On the origin of sensory errors: Contrast discrimination under temporal constraint

Jonathan R. Flynn

Harel Z. Shouval

Estimation of perceptual variables is imprecise and prone to errors. Although the properties of these perceptual errors are well characterized, the physiological basis for these errors is unknown. One previously proposed explanation for these errors is the trial-by-trial variability of the responses of sensory neurons that encode the percept. In order to test this hypothesis, we developed a mathematical formalism that allows us to find the statistical characteristics of the physiological system responsible for perceptual errors, as well as the time scale over which the visual information is integrated. Crucially, these characteristics can be estimated solely from a behavioral experiment performed here. We demonstrate that the physiological basis of perceptual error has a constant level of noise (i.e., independent of stimulus intensity and duration). By comparing these results to previous physiological measurements, we show that perceptual errors cannot be due to the variability during the encoding stage. We also find that the time window over which perceptual evidence is integrated lasts no more than \( \sim 230 \) ms. Finally, we discuss sources of error that may be consistent with our behavioral measurements.

Introduction

In order to interact with the world, all animals must estimate the magnitude of external stimuli; it is well established that this process is prone to errors (Coren, Ward, & Enns, 2003; Dayan & Abbott, 2005; Kandel, Schwartz, Jessell, Siegelbaum, & Hudspeth, 2012). These errors form the foundation of the science of psychophysics, and have been studied as far back as the 19th century (Weber, 1834). However, while the statistical properties of these estimation errors have been rigorously studied in multiple sensory systems (Coren et al., 2003; Kandel et al., 2012), their physiological source is unknown.

One possible physiological source of perceptual errors is the innate variability of sensory encoding neurons. When the same sensory percept is presented repeatedly, the number of action potentials generated by a sensory neuron varies from trial-to-trial (Stein, Gossen, & Jones, 2005). The statistical properties of this variability have been studied extensively (Churchland et al., 2010; Dean, 1981; Mainen & Sejnowski, 1995). Theoretical examinations have previously assumed that such variability in sensory neuron responses contribute to behavioral variability (Goris, Wichmann, & Henning, 2009; May & Solomon, 2015a, 2015b; Mazurek, Roitman, Ditterich, & Shadlen, 2003; Shouval, Agarwal, & Gavornik, 2013). Moreover, several experiments that have related the trial-by-trial variability of sensory neurons to behavior (Britten, Shadlen, Newsome, & Movshon, 1993; Britten et al., 1996; Cohen & Newsome, 2009; Shadlen, Britten, Newsome, & Movshon, 1996) seem to bolster the hypothesis that it is the origin of behavioral variability, though alternative interpretations are possible (Pelli, 1985).

Initially, it would seem that testing the above hypothesis might require very advanced experimental techniques in which a large number of neurons can be both imaged and manipulated. However, we find that we can gain significant insight by using a theoretically motivated psychophysical experiment. The lower limit of behavioral noise depends on the statistics of the encoding neurons via the inverse of the Fisher information (Paradiso, 1988; Seung & Sompolinsky, 1993). Using this measure, we find that we can infer key statistical properties of the neural noise that generates perceptual errors simply by altering the viewing time of the stimuli (see mathematical analysis below). Our
results indicate that the properties of perceptual errors are inconsistent with a noise source arising from spike count variability in the perceptual neurons. These surprising results contradict our original hypothesis and indicate that errors in the encoding stage are not the primary source of perceptual errors.

The current experiment uses a contrast discrimination paradigm to determine how perceptual errors increase as we decrease the time the stimulus is displayed. The psychophysical literature has seen seemingly similar experiments (Bloch, 1885; Gorea, 2015; Legge, 1978; Watson, 1979), but there are key distinctions between those experiments and ours. The famous psychophysicist Bloch conducted one of the first experiments that controlled stimulus duration (Bloch, 1885; Gorea, 2015). His results showed that, for durations shorter than 50 ms, subjects confuse duration and brightness such that a short duration, high brightness stimulus is perceptually indistinguishable from a longer duration, lower brightness stimulus. The physiological explanation for Bloch’s law arises from the properties of retinal neurons, and is valid only for limited durations. In contrast, here we use stimuli with durations longer than 50 ms, for which subjects can estimate both magnitude and duration separately. Other previous experiments have examined the effect of contrast stimulus duration over intervals longer than 50 ms (Legge, 1978; Watson, 1979). However, these experiments were not based on the theoretical foundation developed here, and used different experimental techniques that yielded results not ideally suited to testing our theory’s predictions. Nevertheless, when interpreted with our theoretical framework, their results are consistent with our observations and are discussed in the Results section of this paper.

Methods

Mathematical analysis

Our experiment is based on the intuitive notion that, by changing the time window in which a subject perceives a stimulus, we can change their perceptual precision. By measuring this change, we can determine key statistical properties of the noise that drives perceptual errors. Models that relate behavioral errors to the physiological source of the error have been proposed previously (Drugowitsch, Wyart, Devachelle, & Koechlin, 2016; Graham, 1989). The theoretical framework for our experiment requires the three following assumptions: (a) Perceptual judgment is obtained based on the number of spikes in a time window \( \tau \) (Figure 1A); (b) The spike-count tuning curve, \( R(\theta, \tau) \), is monotonic with respect to both \( \theta \) and \( \tau \), where \( \theta \) is the perceptual variable; note that this tuning curve could describe either single neurons or a combined variable due to a population of neurons; (c) The variability in the spike count (\( \sigma \)) can be approximated by \( \sigma = \beta \cdot R(\theta, \tau)^\rho \), where \( \beta \) and \( \rho \) are empirically determined constants (Dean, 1981). For example, a value of \( \rho = 0.5 \) would represent a Poisson-like relationship between firing rate and noise, whereas a value of \( \rho = 0 \) would represent a constant noise level, independent of the firing rate.

Using the above assumption, it can be shown that the ability to distinguish two stimuli of equal durations depends on the spike count distributions arising from these stimuli. The overlap between the distributions (Figure 1B, magenta regions) determines the probability that the subject will mistakenly believe that the lower magnitude stimulus is actually the higher one. This overlap is related to the traditional Just-Noticeable-Difference (JND) psychometric function (Figure 1C; see Klein, 2001, for background). As \( \tau \) is reduced, the spike count is reduced, the distributions are altered (Figure 1B), and the overlap region (and thus JND) increases. The increase in the overlap depends on the change in the mean spike count and on the variance of the spike count. Therefore, for different statistical models of noise, the change in overlap will be quantifiably different. In the example shown in Figure 1, we observe that decreasing \( \tau \) has a more deleterious effect on discrimination ability in the \( \rho = 0 \) case than in the \( \rho = 0.5 \) case (Figure 1C). The relationship between \( \tau \) and the JND can be mathematically formalized with the following derivations.

