

# Spike Count Reliability and the Poisson Hypothesis

Asohan Amarasingham, Ting-Li Chen, Stuart Geman, Matthew T Harrison  
Division of Applied Mathematics  
Brown University  
Providence, Rhode Island 02912, USA

David L Sheinberg  
Department of Neuroscience  
Brown University  
Providence, Rhode Island 02912, USA

Section: Behavioral/systems/cognitive, Senior Editor: Laurence Trussell

Abbreviated title: Spike count reliability and the Poisson hypothesis

Corresponding Authors: Asohan Amarasingham, Center for Molecular & Behavioral Neuroscience, Rutgers University, 197 University Ave, Newark NJ 07102, asohan@dam.brown.edu; David L. Sheinberg, Department of Neuroscience, Brown University, Box 1953, Providence, RI 02912, David.Sheinberg@brown.edu

Number of Figures: 6

Number of Pages: 38

Keywords: Inferotemporal cortex, spike trains, temporal coding, spike train analysis, trial-to-trial variability, regularity, Poisson hypothesis test, non-Poisson spiking, Vision, Object Recognition

Acknowledgments: AA was supported in part by a National Science Foundation Postdoctoral Fellowship in Biological Informatics. SG was supported in part by the Army Research Office under Contract DAAD19-02-1-0337 and the National Science Foundation under Grants DMS-0074276 and IIS-0423031. DLS was supported by grants from the Keck Foundation, the James S McDonnell Foundation, and NIH-EY014681. We thank G Buzsáki, K Diba, D Robbe, and S Montgomery for comments on the manuscript.

*Journal of Neuroscience*, in press ... Do not cite.

## Abstract

The variability of cortical activity in response to repeated presentations of a stimulus has been an area of controversy in the ongoing debate regarding the evidence for fine temporal structure in nervous system activity. We present a new statistical technique for assessing the significance of observed variability in the neural spike counts with respect to a minimal Poisson hypothesis, which avoids the conventional but troubling assumption that the spiking process is identically-distributed across trials. We apply the method to recordings of inferotemporal cortical neurons of primates presented with complex visual stimuli. On this data, the minimal Poisson hypothesis is rejected: the neuronal responses are too reliable to be fit by a typical firing-rate model, even allowing for sudden, dramatic, time-varying, and trial-dependent rate changes following stimulus onset. The statistical evidence favors a tightly regulated stimulus response in these neurons, close to stimulus onset, though not further away.

The variability of spike trains bears on theories of neural coding. A range of hypotheses have been offered. At one extreme, the existence and precise temporal location of every spike is significant. At another extreme, a spike train is a stochastic process essentially characterized by a slowly-changing rate. The latter hypothesis leads naturally to modeling spikes as the events of a slowly-varying inhomogeneous Poisson process (the “Poisson hypothesis”). Fine-temporal coding would be more likely to yield highly regular spike counts from repeated trials, whereas randomness in the Poisson hypothesis limits the degree of regularity across trials.

A statistic commonly employed to assay variability in spike trains is the empirical variance-mean ratio of the spike counts (the Fano factor) across trials (Softky and Koch, 1993; Shadlen and Newsome, 1998; Tolhurst, Movshon, and Dean, 1983; Oram et al., 1999; Kara, Reinagel, and Reid, 2000). For example, the spike counts for the inhomogeneous Poisson process are distributed as a discrete Poisson random variable, for which the mean and variance are identical. By this measure, the spike counts of *in vivo* cortical spike trains (in contrast perhaps to subcortical structures) have generally been thought to be as variable as Poisson processes and perhaps even more so (Shadlen and Newsome, 1998; Koch, 1999). As Shadlen and Newsome have written: “When an identical visual stimulus is presented for several repetitions, the variance of the neural spike count has been found to exceed the mean spike count by a factor of 1-1.5 wherever it has been measured.”

