5C, showing the height of the peak in decoding precision as a function of $r_{\max} \times K$ for a variety of neuronal populations, all with $q = 2$ and $c_{1/2} = 0.1$. The straight line plots $\max[\tilde{\tau}(x)]$, calculated according to Equation (18) with $v = 2$.

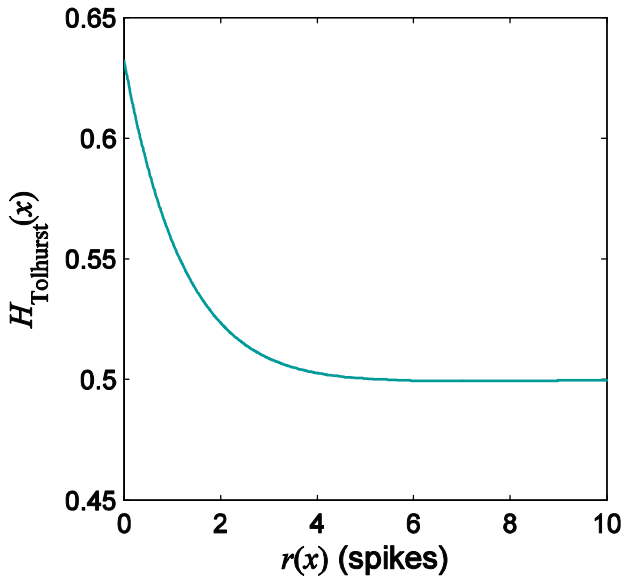

Figure 4. $H_{\text{Tolhurst}}(x)$, as defined in Equation (20), plotted as a function of $r(x)$. This function approaches $(1 - 1/e)$ as $r(x) \to 0$, and approaches 0.5 as $r(x) \to \infty$.

For the Naka-Rushton function with $r_0 = 0$, we can expand Equation (19) in a similar way to Equation (12):

$$\tau_{\text{Tolhurst}}(x) = \sum_{j=1}^{K} H_{\text{Tolhurst}}(x) \times \frac{r_{\max} (q \ln b)^2 b^{q(2z+x)}}{\left(b^{qz} + b^{qx}\right)^3}. \quad (21)$$

Unlike $\tilde{\tau}(x)$, $\tau_{\text{Tolhurst}}(x)$ is specific to the Tolhurst spiking process, rather than being a general approximation of the decoding precision that applies to several different spiking processes.

In each panel of Figure 5, the solid line is the same as in Figure 3, while the dashed line plots the maximum value of $\tau_{\text{Tolhurst}}(x)$. As $r_{\max}$ decreases, $\tau_{\text{Tolhurst}}(x)$ and $\tilde{\tau}(x)$ start to diverge, by a factor that approaches $(1-1/e)/0.5 = 1.264$, i.e. the ratio of the maximum to minimum values of $H_{\text{Tolhurst}}(x)$.

This inexact trade-off between $r_{\max}$ and K applies to the Tolhurst process, but not to the Poisson or Goris processes: for these, $\tilde{\tau}(x)$ is derived from the exact expression for the Fisher information, so the trade-off that we derived for $\tilde{\tau}(x)$ applies to these processes exactly.

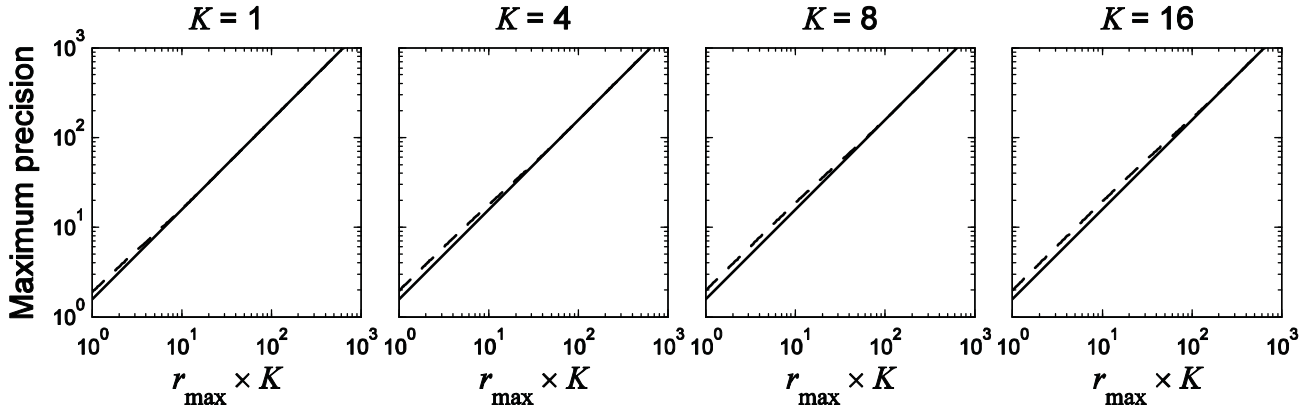

Figure 5. The inexact nature of the trade-off between $r_{\max}$ and K for the true decoding precision of the Tolhurst process. In each panel, the solid line is the same as that in Figure 3, and shows the maximum value of $\tilde{\tau}(x)$; for this approximation of the decoding precision, the trade-off is exact. The dashed curve shows the maximum value of $\tau_{\text{Tolhurst}}(x)$, calculated numerically from Equation (21).; here, the trade-off is only approximate, and the peak of Fisher information does depend slightly on the individual values of $r_{\max}$ and K, rather than being simply a function of their product.

Failure of Clatworthy et al.'s observations for single neurons at low spike rates

When $r_{\max}$ (or, for populations, $r_{\max} \times K$) is substantially less than 50 spikes, the peak of precision does not coincide closely with that of $\tilde{\tau}(x)$ or even the true Fisher information, so the explanations of Clatworthy et al.'s findings given above are no longer valid. However, most of these findings do not apply for these low $r_{\max}$ values either. The failure of Observations 1 and 2 at low spike rates is clear from Figure 1, the failure of Observation 5 at low spike rates is shown in Figure 3, and the failure of Observation 4 at low spike rates is demonstrated in Figure 6.

The median $r_{\max}$ for a 200 ms stimulus is only around 5.7 spikes for V1 neurons (Geisler & Albrecht, 1997), suggesting that, with single-neuron models, the spike count has to be implausibly high for the decoding precision to be well approximated by the Fisher information. However, as shown by Clatworthy et al. (2003), Chirimuuta and Tolhurst (2005a) and our Figure 3, $r_{\max}$ can be approximately traded off against the number of neurons, so that what is important is the average spike count of the population, rather than of the individual neurons. With a population code, it is possible to achieve a high total population spike count while keeping the spike counts for the individual neurons at a plausible level, and this makes the Fisher information more relevant to understanding population coding models than coding schemes based on a single neuron. So far, we have only considered populations of identical neurons. We now turn to populations of differently tuned neurons.

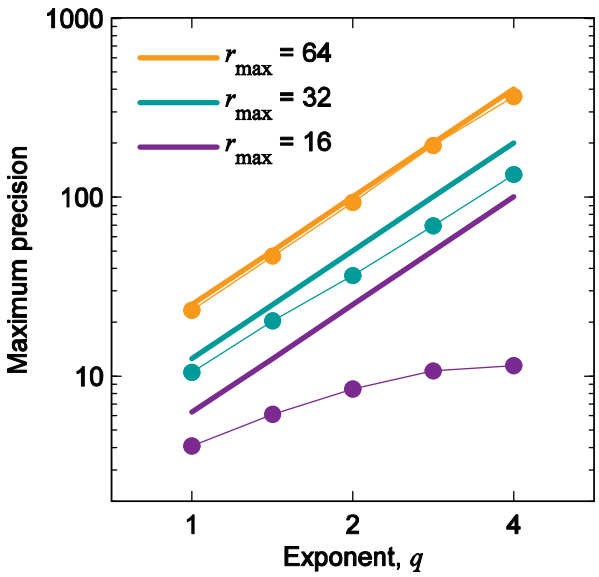

Figure 6. The effect of r_{max} on the relationship between the maximum accuracy and the Naka-Rushton exponent. Results of an investigation the same as that shown in Figure 2, but with three different r_{max} values (indicated by different colours), and a wider range of exponents, q. Thick, solid lines plot $\max\left[\tilde{\tau}(x)\right]$, calculated using Equation (18) with $K = 1$ and $v = 2$; symbols (joined by thin lines) plot the height of the peak of decoding precision from Monte Carlo simulations. At high spike rates, the relationship between maximum precision and q is a straight line on log-log coordinates that is well predicted by $\max\left[\tilde{\tau}(x)\right]$; at low spike rates, the relationship between maximum precision and q becomes a shallow curve.

Population coding of contrast

Clatworthy et al. (2003) applied maximum-likelihood decoding to three different populations of 18 model Tolhurst-spiking neurons, each with $r_{max} = 10$, $r_0 = 0$, and $q = 2$. One population had $\log_{10}(c_{1/2})$ values uniformly distributed between –3 and 0.1, and the other two had $c_{1/2}$ values distributed according to the recorded values in either cat or monkey populations, found by arranging the neurons in ascending order of $c_{1/2}$ and then sampling the population at equal percentile intervals. We did not have access to the exact sets of cat or monkey $c_{1/2}$ values that they used, but we estimated them by fitting functions to the cat or monkey $c_{1/2}$ distributions given in Clatworthy et al.'s Figure 6, and then sampling these distributions in equal percentile steps (see Supplementary Appendix I for details). The advantage of our method is that it can easily be extended to neuronal populations of any size. Having set up the populations of neurons, we then calculated decoding precision scores using Clatworthy et al.'s Monte Carlo methods (see Supplementary Appendix G for details). Figure 7 shows that our decoding precision scores are very similar to those of Clatworthy et al., confirming that our method of generating the sets of $c_{1/2}$ values is a good approximation to that of Clatworthy et al. The coloured curves in Figure 7 show $\tilde{\tau}(x)$, calculated using Equation (12) with $v = 2$. Since $r_{max} = 10$ in these simulations, $H_{Tolhurst}(x) \approx 0.5$ for the most informative neurons, so $\tau_{Tolhurst}(x) \approx \tilde{\tau}(x)$, and there was no advantage in using the closer, but more complex, approximation. Figure 7 shows that, even for these small populations of neurons, with r_{max} of only 10 spikes, the decoding precision from the Monte Carlo simulations is very close to $\tilde{\tau}(x)$ over a wide range of contrasts.

In Figure 7, the only substantial deviation of decoding precision from Fisher information occurs at the ends of the contrast range, where the precision scores shoot up. This is an artifact, discussed by Clatworthy et al., caused by the fact that, in the simulations, the likelihood functions were calculated over a finite range of contrasts. This means that, on trials where the inferred contrast would have fallen beyond the ends of the contrast range, the inferred contrasts instead pile up on the ends of the range, artificially boosting the number of trials on which the inferred contrast takes those values; thus, when the stimulus contrast really is at or close to one of the ends of the range, it is likely to be close to the inferred contrast, so the precision is artificially high. The Fisher information does not show this effect because it is a purely local measure, derived from the contrast-response functions at each point along the contrast axis, so it is unaffected by any bounds on the range of contrasts.

$\tilde{\tau}(x)$ is a good approximation of the decoding precision for the Tolhurst process as long as r_{max} is not too low. For low values of r_{max}, $\tilde{\tau}(x)$ tends to underestimate the true Fisher information, and, as noted earlier, this means that there is not an exact trade-off between r_{max} and the number of neurons, K. For low values of r_{max}, it is better to use $\tau_{Tolhurst}(x)$ to accurately predict decoding precision. This is demonstrated in Figure 8.

There are at least two reasons why we will sometimes need an expression for the decoding precision that remains a close approximation down to very low spike rates. Firstly, the median spike count for a 200 ms stimulus presentation is only around 5.7 spikes for a V1 neuron tuned to the stimulus (Geisler & Albrecht, 1997), and observers can make complex judgements based on even shorter exposures than that (Thorpe, Fize, & Marlot, 1996). Secondly, in many situations, the majority of neu-

rons will not be well tuned to the stimulus, but may still contribute to task performance due to their abundance; these neurons would be expected to have a very low average spike count over the course of a stimulus presentation.

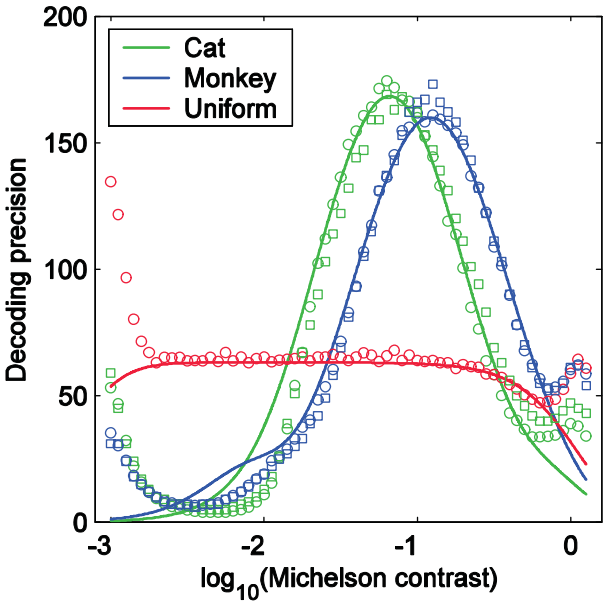

Figure 7. Population decoding precision for three different populations of $K = 18$ neurons. Within each population, the $\log_{10}(c_{1/2})$ values were uniformly distributed between log contrasts of –3 and 0.1 (red), or distributed according to recorded cat (green) or monkey (blue) populations (see Supplementary Appendix I for details). For each neuron, $r_{\max} = 10$, $r_0 = 0$, and $q = 2$. Squares show Clatworthy et al.'s (2003) decoding precision data, read off from their Figure 7; circles show the decoding precision data from our replication of their Monte Carlo methods. The solid curves show $\tilde{\tau}(x)$, calculated using Equation (12) with $v = 2$. Note that, because our method of selecting the $c_{1/2}$ values differed slightly from that of Clatworthy et al., our decoding precision scores show a very slight systematic difference from theirs, which is more noticeable in the cat data. $\tilde{\tau}(x)$ was calculated using our $c_{1/2}$ values, and so shows a slightly better match to our precision scores than to those of Clatworthy et al.

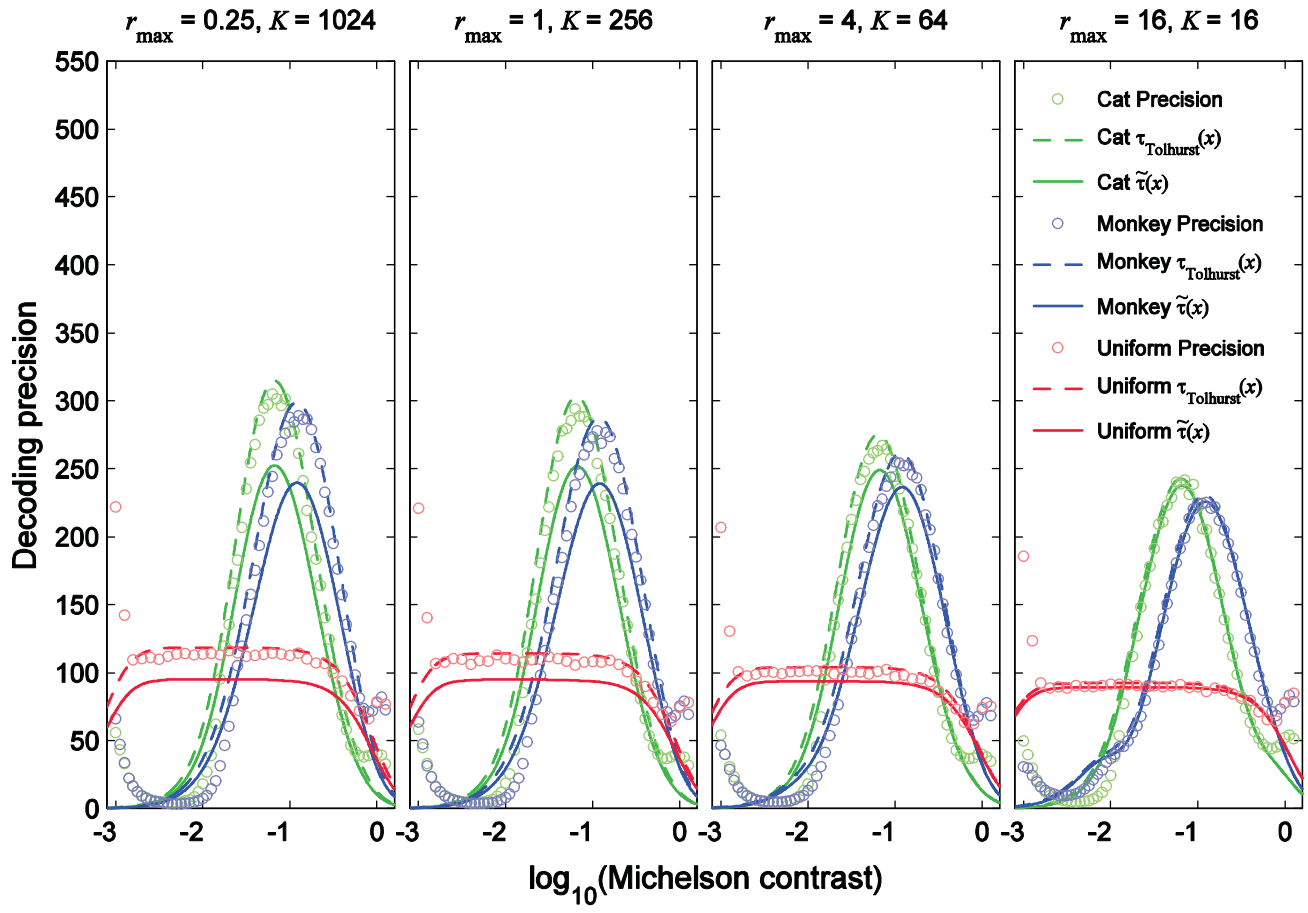

Figure 8. Results of four simulations the same as those in Figure 7, except with different values of r_{max} and K, indicated at the top of each panel. As in Figure 7, circles plot decoding precision from the Monte Carlo simulations, and solid curves plot $\tilde{\tau}(x)$, calculated using Equation (12) with $\nu = 2$. Dashed curves plot $\tau_{\text{Tolhurst}}(x)$, calculated using Equation (21). For $r_{max} = 16$, $\tilde{\tau}(x)$ and $\tau_{\text{Tolhurst}}(x)$ are very well matched to each other, and to the true decoding precision. For $r_{max} = 4$ and below, only $\tau_{\text{Tolhurst}}(x)$ is well matched to the true decoding precision. This figure also illustrates the breakdown of the linear trade-off between r_{max} and K when r_{max} is low. Each panel has the same product $r_{max} \times K$, but the decoding precision falls with increasing r_{max}. This breakdown of the linear trade-off applies to the Tolhurst process shown here, but not the ordinary Poisson process or the Goris process.