Our analysis assumes that we are comparing spike counts in time bins of equal duration, and this is the way our experiments are carried out as well. Additionally, the validity of our analysis is restricted to temporal windows that are large enough such that subjects do not confuse duration and brightness—in other words, durations must be outside the range of Bloch’s law, \( \sim 50 \) ms (Bloch, 1885; Gorea, 2015). For temporal windows that adhere to these assumptions, one can use spike rate and spike count interchangeably. We use spike counts here to simplify some of the analysis, but to deal with stimuli that have different durations, one would have to reformulate some of the equations in terms of spike rates. This, and thoughts on additional errors this might generate, are described in the “Mathematical derivations” section of the Appendix. Note also that our third assumption, which assumes a power-law form for the firing rate variability, is used because it approximates much of the data (Churchland et al., 2010; Dean, 1981), and because it is convenient in that it produces analytical predictions here. However, it could be replaced by other functional forms if necessary without qualitatively affecting the results.
For the sake of clarity, we will also temporarily incorporate two additional assumptions. First, for any given stimuli, there is a time \( t_{\text{sat}} \) such that for observation times \( t > t_{\text{sat}} \), performance at sensory discrimination will not improve. This assumption is motivated by experimental data, both ours and others (G. E. Legge, 1978; Watson, 1979), which exhibit performance saturation as a robust phenomenon. However, were our data sets not to include time points beyond \( t_{\text{sat}} \), our current theory would still be able to determine \( q \).

Second, we will assume that the firing rate is constant and does not depend on time. While this is clearly a gross simplification (Albrecht, Geisler, Frazor, & Crane, 2002; Heller, Hertz, Kjaer, & Richmond, 1995), making this assumption allows for easier demonstration of the equations that we wish to discuss. In the Appendix, we show that our conclusions can be generalized to include biologically realistic firing patterns.

From these assumptions, we can derive a relationship between the variability in stimulus estimation and the spiking statistics. We start by relating the behavioral estimation error to the statistics of the encoding signal, using the equation:

\[
\sigma_{\theta}^2 = \frac{\sqrt{\sigma_E^2 + \sigma_D^2}}{(R')^2} \quad (1)
\]

where \( \sigma_{\theta} \) is the standard deviation of the stimulus estimation, \( \sigma_E \) is the standard deviation of the encoding variable \( R(\theta, \tau) \), \( \sigma_D \) is the standard deviation of the decoding variable, and \( R' \) is the derivative of \( R \) with respect to \( \theta \). Note that \( R(\theta, \tau) \) is a tuning curve in terms of spike count, and depends both on the encoded parameters \( \theta \), and on the duration of the stimulus presentation \( \tau \). \( R \) can be either a single neuron-tuning curve or a population coding based tuning curve. When \( \sigma_D = 0 \), Equation 1 is exactly the inverse of the Fisher information for many forms of encoding statistics (e.g., Gaussian, Poisson); for others, it is a close approximation (Paradiso, 1988; Seung & Sompolinsky, 1993). Due to the Cramer-Rao lower bound, this sets the lower limit of the accuracy of stimulus estimation, and for many distributions the optimal decoder is efficient and can attain the lower bound (Wijsman, 1973). Therefore, any significant

![Figure 1. Theoretical foundation. (A) Neural responses vary depending on the intensity of a stimulus. For the same \( \tau \), a low magnitude stimulus (blue) will, on average, generate fewer spikes than a high magnitude stimulus (red). If the presentation time of the stimulus is truncated (\( \tau < t_{\text{sat}} \); dashed regions), fewer spikes will be counted on average. (B) Spike count distributions for different \( \rho \) and \( \tau \) conditions are represented here (arbitrarily) using gamma distribution functions. The size of the magenta region (i.e., the overlap coefficient) is directly related to discrimination ability. (C) From the overlap coefficients, we can calculate psychometric curves, and thus JNDs, for different conditions. Note how changes in \( \tau \) affect the \( \rho \) conditions differently. (D) Behavioral values of JND cannot be derived axiomatically. However, using Equation 4, we can predict the ratio between JND and JNDS. Note when \( \tau \geq t_{\text{sat}} \), LJR = 0.](http://arvojournals.org/)
behavioral errors in excess of this lower bound are decoding errors.

We do not know the true function $R(\theta, \tau)$. However, for simplicity the function can be decomposed into $R(\theta) = \tau \cdot r(\theta)$, where $\tau$ is the duration of the stimulus, and $r(\theta)$ is the firing rate function. Here we complete the derivation with the assumption that the firing rate is static, but we show in the Appendix that the following conclusions hold for dynamic firing. The need to know the exact details of $r(\theta)$ can be eliminated if we make a ratio of $\sigma_R$ at two values of $\tau$. For simplicity sake, we use $\tau_{sat}$ (as described above) as one of these values, and $\tau_1$ as the other, where $\tau_1 < \tau_{sat}$. This gives us

$$\frac{\sigma_R(\tau_1)}{\sigma_R(\tau_{sat})} = \left(\frac{\tau_{sat}}{\tau_1}\right)^{-\frac{1}{\rho}}$$

(2)

While we know that there is noise in the firing rate of neurons, and that it is substantial, the noise in the decoding process is harder to pin down. We can analyze three cases—the decision noise is substantially less than the firing rate noise, the decision noise is comparable to the firing rate noise, or the decision noise is substantially larger than the firing rate noise. In the case that the dominant source of noise is the variability of firing rates ($\sigma_T \ll \sigma_R$) our analysis results in predictions that are testable at the behavioral level. It is these predictions that are central to the experiment performed here.

Using assumption 3 from the beginning of this section, we can insert a power law $\sigma = \beta \cdot R(\theta, \tau)^\rho$ into Equation 2. Once simplified, this results in the following equation:

$$\frac{\sigma_R(\tau_1)}{\sigma_R(\tau_{sat})} = \left(\frac{\tau_1}{\tau_{sat}}\right)^{-\frac{1}{\rho}}$$

(3)

$\sigma_R$ can be considered the equivalent of the traditional JND used in psychometric experiments. Here the JND measured will depend on the presentation times ($\tau$). With this in mind, we now define the variable LJR, which is the logarithm of the above equation:

$$LJR = \log\left(\frac{JND_\tau}{JND_{sat}}\right) = (\rho - 1) \cdot \log(\tau_1/\tau_{sat})$$

(4)

where LJR is the Log-JND-Ratio (defined here), $JND_\tau$ is the JND for a time window of length $\tau$, and $JND_{sat}$ is the JND for $\tau \geq \tau_{sat}$. Equation 4 is limited to the condition $\tau \geq \tau_{sat}$. By plotting $LJR$ against $\log(\tau)$, we arrive at a line whose slope and intercept can be used to calculate $\rho$ and $\tau_{sat}$. This function is plotted in Figure 1D for two specific cases: $\rho = 0$ (constant noise) and $\rho = 0.5$ (Poisson-like noise). From this formulation, we have designed a contrast discrimination experiment with a range of integration times ($\tau$) that can determine both $\tau_{sat}$ and $\rho$ for contrast perception.

**Experimental setup**

All work was carried out in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki), and informed consent was obtained from all volunteer subjects.