Inferring a *lack* of precision in the neural code from observations of variability is tricky. One line of reasoning is that the existence of variability under identical conditions, when the neuron is presumably signaling the same event, reflects noise (e.g. Poisson noise) in the signaling process itself. However, it is certainly plausible that a significant source of variability is the experimenter’s own uncertainty

about hidden contextual variables which the neuron is encoding (and which may be, for example, internal to the brain), such as attention, or the states of other neurons. Furthermore, the variability of overt behaviors that are difficult, but not impossible to measure, like precise eye position, have been shown to contribute to at least some of the variability commonly reported in cortical responses (Gur, Beylin, and Snodderly, 1997). As Barlow wrote about neural responses in 1972: “their apparently erratic behavior was caused by our ignorance, not the neuron’s incompetence.”

Regularity, on the other hand, cannot be explained away so easily, and in this light, *evidence* of finer temporal structure, particularly in higher order cortical areas, is intriguing. Recently, Muller et al. (Muller et al., 2001) have observed, from recordings of V1 cells in primates presented with sinusoidal gratings, that near stimulus onset the empirical variance-mean ratio is strikingly smaller than unity (in apparent contradiction to the Poisson hypothesis). This is in line with other reports which suggest a greater reliability in cortical responses than might be expected from a Poisson model (Kara, Reinagel, and Reid, 2000; Gershon et al., 1998; Gur, Beylin, and Snodderly, 1997).<sup>1</sup>

Our purpose in this paper is twofold. First, we derive a simple and exact test for a “minimal Poisson hypothesis” that utilizes the variability of spike counts across trials. The minimal Poisson hypothesis includes Poisson processes that have rapidly (in fact arbitrarily) changing spiking rates, possibly differing from trial to trial. Second, we present evidence consistent with these earlier studies of low variance-mean ratios particularly near stimulus onset, gathered here in cells from anterior IT of primates presented with more complex visual stimuli. Using this test on the IT recordings, we are able to reject the minimal Poisson hypothesis, with a notably small amount of data, predominantly near stimulus onset. We are able to argue, furthermore, that these results are not likely to be due to the effects of refractory period *per se*, but rather reflect regularity in the neural response.

## 1 Methods

### 1.1 Subjects and Materials

The basic behavioral and surgical methods employed in this study have been described previously (Sheinberg and Logothetis, 1997; Sheinberg and Logothetis,

---

<sup>1</sup>Some skepticism has been expressed regarding how widely these results would generalize in cortex. For example, commenting on the report of Kara et al. of reliability in layer 4 V1 responses, Movshon (2000) writes: “one may doubt that such high reliability will be found for most neurons lying outside thalamic recipient layers in cortex.”

2001). In brief, following initial behavioral training, two rhesus monkeys underwent aseptic surgery for the placement of a head restraint and a scleral search coil. All surgical procedures were carried out in accordance with the NRC *Guide for the Care and Use of Laboratory Animals*. Following surgery, the monkeys were trained to fixate a small yellow spot (0.25 degrees) appearing in the center of a CRT monitor. After acquiring the spot, between three and five visual images, 4deg on a side, were flashed behind the spot for either 800 or 1100 ms. Inter-stimulus intervals (during which only the spot was visible) were set to the same duration as the stimulus presentation duration (Figure 1). Visual stimuli were selected from a set of commercially available stock photographs of animals, natural scenes, and man made objects (Corel Corp).

During stimulus presentation, the monkeys were required to maintain fixation within a virtual square region 2deg on a side. During data collection, the digitized eye position was stored to disk every 5 ms (200Hz). In addition to keeping their gaze directed within the virtual window, the monkeys learned to follow the spot as it jumped from the center of the display to a new position. Upon refixation of the spot, they were rewarded with a drop of apple juice.