The psychometric function for contrast detection

Chirimuuta and Tolhurst (2005a) used their contrast coding model to simulate a 2-alternative forced-choice (2AFC) contrast detection task, where the observer is presented with a zero-contrast nontarget stimulus, and an above-zero-contrast target stimulus, and has to pick the target. Chirimuuta and Tolhurst measured the model's proportion of correct responses as a function of the target contrast, and fitted a Weibull psychometric function to the data.

For 2AFC tasks, the Weibull function can be defined as

$$P(\text{correct}) = 1 - 0.5\exp\left[-(c/\alpha)^{\beta}\right]. \quad (22)$$

α is the threshold, i.e., the target contrast that gives $P(\text{correct}) = 1 - 0.5/e = 0.816...$, and β controls the function's shape on linear axes, or slope on log axes (see May & Solomon, 2013, for an in-depth analysis of the Weibull function). In psychophysical contrast detection experiments with human observers, β usually takes a value of about 3 (Foley & Legge, 1981; Nachmias, 1981; Mayer & Tyler, 1986; Meese, Georgeson, & Baker, 2006; Wallis, Baker, Meese, & Georgeson, 2013).

When Chirimuuta and Tolhurst's model had a standard Naka-Rushton contrast-response function with $r_0 = 0$ and $q = 2$, Weibull β for detection varied from 1.75 to 1.99 as the number of neurons increased from 1 to 23, but the model's β never reached the normal human level of around 3. Chirimuuta and Tolhurst then introduced a threshold to the Naka-Rushton function by subtracting a constant value from the output and setting negative values to zero. With a threshold on the Naka-Rushton function, β ranged from 2.25 to 4.20, providing a better match to psychophysically obtained values. Chirimuuta and Tolhurst assumed that their failure to obtain sufficiently high Weibull β values with the standard Naka-Rushton function had been caused by the lack of a threshold, and they suggested that the standard, unthresholded Naka-Rushton function "may be crucially inadequate" as a model of the neuronal contrast-response function (Chirimuuta & Tolhurst, 2005a, p. 2956). However, we will now show that, when $r_0 = 0$, the model's psychometric function is close to a Weibull function with $\beta = q$. Thus, the real reason that Chirimuuta and Tolhurst always obtained a β of about 2 with the standard Naka-Rushton function is that they always kept q at 2 in these simulations. We will show that one can obtain any Weibull β by setting q close to the required β value.

Proof that, when $r_0 = 0$, the model's psychometric function for 2AFC contrast detection is close to a Weibull function with $\beta = q$

At the low contrasts corresponding to the model's contrast detection threshold, the population spike rate is typically so low that the Fisher information does not provide a useful approximation of performance, so we need to use other methods. For contrast detection, the model's performance can be derived straightforwardly from basic probability theory.

Because $r_0 = 0$, there is zero response to zero contrast: The nontarget stimulus can never elicit a single spike. If the target elicits at least one spike, the model will respond correctly. If the target fails to elicit any spikes, the model has to guess, and will be correct half the time on a 2AFC task. In summary, the model will be correct on all 2AFC trials except half of those on which the target failed to elicit any spikes. This statement can be formalized as follows:

$$P(\text{correct}) = 1 - 0.5P(\text{no spikes}), \quad (23)$$

where $P(\text{no spikes})$ is the probability of getting no spikes in response to the target. We can already see that the model's psychometric function has a similar form to the Weibull function: $P(\text{correct}) = 1 - 0.5 \times \text{something}$.

From Equation (E.4) of Supplementary Appendix E, we see that, for a single Tolhurst-spiking neuron,

$$P(\text{no spikes}) = P_{\text{Tolhurst}}(N = 0 \mid R = r(c))$$

$$= \exp\left[(1/e - 1)r(c)\right]. \quad (24)$$

Note that, since the Weibull function is defined as a function of contrast in Michelson (linear) units, we express the mean response, $r(\cdot)$, as a function of Michelson contrast in Equation (24), rather than log contrast. For a population of K statistically independent neurons,

$$P(\text{no spikes}) = \prod_{j=1}^{K} \exp\left[(1/e - 1)r_j(c)\right]$$

$$= \exp\left((1/e - 1)\sum_{j=1}^{K} r_j(c)\right), \quad (25)$$

where $r_j(c)$ is the contrast-response function of neuron j. Using Equation (1) to expand $r_j(c)$, we have

$$P(\text{no spikes}) = \exp\left((1/e - 1)\sum_{j=1}^{K} \frac{(r_{\max})_j c^q}{(c_{1/2})_j{}^q + c^q}\right), \quad (26)$$

where $(r_{\max})_j$ and $(c_{1/2})_j$ are the $r_{\max}$ and $c_{1/2}$ parameters of neuron j. Using Equation (26) to substitute for $P(\text{no spikes})$ in Equation (23), we get an exactly correct expression for the model's psychometric function for contrast detection:

$$P(\text{correct}) = 1 - 0.5\exp\left((1/e - 1)\sum_{j=1}^{K} \frac{(r_{\max})_j c^q}{(c_{1/2})_j{}^q + c^q}\right). \quad (27)$$

If all the neurons in the population being monitored by the observer have the same contrast-response function, then Equation (27) reduces to

$$P(\text{correct}) = 1 - 0.5\exp\left(\frac{(1/e - 1)r_{\max} K c^q}{c_{1/2}{}^q + c^q}\right). \quad (28)$$

In this case, Equation (28) shows that there is an *exact* linear trade-off between $r_{\max}$ and K in the psychometric function for 2AFC detection, rather than the approximate trade-off that we get with the Fisher information.

Equations (27) and (28) are precisely correct, but it is illuminating to derive an approximation. As long as there are enough neurons, or $r_{\max}$ is sufficiently high, the contrast detection threshold will be somewhat lower than the lowest $c_{1/2}$ in the population. Thus, at threshold (i.e., around the middle of the psychometric function), the denominators of Equations (26) to (28) will become dominated by the ${c_{1/2}}^q$ term, and the c^q term will make little difference. Dropping c^q from the denominator of Equation (27), we get

$$P(\text{correct}) \approx 1 - 0.5\exp\left[-\left(c/\alpha\right)^q\right], \quad (29)$$

where α is a constant, given by

$$\alpha = \left((1-1/e)\sum_{j=1}^{K} \frac{(r_{\max})_j}{{(c_{1/2})_j}^q} \right)^{-1/q}. \quad (30)$$

Equation (29) has the form of a Weibull function with $\beta = q$. The near-equality in Relation (29) approaches equality as *K* or $r_{\max}$ increase. □

The derivation of Relation (29) is similar to the derivation of the psychometric function for Quick's (1974) vector-magnitude model of contrast detection. The performance of the vector-magnitude model is exactly equivalent to that of a model in which the observer monitors a number of detectors that each has a detection probability that can be described by a Weibull function, with the same β for all detectors. Nachmias (1981) called this assumption of identical β for each detector the "homogeneity assumption", and it is largely equivalent to the implicit assumption in the above proof that all neurons have the same *q*. In the Discussion we expand on the links between our analysis and that of Quick.

Another attribute of the model's psychometric function: Lapse rate

Note that Relation (29) is a Weibull function with zero lapse rate, i.e. it predicts that the model's performance will asymptote to perfect performance ($P(\text{correct}) = 1$) with increasing target contrast. For most instantiations of the model, this is very close to the truth. However, when $r_{\max}$ and *K* are both very low, the model's asymptotic performance is far below 1. This is because, as *c* increases, the denominator of Equation (26) will become more and more dominated by the c^q term, and the ${c_{1/2}}^q$ term will make less and less difference, so

$$\lim_{c\to\infty} P(\text{no spikes}) = \exp\left((1/e-1)\sum_{j=1}^{K} (r_{\max})_j \right). \quad (31)$$

For very low $r_{\max}$ and *K*, this value can be substantially above zero, so that, even for infinite contrast, the model has to guess on a significant proportion of 2AFC trials.

This behaviour can be accommodated by including a "lapse rate" parameter, λ, in the definition of the Weibull function:

$$P(\text{correct}) = (1-\lambda) - (0.5-\lambda)\exp\left[-\left(c/\alpha\right)^\beta\right]. \quad (32)$$

This equation approaches an asymptote of $P(\text{correct}) = (1-\lambda)$ as $c \to \infty$. When $\lambda = 0$, Equation (32) reduces to Equation (22). We now derive an expression for the model's "lapse rate" parameter (strictly speaking, the model never "lapses" – a low asymptotic performance level arises from a low spike rate, rather than a finger error or a failure to look at the stimuli on some 2AFC trials).

Using Equation (31) to substitute for $P(\text{no spikes})$ in Equation (23), we obtain the asymptotic value of $P(\text{correct})$ given by

$$\lim_{c\to\infty} P(\text{correct}) = 1 - 0.5\exp\left((1/e-1)\sum_{j=1}^{K} (r_{\max})_j \right). \quad (33)$$

Since this asymptotic value of $P(\text{correct})$ is $(1-\lambda)$, we have

$$\lambda = 0.5\exp\left((1/e-1)\sum_{j=1}^{K} (r_{\max})_j \right). \quad (34)$$

This expression for λ quickly approaches zero as $r_{\max}$ or *K* increase.

Verifying our equations using Monte Carlo simulations

We simulated the 2AFC contrast detection experiments for a range of model parameterizations (see Supplementary Appendix J for details of the methods). For each parameterization, each neuron in the population had an identical contrast-response function. We fitted the 3-parameter Weibull function (Equation (32)) to the data for each model parameterization. The fitted values of α, β

and λ are plotted as symbols in Figures 9, 10, and 11, respectively. The solid lines plot the corresponding analytical expressions (α given by Equation (30), β given by q, and λ given by Equation (34)).

Equation (34) is the model's true "lapse rate" parameter (not an approximation), so it is not surprising that it matches the fitted λ values extremely well in Figure 11. Our analytical expressions for α and β are approximations that become increasingly accurate as K or $r_{\max}$ increase.

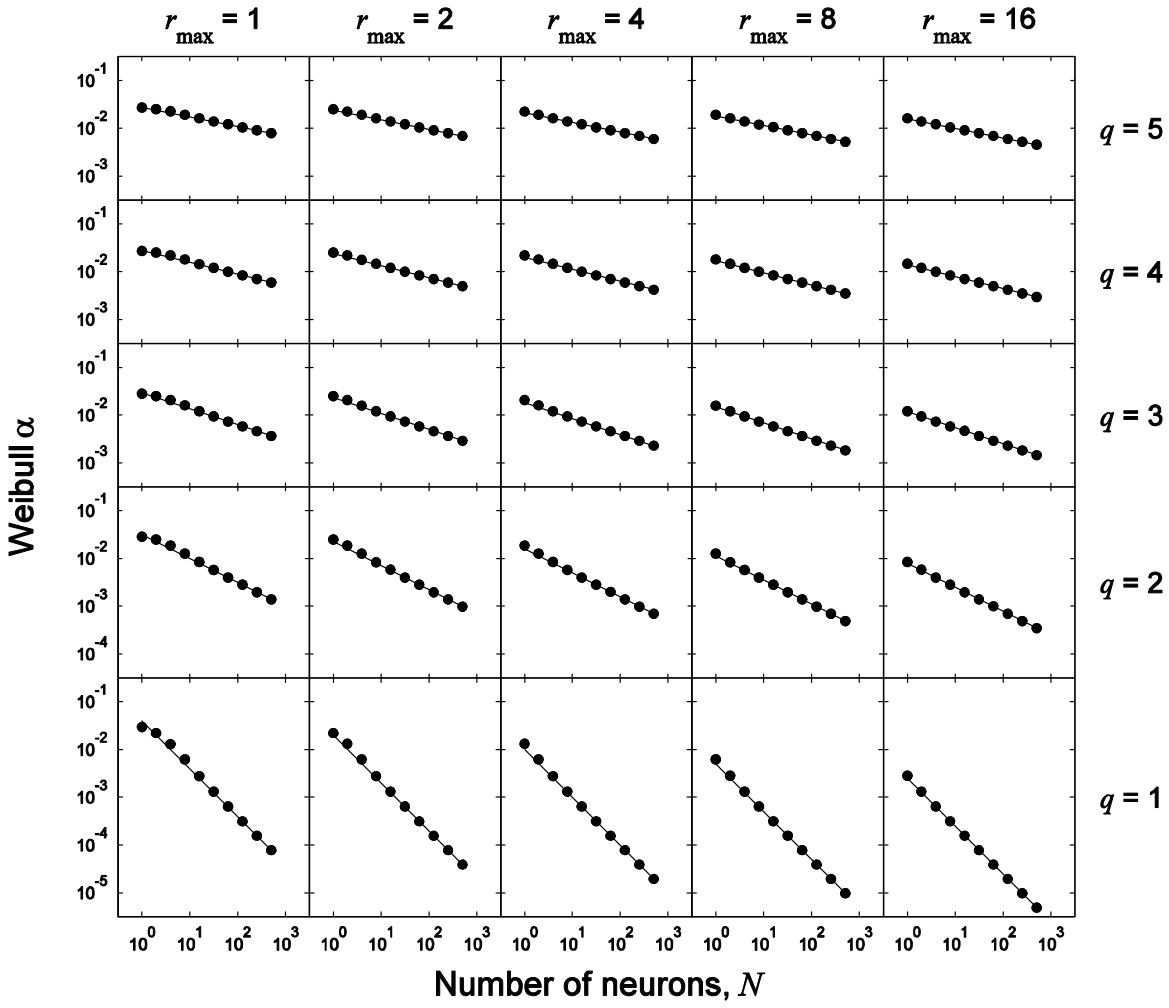

Figure 9. Filled circles show Weibull α parameters fitted to the 2AFC contrast detection simulation data. Each column of panels shows a different $r_{\max}$; each row shows a different q. The solid lines plot the approximation of α given by Equation (30).