A group of nine visually normal subjects, two of whom were disqualified for poor performance, were placed in a darkened room and instructed to place their chin on a chinrest 0.91 m from the screen. Stationary gratings were presented with a randomized phase and a spatial frequency of 1 cycle/° of vision. Each trial consisted of a sequence of six images: a randomized visual mask, a reference contrast grating, another mask, a test contrast grating, a final mask, and a prompt (see Figure 2A). During the prompt part of a trial, subjects were instructed to answer the question “Did the second grating have higher contrast than the first?” Upon answering, subjects received feedback on whether their response was correct, followed immediately by the next trial. To prevent users from reliably using information from visual afterimages, checkerboard masks were presented for 0.5 s both prior to and following the stimuli, independent of the timing condition. The high contrast checkerboards had the additional benefit of keeping the subject in approximately the same contrast adaptation state throughout every trial (Heinrich & Bach, 2001).

Each condition was defined by the presentation time and the reference contrast. Reference and contrast stimuli were presented for either 600, 300, 150, 125, 100, 75, or 50 ms. The reference contrast was presented as a random phase sinusoidal grating with a median brightness of 26.80 lumens, and a Michelson contrast of 0.05, 0.125, 0.275, and 0.35. For each trial within a particular condition, the test stimulus contrast intensity was chosen using an adaptive algorithm.

**Adaptive Bayesian algorithm**

For this experiment, using traditional techniques to estimate the psychometric function would have required upwards of 30,000 trials for each subject. To reduce the number of trials needed, an adaptive Bayesian algorithm was used to efficiently find the JND in each experimental condition. The principles of the algorithm were primarily drawn from Kontsevich and Tyler (1999), with small alterations to fit this specific task (see the section entitled “Bayesian adaptation algorithm alterations” in the Appendix for details). To summarize the method, an extremely wide prior probability was established for the psychometric function’s slope in each condition, with a mean based on preliminary data. From this prior distribution, a test stimulus was chosen via an entropy-based cost function.
in order to maximize the amount of information gained from the subject’s response. Upon response, the priors were updated with the new information, and a new test stimulus was chosen. This was repeated for 45 trials per condition (see Figure 2B for a graphical representation), with each condition being repeated in a counterbalanced version of itself (i.e., where the reference and test stimuli were switched). This resulted in approximately 7,200 trials being presented to each subject, broken up into 25 min blocks. Further details on this and the experimental setup can be found in the first section of the Appendix.

Calculating JND-related measures

For each condition combination, the Bayesian algorithm converged on the correct slope and threshold of the psychometric function. To determine JND, we found the stimulus value that would correspond to 84% (1σ) on the determined psychometric function. JNDsat (necessary for solving Equation 4) was found by averaging the JND for the 300- and 600-ms conditions. These time values are presumed to be above τsat because JND300 and JND600 are statistically indistinguishable from one another when combined across all subjects (unpaired t test, p = 0.28). JNDsat was calculated individually for each subject.

Once established, the value of JNDsat permits us to calculate the LJR (see Equation 4 above) for each subject and condition (Figure 3). For each subject, we used an iterative least square estimation method (as implemented with the “nlinfit” function within MATLAB) to do a nonlinear fit of the data to Equation 4 and extract the parameters τsat and ρ. Several distinct fitting methods were used and yielded similar results.

Results

Our theoretical analysis (above) relates the statistics of perceptual errors to the statistics of the neural code. Specifically, we noted how perceptual errors would vary with changes to the window of temporal integration given different models of noise. On the basis of this theory, we designed an experiment to test if the source of perceptual errors is indeed the variability of the encoding sensory neurons. The hypothesis that behavioral variability depends on the variability of the encoding neurons underlies a significant body of scientific work (Britten et al., 1993; Britten et al., 1996; Cohen & Newsome, 2009; Mazurek et al., 2003; Shadlen et al., 1996; Shouval et al., 2013).

The perceptual errors of our subjects were tested in a contrast discrimination task for different durations of stimulus presentation and different contrasts. The results of each subject (Shown in Table 1 and Figure 3) were fit separately to Equation 4 (dashed red line in Figure 3). Taking the intersubject statistics (i.e., averaging across the values calculated for each subject), we find the average of the slope on the log-log plot to be approximately −1.2, which implies that ρ = −0.20 ± 0.31. This value of ρ is significantly different (ρ ≈
0.00014) from the Poisson-like value of $\rho = 0.58$ found through an electrophysiology experiment (Dean, 1981). However, it is statistically indistinguishable from a constant noise condition ($\rho \cong 0$). We also find that $\tau_{\text{sat}} = 232 \pm 83$ (Figure 3D and Table 1). Note that this value of $\tau_{\text{sat}}$ is consistent with the estimation of JND$_{\text{sat}}$ done at the end of the Methods section, which showed $\tau_{\text{sat}}$ should be less than 300 ms. Similar results are also derived and presented in both Table 1 and Figure 3 under “Combined data” by combining all the data from every subject, and also finds the best fit for Equation 4.

Further, we have done analysis across the different reference contrast levels to tell if there is any trend relating the reference contrast stimulus to the resulting $\rho$ or $\tau_{\text{sat}}$ values. Largely, we find that no such effect exists, with the exception of our lowest contrast level (see Figure 4). At extremely low contrast intensities, subjects see a significant reduction in both $\rho$ and $\tau_{\text{sat}}$. We interpret this as a possible ramification of the dipper function, the well described phenomenon of nonlinear thresholds at low contrast values (Bradley & Ohzawa, 1986; Legge & Foley, 1980; Tolhurst & Barfield, 1978), although this theory would take additional data collection to confirm.

Our analysis above was based on the assumption that spike counts are the primary feature being decoded, and that the spike counts for the reference and test stimuli are being compared during decoding. However, as we are able to distinguish between a short but intense grating and a long but weak grating (beyond 50 ms), the time of the stimulus presentation must also be known or inferred. If the decoding mechanism is provided with an estimate of stimulus duration, it can infer the contrast intensity from the spike count. However, it is well established that timing perception is also prone to errors and that these errors increase linearly with the magnitude of the stimulus in an effect called scalar timing (Buhusi & Meck, 2005; Church, 2003; Gibbon,

<table>
<thead>
<tr>
<th>Subject</th>
<th>$\tau_{\text{sat}}$ (ms)</th>
<th>$SE_{\tau}$ (ms)</th>
<th>$\rho$</th>
<th>$SE_{\rho}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject 1</td>
<td>169.22</td>
<td>13.17</td>
<td>-0.2180</td>
<td>0.1551</td>
</tr>
<tr>
<td>Subject 2</td>
<td>339.73</td>
<td>38.55</td>
<td>0.0496</td>
<td>0.0984</td>
</tr>
<tr>
<td>Subject 3</td>
<td>165.09</td>
<td>9.50</td>
<td>-0.6990</td>
<td>0.1492</td>
</tr>
<tr>
<td>Subject 4</td>
<td>182.85</td>
<td>14.98</td>
<td>-0.5598</td>
<td>0.1815</td>
</tr>
<tr>
<td>Subject 5</td>
<td>148.42</td>
<td>12.27</td>
<td>-0.0571</td>
<td>0.1486</td>
</tr>
<tr>
<td>Subject 6</td>
<td>330.60</td>
<td>29.57</td>
<td>-0.0661</td>
<td>0.0949</td>
</tr>
<tr>
<td>Subject 7</td>
<td>284.93</td>
<td>71.17</td>
<td>0.1184</td>
<td>0.1976</td>
</tr>
<tr>
<td>Combined data</td>
<td>211.20</td>
<td>12.19</td>
<td>-0.1057</td>
<td>0.0774</td>
</tr>
</tbody>
</table>

Table 1. Summarized data from each subject. Notes: Here, we see the calculated $\tau_{\text{sat}}$ and $\rho$, along with their standard error, for each subject. To do analysis of the data, we have taken two approaches. First, under Combined data, we have combined the derived LJR for every subject, time, and reference condition; following that, we fit Equation 4 to this cluster of data and extracted $\tau_{\text{sat}}$ and $\rho$. Second, under Intersubject statistics, we simply averaged the values of $\tau_{\text{sat}}$ and $\rho$ found for each subject.