Figure 1 about here

## 1.2 Single Cell Recordings

Once the animals were trained in the fixation task, a ball and socket chamber housing an 18 gauge guide tube was placed directly above the anterior temporal lobes (AP: +18, ML: +19). Single unit recordings were made by lowering glass coated platinum-iridium electrodes by microdrive (Kopf Model 650) into the lower bank of the superior temporal sulcus (STS) and the lateral convexity of the inferior temporal gyrus, just posterior to the anterior middle temporal sulcus (AMTS). The neural signal was amplified using a Bak A-1 amplifier with remote head stage, and the conditioned signal was fed into a time-amplitude window discriminator. Individual cells were isolated while the animals performed the fixation task. Each cell was tested with at least eight stimuli, but often with many more (up to 80). Note that a significant effort was made to subselect visual stimuli from the relatively large test set that effectively elicited robust neuronal responses, as judged by online rasters. Furthermore, we restricted the data analysis only to recordings of neuronal activity characterized by high signal-to-noise ratios, indicating high single unit isolation quality (Figure 2).

Figure 2 about here

### 1.3 A Statistical Test for Minimally Poisson Spike Trains

Consider a series of spike trains from a single neuron obtained from  $n$  separate trials (each involving, for example, presentation of the same stimulus):

$$\{t_1^1, t_2^1, t_3^1, \dots, t_{m_1}^1\}, \{t_1^2, t_2^2, t_3^2, \dots, t_{m_2}^2\}, \dots, \{t_1^n, t_2^n, \dots, t_{m_n}^n\},$$

where there are  $m_j$  spikes in trial  $j$ , and  $t_i^j$  is the time of occurrence of the  $i$ 'th spike in the  $j$ 'th trial, relative to a stimulus onset at time 0.

A *homogeneous* Poisson process of rate  $\lambda$  is a statistical model of the spike train which is characterized by two conditions: i) the events in non-overlapping time intervals are independent, and ii) the expected number of spikes in a time interval of length  $T$  is  $\lambda T$ . An *inhomogeneous* Poisson process is the nonstationary analogue of this: events in non-overlapping time intervals are independent, but the expected number of spikes is governed by a *time-varying* rate  $\lambda(t)$ . An implication of these assumptions, for both homogeneous and inhomogeneous Poisson processes, is that the total number of spikes in any time interval is distributed as a Poisson random variable.

The mean  $\mu$  and variance  $\sigma^2$  of any Poisson random variable are identical, and therefore the variance-mean ratio  $\sigma^2/\mu$ , called the *Fano factor*, is equal to one. Perhaps for this reason, the Fano factor is employed frequently in neuroscience as a rate-controlled measure of the reliability of spike counts of neural responses.

The mean  $\mu$  and variance  $\sigma^2$  must be estimated from data. Usually the empirical statistics are employed:

$$\hat{\mu} := \frac{1}{n} \sum_{i=1}^n m_i \quad \hat{\sigma}^2 := \frac{1}{n-1} \sum_{i=1}^n (m_i - \hat{\mu})^2.$$

(Here we use  $\hat{\mu}$  to signify the estimate of  $\mu$ , et cetera; see Kass et al., 2004, for an introductory overview of statistical concepts in the context of neuroscience.) The estimates  $\hat{\mu}$  and  $\hat{\sigma}^2$  will converge in a sense of probability to the correct distributional quantities,  $\mu$  and  $\sigma^2$ , as the number of trials  $n$  increases, provided that  $m_1, \dots, m_n$  are independent and identically-distributed.

Typically the Fano Factor and Poisson standards are evaluated by comparing the relationship between the empirical mean  $\hat{\mu}$  and the empirical variance  $\hat{\sigma}^2$  of spike counts across a population of neurons and response conditions. This is done visually with a scatterplot and/or analytically, either by fitting a population of estimated mean and variance pairs  $(\hat{\mu}, \hat{\sigma}^2)$  to a power law and examining the fitted coefficients, or by reporting the average of empirical Fano Factors  $\hat{\sigma}^2/\hat{\mu}$  across a population (see McAdams and Maunsell, 1999, Berry et al., 1997, Buracas et al., 1998, Gur et al., 1997, and Muller et al., 2001, for some representative examples). Thus, estimated quantities (i.e., the estimated mean  $\hat{\mu}$ , estimated variance  $\hat{\sigma}^2$ , and estimated Fano Factor  $\hat{\sigma}^2/\hat{\mu}$ ) are treated interchangeably with their associated distributional quantities (i.e., the mean  $\mu$ , variance  $\sigma^2$ ,