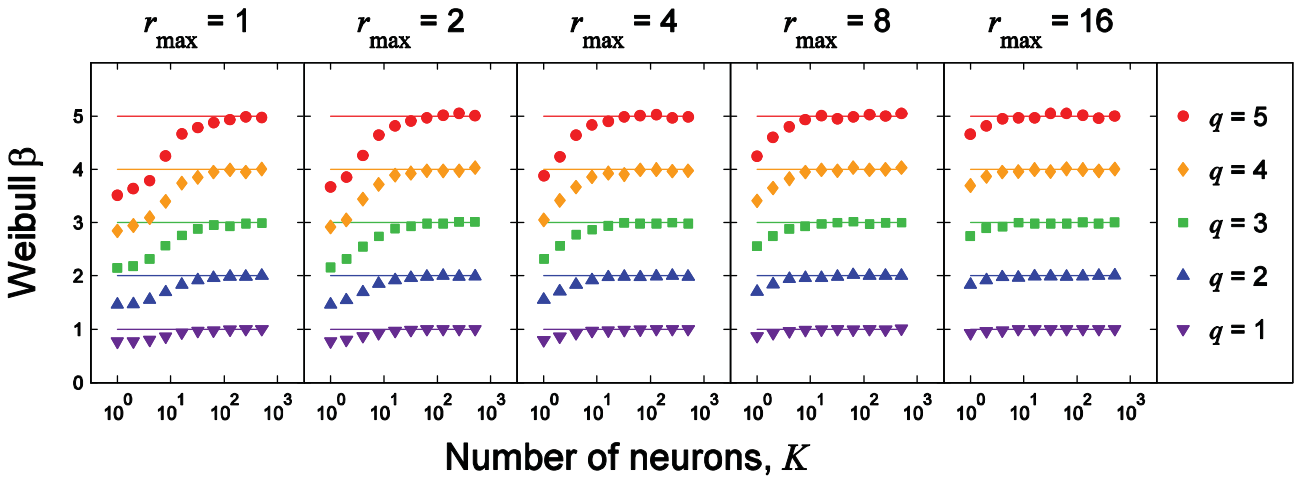

Figure 10. Symbols show Weibull β parameters fitted to the 2AFC contrast detection simulation data. Each panel shows a different $r_{\max}$. Different values of q are indicated by different symbols and colours, according to the legend on the right. The solid lines plot the predicted values of β (= q) for our approximation of the psychometric function (Relation (29)).

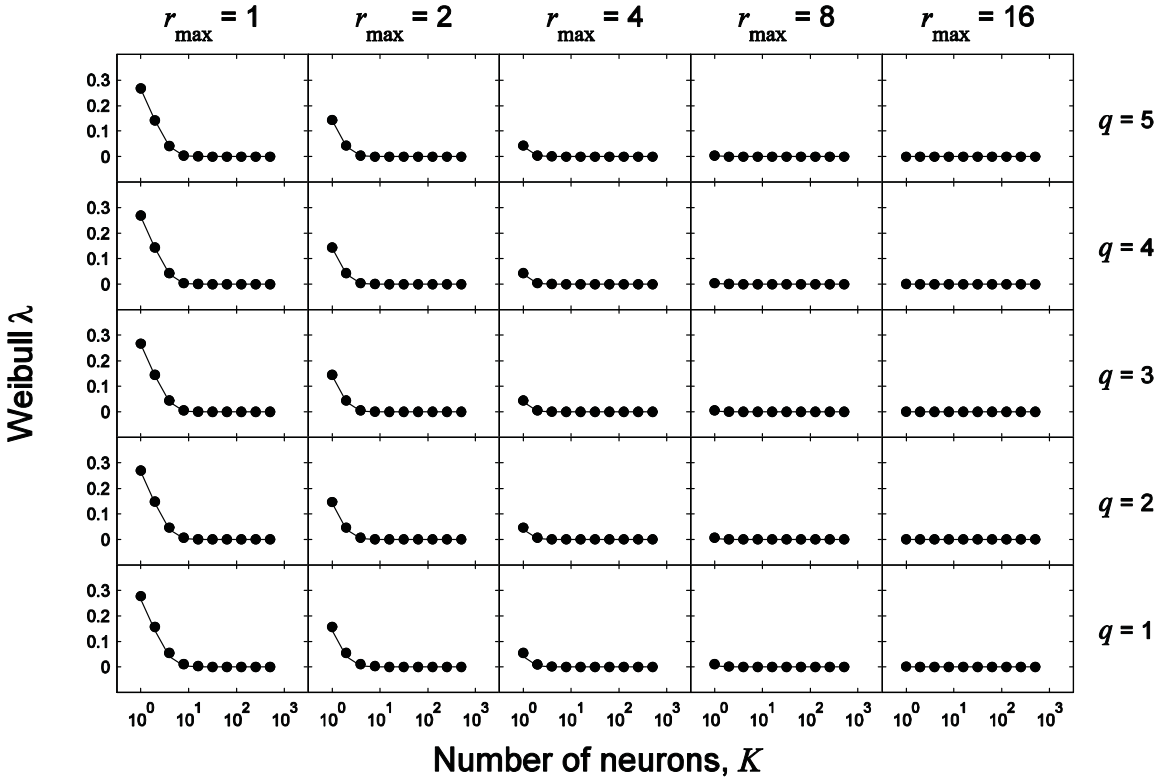

Figure 11. Filled circles show Weibull λ parameters fitted to the 2AFC contrast detection simulation data. Each column of panels shows a different $r_{\max}$; each row shows a different q. The solid curves plot λ given by Equation (34). As shown in Equation (34) (and borne out by our simulations), λ depends only on K and $r_{\max}$, and is independent of q.

Study	No. of neurons	Mean q	Median q
Albrecht & Hamilton (1982) – cat	127	2.5	N/A
Albrecht & Hamilton (1982) – monkey	98	3.4	N/A
Geisler & Albrecht (1997) – cat	247	3.0	2.8
Geisler & Albrecht (1997) – monkey	71	2.5	2.0
Sclar et al (1990) – monkey	85	2.65	2.4
Overall mean		2.9	

Table 1. Mean and median Naka-Rushton exponent, q, from three different physiological studies of V1. Albrecht and Hamilton (1982) did not report the median, so these are omitted. Sclar et al. (1990) did not report the mean – the value given in this table was calculated from the histogram of q values shown in their Figure 6. The centres of the bars of their histogram ranged from 0.875 to 6.375 in steps of 0.25. The numbers of neurons in the corresponding bins were 8, 5, 2, 12, 7, 6, 4, 3, 9, 3, 3, 5, 4, 4, 0, 4, 1, 1, 1, 2, 0, 0, and 1. Taking the q of each neuron in a histogram bin to be the central value of the bin (a reasonable approximation given the narrowness of their bins), this gives a mean value of 2.65, calculated from 85 neurons. Sclar et al.'s study included one further V1 neuron, with $q > 8$, but, since they were no more specific about its q value than that, we excluded it from our analysis. The overall mean was the average of the "Mean q" values, weighted by the number of neurons in each study (this is equivalent to pooling the neurons across all three studies, and finding the mean of the pooled sample).

The Consul-Jain spiking process also generates a Weibull psychometric function for detection

The analytical expressions for the psychometric function for detection, derived above, apply only to the Tolhurst spiking process. We can also derive analogous expressions for the Consul-Jain process (which includes the ordinary Poisson, when $F = 1$). Equation (E.17) of Supplementary Appendix E states that, for a single Consul-Jain-spiking neuron,

$$P(\text{no spikes}) = \exp\left(-r(c)/\sqrt{F}\right), \tag{35}$$

where F is the Fano factor. If we follow a series of mathematical steps analogous to (23) to (29) above, but using Equation (35), instead of Equation (24), to express the probability of a single neuron not spiking, we obtain an approximation of the psychometric function with the same form as Relation (29), but with

$$\alpha = \left(\sum_{j=1}^{K} \frac{(r_{\max})_j}{(c_{1/2})_j{}^q \sqrt{F_j}} \right)^{-1/q}, \tag{36}$$

where F_j is the Fano factor or neuron j. Similarly, if we follow a series of steps analogous to those in Equations (31) to (34), but using Equation (35), instead of Equation (24), to express the probability of a single neuron not spiking, we obtain the following expression for the "lapse rate" parameter:

$$\lambda = 0.5 \exp\left(-\sum_{j=1}^{K} \frac{(r_{\max})_j}{\sqrt{F_j}} \right). \tag{37}$$

Discussion

The purpose of this study was to explain a number of empirical modelling results reported by Tolhurst and colleagues. Their results were obtained from Monte Carlo simulations using models of spiking neurons with Naka-Rushton contrast-response functions and the doubly stochastic Poisson spiking process defined in Equation (5). The numerical nature of the simulations meant that it was not clear why the results occurred, or how they generalized across parameter space.

We addressed these problems by deriving equations to explain the model's performance. This kind of analysis is essential if we are to understand the brain: If we use realistic models to simulate brain processes, but do not understand why the models behave in the way that they do, then we have not really explained anything – we have just shifted the problem from understanding the brain to understanding the model.

We began by deriving a closed-form expression for the Tolhurst likelihood function that was more tractable than the infinite series that originally defined this function. This expression played a role in understanding two facets of the model's performance: (1) the relationships between decoding precision and the neuronal parameters; (2) the form of the psychometric function for contrast detection.

Decoding precision

To explain how the neuronal parameters map onto decoding precision, we derived an analytical approximation of the decoding precision, $\tilde{\tau}(x)$, which can be adjusted to apply to a variety of different spiking processes by setting the value of a single scalar parameter, v. For the Tolhurst and Consul-Jain processes, v should be set to the Fano factor. For the Goris process, v should be set to $1/(1-\sigma_G^2)$, where σ_G^2 is the variance of the gain signal. For the Tolhurst and Consul-Jain processes, $\tilde{\tau}(x)$ is an estimate of the Fisher information. For the Goris process, the Fisher information varies across trials, and $\tilde{\tau}(x)$ is its modal value.

The expression for $\tilde{\tau}(x)$ reveals some surprisingly simple relationships between decoding precision and the neuronal parameters, and explained the five observations of Clatworthy et al. (2003) that we investigated. For example, Equation (18) shows that, for a population of identical, statistically independent neurons, the height of the peak of decoding precision is approximately proportional to $r_{\max} \times K \times q^2$, and independent of $c_{1/2}$. The expression for $\tilde{\tau}(x)$ also revealed that, to a good approximation, the contrast most accurately encoded by the neuron is that for which the mean response is $r_{\max}/3$; we call this contrast $c_{1/3}$. Figure 7 shows that $\tilde{\tau}(x)$ matches the decoding precision very closely for a population of only 18 Tolhurst-spiking neurons with $r_{\max}$ of only 10 spikes.

$\tilde{\tau}(x)$ is an exact expression for the modal Fisher information for the Goris process, and gives the exact expression for the Fisher information for the Poisson process, which is the Consul-Jain process with a Fano factor of 1. However, for the Tolhurst process and the Consul-Jain process with $F > 1$, $\tilde{\tau}(x)$ is an approximation of the Fisher information, and is less accurate when the mean spike rate of the most informative neurons is very low. In Supplementary Appendix E, we derived expressions, $\tau_{\text{Tolhurst}}(x)$ and $\tau_{\text{C-J}}(x)$, which are close approximations of the Fisher information of the Tolhurst and Consul-Jain processes across all parameter values. These expressions reveal more complicated relationships between decoding precision and the neuronal parameters that hold when the spike rate is very low. Figure 8 shows the superiority of $\tau_{\text{Tolhurst}}(x)$ over $\tilde{\tau}(x)$ at very low spike rates.

The psychometric function for contrast detection

To explain Chirimuuta and Tolhurst's (2005a) finding regarding the slope of the psychometric function for contrast detection, we derived an analytical approximation of the model's psychometric function, which we showed approaches a Weibull function with $\beta = q$ as $r_{\max}$ or K increase. We thus refuted Chirimuuta and Tolhurst's conclusion that the standard Naka-Rushton function is unable to give Weibull β values that are high enough to match those of human observers. To obtain β values of around 3, typical of human observers, we just need a Naka-Rushton exponent of around 3. Such levels are not atypical, and indeed Table 1 shows that the mean q over 628 cells from three physiological studies of V1 is 2.9.

This example illustrates the power of the analytical approach taken here, compared with Monte Carlo simulations. Because of the considerable time that it takes to run the simulations, Chirimuuta and Tolhurst (2005a) could not feasibly explore every corner of parameter space in each investigation, so they took the entirely reasonable step of fixing the Naka-Rushton exponent, q, at a physiologically plausible level of 2. However, this turned out to be a fatal decision, because it prevented them from ever finding a parameterization of the model that gave a sufficiently high Weibull β with the standard Naka-Rushton tuning function. Using the analytical approach, Relation (29) makes it immediately clear that the Naka-Rushton exponent, q, is the key parameter for controlling β, and that the model's psychometric function approximates a Weibull function with β actually equal to q.

It should be noted that our derivation of a psychometric function with the form of a Weibull function applies only to the case of $r_0 = 0$. With nonzero r_0, the analytical form of the psychometric function is different (we have analyzed this more general case, and will present it in another paper). It is implausible that, in human vision, contrast detection is mediated entirely by neurons with zero spontaneous firing rate. The assumption that there is no neuronal response to zero contrast is often called the "high-threshold assumption"[2]: Under the conventional assumption of additive, stimulus-independent, noise, the lack of response to zero contrast implies that there is a threshold on the output of the sensory units, which lies enough standard deviations above the mean of the noise for there to be a negligible probability of a response to zero contrast. Because Chirimuuta and Tolhurst's model has no sensory response to zero contrast, it is formally equivalent to a high-threshold model, even though (with the standard Naka-Rushton function) it does not actually contain a sensory threshold. In high-threshold theory (and

in Chirimuuta and Tolhurst's model), detection errors are always unlucky guesses on 2AFC trials that failed to elicit a response, whereas there is plenty of psychophysical evidence that incorrect responses are caused at least partly by "hallucinations" due to noise in the nontarget interval, rather than entirely by unlucky guesses (Tanner & Swets, 1954; Swets, Tanner, & Birdsall, 1961; Nachmias, 1981; Solomon, 2007; Laming, 2013).

Despite the implausibility of Chirimuuta and Tolhurst's model with $r_0 = 0$ as a model of contrast detection, we presented our analysis of it for three reasons:

1. It allows us to fully understand why Chirimuuta and Tolhurst's (2005a) contrast detection simulations resulted in a fitted Weibull β that approached 2 with increasing number of neurons; this in turn allows us to refute their conclusion that the Naka-Rushton function requires a threshold to make it a plausible model of the neuronal contrast-response function;
2. neurons with zero or negligible spontaneous firing rates do exist (e.g. see Dean's (1981a) Figure 2) so it is not inconceivable that there are some organisms or experimental situations to which our analysis applies;
3. it allows us to address Tyler and Chen's (2000) claim that high-threshold analysis of probability summation is "fundamentally flawed"; this idea is explored in the next subsection.

High-threshold probability summation

Our derivation of the model's psychometric function for detection is an example of *high-threshold probability summation*. When $r_0 = 0$, each neuron acts as a "detector"; the observer detects the target if at least one neuron responds during the target presentation, and has to guess the correct answer otherwise. The more neurons the observer is monitoring, the greater the chance that at least one neuron will respond. The term "probability summation" refers to this increase in detection probability due to an increase in the number of detectors. The psychometric function in this case gives the probability that either at least one neuron responds during the target presentation or, if none responds, that the observer guesses correctly.

Quick (1974) showed that, if the detectors are statistically independent, and each detector has a detection probability that is a Weibull function with the same β, then the observer's psychometric function will be a Weibull function with that β-value, and with the detection threshold parameter, α, given by

$$\alpha = \left(\sum_{j=1}^{K} \alpha_j^{-\beta} \right)^{-1/\beta}, \tag{38}$$

where α_j is the detection threshold parameter of detector j. For a very clear derivation of Equation (38), see Nachmias (1981), but note that his Equation (4) has a typographical error: It is missing the minus sign on the exponent, β. If all the detectors are identical (but statistically independent) then they all have the same α_j, and Equation (38) reduces to

$$\alpha = K^{-1/\beta} \alpha_j . \tag{39}$$

Thus, in Quick's model, the Weibull β parameter controls how much of a reduction in detection threshold we achieve by increasing the number of detectors, K. If β is low, then the detection threshold, α, decreases quickly as K increases; if β is high, then the detection threshold decreases more slowly as K increases.

Equation (38) gives the observer's detection threshold according to Quick's model, while Equation (30) gives the observer's approximate detection threshold according to Chirimuuta and Tolhurst's (2005a) model. Equation (38) has the exact form of Equation (30) if $\beta = q$ and

$$\alpha_j = \frac{(c_{1/2})_j}{\left[(1 - 1/e)(r_{\max})_j \right]^{1/q}} . \tag{40}$$

Thus, there is a near-equivalence between Quick's model and that of Chirimuuta and Tolhurst. If all the neurons in Chirimuuta and Tolhurst's model have the same contrast-response function, then the model's detection threshold approximation given by Equation (30) reduces to

$$\alpha = K^{-1/q} \frac{c_{1/2}}{\left[(1 - 1/e) r_{\max} \right]^{1/q}} . \tag{41}$$

Therefore, since $\beta \approx q$ in Chirimuuta and Tolhurst's model, their model shows approximately the same probability summation effects as Quick's model, with detection threshold proportional to $K^{-1/\beta}$. For the modelling in Figure 9, all the neurons were identical, so Equation (41) is equivalent to Equation (30) (which was used to generate the solid lines in Figure 9), and it is clear that this equation does accurately predict the detection threshold of Chirimuuta and Tolhurst's model, particularly for the higher values of $r_{\max}$ or K.