Figure 3. Individual and combined Log-JND-Ratio plots. Above, we have plotted the individual subject data along with the best fit line for Equation 4. The calculated $\tau_{\text{sat}}$ is denoted by the dashed blue line, and $\rho$ (the noise term) is equivalent to the slope of the initial segment of the line plus one. The final graph, “Combined data,” brings together all of the data from every subject, and also finds the best fit for Equation 4.
to the total errors in perception (the time window, scalar timing errors could contribute the spike count and inferring rate using an estimate of variables). In other words, if the decoder is measuring into contrast estimates (as well as other perceptual to infer contrast, errors in estimating time will propagate Church, & Meck, 1984). If estimates in timing are used 

\[ \tau_{\text{sat}} = \frac{\text{variance in contrast estimates}}{\text{duration of the sensory variable}} \]

Figure 4. \( \rho \) and \( \tau_{\text{sat}} \) values across reference contrast conditions. In both charts, contrast in measured via Michelson Contrast. (A) The dependence of \( \tau_{\text{sat}} \) on the reference contrast magnitudes. (B) The dependence of \( \rho \) on the reference contrast magnitudes. Values were obtained by taking the mean across all subjects within a reference condition. Significant differences (\( p < 0.05 \)) were only found by comparing to the lowest reference contrast level, and are denoted by a *.

Our experimental results show that (a) the brain uses a limited temporal integration window (\( \approx 230 \) ms) to form percepts of contrast, and (b) that the noise affecting contrast discrimination is roughly constant and independent of the magnitude or temporal duration of the sensory variable. This second result contradicts the hypothesis that perceptual errors arise from the variability of the sensory neurons that encode the percept. Below, we discuss how these results compare to previous publications, the theoretical explanations for these results, and the testable ramifications of these theories.

Our first result is that the length of the integration window for discrimination of contrast stimuli is approximately 232 ms. This result is somewhat similar to that of Legge (1978), who also varied the presentation time in a contrast dependent behavioral task. In that experiment, \( \tau_{\text{sat}} \) was found to range from 50 to 1000 ms, depending on the spatial frequency of the stimuli. For the spatial frequency closest to ours (0.75 cycles/° of vision), \( \tau_{\text{sat}} \) was found to be approximately 100 ms. Whereas the results are somewhat different, there are significant distinction between that experiment and our own: It used a staircase procedure (Wetherill & Levitt, 1965) to estimate the parameters of the psychometric curve rather than Bayesian adaption (Kontsevich & Tyler, 1999), it had a smaller number of subjects, and it used a contrast detection task rather than a discrimination task. This last fact may have contributed significantly to the differences observed. In Legge (1978), stimuli were determined to be present or absent compared to a blank screen. To compare, our study displayed similar gratings, and subjects chose the one with the higher contrast. This means our study captures decision behavior over a wider array of comparisons, which have established nonlinear effects—e.g., the dipper function (Bradley & Ohzawa, 1986; Legge & Foley, 1980; Tolhurst & Barfield, 1978). Additionally, we used a substantially higher contrast mask (\( \sim 1 \) vs. 0.2) than Legge (1978). Within his experiment, increasing the magnitude of the contrast mask appeared to increase \( \tau_{\text{sat}} \); however, it is difficult to generalize this relationship due to limited data within both experiments.

We propose that the integration window is consistent with established cellular responses to contrast stimuli. It has been shown that contrast-detecting neurons have a highly transient firing rate response whose information content returns to baseline levels after approximately 200 ms (Heller et al., 1995). Given the properties of such sensory neurons, it might not be possible to extract useful information from longer presentation times. While it may be surprising that our visual system is unable utilize temporal windows longer than 230 ms for contrast perception, it is intriguing to note that this is approximately the same length of time as a standard intersaccade interval (Carpenter, 1988). We postulate that, if the eyes rarely stay in a single position for

**Discussion**

Our experimental results show that (a) the brain uses a limited temporal integration window (\( \approx 230 \) ms) to form percepts of contrast, and (b) that the noise affecting contrast discrimination is roughly constant and independent of the magnitude or temporal duration of the sensory variable. This second result contradicts the hypothesis that perceptual errors arise from the variability of the sensory neurons that encode the percept. Below, we discuss how these results compare to previous publications, the theoretical explanations for these results, and the testable ramifications of these theories.

Our first result is that the length of the integration window for discrimination of contrast stimuli is approximately 232 ms. This result is somewhat similar to that of Legge (1978), who also varied the presentation time in a contrast dependent behavioral task. In that experiment, \( \tau_{\text{sat}} \) was found to range from 50 to 1000 ms, depending on the spatial frequency of the stimuli. For the spatial frequency closest to ours (0.75 cycles/° of vision), \( \tau_{\text{sat}} \) was found to be approximately 100 ms. Whereas the results are somewhat different, there are significant distinction between that experiment and our own: It used a staircase procedure (Wetherill & Levitt, 1965) to estimate the parameters of the psychometric curve rather than Bayesian adaption (Kontsevich & Tyler, 1999), it had a smaller number of subjects, and it used a contrast detection task rather than a discrimination task. This last fact may have contributed significantly to the differences observed. In Legge (1978), stimuli were determined to be present or absent compared to a blank screen. To compare, our study displayed similar gratings, and subjects chose the one with the higher contrast. This means our study captures decision behavior over a wider array of comparisons, which have established nonlinear effects—e.g., the dipper function (Bradley & Ohzawa, 1986; Legge & Foley, 1980; Tolhurst & Barfield, 1978). Additionally, we used a substantially higher contrast mask (\( \sim 1 \) vs. 0.2) than Legge (1978). Within his experiment, increasing the magnitude of the contrast mask appeared to increase \( \tau_{\text{sat}} \); however, it is difficult to generalize this relationship due to limited data within both experiments.

We propose that the integration window is consistent with established cellular responses to contrast stimuli. It has been shown that contrast-detecting neurons have a highly transient firing rate response whose information content returns to baseline levels after approximately 200 ms (Heller et al., 1995). Given the properties of such sensory neurons, it might not be possible to extract useful information from longer presentation times. While it may be surprising that our visual system is unable utilize temporal windows longer than 230 ms for contrast perception, it is intriguing to note that this is approximately the same length of time as a standard intersaccade interval (Carpenter, 1988). We postulate that, if the eyes rarely stay in a single position for
longer than ~200 ms, there would be little reason for the brain to be able to integrate over longer time intervals. Both the short intersaccade intervals and integration time window might have arisen because of natural evolutionary pressures favoring fast reaction times over very precise stimulus estimation.