and Fano Factor  $\sigma^2/\mu$  of the distribution presumed to generate the data). This poses fewer problems when these quantities are used solely for comparative analysis (for example, to compare the responses of different classes of neurons, Kara et al., 2000; or different stimulus conditions, McAdams and Maunsell, 1999), particularly when nonparametric methods are used. As an evaluation of the Poisson model, however, ignoring the sampling variability of these estimators could potentially introduce approximation error into the final conclusions. Standard descriptions of methodology (Teich et al., 1997; Gabbiani and Koch, 1998; Koch, 1999; Dayan and Abbott, 2001) do not discuss or account for this source of error.

Such issues are moot if there *are* no such true distributional quantities, as when the assumptions which postulate such quantities are not valid. Thus, while we do address this aforementioned gap, the issue of sampling variability is not our major concern here. Rather, we are concerned with the assumption that, even under identical experimental conditions, the responses of a neuron can be reasonably modeled as repeatable quantities. Indeed, the bulk of statistical theory and methods applies to sequences of observations that can reasonably be modeled as identically distributed. It is therefore not surprising that the bulk of applications of statistics to neurophysiological data are made under the assumption, sometimes implicit, that repeated measurements from the same neuron or collection of neurons under the same experimental conditions produces a sequence of identically distributed observations. Thus, a large value for the Fano factor (see Shadlen & Newsome, 1998; Dayan & Abbott, 2001; Koch, 1999; Kara et al., 2000, and references therein) is surprising and meaningful only in proportion to our willingness to believe in the proposition of trial-to-trial statistical stationarity. And sophisticated point-process models of spike trains that accommodate inhomogeneous firing rates and/or spike-to-spike dependencies (Berry and Meister, 1998; Kass and Ventura, 2001) are useful in proportion to our ability to estimate time-varying rates and other high-dimensional parameters. Yet in the absence of statistically repeatable observations the estimation of such quantities is difficult, and even raises issues of identifiability.

Given the magnitude of unknown and unmeasurable internal and external variables, assumptions of repeatability warrant concern. We derive here an exact statistical test of the Poisson nature of the spike train process, broadly defined, that makes no assumption about trial-to-trial stationarity, and even allows for some measure of trial-to-trial statistical dependence (see §2.1). On these terms, one could hardly expect to find a test that gives evidence for excess *variability* in the spike-generation process. But on the other hand, it is possible to reject a broad class of models in the direction of excess *regularity* without appealing to an assumption of trial-to-trial stationarity. Loosely speaking, allowing for additional variability in the form of trial-to-trial variation does not make observations of regularity any less surprising.

Our null hypothesis ( $H_0$ ) is that  $m_1, m_2, \dots, m_n$ , are independent (but not necessarily identically distributed) Poisson random variables. This is a *minimal* Poisson hypothesis in the sense that it contains, as a special case, the hypothesis that the recordings come from independent, possibly inhomogeneous, and possibly differing from trial to trial, Poisson processes.

In §2.1 we will derive the following exact statistical test, the “Poisson variability test,” for  $H_0$ :

$$\text{Reject } H_0 \text{ at level } \alpha \text{ if } \sum_{i=1}^n m_i^2 \leq f(n, \alpha, \hat{\mu})$$