As noted above, there is psychophysical evidence against Quick's "high-threshold" model. Tyler and Chen (2000) went further, arguing, not just that high-threshold theory has evidence against it, but that the analysis of

probability summation using high-threshold theory is "fundamentally flawed". Their analysis of high-threshold theory made several basic assumptions about the high-threshold model, including the following five:

1. the observer monitors a set of channels;
2. within each channel, there is a continuous signal that increases linearly with the stimulus strength;
3. noise is added to this signal;
4. the noise might be additive (standard deviation independent of the signal level), or multiplicative (standard deviation proportional to a power function of the signal level), or the noise could be the sum of an additive and a multiplicative component;
5. a sensory threshold is applied to the noisy internal signal in each channel, so that a stimulus is detected if and only if the noisy signal falls above threshold in at least one of the channels.

We should emphasize that Tyler & Chen were not arguing in support of this model; they argued that it was fundamentally flawed.

Tyler and Chen noted that, if the internal noise in this model were fully multiplicative, then the internal signal-to-noise ratio would be unchanged by a change in stimulus strength. Thus, detection performance would remain the same for any stimulus level, and measurement of detection thresholds would be impossible! They concluded that, even if there is a multiplicative component to the noise, detection performance must be limited by an additive component.

Tyler and Chen derived the shape of the additive noise distribution that would give rise to a Weibull psychometric function for the high-threshold model outlined above[3]. They showed that, for most Weibull β values, the noise distribution deviated markedly from a Gaussian. They found this unacceptable because, according to the Central Limit Theorem, the sum of a large number of non-Gaussian-distributed random variables is asymptotically Gaussian-distributed, and so, if there are many different sources of noise from the stimulus to the neuronal decision mechanism, the noise is likely to be Gaussian at the decision mechanism. They then showed that high-threshold probability summation fails for additive Gaussian noise. They showed that, if the sensory threshold were low enough to be exceeded by the noisy signal in one channel 75% of the time, then the noisy signal would exceed the sensory threshold in at least one of 100 channels almost all the time, even when the stimulus intensity was reduced to zero. Thus, if the stimulus area or number of components increased, so that the observer was monitoring many more channels, the observer would almost always be in a "detect" state, even when the stimulus was absent. The signal would have to be reduced to a physically unachievable negative contrast for the observer to be in a "detect" state 75% of the time. This problem does not occur with the noise distribution implied by the Weibull function, because the Weibull probability density function falls to zero at the sensory threshold (see Supplementary Appendix K, especially Figure K.1), and so, when the stimulus intensity is zero, no detector's noisy signal will exceed its sensory threshold; but, as already noted, Tyler and Chen rejected that distribution because of its "bizarre" non-Gaussian nature when β is not close to 4 (Tyler & Chen, 2000, p. 3127). They therefore concluded that probability summation with high-threshold theory is fundamentally flawed.

Tyler and Chen's arguments are perfectly valid, but they apply only to the set of high-threshold models defined by their assumptions, i.e. that each channel contains a continuous signal to which noise is added. With a detection model based on spiking neurons, the signal is discrete. When the spontaneous firing rate is zero, as it was in Chirimuuta and Tolhurst's model, detection occurs when a single neuron produces a single spike. In these circumstances, it is not appropriate to approximate the neuronal signal as a continuous signal to which noise is added. It is perfectly legitimate to apply probability summation to find the probability that at least one neuron gives at least one spike. When we do this, we find that the model's threshold is very close to being proportional to $K^{-1/\beta}$, as in standard high-threshold Weibull analysis. The contrast threshold for detection never reaches zero, however many neurons the observer is monitoring, but we do not have do rely on bizarre model characteristics to achieve this: Aside from the zero spontaneous firing rate, the neurons in Chirimuuta and Tolhurst's model have contrast-response functions and noise distributions that are physiologically plausible to a reasonable extent. One could ask what is the equivalent "continuous signal plus additive noise" model. Since Chirimuuta and Tolhurst's model produces a psychometric function that closely approximates a Weibull function, the equivalent "continuous signal plus additive noise" model is closely approximated by the one derived by Chen and Tyler (and outlined in Supplementary Appendix K) with the "bizarre" noise distributions. However, this bizarreness comes from forcing the model into the Procrustean bed of "continuous signal plus additive noise", rather than being an implausible characteristic of the model itself.

Conclusions

We derived equations that explained the performance of the contrast coding model described by Tolhurst and colleagues (Clatworthy et al., 2003; Chirimuuta et al.,

2003; Chirimuuta & Tolhurst, 2005a, 2005b). These equations gave a deep insight into the results of their Monte Carlo simulations.

As long as the stimulus contrast is high enough (and the neuronal population large enough) to generate a sufficiently high population spike rate, the decoding precision can be closely approximated by the Fisher information. We derived an estimate of decoding precision, $\tilde{\tau}(x)$, which approximates the Fisher information for a population of neurons with Tolhurst's spiking process, as long as the mean spike count of the most informative neurons is around 5 spikes or more. $\tilde{\tau}(x)$ is also an estimate of the Fisher information for the Consul-Jain spiking process, and gives the exact Fisher information for the Poisson process, and the exact modal value of the Fisher information for the Goris process when the decoder has access to the gain signal in Goris et al.'s model of neuronal spiking. Our expression for $\tilde{\tau}(x)$ revealed simple relationships between the properties of the neurons and the decoding precision that hold to a good approximation when the mean spike rate is sufficiently high. We used this expression to explain five relationships between decoding precision and the neuronal parameter values that Clatworthy et al. (2003) observed from their Monte Carlo simulations.

For the case of contrast detection, the spike rate is too low for the Fisher information to match decoding precision. To analyze the performance of Chirimuuta and Tolhurst's (2005a) model in a 2AFC contrast detection task, we used basic probability theory. We derived an expression for the model's psychometric function for contrast detection, and showed that, as K or $r_{\max}$ increases, the psychometric function asymptotically approaches a Weibull function with $\beta = q$. Our work therefore reveals a previously unknown theoretical connection between two of the most widely used functions in vision science: the Weibull psychometric function and the Naka-Rushton contrast-response function. This relationship explained why Chirimuuta and Tolhurst always obtained a Weibull β of about 2 in their modelling (they always had $q = 2$ in their assessments of the model's Weibull β), and allowed us to refute their conclusion that it is necessary to have a threshold on the Naka-Rushton function to achieve Weibull β values that match those found with human observers. Their threshold on the Naka-Rushton function had a similar effect to increasing q, as it made the spike rate increase more abruptly with increasing contrast.

Acknowledgments

This work was supported by EPSRC grant EP/H033955/1 to Joshua Solomon. The authors would like to thank Christopher Tyler for giving detailed feedback on an earlier draft of this paper. We also thank Wei Ji Ma and an anonymous reviewer for insightful comments on our manuscript.

Commercial relationships: none.
Corresponding author: Keith A. May.
Email: keith@keithmay.org
Address: UCL Department of Computer Science, Gower Street, London, WC1E 6BT, UK.

Footnotes

[1]In this paper, we use the word "trial" in two ways. Firstly, we use it in the way a physiologist would, to mean a stimulus presentation. Secondly, we use it to mean a trial on a 2-alternative forced-choice psychophysical experiment, on which the observer is presented with two stimuli, and has to make a response. To distinguish these two meanings, we always refer to the latter type of trial as a "2AFC trial".

[2]In this discussion, we use the word "threshold" in two ways: (1) to refer to an internal threshold on the sensory signal; (2) to refer to the stimulus contrast corresponding to a particular level of detection performance. We have tried to make the meaning clear in each case, by using the term "sensory threshold" for the former case, and "detection threshold" for the latter.

[3]As pointed out by Mortensen (2002), Tyler and Chen's published equation for the probability density function (PDF) of the noise (Tyler and Chen's Equation 4b) contains errors. However, Tyler and Chen's plots of the noise PDFs (shown in their Figure 2b) are correct, so we assume that the errors in Tyler and Chen's Equation 4b are typographical rather than fundamental problems with their analysis. To clarify matters, we present in Supplementary Appendix K a derivation of the PDF, based on Mortensen's derivation, but somewhat easier to follow than Mortensen's derivation or that of Tyler and Chen.

References

Albrecht, D. G. & Hamilton, D. B. (1982). Striate cortex of monkey and cat: contrast response function. *Journal of Neurophysiology, 48*, 217-237.

Bair, W. & O'Keefe, L. P. (1998). The influence of fixational eye movements on the response of neurons in area MT of the macaque. *Visual Neuroscience, 15*, 779-786.

Bradley, A., Skottun, B. C., Ohzawa, I., Sclar, G., & Freeman, R. D. (1987). Visual orientation and spatial frequency discrimination: A comparison of single neurons and behavior. *Journal of Neurophysiology, 57*, 755-772.

Britten, K. H., Shadlen, M. N., Newsome, W. T., & Movshon, J. A. (1993). Responses of neurons in macaque MT to stochastic motion signals. *Visual Neuroscience, 10*, 1157-1169.

Buracas, G. T., Zador, A. M., DeWeese, M. R., & Albright, T. D. (1998). Efficient discrimination of temporal patterns by motion-sensitive neurons in primate visual cortex. *Neuron, 20*, 959-969.

Chirimuuta, M., Clatworthy, P. L., & Tolhurst, D. J. (2003). Coding of the contrasts in natural images by visual cortex (V1) neurons: a Bayesian approach. *Journal of the Optical Society of America, A, 20*, 1253-1260.

Chirimuuta, M. & Tolhurst, D. J. (2005a). Does a Bayesian model of V1 contrast coding offer a neurophysiological account of human contrast discrimination? *Vision Research, 45*, 2943-2959.

Chirimuuta, M. & Tolhurst, D. J. (2005b). Accuracy of identification of grating contrast by human observers: Bayesian models of V1 contrast processing show correspondence between discrimination and identification performance. *Vision Research, 45*, 2960-2971.

Clatworthy, P. L., Chirimuuta, M., Lauritzen, J. S., & Tolhurst, D. J. (2003). Coding of the contrasts in natural images by populations of neurons in primary visual cortex (V1). *Vision Research, 43*, 1983-2001.

Consul, P. C. & Jain, G. C. (1973). A generalization of the Poisson distribution. *Technometrics, 15*, 791-799.

Cramér, H. (1946). *Mathematical Methods of Statistics*. Princeton, NJ: Princeton University Press.

Dayan, P. & Abbott, L. F. (2001). *Theoretical Neuroscience: Computational and mathematical modeling of neural systems*. Cambridge, MA: MIT Press.

Dean, A. F. (1981a). The relationship between response amplitude and contrast for cat striate cortical neurones. *Journal of Physiology, 318*, 413-427.

Dean, A. F. (1981b). The variability of discharge of simple cells in the cat striate cortex. *Experimental Brain Research, 44*, 437-440.

Durant, S., Clifford, C. W. G., Crowder, N. A., Price, N. C. S., & Ibbotson, M. R. (2007). Characterizing contrast adaptation in a population of cat primary visual cortical neurons using Fisher information. *Journal of the Optical Society of America, A, 24*, 1529-1537.

Foley, J. M. & Legge, G. E. (1981). Contrast detection and near-threshold discrimination in human vision. *Vision Research, 21*, 1041-1053.

Geisler, W. S. & Albrecht, D. G. (1997). Visual cortex neurons in monkeys and cats: Detection, discrimination, and identification. *Visual Neuroscience, 14*, 897-919.

Goris, R. L. T., Movshon, J. A., & Simoncelli, E. P. (2014). Partitioning neuronal variability. *Nature Neuroscience, 17*, 858-854.

Laming, D. (2013). Probability summation-a critique. *Journal of the Optical Society of America, A, 30*, 300-315.

May, K. A. & Solomon, J. A. (2013). Four theorems on the psychometric function. *PLoS ONE, 8*(10):e74815.

May, K. A. & Solomon, J. A. (2014). Connecting psychophysical performance to neuronal response properties I: Discrimination of suprathreshold stimuli. Submitted.

Mayer, M. J. & Tyler, C. W. (1986). Invariance of the slope of the psychometric function with spatial summation. *Journal of the Optical Society of America, A, 3*, 1166-1172.

McAdams, C. J. & Maunsell, J. H. R. (1999). Effects of attention on the reliability of individual neurons in monkey visual cortex. *Neuron, 23*, 765-773.

Meese, T., Georgeson, M. A., & Baker, D. H. (2006). Binocular contrast vision at and above threshold. *Journal of Vision, 6*, 1224-1243.

Mortensen, U. (2002). Additive noise, Weibull functions and the approximation of psychometric functions. *Vision Research, 42*, 2371-2393.

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

Naka, K. I. & Rushton, W. A. H. (1966). S-potentials from colour units in the retina of fish (cyprinidae). *Journal of Physiology, 185*, 536-555.

Oram, M. W., Weiner, M. C., Lestienne, R., & Richmond, B. J. (1999). Stochastic nature of precisely timed spike patterns in visual system neuronal responses. *Journal of Neurophysiology, 81*, 3021-3033.

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

Rao, C. R. (1945). Information and accuracy obtainable in the estimation of statistical parameters. *Bulletin of the Calcutta Mathematical Society, 37*, 81-89.

Sakata, S. & Harris, K. D. (2009). Laminar structure of spontaneous and sensory-evoked population activity in auditory cortex. *Neuron, 64*, 404-418.

Sclar, G., Maunsell, J. H. R., & Lennie, P. (1990). Coding of image contrast in central visual pathways of the macaque monkey. *Vision Research, 30*, 1-10.

Scobey, R. P. & Gabor, A. J. (1989). Orientation discrimination sensitivity of single units in cat primary visual cortex. *Experimental Brain Research, 77*, 398-406.

Skottun, B. C., Bradley, A., Sclar, G., Ohzawa, I., & Freeman, R. D. (1987). The effects of contrast on visual orientation and spatial frequency discrimination: a comparison of single cells and behaviour. *Journal of Neurophysiology, 57*, 773-786.

Smith, S. W. (1999). *The Scientist and Engineer's Guide to Digital Signal Processing, 2nd Edition*. San Diego: California Technical Publishing.

Snowden, R. J., Treue, S., & Andersen, R. A. (1992). The response of neurons in areas V1 and MT of the alert rhesus monkey to moving random dot patterns. *Experimental Brain Research, 88*, 389-400.

Softky, W. R. & Koch, C. (1993). The highly irregular firing of cortical cells is inconsistent with temporal integration of random EPSPs. *Journal of Neuroscience, 13*, 334-350.

Solomon, J. A. (2007). Intrinsic uncertainty explains second responses. *Spatial Vision, 20*, 45-60.

Swets, J. A., Tanner, W. P., & Birdsall, T. G. (1961). Decision processes in perception. *Psychological Review, 68*, 301-340.

Swindale, N. V. & Mitchell, D. E. (1994). Comparison of receptive field properties of neurons in area 17 of normal and bilaterally amblyopic cats. *Experimental Brain Research, 99*, 399-410.

Tanner, W. P. & Swets, J. A. (1954). A decision-making theory of visual detection. *Psychological Review, 61*, 401-409.

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature, 381*, 520-522.

Tolhurst, D. J., Movshon, J. A., & Dean, A. F. (1983). The statistical reliability of signals in single neurons in cat and monkey visual cortex. *Vision Research, 23*, 775-785.

Tolhurst, D. J., Movshon, J. A., & Thompson, I. D. (1981). The dependence of response amplitude and variance of cat visual cortical neurones on stimulus contrast. *Experimental Brain Research, 41*, 414-419.

Tyler, C. W. & Chen, C.-C. (2000). Signal detection theory in the 2AFC paradigm: Attention, channel uncertainty and probability summation. *Vision Research, 40*, 3121-3144.

Vogels, R., Spileers, W., & Orban, G. A. (1989). The response variability of striate cortical neurons in the behaving monkey. *Experimental Brain Research, 77*, 432-436.

Wallis, S. A., Baker, D. H., Meese, T. S., & Georgeson, M. A. (2013). The slope of the psychometric function and non-stationarity of thresholds in spatiotemporal contrast vision. *Vision Research, 76*, 1-10.

Xie, X. (2002). Threshold behaviour of the maximum likelihood method in population decoding. *Network: Computation in Neural Systems, 13*, 447-456.

Supplementary appendices for "Connecting psychophysical performance to neuronal response properties II: Contrast decoding and detection"

Keith A. May

Centre for Applied Vision Research, City University London, London, UK.