Our initial hypothesis was that a likely source for behavioral variability is the trial-by-trial spike count variability (Shouval et al., 2013). Our experimental results are inconsistent with this hypothesis since the behaviorally determined $\rho$ value is significantly different ($\rho \approx 0.00014$) from the previously determined physiological value of $\rho = 0.58$ for single contrast encoding neurons in V1 (Dean, 1981).

We have approximated the value of $\rho$ from the most comparable portion of the data set obtained by Legge (1978, i.e., spatial frequency of 0.75), and found it to be approximately 0.25. Although this differs from our estimate, due to the difference in the actual experiments we cannot determine the origin of this difference and cannot even determine if it is statistically different from our result.

The value of the $\rho$ that we find in this experiment, as well as that just estimated from previous data (Legge, 1978), is very different than the value one expects from single neurons that encode contrast (Churchland et al., 2010; Dean, 1981). One possible explanation of this discrepancy is that, since the stimulus is encoded by an ensemble of correlated neurons (and not by a single neuron), some forms of averaging over the population may give rise to a constant noise distribution across the entire ensemble. There is some empirical support for this notion (Chen, Geisler, & Seidemann, 2006). However, such an interpretation is very puzzling if one assumes that perceptual noise ($\sigma_n$) were to indeed arise from the encoding neurons. Although a combination of correlated Poisson-like neurons could have statistics that differ significantly from Poisson statistics, there seems to be no way of combining such noisy signals to obtain a constant noise source. This would require a situation where the neural signal accumulates over time while noise does not, which seems to be a logical impossibility. Therefore, it is highly unlikely that the behavioral errors in contrast perception arise from the variability of the encoding sensory neurons in V1.

An alternative explanation is that the source of perceptual errors arises from a suboptimal decoding process (possibly involving noisy decision or memory operations) with a constant noise level ($\rho = 0$). This suboptimal decoding process would overwhelm any noise from spike count or timing variability (Johnson, 1980). Such noise added during decoding would be independent of the stimulus duration or contrast. This interpretation is consistent with observations that perceptual errors are far greater than expected from averaging over many sensory neurons (Britten et al., 1993; Tolhurst, Movshon, & Dean, 1983). However, this alternative seems inconsistent with experiments that report significant choice probabilities in sensory cortical neurons (Britten et al., 1996). One possible explanation is that significant choice probabilities arise from sensory neurons that get top-down feedback from the decoding neurons. In such a case, the behavioral variability does not arise from the variability of sensory neurons; rather, the decoding process affects the variability of the sensory neurons. There exists some evidence for this (Cumming & Nienborg, 2016). Further experiments are needed to resolve this apparent contradiction between our results and previous results regarding choice probabilities of sensory neurons.

Criticism could be applied to our model’s third assumption, which requires that the spike count distribution be approximated with a power law. For example, recent results (Goris, Movshon, & Simoncelli, 2014) have suggested a different functional form for single neuron spike count variability. In this paper, the authors postulated that noise can be best described by a double stochastic process in which the gain of each neuron varies from trial to trial. This would result in a spike count having a Poisson distribution with a rate that depends both on the stimulus and a stochastic gain. More importantly for our purpose, the fluctuations of gain are correlated between neurons but the Poisson process is independent. Averaging could reduce the Poisson portion of the variability, but, due to correlations, the stochastic gain component will not be eliminated. However, since the gain dependent variability is stimulus dependent, this cannot account for our results. Whereas this particular claim may not disrupt our model, it remains possible that future work will uncover further patterns in the spike rate noise that are problematic. Given the robustness of the power law fit for contrast neurons (Churchland et al., 2010; Dean, 1981), we do not expect this to occur, but it admittedly is not inconceivable.

The subfield of molecular psychophysics offers another potential source of noise that we have not controlled for. In some of these papers (Green, 1964; Li, Klein, & Levi, 2006; Neri, 2010), there is evidence that there are trial-to-trial correlations in response. To put it simply, a response to trial $n$ biases the response to trial $n + 1$, often in such a way that cannot be detected by averaging across the entire data set (as we do in Appendix section “Analyzing for bias and order effects”). It is possible that this could leave a noise signature in the behavior, although extracting this bias is difficult. Methods to isolate this bias (Neri, 2010) are not compatible with adaptive tests, and further experiments would need to be pursued to determine the extent that this effects our calculated $\rho$.

A recently published paper obtained results that are consistent with our findings and that reinforce their
generality (Drugowitsch et al., 2016). Like our own research, the authors use analytical methods to find the behavioral signatures of noise characteristics. That paper comes to similar conclusions as we have here, namely that neural stochasticity at the sensory encoding level is unable to explain the measured perceptual errors. Several things distinguish our methods from those of Drugowitsch et al. While their experiment did use a contrast stimuli, it used it within a probabilistic cue combination task. They compared it to the Bayes-optimal integration of information, whereas we compare our more traditional psychophysical task to the discrimination threshold predicted using Fisher information. We suggest that our method can complement theirs by providing a simpler analytical model to work with, and a more tractable experimentation protocol.

To summarize, we have shown that the primary introduction of noise is not in the encoding phase (i.e., stochastic processes translating time and stimulus intensity into firing rates), but rather in the decoding phase (e.g., decision, memory storage and retrieval, and comparison). In future experiments, we will isolate aspects of the decoding process, and determine which one is the primary contributor to our sensory errors.

**Keywords:** contrast sensitivity, encoding/decoding, neuron, noise, detection/discrimination

## Acknowledgments

The authors would like to acknowledge Dr. Anthony Wright and Dr. Daniel Felleman at UT Health Science Center at Houston for contributions of materials and expertise in the development of this experiment.

Commercial relationships: none.
Corresponding author: Harel Shouval.
Email: harel.shouval@uth.tmc.edu.
Address: Department of Neurobiology, and Anatomy, UT Health Science Center at Houston Lab of Harel Shouval, Houston, TX, USA.

## References


Shouval, H. Z., Agarwal, A., & Gavornik, J. P. (2013). Scaling of perceptual errors can predict the shape of


**Appendix**

**Materials and methods**

**Hardware setup**

Subjects were seated in a darkened room in front of a Hitachi SuperScan 21 Supreme CRT Monitor (Hitachi, Tokyo, Japan). The monitor provided the only source of illumination in the room. A Pentium 4, 3 Ghz Windows XP SP2 computer controlled the screen with a Cambridge Research Systems VSG2/5 graphics card (Cambridge Research Systems, Rochester, UK), which refreshed the monitor at a rate of 80 Hz and a resolution of 769 × 1024. The screen’s width was 34.8 cm. A black cardboard sheet was placed in front of the screen with a 12-in diameter circle cut out of it in order to avoid subjects seeing the edge of the screen. This is due to distortion effect at the edge CRT monitors that may otherwise alter the stimuli. The experiment was written and executed within MATLAB 2011B, and made heavy use of PsychToolbox-3.

To assure that subjects maintained a constant view of the screen, a headrest was placed such that their eyes were 0.91 m away from the screen. To deal with height differences among subjects, the headrest was vertically adjustable, and subjects were permitted to adjust it as needed during the experiment.