The critical value,  $f(n, \alpha, \hat{\mu})$ , can be explicitly computed and tabulated (Amarasingham, 2004). But in practice the actual significance of a *given* data set is more to the point: for each  $n$  (number of trials),  $N$  (total number of spikes,  $n\hat{\mu}$ ), and  $S$  (the *statistic*,  $\sum_{i=1, \dots, n} m_i^2$ ), we compute the  $p$ -value  $\alpha(n, N, S)$ , which is the lowest probability at which the Poisson variability test rejects the null.  $P$ -values can be computed exactly by dynamic programming or approximately by Monte Carlo sampling.<sup>2</sup> Dynamic programming is computationally feasible for smaller values of  $N$  and  $S$ , and was used in all of the experiments reported here. The Monte Carlo approach is feasible in essentially any situation and has the virtue of simplicity, as is evident from the code listing provided in Appendix A. One needs to choose the number of Monte Carlo samples to be consistent with the desired accuracy, but this requires only a simple application of the binomial distribution. As a benchmark, 10,000 Monte Carlo samples yield an alpha value with 95% confidence interval no larger than  $\pm 0.01$ . Matlab code for running either or both algorithms is available for download at <http://www.dam.brown.edu/ptg/REPORTS/amarasingham/pvt.html>, as is a copy of Amarasingham (2004), which contains a complete discussion of both algorithms.

Under the constraint  $m_1 + m_2 + \dots + m_n = n\hat{\mu}$ , the sum of squares  $m_1^2 + m_2^2 + \dots + m_n^2$  is smallest when the counts  $m_1, m_2, \dots, m_n$  are most nearly equal. (This is in fact a corollary of the more familiar statement that the variance measures the “spread” of a distribution, since  $\frac{1}{n} \sum_{i=1}^n (m_i - \hat{\mu})^2 = (\frac{1}{n} \sum_{i=1}^n m_i^2) - \hat{\mu}^2$ .) Therefore, the test has power towards alternative hypotheses that would predict more repeatable observations of spike counts than the minimal Poisson hypothesis.

---

<sup>2</sup>We do not explore here a third approach, which is to approximate the multinomial probability which arises in §2.1 classically, with the  $\chi^2$  distribution.

## 2 Results

### 2.1 Derivation of the Minimal Poisson Variability Test

Recall that the null hypothesis ( $H_0$ ) is that  $m_1, m_2, \dots, m_n$  are independent Poisson random variables. Let us designate the (unknown) means as  $\lambda_1, \lambda_2, \dots, \lambda_n$ . Define the empirical mean and variance statistics as usual:

$$\hat{\mu} := \frac{1}{n} \sum_{i=1}^n m_i \quad \hat{\sigma}^2 := \frac{1}{n-1} \sum_{i=1}^n (m_i - \hat{\mu})^2$$

A hypothesis test specifies a region of the space of outcomes (called the *critical region*), which is unlikely for all distributions in the null hypothesis. The null hypothesis is “rejected” when observations fall in the critical region. The *power* of a test with respect to an alternative distribution is the probability that the alternative distribution assigns to the critical region, i.e. the probability that the null hypothesis is rejected when the alternative is true. We seek a hypothesis test under our null which has higher power for alternative distributions in which  $\hat{\sigma}^2/\hat{\mu}$  (i.e. the empirical Fano factor) tends to be small, or more to the point, such that  $\hat{\sigma}^2$  is small, given  $\hat{\mu}$ . It is equivalent to use  $\sum_{i=1}^n m_i^2$  in place of  $\hat{\sigma}^2$ , since they preserve the same order when  $\hat{\mu}$  is fixed.

One way to proceed is to form a partition of the sample space based on  $\hat{\mu}$ . We seek  $f(n, \alpha, \hat{\mu})$  such that

$$P \left( \sum_{i=1}^n m_i^2 \leq f(n, \alpha, \hat{\mu}) \middle| \hat{\mu} \right) \leq \alpha \quad \text{for all } \hat{\mu} \quad \text{for all } P \in H_0. \quad (1)$$

Then the event  $\{\sum_{i=1}^n m_i^2 \leq f(n, \alpha, \hat{\mu})\}$  will be an  $\alpha$ -level hypothesis test, since:

$$\begin{aligned} P \left( \sum_{i=1}^n m_i^2 \leq f(n, \alpha, \hat{\mu}) \right) &= \sum_{\hat{\mu}} P \left( \sum_{i=1}^n m_i^2 \leq f(n, \alpha, \hat{\mu}) \middle| \hat{\mu} \right) P(\hat{\mu}) \leq \sum_{\hat{\mu}} \alpha P(\hat{\mu}) \\ &= \alpha \quad \text{for all } P \in H_0 \end{aligned} \quad (2)$$

To derive  $f(n, \alpha, \hat{\mu})$ , it is straightforward to apply the following proposition.