Joshua A. Solomon

Centre for Applied Vision Research, City University London, London, UK

Appendix A: Main symbols

b	Base of logarithm when Michelson contrast, c, is expressed as log contrast, x
c	Michelson contrast
$c_{1/2}$	Naka-Rushton function parameter: semi-saturation contrast
$c_{1/3}$	The Michelson contrast for which the output of the Naka-Rushton function exceeds r_0 by $r_{\max}/3$
H	A function (used with subscripts) that appears in the approximations of the Fisher information (see Appendix E)
J	Fisher information
j	Integer index of the neurons in a population
K	The number of neurons being monitored by the observer
N	Random variable representing the number of spikes produced by a neuron
n	The value of N on a particular stimulus presentation
$\mathbf{N}$	Vector of random variables representing the number of spikes produced by each neuron in the population
$\mathbf{n}$	The value of $\mathbf{N}$ on a particular stimulus presentation
P	Probability
q	Naka-Rushton function parameter: exponent, which determines tuning sharpness
R	Random variable representing the mean response of a neuron
$r(x)$	The neuron's tuning function, which gives the value of R on a particular stimulus presentation
r_0	Naka-Rushton function parameter: spontaneous firing rate
$r_{\max}$	Naka-Rushton function parameter: maximum increment from r_0
T	Number of trials
u_n	Parameter of the "approximate" probability distribution used in Appendix E
v	Parameter that appears in the expressions for the "approximate" probability distribution (Appendix E) and $\tilde{\tau}(x)$
X	Random variable representing the stimulus log Michelson contrast
x	The value of X on a particular stimulus presentation
$\hat{X}$	Random variable representing the estimated stimulus level after decoding the spike counts
$\hat{x}$	The value of $\hat{X}$ for a particular stimulus presentation
z	Naka-Rushton function parameter: log semi-saturation contrast
α	Weibull "threshold" parameter
β	Weibull "slope" or "shape" parameter
λ	Weibull "lapse rate" parameter

σ_G	Standard deviation of the gain signal in Goris et al.'s neuronal spiking model
$\tilde{\tau}(x)$	General estimate of the precision for decoding stimulus x, which can be parameterized to apply to the Poisson, Goris, Tolhurst, or Consul-Jain spiking processes
$\tau_{\text{Tolhurst}}(x)$	Analytical estimate of the decoding precision for the Tolhurst spiking process, which can be substantially more accurate than $\tilde{\tau}(x)$ when the mean spike count of the most informative neurons is very low
$\tau_{\text{C-J}}(x)$	Analytical estimate of the decoding precision for the Consul-Jain spiking process, which can be substantially more accurate than $\tilde{\tau}(x)$ when the mean spike count of the most informative neurons is very low
Ω	Recursive function used in the closed-form expression for the Tolhurst likelihood function

Appendix B: Finite series expansion of the Tolhurst likelihood function

The finite series expansion of the Tolhurst likelihood function is derived in Theorem 1, below. As a prelude to this theorem, we will find it useful to define a function, $\Omega_n(\xi)$, given by

$$\Omega_n(\xi) = \frac{1}{e^{\xi}} \sum_{\mu=0}^{\infty} \frac{\mu^n}{\mu!} \xi^{\mu} . \tag{B.1}$$

First, consider the case of $n = 0$. In this case, we have

$$\Omega_0(\xi) = \frac{1}{e^{\xi}} \sum_{\mu=0}^{\infty} \frac{1}{\mu!} \xi^{\mu} . \tag{B.2}$$

The summation in Equation (B.2) is the series expansion of e^{ξ}, and so we have

$$\Omega_0(\xi) = 1 . \tag{B.3}$$

Lemma 1 derives a recursive expression for $\Omega_n(\xi)$ for any integer $n > 0$.

Lemma 1. *For all integers* $n > 0$,

$$\Omega_n(\xi) = \xi \sum_{k=0}^{n-1} \binom{n-1}{k} \Omega_k(\xi), \text{ where } \binom{n-1}{k} \text{ is the kth coefficient of the binomial expansion of degree } n-1 .$$

Proof. For $n > 0$, the first term of the infinite series expansion of $\Omega_n(\xi)$ given by Equation (B.1) is 0, because this term includes multiplication by μ^n, which is 0 for the first term (since $\mu = 0$). Thus, we can start the series at $\mu = 1$:

$$\Omega_n(\xi) = \frac{1}{e^{\xi}} \sum_{\mu=1}^{\infty} \frac{\mu^n}{\mu!} \xi^{\mu} . \tag{B.4}$$

Now, since $\mu \neq 0$ in Equation (B.4), we can divide the top and bottom of each term in the series by μ to give

$$\Omega_n(\xi) = \frac{1}{e^{\xi}} \sum_{\mu=1}^{\infty} \frac{\mu^{n-1}}{(\mu-1)!} \xi^{\mu} . \tag{B.5}$$

Let us define ν as

$$\nu = \mu - 1 . \tag{B.6}$$

Then,

$$\Omega_n(\xi) = \frac{1}{e^{\xi}} \sum_{\nu=0}^{\infty} \frac{(\nu+1)^{n-1}}{\nu!} \xi^{\nu+1} \tag{B.7}$$

$$= \frac{\xi}{e^{\xi}} \sum_{\nu=0}^{\infty} \frac{(\nu+1)^{n-1}}{\nu!} \xi^{\nu} \tag{B.8}$$

Since $n > 0$, we can expand $(\nu+1)^{n-1}$ in Equation (B.8) using the binomial expansion:

$$\Omega_n(\xi) = \frac{\xi}{e^{\xi}} \sum_{\nu=0}^{\infty} \frac{\sum_{k=0}^{n-1} \binom{n-1}{k} \nu^k}{\nu!} \xi^{\nu} . \tag{B.9}$$

Applying the distributive rule [i.e. $a(b+c) = ab + ac$] to Equation (B.9), we get

$$\Omega_n(\xi) = \frac{\xi}{e^{\xi}} \sum_{\nu=0}^{\infty} \sum_{k=0}^{n-1} \frac{\binom{n-1}{k} \nu^k}{\nu!} \xi^{\nu} . \tag{B.10}$$

Rearranging the order of summation gives

$$\Omega_n(\xi) = \frac{\xi}{e^{\xi}} \sum_{k=0}^{n-1} \sum_{\nu=0}^{\infty} \frac{\binom{n-1}{k} \nu^k}{\nu!} \xi^{\nu} . \quad \text{(B.11)}$$

Applying the distributive rule again gives

$$\Omega_n(\xi) = \xi \sum_{k=0}^{n-1} \left[\binom{n-1}{k} \frac{1}{e^{\xi}} \sum_{\nu=0}^{\infty} \frac{\nu^k}{\nu!} \xi^{\nu} \right] \quad \text{(B.12)}$$

$$= \xi \sum_{k=0}^{n-1} \binom{n-1}{k} \Omega_k(\xi) . \square \quad \text{(B.13)}$$

The finite series expansion of each $\Omega_n(\xi)$ is a polynomial of degree n. As examples, Equations (B.14) to (B.19) give full, non-recursive expressions for $\Omega_n(\xi)$ for $n = 0$ to 5.

$$\Omega_0(\xi) = 1 \quad \text{(B.14)}$$

$$\Omega_1(\xi) = \xi \quad \text{(B.15)}$$

$$\Omega_2(\xi) = \xi^2 + \xi \quad \text{(B.16)}$$

$$\Omega_3(\xi) = \xi^3 + 3\xi^2 + \xi \quad \text{(B.17)}$$

$$\Omega_4(\xi) = \xi^4 + 6\xi^3 + 7\xi^2 + \xi \quad \text{(B.18)}$$

$$\Omega_5(\xi) = \xi^5 + 10\xi^4 + 25\xi^3 + 15\xi^2 + \xi . \quad \text{(B.19)}$$

The polynomial coefficients for any $\Omega_n(\xi)$ can be calculated using the MATLAB program, `OmegaCoefficients`, given in Appendix C.

Lemma 1 shows how to express the infinite series expansion in Equation (B.1) as a finite series, given in Equation (B.13). We now prove Theorem 1, which uses Lemma 1 to express the Tolhurst likelihood function as a finite series. To reduce notational clutter, we will use the single letter r, instead of $r(x)$, to represent the mean spike rate.

Theorem 1.

$$P_{\text{Tolhurst}}(N = n \mid R = r) = \frac{e^{r(1/e-1)}}{n!} \Omega_n(r/e) .$$

Proof. Using Equation (4) of the main paper to expand the expression for the Poisson distribution in Equation (5), we have

$$P_{\text{Tolhurst}}(N = n \mid R = r)$$

$$= \sum_{\mu=0}^{\infty} \frac{\mu^n}{n!} e^{-\mu} \frac{r^{\mu}}{\mu!} e^{-r} \quad \text{(B.20)}$$

$$= \frac{e^{-r}}{n!} \sum_{\mu=0}^{\infty} \frac{\mu^n}{\mu!} (r/e)^{\mu} \quad \text{(B.21)}$$

$$= \frac{e^{-r} e^{r/e}}{n!} \times \frac{1}{e^{r/e}} \sum_{\mu=0}^{\infty} \frac{\mu^n}{\mu!} (r/e)^{\mu} \quad \text{(B.22)}$$

$$= \frac{e^{r(1/e-1)}}{n!} \Omega_n(r/e) . \square \quad \text{(B.23)}$$

Equation (B.23) can be evaluated for any r and n using the MATLAB program, `PTolhurst`, given in Appendix D.

Appendix C: Software for calculating the coefficients of Ω_n

In this appendix, we give two MATLAB programs for calculating the coefficients of the Ω function, $\Omega_n(\xi)$. The programs both implement the recursive definition of $\Omega_n(\xi)$, given in Equation (B.13).

The first program, `OmegaCoefficientsRecursive`, uses a recursive algorithm that outputs the coefficients of $\Omega_n(\xi)$. The program is easy to follow because it is exactly analogous to the recursive definition of the Ω function. However, for large n, it is epically inefficient, as it keeps recalculating the coefficients of Ω functions of lower degree.

The second program, `OmegaCoefficients`, uses a non-recursive algorithm that calculates each set of coefficients once, and then stores them for future use. This program is less easy to follow, as it contains an outer loop, and is not exactly analogous to the recursive definition of the Ω function, but it is much more efficient. Because the program stores the coefficients of each Ω function from $\Omega_0(\xi)$ to $\Omega_n(\xi)$, it outputs the coefficients of all of these Omega functions, rather than just the coefficients of $\Omega_n(\xi)$.

```
function C = OmegaCoefficientsRecursive(n)
% Returns C, the vector of coefficients of the Omega function of degree n.
% C has n+1 terms, as the Omega function is a polynomial of degree n.
% n must be a scalar integer.

if n == 0
    % Omega(0, x) = 1 for all x, i.e. a polynomial of degree 0 with coefficient 1
    C = 1;
else
    % set up one coefficient for every term in the n-degree polynomial
    C = zeros(1,n+1);

    for k = 0:(n-1)

        % For each k in the summation in Equation (33), find the
        % coefficients of the Omega function of degree k.
        Ck = OmegaCoefficientsRecursive(k);

        % Then shift the coefficients along one position in the polynomial
        % (i.e. one power of x, because the omega function is multiplied by
        % x in Equation (33), so the coefficient of the mth term from the
        % right in in the old Omega function becomes the coefficient of the
        % (m+1)th term from the right in the new Omega function.
        % Also, pad out slots to the left to make n+1 coefficients.
        % Then multiply by the binomial coefficient, as in Equation (33),
        % and add to the total value of each coefficient.
        C = C + (factorial(n-1) / (factorial(k) * factorial(n-1-k))) * ...
                                                [zeros([1,n-1-k]), Ck, 0];

    end
end


function C = OmegaCoefficients(n)
% Returns C, the matrix of coefficients of Omega functions of degree 0 to n.
% Row m+1 of C gives the coefficients for the Omega function of degree m
% C has n+1 columns, as the highest-degree Omega function is a polynomial of
% degree n.
% n must be a scalar integer.

C = zeros(n+1,n+1); % n+1 Omega functions (rows) and n+1 coefficients (columns)

% Omega(0, x) = 1 for all x, i.e. a polynomial of degree 0, with coefficient 1
C(1,n+1) = 1;

% calculate coefficients of each omega function of degree 1 to n
for m = 1:n
    for k = 0:(m-1)

        % For each k in the summation in Equation (33), find the
        % coefficients of the Omega function of degree k
        Ck = C(k+1,:);

        % Then shift the coefficients along one position in the polynomial
        % (i.e. one power of x, because the omega function is multiplied by
        % x in Equation (33), so the coefficient of the mth term from the
        % right in in the old Omega function becomes the coefficient of the
        % (m+1)th term from the right in the new Omega function.
        % Then multiply by the binomial coefficient, as in Equation (33),
        % and add to the total value of each coefficient.
        C(m+1,:) = C(m+1,:) + ...
            (factorial(m-1) / (factorial(k) * factorial(m-1-k))) * ...
                                                    [Ck(2:end), 0];

    end
end
```

Appendix D: Software for calculating the Tolhurst likelihood function

In this appendix, we give two MATLAB programs for calculating the Tolhurst likelihood function, $P_{\text{Tolhurst}}(N = n \mid R = r)$. Each program calls a program that calculates the relevant Ω function and then calculates the Tolhurst likelihood function from this as in Equation (B.23) (strictly speaking these "programs" are MATLAB *functions*, but we use the term "program" here to avoid confusion with the mathematical functions that are being calculated). As with the programs for calculating the coefficients of the Ω function, we present recursive and non-recursive versions. Again, the recursive version is easier to follow, but too inefficient to be of practical use other than for small values of n. The recursive version, `PTolhurstRecursive(n,r)`, gives $P_{\text{Tolhurst}}(N = n \mid R = r)$ for a scalar integer n, and any size matrix of r values. The non-recursive version, `PTolhurst(n,r)`, calculates $P_{\text{Tolhurst}}(N = m \mid R = r)$ for all integers, m, between 0 and n, inclusive. n must be a scalar integer. r must be a real scalar or row vector. Row m+1 of the output gives $P_{\text{Tolhurst}}(N = m \mid R = r)$ for each element of the input argument, r.

```
function P = PTolhurstRecursive(n,r)
% Returns probability of n for the Tolhurst process, given a mean of r.
% n must be a scalar integer.
% r can be a real scalar or matrix of any size or number of dimensions.
% The output, P, has the same size and number of dimensions as r.
e = exp(1);
P = (exp(r*(1/e - 1))/factorial(n)) .* OmegaRecursive(n, r/e);
function y = OmegaRecursive(n,x)
if n == 0
    y = 1;
else
    y = 0;
    for k = 0:(n-1)
        y = y + (factorial(n-1) / (factorial(k) * factorial(n-1-k))) * OmegaRecursive(k,x);
    end
    y = x .* y;
end
```

```
function P = PTolhurst(n,r)
% Returns probability of 0 to n for the Tolhurst process, given a mean r.
% n must be a scalar integer.
% r can be a real scalar or row vector.
% The (m+1)th row of P gives P(N = m | R = r) for each element in the input argument, r.
e = exp(1);
P = exp(repmat(r,[n+1,1])*(1/e - 1)) ./ ...
            repmat(factorial([0:n]'),[1,length(r)]) .* Omega(n,r/e);
function y = Omega(n,x)
% Note, Omega(m,x) is stored in row m+1 or Omegas
Omegas = zeros(n+1,length(x));
Omegas(1,:) = ones(1,length(x));
% calculate Omega(m,:) for each m = 1:n
for m = 1:n
    k = [0:(m-1)]';
    % Calculate binomial coefficients
    C = (factorial(m-1) ./ (factorial(k) .* factorial(m-1-k)));
    C = repmat(C, [1,length(x)]);
    Omegas(m+1,:) = x .* sum(C .* Omegas([0:(m-1)]+1,:),1);
end
y = Omegas;
```

Appendix E: Approximations of the Fisher information for non-Poisson-spiking neurons

A derivation of an exact formula for the Fisher information of a neuron with a Tolhurst or Consul-Jain spiking process turned out to be intractable, so we closely approximated both of these distributions using a different expression, for which an analytical expression for the Fisher information could be derived. We use the term "approximate distribution" to refer to the expression that we used to approximate the Tolhurst and Consul-Jain distributions. The approximate distribution is given by

$$P_{\text{Approx}}(N=n \mid R=r(x))$$

$$= u_n P_{\text{Poisson}}\left(N=n/v \mid R=r(x)/v\right). \quad \text{(E.1)}$$

$$= u_n \frac{\left[r(x)/v\right]^{(n/v)}}{\Gamma(n/v+1)} \exp(-r(x)/v). \quad \text{(E.2)}$$

In Equation (E.2), we have to express the Poisson distribution using the continuous gamma function, $\Gamma(n/v+1)$, instead of the factorial, $(n/v)!$, so that it can be evaluated for non-integer values of n/v. This approximate distribution has a parameter v, and an infinite set of parameters, u_n, one for each possible value of n (the integers 0 to ∞). However, it turns out that the values of the u_n parameters have no effect on the Fisher information (this comes about because the Fisher information is the average negative 2nd derivative of the log-likelihood function – the log function converts multiplicative constants, u_n, into additive constants, which have no effect on the derivative). Thus, we can proceed as if P_{Approx} was parameterized only by v. The only purpose of fitting the u_n parameters is to verify that P_{Approx} is good approximation. Note that, for many values of v, it may be impossible to find a set of parameters, u_n, for which $\sum_{n=0}^{\infty} P_{\text{Approx}}(N=n \mid R=r(x))$ is exactly equal to 1 for all $r(x)$, and so P_{Approx} is not technically a probability distribution. However, it can provide a sufficiently close fit to the true Tolhurst and Consul-Jain distributions for it to generate a good approximation of the Fisher information.