**Training**

Subjects were trained prior to the task using a PowerPoint demonstration to familiarize themselves with what they were going to see. Following this initial training, they were given a shortened form of the normal task and asked to demonstrate that they understood what they were being asked to do. Subjects who performed below 70% accuracy on the training task were assumed not to have understood the directions, and given further instruction until they performed satisfactorily.

**!**

**Stimuli**

Subjects were presented with two sinusoidally defined gratings on the monitor, and asked to determine which one had a higher contrast. The gratings were presented as having 1 cycle/° of vision. Luminance of the gratings could range from 9.32 to 44.27 lumens, and the average intensity was 26.80 lumens, for all gratings. The full range of the monitor (0–46.6 lumens) was not used due to empirically determined distortion effects at these extended levels. The gratings were always presented vertically, and a random phase was chosen for each presentation.

To prevent subjects from having afterimages of the gratings (and therefore potentially integrating additional information after the grating was no longer presented), a randomized checkerboard pattern was used to mask the grating. The checkerboards were presented at the start of the trial, and presented for 500 ms each after the display of the reference and test gratings. Every presentation was independently randomized (i.e., each square could either be black or white, and was determined by a random number generator) in order to prevent subjects from being able to predict what they were going to see at each spot.

Throughout the whole experiment, including the checkerboard masks, the average luminance of the screen was designed to be 26.80 lumens. This was to maintain the same level of light adaptation throughout. The monitor brightness levels were verified several times throughout the data collection phase using a Tektronix J17 LumaColor meter with a J1803 luminance head, and did not vary substantially from month to month.
The stimulus was varied over two conditions: reference magnitude and display time. The reference gratings, which fluctuated sinusoidally around the mean intensity of 26.80 lumens, had a Michelson contrast of 0.05, 0.125, 0.2, 0.275, or 0.35. Both the reference and test stimuli were presented at 600, 300, 150, 125, 100, 75, and 50 ms. The experiment was performed in blocks of 45 trials, each with a fixed reference contrast and fixed duration. Each subject was exposed to approximately 7,200 trials in total, with some variability due to small alterations in the experimental protocol. A Bayesian adaptive algorithm, described in detail below, was used during each trial to determine the appropriate magnitude of the test stimulus. Each block was performed at least twice to counter-balance any order effects—one section had the reference stimuli prior to the test stimuli, and the other reversed this. Later in the Appendix, we analyze this counterbalancing and show there is no significant order effect.

We made an exception to the counterbalancing guideline in the 600-ms condition. It was added later in the testing protocol, and was only tested in the “reference first” variation because previous testing had not suggested any choice bias. The order of these condition sections was randomized to prevent behavioral variation between data collection days from affecting global analysis.

Bayesian adaption algorithm alteration

In order to speed up data collection in our experiment, we used an adaptive algorithm (Kontsevich & Tyler, 1999). However, in order to more easily connect with the theory developed here (main paper and the “Mathematical derivations” section in the Appendix), the range of the sigmoid JND function was adjusted from [0.5 1] to [0 1]; this can be interpreted as changing the y axis from percent correct to percent perceived as higher. See Figure A1 for a description and graphical representation of the differences between these two methods. Both methods are used somewhat interchangeably within the psychophysical literature, which can cause significant confusion. The following equation represents the function that our algorithm optimizes for:

\[
\Psi(\theta) = \frac{1 + \text{erf}(\beta \ast (\theta - T))}{2} \tag{A1}
\]

where the psychometric function \(\Psi\) represents percent perceived as higher, and \(\theta\) is the magnitude of the stimulus. The two variables \(\beta\) (the slope) and \(T\) (the threshold) are the variables that the adaptive algorithm changes to better fit the psychometric function of the test subject. Prior probability distributions were constructed for \(\beta\) and \(T\). An analysis of our data (shown in “Analyzing for bias and order effects”) suggested little
to no bias in user responses (i.e., the user was equally likely to be correct for test stimuli both above and below the reference stimuli), so the prior for the threshold was chosen to be very narrow. Since our primary goal was to determine the slope of the curve, a very broad, nearly flat distribution was used. A copy of the code used, as well as the raw data from subjects, is available upon request.

Convergence

Within their paper, Kontsevich and Tyler (1999) provided a great deal of support for their work, which will not be reproduced here. However, we have made modifications to their suggested algorithm; while these justifications are well founded, it is prudent test the algorithm to show that there are no unexpected consequences of our modified assumptions.

As described in the previous section, the main features that our algorithm adapts for are the slope ($\beta$) and threshold ($T$) of the psychometric function (Equation A1). For clarity, we will call the true values of these variables $\beta_{\text{true}}$ and $T_{\text{true}}$. At the start of each experimental condition, the adaptive algorithm is given a prior probability distribution for $\beta$ and $T$, which we will denote as $\beta_{\text{prior}}$ and $T_{\text{prior}}$. These distributions can be interpreted as the adaptive algorithms initial “belief” (and confidence about that belief) about what $\beta_{\text{true}}$ and $T_{\text{true}}$ are (see Figure 3.1B for a graphical representation). Over the course of the experiment, the algorithm updates the prior distributions based on the responses of the subject. Essentially, the algorithm starts with a (potentially wrong) belief, and converges onto the correct answer. Here, we developed a Monte Carlo simulation to test for this convergence in a variety of conditions.

To build this simulation, we took the code developed for the experiment and inserted a simulated subject. This simulated subject was essentially a psychometric for the experiment and inserted a simulated subject. conditions. A known issue with adaptive tests is that, if subjects fail to pay attention to the task (i.e., answer randomly), there can be a failure to converge within the number of trials allotted. To deal with this, test sections where performance was below 80% or above 95% accuracy were excluded from analyses. Two subjects who performed below the 80% threshold in over half their sections were excluded entirely from the analysis under the assumption that they were guessing randomly for large parts of the experiment.

JND value

Typically, within the psychophysical literature, JND is defined as the difference between the reference stimulus and the test stimulus that the subject perceives as higher 75% of the time. The 75% point is arbitrary, and has been eschewed in this paper to make the math easier. Discussion of JND throughout the rest of the paper is defined as 84%, which corresponds to the first standard deviation above and below the reference stimuli for a Gaussian distribution. Since the Bayesian adaptive algorithm determines the slope and threshold of the psychometric function, rather than a specific point, it would be computationally trivial to switch between the 75% and 84% JND if it were necessary for future comparisons.

Analyzing for bias and order effects

We have analyzed the response bias and order effects to make sure that our experimental design did not affect the responses of our subjects. To test for bias, we wished to make sure that subjects were not more inclined to select “yes” rather than “no” when given our prompt (“Was the second stimulus higher than the first?”). In Table A1, we can see the probability that each subject responded “yes” in both the normal and counterbalance conditions. Testing with a one-sample $t$ test, we failed to reject the null hypothesis (i.e., that the data is not biased in a statistically significant way) in the normal condition ($p = 0.93$), the reverse condition ($p = 0.19$), or a combination of the two conditions ($p = 0.37$). This failure to reject suggests that there is no bias in response.

To test for order preference (i.e., the possibility that subjects’ perception was altered by which stimulus was presented first), we counterbalanced our experimental design so that in one condition the reference stimulus was displayed before the test stimulus; in the other condition the reverse was true. In both conditions, the
subject still responded to the same prompt (“Did the second stimulus have a higher contrast than the first?”), and the subject was not informed of the order switch. Calling back to Table A1, we used a two-sided $t$ test to test between the normal and reverse conditions, and failed to reject the null hypothesis (i.e., that the two conditions are statistically indistinguishable) at a $p = 0.26$. This lends credence to the hypothesis that there is no order effect.

Beyond this, we can look at the calculated JND across all subjects. If there is an order effect, we would expect to see differences between the normal and reverse presentations. In Table A2, we have recorded the average and standard deviation of the JND in every time and reference condition, collapsed across subjects. When the normal and reverse conditions of the same Time X Reference combination are compared to one another with a two-sided $t$ test, no condition reaches statistical significance when corrected for multiple comparisons. When left uncorrected, one condition (150 ms $\times$ 0.125 Michelson Contrast) exceeds a traditional significance of $a = 0.5$; this is to be expected via random chance. Another test that can be performed is comparing all the JND values directly rather than averaging across subjects first. To do this, we used a paired-sample $t$ test, which makes a direct comparison for every Subject $\times$ Time $\times$ Reference condition in the normal and reversed presentation condition. Once again, we fail to reject the null hypothesis ($p = 0.40$), further reinforcing the notion that there is no order effect.

<table>
<thead>
<tr>
<th>Subject</th>
<th>Normal</th>
<th>Reverse</th>
<th>Combined</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject 1</td>
<td>0.50</td>
<td>0.52</td>
<td>0.50</td>
</tr>
<tr>
<td>Subject 2</td>
<td>0.49</td>
<td>0.49</td>
<td>0.50</td>
</tr>
<tr>
<td>Subject 3</td>
<td>0.49</td>
<td>0.49</td>
<td>0.50</td>
</tr>
<tr>
<td>Subject 4</td>
<td>0.48</td>
<td>0.53</td>
<td>0.50</td>
</tr>
<tr>
<td>Subject 5</td>
<td>0.54</td>
<td>0.53</td>
<td>0.50</td>
</tr>
<tr>
<td>Subject 6</td>
<td>0.50</td>
<td>0.56</td>
<td>0.50</td>
</tr>
<tr>
<td>Subject 7</td>
<td>0.49</td>
<td>0.49</td>
<td>0.50</td>
</tr>
<tr>
<td>Combined</td>
<td>0.50</td>
<td>0.52</td>
<td>0.51</td>
</tr>
</tbody>
</table>

Table A1. Measurements of bias. Note: The table above records the probability that a subject will respond “Yes” to the prompt “Does the second grating have a higher contrast than the first?” Every subject responds in an unbiased manner.
Mathematical derivations

The implications of variable firing rates

To make the math simpler in the above section, we assumed that firing rate, \( r(\theta) \), was a static function (i.e., the firing rate does not change over time). However, this is not a biologically realistic assumption; firing rates often have temporal dynamics. In this section, we demonstrate why this does not matter for our model.

Let us assume that the spike count function, \( R(\theta) \), is separable, and can be written as a multiplication of two functions: \( g(\tau) \) that accounts for the spike count for a given window \( \tau \), and \( f(\theta) \) that modulates the spike count for different stimulus intensities. Further, we can say that for presentation times over approximately 75 ms, we can assume that the firing rate is falling [See Heller et al. (1995) or figure 8 in Albrecht et al. (2002) for evidence.] This means that \( g(\tau) \) is less than linear, and can be approximated with the following equation:

\[
g(\tau) = \tau^\eta \quad \text{for } 0 < \eta < 1 \quad (A2)
\]

With these assumptions, we can begin to show that a variable firing rate will not effect the conclusion in the main paper. Inserting our separable \( R(\theta) \) function into Equation 1 from the main text, we arrive at

\[
\sigma(\tau) = \frac{\sigma_R}{g(\tau)f'(\theta)} \quad (A3)
\]

Following the same process we did in the previous section where we created a ratio between \( \sigma_0(\tau_1) \) and \( \sigma_0(\tau_{\text{sat}}) \), we arrive at

\[
\frac{\sigma_0(\tau_1)}{\sigma_0(\tau_{\text{sat}})} = \frac{\sigma_R(\tau_1)}{g(\tau_1)f'(\theta)} / \frac{\sigma_R(\tau_{\text{sat}})}{g(\tau_{\text{sat}})f'(\theta)} \quad (A4)
\]

Rearranging and combining with the power law in assumption 3 from the main paper,

\[
\frac{\sigma_0(\tau_1)}{\sigma_0(\tau_{\text{sat}})} = \frac{JND_1}{JND_{\text{sat}}} = \frac{g(\tau_{\text{sat}})}{g(\tau_1)} \cdot \beta \cdot g(\tau_1)^\rho f(\theta)^\rho \quad (A5)
\]

\[
\frac{JND_1}{JND_{\text{sat}}} = \left( \frac{g(\tau_{\text{sat}})}{g(\tau_1)} \right)^{1-\rho} \quad (A6)
\]

Plugging Equation A2 in to Equation A6,

\[
\frac{JND_1}{JND_{\text{sat}}} = \left( \frac{\tau_1}{\tau_{\text{sat}}} \right)^{\eta(\rho-1)} \quad (A7)
\]

Taking the logarithm of both sides allows us to compare to Equation 4 from the main paper,
\[ LJR = \log \left( \frac{JND}{JND_{sat}} \right) = \eta \cdot (\rho - 1) \cdot \log(\tau_1 / \tau_{sat}) \]

for \( \tau \leq \tau_{sat} \)

(A8)

Just like with Equation 4, if we plot \( LJR \) against \( \log(\tau_1) \), the slope \( (m) \) of this line will be equivalent to the coefficient of the above equation:

\[ m = \eta \cdot (\rho - 1) \quad (A9) \]

Experimentally, we have found \( m = -1.11 \pm 0.09 \). From Equation A9, \( \eta \) would have to equal 2.64 for our original hypothesis of \( \rho = 0.58 \) to be true. As stated when we introduced the \( \eta \) variable, a biologically reasonable approximation for \( \eta \) would be \( 0 < \eta < 1 \). This implies that, even assuming a dynamically varying firing rate, we would still estimate \( \rho < 0 \). Thus, for physiologically plausible firing rate dynamics, the behavioral statistics are inconsistent with the premise of spike count variability as the source of behavioral noise.

The impact of errors in estimating temporal intervals

In the above analysis, we have used spike counts \( (S) \) as a proxy for spike rate \( (r) \), since we are comparing stimuli that have the same duration. However, this approximation hides an assumption that the time window is perfectly estimated, which may not be the case. Certainly at the behavioral level, there is substantial literature on errors in temporal estimation (Buhusi & Meck, 2005; Church, 2003; Gibbon et al., 1984). In this section, we analyze the impact of these temporal estimation errors on stimulus magnitude estimation.

By definition, \( r = S / \tau \), where \( \tau \) is the presentation time of the stimulus. Assume that there are two sources of noise: noise in the spike count \( (\Delta S) \), and noise in the estimation of the temporal interval \( (\Delta \tau) \). Therefore, on each trial \( S = \langle S \rangle + \Delta S \) and \( \tau = \langle \tau \rangle + \Delta \tau \), where \( \langle \rangle \) represents the expectation value. For simplicity, these calculations assume that \( \Delta \tau \ll \tau \) and \( \Delta S \ll S \). We will mostly consider here the case where \( \Delta S \ll \Delta \tau \).