Appendix F shows that, assuming P_{Approx} *is* a probability distribution generating a neuron's spikes, the Fisher information, J, is given by

$$J = \frac{1}{v} \times \frac{r'(x)^2}{r(x)}. \quad \text{(E.3)}$$

Fisher information for the Tolhurst process

Using Equation (B.14) of Appendix B to substitute for $\Omega_n(\cdot)$ in Equation (B.23) when $n=0$, we obtain

$$P_{\text{Tolhurst}}(N=0 \mid R=r(x))$$

$$= e^{(1/e-1)r(x)} \quad \text{(E.4)}$$

$$= P_{\text{Poisson}}(N=0 \mid R=(1-1/e)r(x)), \quad \text{(E.5)}$$

which has the form of P_{Approx} with $v=1/(1-1/e)$. For $n>0$, we cannot express the Tolhurst distribution exactly in the form of P_{Approx}, but we can closely approximate it for all n except $n=1$. For $n>1$, we have found that

$$P_{\text{Tolhurst}}(N=n \mid R=r(x)) \approx$$

$$u_n P_{\text{Poisson}}\left(N=n/2 \mid R=r(x)/2\right), \quad \text{(E.6)}$$

n	u_n
2	0.5007
3	0.5024
4	0.5018
5	0.5015
6	0.5014
7	0.5012
8	0.5011
9	0.5010
10	0.5009
11	0.5008

Table E.1. Parameters that minimize the error in the approximation of the Tolhurst distribution given by Relation (E.6). These parameters were fitted by using Equation (E.2) to evaluate the right hand side of Relation (E.6) over a range of $r(x)$ from 0 to 100 in steps of 0.01, and comparing with the true Tolhurst distribution (the left hand side of Relation (E.6), given by Equation (B.23)) for the same values of $r(x)$. We found the u_n parameters that minimized the sum of squared differences between the left and right sides of Relation (E.6). All decimal expansions are shown to an accuracy of 4 significant figures.

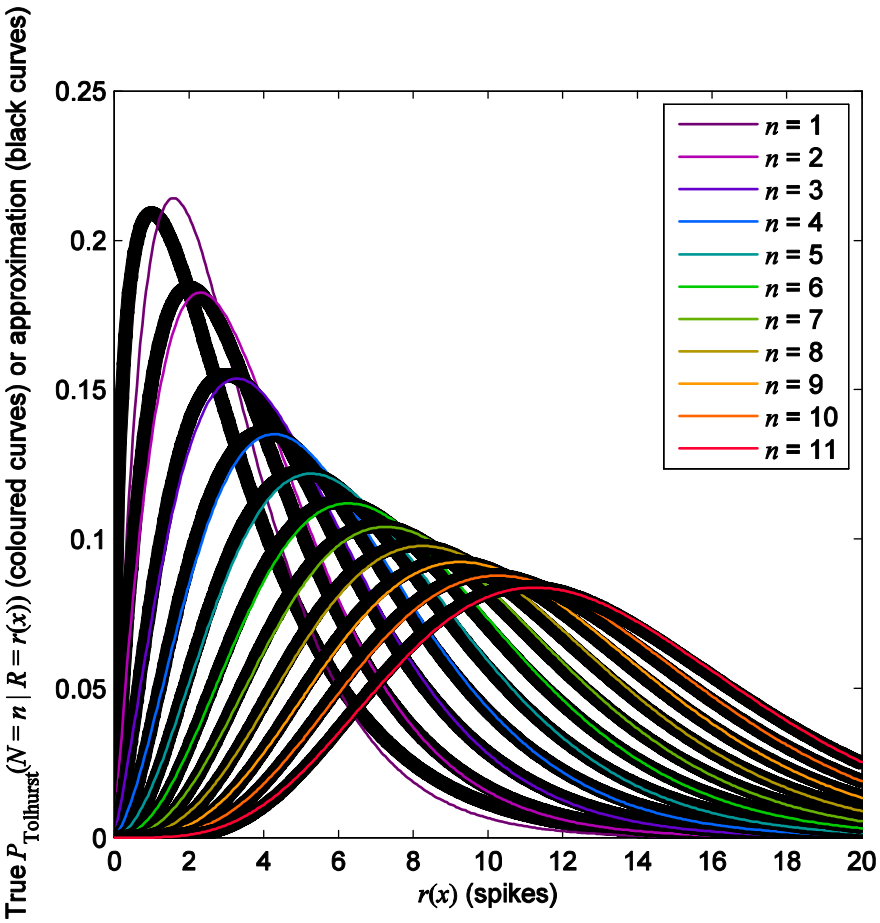

Figure E.1. The Tolhurst likelihood functions (thin, coloured curves), plotted against the best-fitting approximations (thick, black curves). $r(x)$ is the mean number of spikes elicited by stimulus x. For $n = 1$, the approximation is described by Relation (E.7). For $n > 1$, the approximation is described by Relation (E.6) with parameter values, u_n, given in Table E.1.

with best-fitting u_n values given in Table E.1 for $n = 2$ to 11. This has the form of P_{Approx} with $v = 2$. For $n = 1$, there is no really good approximation of the Tolhurst process that has the form of P_{Approx}; the best (least-squares) approximation is

$$P_{\text{Tolhurst}}(N = 1 \mid R = r(x)) \approx$$

$$0.4106 \times P_{\text{Poisson}}\left(N = 1/2.306 \mid R = r(x)/2.306\right), \quad \text{(E.7)}$$

which has the form of P_{Approx} with $v = 2.306$. Figure E.1 plots the true Tolhurst likelihood functions against the approximations given by Relations (E.6) and (E.7).

So, for a proportion $P(N = 0)$ of trials, the Tolhurst distribution is P_{Approx} with $v = 1/(1 - 1/e)$, for which the Fisher information (given by Equation (E.3)) is $(1 - 1/e)r'(x)^2 / r(x)$; for a proportion $P(N = 1)$ of trials, the Tolhurst distribution is reasonably well approximated by P_{Approx} with $v = 2.306$, for which the Fisher information is $r'(x)^2 / [2.306 r(x)]$; for the remaining proportion of trials, $[1 - P(N = 0) - P(N = 1)]$, the Tolhurst distribution is well approximated by P_{Approx} with $v = 2$, for which the Fisher information is $r'(x)^2 / [2r(x)]$. Since the Fisher information is a trial-averaged quantity (the average negative 2nd derivative of the log-likelihood function), we can calculate it separately for these three different scenarios ($n = 0$, $n = 1$ and $n > 1$), and then average them, weighted according to their probabilities, to give a close approximation of the Fisher information for a single Tolhurst-spiking neuron:

$$J \approx \left((1 - 1/e)P(N = 0) + \frac{P(N = 1)}{2.306} + \frac{1 - P(N = 0) - P(N = 1)}{2} \right) \frac{r'(x)^2}{r(x)}. \quad \text{(E.8)}$$

Equation (E.4) gives an expression for $P(N=0)$ for the Tolhurst distribution. Using Equation (B.15) of Appendix B to substitute for $\Omega_n(\cdot)$ in Equation (B.23) when $n=1$, we get an expression for $P(N=1)$:

$$P_{\text{Tolhurst}}(N=1 \mid R=r(x)) = r(x)e^{(1/e-1)r(x)-1}. \quad \text{(E.9)}$$

Using Equations (E.4) and (E.9) to substitute for $P(N=0)$ and $P(N=1)$ in Relation (E.8), we obtain

$$J \approx H_{\text{Tolhurst}}(x) \times \frac{r'(x)^2}{r(x)}, \quad \text{(E.10)}$$

where

$$H_{\text{Tolhurst}}(x) = \left(\frac{1}{2} - \frac{1+0.06630 \times r(x)}{e}\right) \times \exp[r(x)(1/e-1)] + \frac{1}{2}. \quad \text{(E.11)}$$

For a population of statistically independent neurons, the Fisher information is the sum of the Fisher information of the individual neurons. This gives us an accurate approximation of the decoding precision for a population of independent Tolhurst-spiking neurons:

$$\tau_{\text{Tolhurst}}(x) = \sum_{j=1}^{K} H_{\text{Tolhurst}}(x) \times \frac{r_j'(x)^2}{r_j(x)}. \quad \text{(E.12)}$$

$H_{\text{Tolhurst}}(x)$ is plotted in Figure 4 of the main paper. For $r(x)$ of around 5 spikes or more, $H_{\text{Tolhurst}}(x) \approx 0.5$, and so $\tau_{\text{Tolhurst}}(x)$ is close to $\tilde{\tau}(x)$ (defined in Equation (11) of the main paper) with $v=2$. Figure 8 of the main paper demonstrates that $\tau_{\text{Tolhurst}}(x)$ is substantially more accurate than $\tilde{\tau}(x)$ for very low spike rates.

Fisher information for the Consul-Jain process

In this subsection, we derive an approximation of the Fisher information for a Consul-Jain-spiking neuron using analogous methods to the previous subsection.

Letting $n=0$ in Equation (6) of the main paper, we obtain

$$P_{\text{C-J}}(N=0 \mid R=r(x)) = \exp\left(-r(x)/\sqrt{F}\right) \quad \text{(E.13)}$$

$$= P_{\text{Poisson}}\left(N=0/\sqrt{F} \mid R=r(x)/\sqrt{F}\right), \quad \text{(E.14)}$$

which has the form of P_{Approx} with $v=\sqrt{F}$. For $n>0$, we have found that, for Fano factors not too far above 1,

$$P_{\text{C-J}}(N=n \mid R=r(x)) \approx u_n P_{\text{Poisson}}\left(N=n/F \mid R=r(x)/F\right), \quad \text{(E.15)}$$

which has the form of P_{Approx} with $v=F$. Table E.2 gives the best-fitting u_n values for $n=1$ to 11 when $F=1.5$; Figure E.2 plots the true Consul-Jain likelihood functions for $n=1$ to 11, and their approximations given by Relation (E.15).

n	u_n
1	0.6948
2	0.6804
3	0.6757
4	0.6734
5	0.6720
6	0.6711
7	0.6704
8	0.6699
9	0.6696
10	0.6693
11	0.6690

Table E.2. Parameters that minimize the error in the approximation of the Consul-Jain likelihood function given by Relation (E.15) for $F=1.5$. These parameters were fitted by using Equation (E.2) to evaluate the right hand side of Relation (E.15) over a range of $r(x)$ from 0 to 100 in steps of 0.01, and comparing with the true Consul-Jain likelihood function (the left hand side of Relation (E.15), given by Equation (6) of the main paper) for the same values of $r(x)$. We found the u_n parameters that minimized the sum of squared differences between the left and right sides of Relation (E.15) when $F=1.5$. All decimal expansions are shown to an accuracy of 4 significant figures.

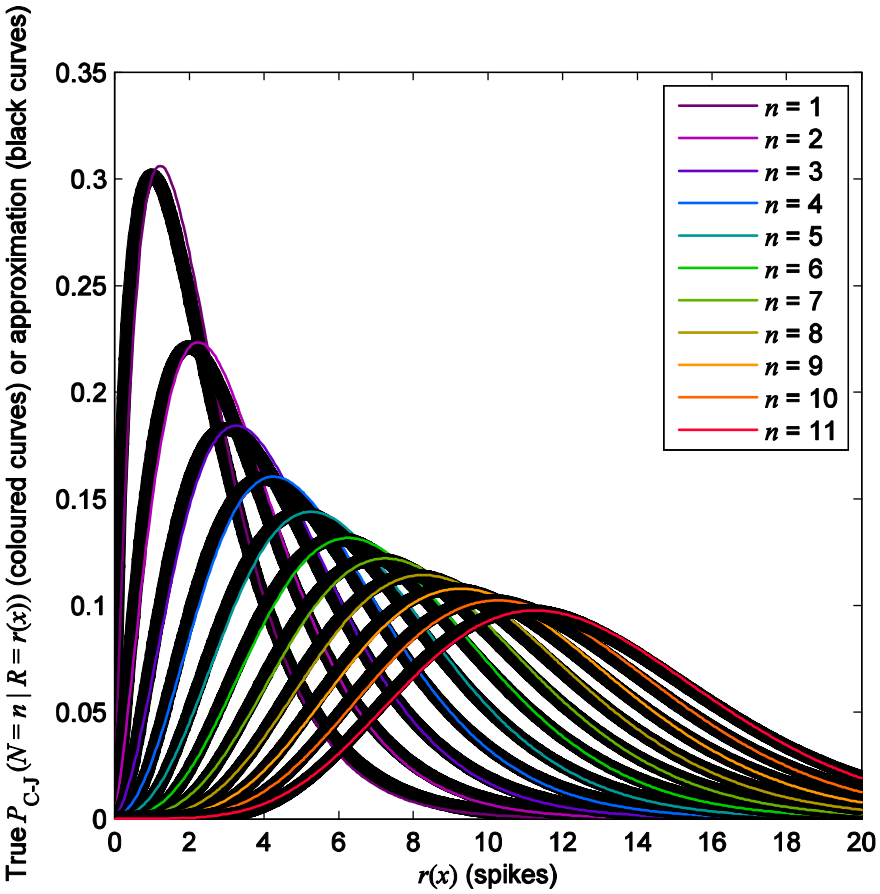

Figure E.2. Consul-Jain likelihood functions with $F = 1.5$, plotted against best-fitting approximations from Relation (E.15). $r(x)$ is the mean number of spikes elicited by stimulus x. The thin, coloured curves plot the true Consul-Jain likelihood functions (Equation (6)) with $F = 1.5$; the thick, black curves show the approximations given by the right hand side of Relation (E.15), with the u_n parameter values in Table E.2.

So, for a proportion $P(N = 0)$ of trials, the Consul-Jain distribution is P_{Approx} with $v = \sqrt{F}$, for which the Fisher information (given by Equation (E.3)) is $r'(x)^2 \big/ \left[\sqrt{F} r(x)\right]$; for the remaining proportion of trials, $\left[1 - P(N = 0)\right]$, the Consul-Jain distribution is well approximated by P_{Approx} with $v = F$, for which the Fisher information is $r'(x)^2 \big/ \left[F r(x)\right]$. Since the Fisher information is a trial-averaged quantity, we can calculate it separately for these two different scenarios ($N = 0$ and $N > 0$), and then average them, weighted according to their probabilities, to give a close approximation of the Fisher information for a single, Consul-Jain-spiking neuron:

$$J \approx \left(\frac{P(N = 0)}{\sqrt{F}} + \frac{1 - P(N = 0)}{F} \right) \frac{r'(x)^2}{r(x)}. \qquad \text{(E.16)}$$

Equation (E.13) shows that, for the Consul-Jain process,

$$P(N = 0) = \exp\left(-r(x) \big/ \sqrt{F} \right). \qquad \text{(E.17)}$$

Using Equation (E.17) to substitute for $P(N = 0)$ in Relation (E.16), we obtain

$$J \approx H_{\text{C-J}}(x) \times \frac{r'(x)^2}{r(x)} \qquad \text{(E.18)}$$

where

$$H_{\text{C-J}}(x) = \frac{\exp\left(-r(x) \big/ \sqrt{F} \right)}{\sqrt{F}} + \frac{1 - \exp\left(-r(x) \big/ \sqrt{F} \right)}{F}$$

$$= \exp\left(-\frac{r(x)}{\sqrt{F}} \right) \left(\frac{1}{\sqrt{F}} - \frac{1}{F} \right) + \frac{1}{F}. \qquad \text{(E.19)}$$

For a population of statistically independent neurons, the Fisher information is the sum of the Fisher information of the individual neurons. This gives us an accurate approximation of the decoding precision for a population of independent Consul-Jain-spiking neurons:

$$\tau_{\text{C-J}}(x) = \sum_{j=1}^{K} H_{\text{C-J}}(x) \times \frac{r_j'(x)^2}{r_j(x)} \quad . \qquad \text{(E.20)}$$

$H_{\text{C-J}}(x)$ is plotted in Figure E.3 for several different Fano factors, F. For sufficiently high $r(x)$, $H_{\text{C-J}}(x) \approx 1/F$, and so $\tau_{\text{C-J}}(x)$ is close to $\tilde{\tau}(x)$ (defined in Equation (11) of the main paper) with $v = F$. For Fano factors around the top of the physiologically plausible range (around 3), the mean spike count only needs to be about 7 spikes for $\tilde{\tau}(x)$ with $v = F$ to be within 1% of $\tau_{\text{C-J}}(x)$.