Therefore, we obtain

\[
\begin{align*}
 \langle r \rangle &= \frac{S}{\tau} \\
 &= \frac{\langle S \rangle + \Delta S}{\langle \tau \rangle + \Delta \tau} \approx \frac{\langle S \rangle + \Delta S}{\langle \tau \rangle + \Delta \tau} \cdot \left(1 - \frac{\Delta \tau}{\langle \tau \rangle} \right) \\
 &\approx \frac{\langle S \rangle}{\langle \tau \rangle} + \frac{\Delta S}{\langle \tau \rangle} + \Delta \tau \frac{\langle S \rangle}{\langle \tau \rangle^2} \\
 &\approx \langle r \rangle + \frac{\Delta S}{\langle \tau \rangle} + \Delta \tau \frac{\langle S \rangle}{\langle \tau \rangle^2} \quad \text{(A10)}
\end{align*}
\]

We will now take the square and then average \( r \) to obtain its variance. First, taking the square of Equation A10:

\[
r^2 \approx \left( \frac{\langle S \rangle}{\langle \tau \rangle} + \frac{\Delta S}{\langle \tau \rangle} + \Delta \tau \frac{\langle S \rangle}{\langle \tau \rangle^2} \right)^2 \\
\approx \frac{\langle S \rangle^2}{\langle \tau \rangle^2} + \Delta S^2 \langle \tau \rangle^2 - \Delta \tau \frac{\langle S \rangle^2}{\langle \tau \rangle^2} - \Delta \tau \frac{\langle S \rangle^2}{\langle \tau \rangle^2} - \Delta S \cdot \Delta \tau \frac{\langle S \rangle}{\langle \tau \rangle^2} + \Delta \tau \frac{\langle S \rangle^2}{\langle \tau \rangle^2} \quad \text{(A11)}
\]

We now take the average. Note that any term linear with respect to \( \Delta \tau \) or \( \Delta S \) must vanish. Additionally, if the distribution of \( \tau \) is independent of the distribution of \( S \), terms with \( \Delta \tau \cdot \Delta S \) as coefficients must also disappear:

\[
\langle r^2 \rangle \approx \frac{\langle S \rangle^2}{\langle \tau \rangle^2} + \Delta S^2 \langle \tau \rangle^2 + \langle \tau \rangle^2 \langle S \rangle^2 \quad \text{(A12)}
\]

If we assume that \( \Delta S \) is small, and define \( \sigma_r^2 = \sigma_T \frac{\langle S \rangle}{\langle \tau \rangle} \), we get the standard deviation of the estimated rate:

\[
\sigma_r = \sigma_T \frac{\langle S \rangle}{\langle \tau \rangle} \quad \text{(A13)}
\]

By invoking the scalar timing law (Weber’s law for temporal estimation; Buhusi & Meck, 2005; Church, 2003; Gibbon et al., 1984), which has the form \( \sigma_T = \alpha \langle \tau \rangle \), we obtain

\[
\sigma_r = \alpha \frac{\langle S \rangle}{\langle \tau \rangle} \quad \text{(A14)}
\]

Let’s assume a simple case of constant firing rate over the period \( \tau \), and that the spike count depends on the parameter through a tuning curve \( f(\theta) \) such that \( \langle S(\theta, \tau) \rangle = \tau \cdot f(\theta) \). Using this, we get

\[
\sigma_r = \alpha \cdot f(\theta) \quad \text{(A15)}
\]

This implies that the noise of the rate variable is independent of the temporal window \( \tau \). We will now use a modification of Equation A4 (which was used in estimating the error of the decoded variable \( 0 \)) where we use the deduced rate rather than the spike count:

\[
\frac{\sigma_\theta(\theta, \tau_1)}{\sigma_\theta(\theta, \tau_2)} = \frac{\sigma_r(\theta, \tau_1) / f'(\theta)}{\sigma_r(\theta, \tau_2) / f'(\theta)} = 1 \quad \text{(A16)}
\]

By taking the log of this expression, we clearly find that the LJR should have a slope of 0.

Now generalize this to a case where \( S(\theta, \tau) = f(\theta) \cdot g(\tau) \), where \( g(\tau) \) is not the density, but rather the cumulative number of spikes between 0 and time \( \tau \). Similar to the linear case, we can retrieve \( f(\theta) \) by dividing the spike count with \( g(\tau) \). Note that in the constant firing rate case \( g(\tau) = \tau \), so this formulation is equivalent to the constant firing case. Let’s define this new variable \( q = S(\tau, \theta) / g(\tau) \), which is an attempt to retrieve \( f(\theta) \). Here, we will only examine the simple case where \( \Delta S \) is very small and can be ignored.
Approximating the second moment,

\[
\langle q^2(\theta, \tau) \rangle = \left( \frac{\langle S(\theta, \tau) \rangle g(\tau)}{g(\tau)} \right)^2 \left( 1 - \Delta \tau \frac{g'(\tau)}{g(\tau)} \right)^2
\]

The last stage is obtained because \( \langle \Delta \tau \rangle = 0 \). Therefore, the variance is

\[
\sigma_q^2 = \langle \Delta \tau^2 \rangle \left( \frac{\langle S(\theta, \tau) \rangle g(\tau)}{g(\tau)} \right)^2 \left( \frac{g'(\tau)}{g(\tau)} \right)^2
\]

By the scalar timing law \( \langle \Delta \tau^2 \rangle = \tau^2 \cdot \tau^2 \), and using \( S(\theta, \tau) = f(\theta) \cdot g(\tau) \) we get that

\[
\sigma_q(\tau) = \frac{\sigma_{q}}{f'(\theta)} = \frac{\tau \cdot g'(\tau)}{f'(\theta) \cdot g(\tau)}
\]

Consequently,

\[
\frac{\sigma_q(\tau_1)}{\sigma_q(\tau_2)} = \frac{\tau_1 \cdot g'(\tau_1) \cdot g(\tau_2)}{\tau_2 \cdot g'(\tau_2) \cdot g(\tau_1)}
\]

For a power-law \( g(\tau) = \tau^\eta \) we get that this ratio is one, just like the constant firing rate case (\( \eta = 0 \)).

The analysis above has been verified with a Monte Carlo simulation in MATLAB (available upon request; results displayed in Figure A3). The simulation also demonstrated that, even when the assumption \( \Delta \tau \ll \tau_{\text{sat}} \) does not hold, our conclusions hold for a wide range of \( \Delta \tau \) values. This include values of \( \Delta \tau \) up to and including \( \tau_{\text{sat}} \).

It is important to note the difference between the simulated results in Figure A3, and the experimental results in Figure 2 of the main paper. The slope of the line here is approximately 0, which would require a \( \rho \approx 1 \). Given that noise from different systems would combine in an additive way, there is no way it could combine (solely) with Poisson neural noise (\( \rho \approx 0.5 \)) to give us our findings (\( \rho \approx 0 \)).