Figure E.4 plots the true Fisher information (thin, coloured lines) as a function of contrast for a single Consul-Jain-spiking neuron with $F = 1.5$ (a typical value for cortical neurons), and compares this against $\tilde{\tau}(x)$ with $v = F$ (thick, black lines). $\tilde{\tau}(x)$ provides an excellent approximation of the true Fisher information for moderate or high spike rates, but starts to get inaccurate when the spike rate gets very low. Figure E.5 plots the true Fisher information against $\tau_{\text{C-J}}(x)$ with the same Fano factor, $F = 1.5$. $\tau_{\text{C-J}}(x)$ provides an excellent match to the Fisher information at all spike rates.

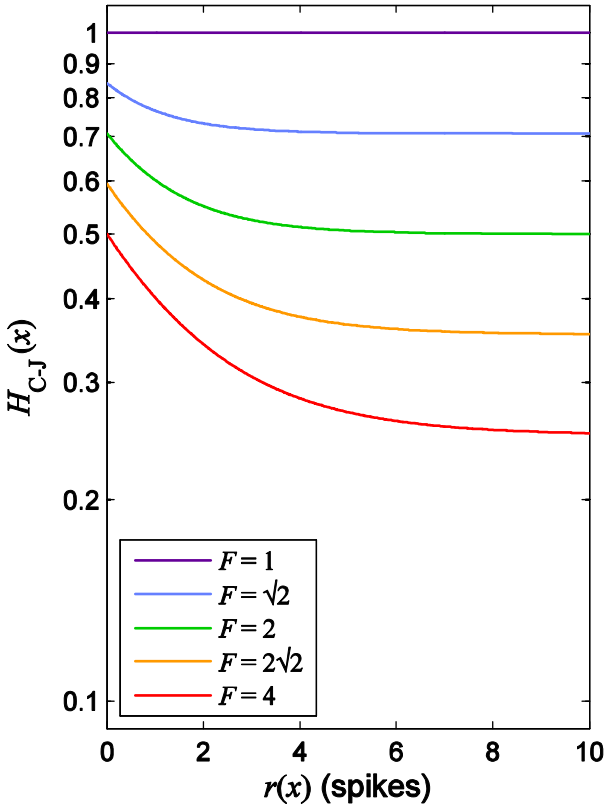

Figure E.3. Each different coloured line plots $H_{\text{C-J}}(x)$ as a function of $r(x)$ for a particular Fano factor, F, according to Equation (E.19). This function approaches $1/\sqrt{F}$ as $r(x) \to 0$, and approaches $1/F$ as $r(x) \to \infty$.

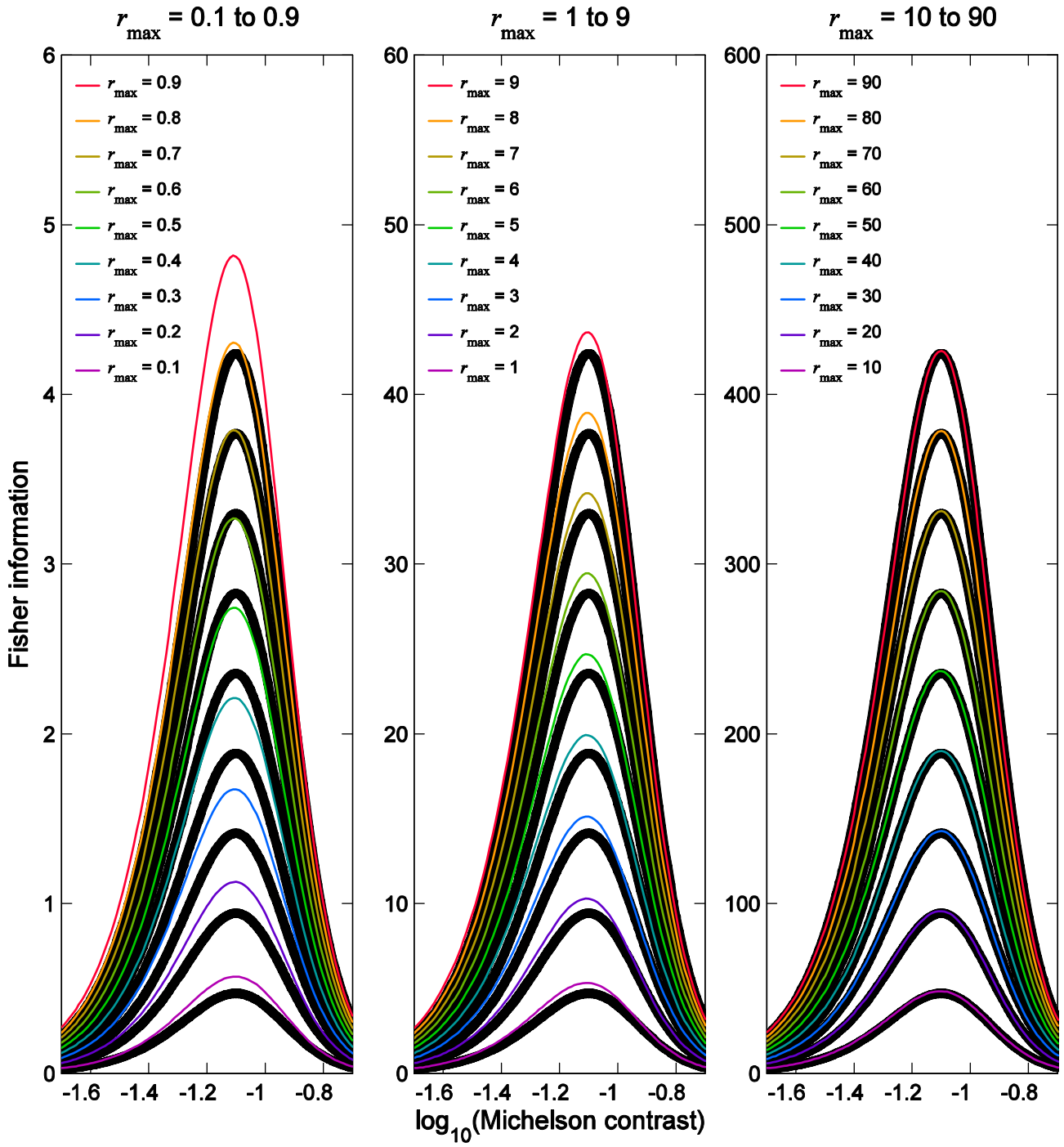

Figure E.4. True Fisher information of a Consul-Jain-spiking neuron with $F = 1.5$, and the closed-form approximation given by $\tilde{\tau}(x)$. The thin, coloured curves plot a single neuron's true Fisher information for decoding contrast, calculated numerically (see Supplementary Appendix H for methods). The neuron had a Consul-Jain spiking process and a Naka-Rushton tuning function with $z = -1$, $q = 3$, and $r_0 = 0$. The corresponding thick, black curves plot the approximations given by $\tilde{\tau}(x)$ with $v = 1.5$ (Equation (11) of the main paper).

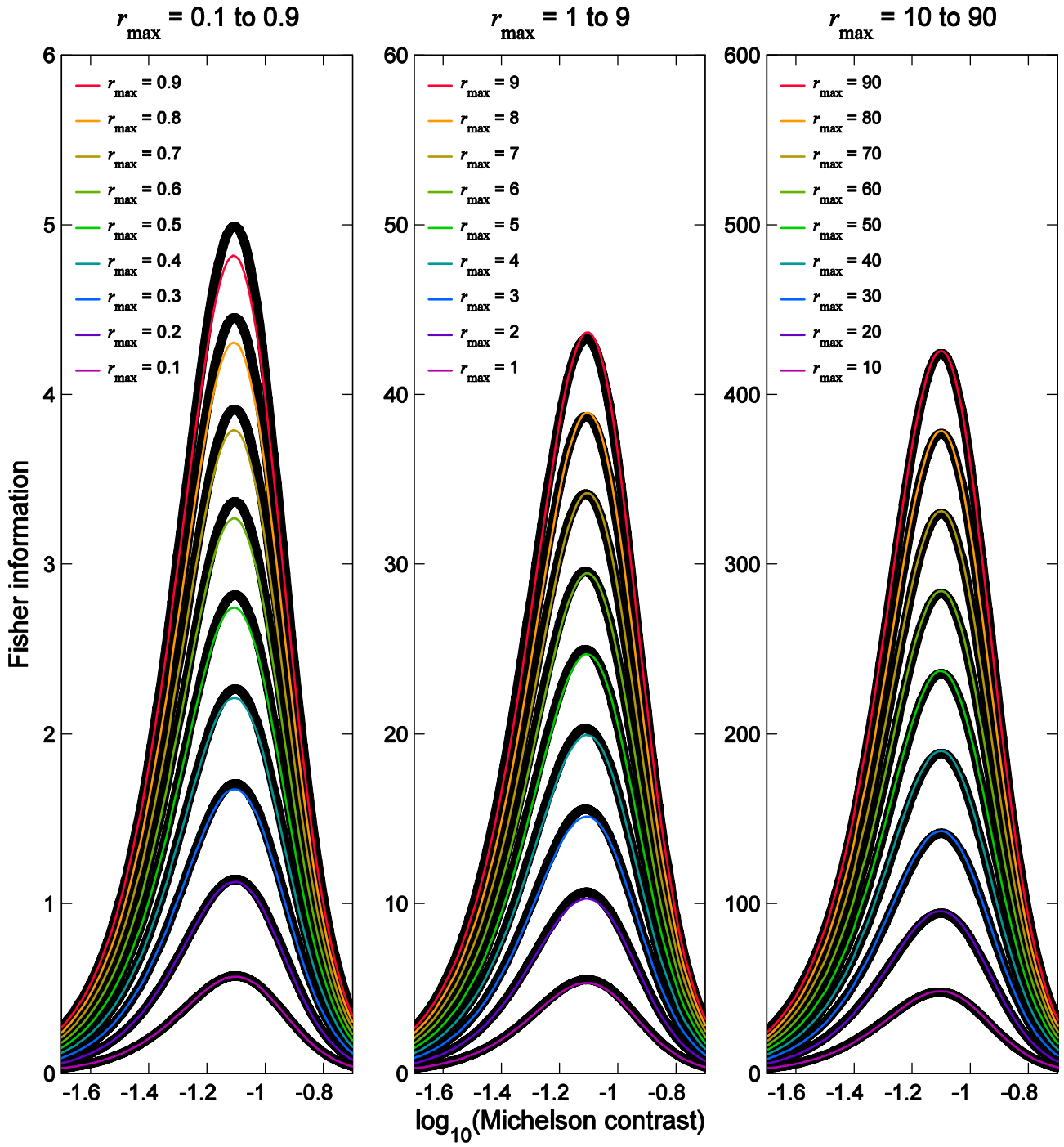

Figure E.5. The same as Figure E.4, except that the thick, black curves plot the approximations given by $\tau_{\text{C-J}}(x)$ (Equation (E.20)).

Appendix F: Fisher information for the approximate spiking distribution

The Fisher information, J, for a single neuron decoding a stimulus of value x is given by the average negative 2nd derivative of the log-likelihood function:

$$J = \left\langle -\frac{d^2 \ln P(N=n \mid X=x)}{dx^2} \right\rangle, \tag{F.1}$$

where $\langle y \rangle$ is the trial-averaged value of y. Since the mean spike rate, $r(x)$, is a deterministic function of x, we have

$$P(N=n \mid X=x) = P(N=n \mid R=r(x)). \tag{F.2}$$

Substituting Equation (F.2) into (F.1), we obtain

$$J = \left\langle -\frac{d^2 \ln P(N=n \mid R=r(x))}{dx^2} \right\rangle. \tag{F.3}$$

Suppose the neuron's true likelihood function was given by $P_{\text{Approx}}(N=n \mid R=r(x))$, defined in Equation (E.2). Then

$$\ln P_{\text{Approx}}(N=n \mid R=r(x))$$

$$= (n/v)\ln(r(x)) - r(x)/v +$$

$$(n/v)\ln(1/v) + \ln\left(\frac{u_n}{\Gamma((n/v)+1)}\right) \tag{F.4}$$

$$= \frac{1}{v}\left[n \ln(r(x)) - r(x)\right] + \text{terms independent of } x. \tag{F.5}$$

Note that the value of u_n is irrelevant to the Fisher information, because it gets absorbed into the "terms independent of x", which disappear when we differentiate.

Differentiating twice with respect to x, we get

$$-\frac{d^2 \ln P_{\text{Approx}}(N=n \mid R=r(x))}{dx^2} =$$

$$\frac{1}{v}\left[r''(x) - n\left(\frac{r''(x)}{r(x)} - \left(\frac{r'(x)}{r(x)}\right)^2\right)\right]. \tag{F.6}$$

The trial-averaged value of this gives us the Fisher information for this distribution:

$$J = \frac{1}{v}\left\langle r''(x) - n\left(\frac{r''(x)}{r(x)} - \left(\frac{r'(x)}{r(x)}\right)^2\right)\right\rangle \tag{F.7}$$

$$= \frac{1}{v}\left[r''(x) - \langle n \rangle\left(\frac{r''(x)}{r(x)} - \left(\frac{r'(x)}{r(x)}\right)^2\right)\right]. \tag{F.8}$$

The step from Equation (F.7) to (F.8) follows from the fact that $\langle an+b \rangle = a\langle n \rangle + b$ for constant a and b (note that, for a given x, the only term on the right hand side of Equation (F.7) that varies across trials is n). Since $r(x)$ is the mean spike count, we have

$$\langle n \rangle = r(x). \tag{F.9}$$

Using Equation (F.9) to substitute for $\langle n \rangle$ in Equation (F.8), we obtain

$$J = \frac{1}{v} \times \frac{r'(x)^2}{r(x)}. \tag{F.10}$$

Note that, for $v=1$ and $u_n = 1$ for all n, the "approximate" probability distribution is exactly the Poisson, so an exact expression for the Fisher information of the Poisson distribution is given by

$$J = \frac{r'(x)^2}{r(x)} \tag{F.11}$$

(see also Dayan & Abbott, 2001, Chapter 3).

Appendix G: Estimating decoding precision

Our estimates of decoding precision were obtained using methods essentially the same as those of Clatworthy et al. (2003) and Chirimuuta et al. (2003). For each neuron, j, in the population, we precalculated $P(N_j = n_j \mid X = x)$ for a large range of stimulus values, x, and spike counts, n_j. For each n_j, x was varied in discrete steps of 0.01 from −3 to 0.1 (to be compatible with Clatworthy et al.). For each x, n_j took all integer values from 0 to a number beyond which the maximum of $P(N_j = n_j \mid X = x)$ over all x was negligible. The precalculated likelihood func-

tions were used both to generate the spikes and to decode them.

In each simulation, the precalculated values of $P(N_j = n_j \mid X = x)$ for each neuron, j, were given by

$$P(N_j = n_j \mid X = x) = P_{\text{Tolhurst}}(N_j = n_j \mid R = r(x)). \quad \text{(G.1)}$$

We used these values to precalculate the cumulative probability distributions as well. On each trial with stimulus level x, each possible value of n_j had a probability of occurrence given by Equation (G.1). This was achieved by generating a random number sampled from a flat distribution between 0 and 1, and then finding the lowest n_j for which the cumulative probability distribution of n_j was greater than or equal to the random number.

After generating the spike counts, we used the precalculated likelihood functions to decode them. On each trial, we needed to find the stimulus level, x, that maximized the stimulus likelihood, $P(\mathbf{N} = \mathbf{n} \mid X = x)$, where $\mathbf{N}$ is a random variable representing the population response, and $\mathbf{n}$ is its value. For statistically independent neurons, $P(\mathbf{N} = \mathbf{n} \mid X = x)$ is the product of $P(N_j = n_j \mid X = x)$ over the population (see Equation (8) of the main paper). For large populations, this product can be too small to represent using floating point values on a standard computer, so instead we maximized $\ln P(\mathbf{N} = \mathbf{n} \mid X = x)$ (which increases monotonically with $P(\mathbf{N} = \mathbf{n} \mid X = x)$, and therefore peaks at the same value of x). $\ln P(\mathbf{N} = \mathbf{n} \mid X = x)$ is given by $\sum_j \ln P(N_j = n_j \mid X = x)$. The stimulus estimate, $\hat{x}$, was the value of x that maximized the likelihood. we repeated this process 10,000 times for each stimulus value, x, and the precision was calculated using Equation (10) of the main paper.

Appendix H: Numerical calculation of true Fisher information

For a single neuron, the general expression for the Fisher information, J, can be written as

$$J = \left\langle \left(\frac{d \ln P(N = n \mid X = x)}{dx} \right)^2 \right\rangle \quad \text{(H.1)}$$

where $\langle y \rangle$ is the trial-averaged value of y (see Dayan & Abbott (2001), p. 109). Since the mean spike rate is a deterministic function of x, we have $P(N = n \mid X = x) = P(N = n \mid R = r(x))$. The expression for $P(N = n \mid R = r(x))$ depends on the spiking process and the tuning function, $r(x)$. To numerically calculate the Fisher information for a single neuron, we first calculated $\ln P(N = n \mid R = r(x))$ for all x from –3 to 0.1 in steps of $\delta x = 0.001$, and for all integers, n, from 0 to M, where M was high enough for the value of $\ln P(N = n \mid R = r(x))$ to be negligible. For each n, we then numerically differentiated $\ln P(N = n \mid R = r(x))$ with respect to x, to give, for each x and n, a close approximation of the derivative, $d \ln P(N = n \mid R = r(x))/dx$. This approximation is given by

$$\frac{\ln P(N = n \mid R = r(x + \delta x)) - \ln P(N = n \mid R = r(x - \delta x))}{2\delta x}. \quad \text{(H.2)}$$

The numerical value, $J^{\text{numerical}}(x)$, of the Fisher information was then calculated from this numerical derivative using a discrete approximation of Equation (H.1):

$$J^{\text{numerical}}(x) = \sum_{n=0}^{M} P(N = n \mid R = r(x)) \times \left(\frac{\ln P(N = n \mid R = r(x + \delta x)) - \ln P(N = n \mid R = r(x - \delta x))}{2\delta x} \right)^2. \quad \text{(H.3)}$$

The weighted sum in Equation (H.3) approximates the mean value of the squared derivative across trials.

Appendix I: Generating semi-saturation contrasts for cat and monkey populations of neurons

Our distributions of cat and monkey z values (i.e. $\log_{10}(c_{1/2})$ values) were based on the histograms in Clatworthy et al.'s (2003) Figure 6, which show the distributions of semi-saturation contrasts obtained from fitting Naka-Rushton functions to the contrast-response functions of V1 neurons in many different physiological experiments. These histograms show the number of neurons falling into each bin of width $\delta z = 0.2$ $\log_{10}$ contrast units. Clatworthy et al.'s Figure 6 plots the number of neurons in each bin as a function of the z-value corresponding to the centre of the bin, for cat and monkey populations, and these distributions are re-plotted as filled

circles in our Figure I.1. We fitted smooth probability density functions (PDFs) to these distributions, using a maximum-likelihood fit.

For the cat data, the fitted PDF, $f_{\text{cat}}(z)$ was a single Gaussian, with two parameters, the mean, μ, and the standard deviation, σ.

$$f_{\text{cat}}(z) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(z-\mu)^2}{2\sigma^2} \right). \qquad \text{(I.1)}$$

Assuming $f_{\text{cat}}(z)$ is the true PDF, and the histogram bin width, δz, is small, the probability, $P(z)$, of a neuron falling in a bin centred on z is closely approximated by $P(z) = f_{\text{cat}}(z)\delta z$. The log likelihood of the parameters (μ, σ) is the log probability of the data given the parameters, i.e. $\sum_z m(z) \ln P(z)$, where summation is over the bin centres, z, and $m(z)$ is the number of neurons in the bin centred on z. We found the parameters of the PDF with the highest log likelihood, and these were given by $\mu = -0.9928$, and $\sigma = 0.3833$. This PDF is plotted as the smooth curve in our Figure I.1A, scaled for the number of neurons in the population.

The Monkey data in Clatworthy et al.'s Figure 6 showed two peaks, and so the distribution was fitted with a PDF, $f_{\text{monkey}}(z)$, formed from the sum of two Gaussians:

$$f_{\text{monkey}}(z) = \frac{f_1(z) + f_2(z)}{\int_{-\infty}^{\infty} f_1(z) + f_2(z) dz}, \qquad \text{(I.2)}$$

where

$$f_1(z) = A \exp\left(-\frac{(z-\mu_1)^2}{2{\sigma_1}^2} \right), \qquad \text{(I.3)}$$

$$f_2(z) = \exp\left(-\frac{(z-\mu_2)^2}{2{\sigma_2}^2} \right), \qquad \text{(I.4)}$$

and

$$\begin{aligned} \int_{-\infty}^{\infty} f_1(z) + f_2(z) dz &= \int_{-\infty}^{\infty} f_1(z) dz + \int_{-\infty}^{\infty} f_2(z) dz \\ &= A\sigma_1\sqrt{2\pi} + \sigma_2\sqrt{2\pi} \\ &= (A\sigma_1 + \sigma_2)\sqrt{2\pi}. \end{aligned} \qquad \text{(I.5)}$$

The monkey PDF was fitted to Clatworthy et al.'s data using the same maximum-likelihood method as for the cat data, except that the monkey PDF had five parameters, rather than two. The fitted values were $A = 3.643$, $\mu_1 = -0.7247$, $\sigma_1 = 0.3985$, $\mu_2 = 0.6747$, and $\sigma_2 = 0.1927$. The scaled PDF is plotted as the smooth curve in our Figure I.1B.

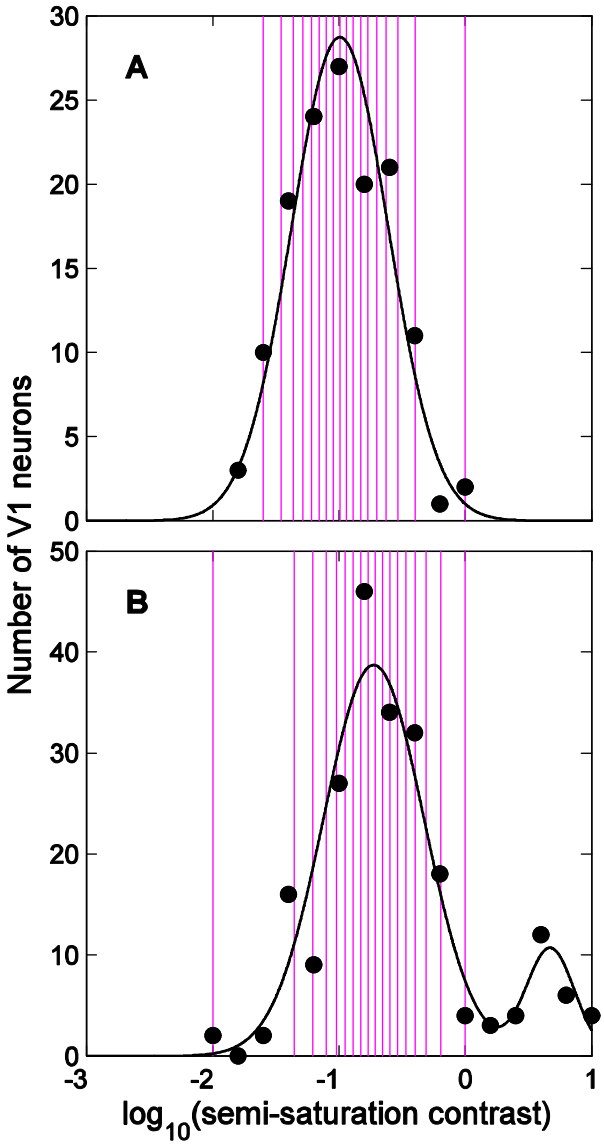

Figure I.1. Distributions of semi-saturation contrasts in V1. (A) Cat data. (B) Monkey data. Filled circles show frequency data from Clatworthy et al.'s (2003) Figure 6. Smooth curves show the PDFs that we fitted to these data, vertically scaled for the size of the population. The vertical scaling was performed by taking the PDF, and then multiplying its height by $M \times \delta z$, where δz is the width of each histogram bin (0.2), and M is the number of neurons in Clatworthy et al.'s histogram (cats: M = 138 neurons; monkey: M = 219 neurons). This scaling meant that the function gave the expected number of neurons in each bin. Magenta vertical lines indicate the 18 semi-saturation contrasts used for each animal in Figure 7 of the main paper.

The smooth curves were truncated at $z = 0$ at the top end, to be compatible with Clatworthy et al, who excluded all neurons with $c_{1/2} > 1$ (i.e. $z > 0$). At the bottom end, the monkey curve was truncated at $z = -2$, and the cat curve was truncated at $z = -1.6$. The lower limit for the monkey data corresponded to the centre of the leftmost nonzero histogram bin in Clatworthy et al.'s Figure 6B. The lower limit for the cat data corresponded to the centre of the second-to-left nonzero histogram bin in Clatworthy et al.'s cat data; the reason for this choice for the cat data was that it gave a better fit to Clatworthy et al.'s accuracy scores than setting it to the centre of the leftmost bin, although both gave a good fit. The amplitudes of these truncated PDFs were rescaled so that they integrated to 1 (making them true PDFs), and then we generated corresponding cumulative distribution functions that varied between 0 and 1. K z-values were then obtained by reading off the z-values corresponding to cumulative probabilities evenly spaced from 0 to 1, in steps of $1/(K-1)$. These z-values are shown as vertical lines in Figure I.1 for the example of $K = 18$, as used in Figure 7 of the main paper.

Appendix J: Simulation methods for 2AFC contrast detection tasks

The model was set up the same as for estimating decoding precision (Appendix G) except that the set of contrasts over which the likelihood functions were precalculated included zero Michelson contrast (i.e., $x = -\infty$). For Naka-Rushton exponent $q = 1$, the likelihood functions were precalculated over values of $x = -\infty$ and $x = -7$ to 0 in steps of 0.01, and the target log contrast ranged from −7 to 0 in steps of 0.05. For q = 2 to 5, the likelihood functions were precalculated over values of $x = -\infty$ and $x = -5$ to 0 in steps of 0.01, and the target log contrast ranged from −5 to 0 in steps of 0.05.

q took values of 1, 2, 3, 4, or 5; $r_{\max}$ took values of 1, 2, 4, 8, or 16; the number of neurons, K, took values of 1, 2, 4, 8, 16, 32, 64, 128, 256, or 512. We simulated a 2AFC detection task with each combination of these parameters. For each combination, all the neurons had identical contrast-response functions, with $c_{1/2} = 0.025$.

On each 2AFC trial, we generated spikes for the given target contrast as described in Appendix G. The zero-contrast stimulus always gave zero spikes, because $r_0 = 0$ in these stimulations. The model responded correctly on 2AFC trials on which the target elicited at least one spike; for each contrast level, we counted up these 2AFC trials, and then added half the remaining 2AFC trials, on which the model would have had to guess (with 0.5 probability of guessing correctly). This gave the model's total number of "correct" responses, which we divided by the total number of 2AFC trials (10,000) to give the proportion correct. For each parameterization of the model, a 3-parameter Weibull function (Equation (32) of the main paper) was fitted to the proportion of correct responses as a function of target Michelson contrast.

Appendix K: The additive noise distribution that gives a Weibull psychometric function

Here, we prove that, assuming that the response of each detector, j, is a linear function of contrast plus a sample of additive, statistically independent noise, and the observer detects the stimulus if at least one detector responds above its sensory threshold, θ_j, then the observer's psychometric function will be a Weibull function with slope β when the noise on each detector has a cumulative distribution function (CDF), F, given by

$$F_{E_j}(\varepsilon) = P(E_j \le \varepsilon) = \begin{cases} 1 & \varepsilon > \theta_j \\ \exp\left[-(\theta_j - \varepsilon)^{\beta}\right] & \varepsilon \le \theta_j \end{cases}. \quad \text{(K.1)}$$

The term E_j in Equation (K.1) is a random variable representing the noise added to detector j. ε is a specific instance of E_j. By assumption, the response, η_j, of each detector, j, is given by

$$\eta_j = c/\alpha_j + E_j \quad \text{(K.2)}$$

where α_j is the reciprocal of that detector's sensitivity to the stimulus. Also, by assumption, the probability that the observer detects the stimulus is the probability that not all the responses fall below threshold:

$$P(\text{detection}) = 1 - P\left(\forall j \ \eta_j \le \theta_j\right) \quad \text{(K.3)}$$

$$= 1 - P\left(\forall j \ c/\alpha_j + E_j \le \theta_j\right) \quad \text{(K.4)}$$

$$= 1 - P\left(\forall j \ E_j \le \theta_j - c/\alpha_j\right) \quad \text{(K.5)}$$

$$= 1 - \prod_{j=1}^{K} P\left(E_j \le \theta_j - c/\alpha_j\right) \quad \text{(K.6)}$$

$$= 1 - \prod_{j=1}^{K} F_{E_j}\left(\theta_j - c/\alpha_j\right), \qquad \text{(K.7)}$$

where K is the number of detectors.

Now, let us assume that the CDF of the noise, F_{E_j}, in Equation (K.7) is given by Equation (K.1). Since both contrast, c, and sensitivity, $1/\alpha_j$, are nonnegative values, we have $c/\alpha_j \geq 0$, and therefore $\theta_j - c/\alpha_j \leq \theta_j$. Substituting $\theta_j - c/\alpha_j$ for ε in Equation (K.1), we obtain

$$F_{E_j}\left(\theta_j - c/\alpha_j\right) = \exp\left[-\left(c/\alpha_j\right)^{\beta}\right]. \qquad \text{(K.8)}$$

Using Equation (K.8) to substitute for $F_{E_j}\left(\theta_j - c/\alpha_j\right)$ in Equation (K.7) gives

$$P(\text{detection}) = 1 - \prod_{j=1}^{K} \exp\left[-\left(c/\alpha_j\right)^{\beta}\right] \qquad \text{(K.9)}$$

$$= 1 - \exp\left[-\left(c/\alpha\right)^{\beta}\right] \qquad \text{(K.10)}$$

where

$$\alpha = \left(\sum_{j=1}^{K} \alpha_j^{-\beta}\right)^{-1/\beta}. \qquad \text{(K.11)}$$

Thus, the psychometric function is a Weibull function with threshold, α, determined from the sensitivities of the individual detectors using exactly the same equation as that of Quick. For application to 2AFC experiments, Equation (K.10) must be corrected for guessing, which yields Equation (22) of the main paper. □

Note that, since sensitivity is nonnegative, the contrast threshold given by Equation (K.11) is always above zero, and so Tyler and Chen's (2000) argument that "high-threshold probability summation fails for additive noise" does not apply when the psychometric function has the form of a Weibull function.

The noise probability density function (PDF) corresponding to the CDF in Equation (K.1) can be found by differentiating Equation (K.1) with respect to ε:

$$f_{E_j}(\varepsilon) = \beta(\theta_j - \varepsilon)^{\beta-1} \exp\left[-(\theta_j - \varepsilon)^{\beta}\right]. \qquad \text{(K.12)}$$

This function is plotted in Figure K.1 for $\beta = 1.3$, 2, 4, and 8, with ε expressed in units such that $\theta_j = 1$. These are the same functions as plotted in Figure 2b of Tyler and Chen.

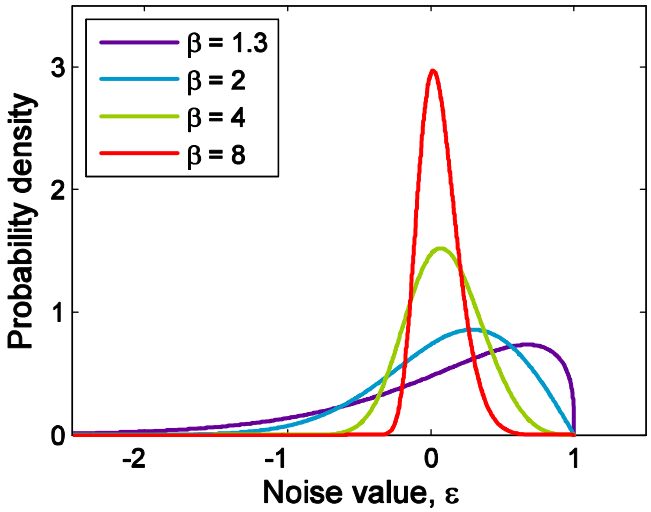

Figure K.1. The Weibull noise PDF, as defined in Equation (K.12) for four values of β. ε is the noise on detector j, which is expressed in units such that the detector's sensory threshold, θ_j, is equal to 1.