**Proposition.** *If  $m_1, m_2, \dots, m_n$  are independent Poisson random variables with means  $\lambda_1, \lambda_2, \dots, \lambda_n$ , respectively, then for all  $r$ , and for all  $\hat{\mu}$ , we have*

$$\max_{\lambda_1, \lambda_2, \dots, \lambda_n} P \left( \sum_{i=1}^n m_i^2 \leq r \middle| \hat{\mu} \right) = P \left( \sum_{i=1}^n X_i^2 \leq r \right) \quad (3)$$

where  $X_1, X_2, \dots, X_n$  are distributed multinomially with parameters  $\{n\hat{\mu}; 1/n, 1/n, \dots, 1/n\}$ .

A multinomial distribution generalizes the binomial distribution to a many-sided die, and has the form

$$P(X_1 = m_1, X_2 = m_2, \dots, X_n = m_n) = \begin{cases} \binom{N}{m_1 m_2 \dots m_n} \prod_{i=1}^n p_i^{m_i} & \text{if } \sum_{i=1}^n m_i = N \\ 0 & \text{otherwise,} \end{cases}$$

where  $p_1, p_2, \dots, p_n$  satisfy  $p_i \geq 0$  and  $\sum_{i=1}^n p_i = 1$ . In shorthand,  $X_1, \dots, X_n \sim \mathcal{M}(N; p_1, p_2, \dots, p_n)$ . This proposition arises from the following observation: *conditioned* on  $\hat{\mu}$ , the average number of spikes, the counts  $m_1, m_2, \dots, m_n$  are distributed multinomially with parameters

$$\left\{ n\hat{\mu}; \frac{\lambda_1}{\sum_{i=1}^n \lambda_i}, \frac{\lambda_2}{\sum_{i=1}^n \lambda_i}, \dots, \frac{\lambda_n}{\sum_{i=1}^n \lambda_i} \right\}. \quad (4)$$

That is, if the firing rates  $\lambda_1, \lambda_2, \dots, \lambda_n$ , and the total spike count are known, one could sample from the minimal Poisson model by assigning each of the  $n\hat{\mu}$  spikes to a trial based on the outcome of an  $n$ -sided die in which each face is biased in proportion to the firing rate ( $\lambda_i$ ) of the trial to which it corresponds. This lends itself to the conjecture that the empirical variance of the  $m_i$ , conditioned on  $\hat{\mu}$ , will be *most* likely to be large exactly when the  $n$ -sided die is unbiased. (An unbiased die is one which produces all sides with equal likelihood). We formalize this with the proposition, which is not difficult to prove. The proof and other technical details are contained in Amarasingham (2004) .

Thus if

$$f(n, \alpha, \hat{\mu}) := \max \left\{ k : P \left( \sum_{i=1}^n X_i^2 \leq k \right) \leq \alpha \right\}$$

where  $X_1, \dots, X_n \sim \mathcal{M}(n\hat{\mu}; 1/n, 1/n, \dots, 1/n)$

then, in light of equations (1), (2), and (3),  $\{\sum_{i=1}^n m_i^2 \leq f(n, \alpha, \hat{\mu})\}$  is an  $\alpha$ -level hypothesis test for  $H_0$ .

In practice, the interest is in the  $p$ -value, given an observed value of  $\sum_{i=1}^n m_i^2$ . For example, if over four trials one observes spike counts of  $\{2, 3, 1, 4\}$ , then  $\sum_{i=1}^n m_i^2 = 2^2 + 3^2 + 1^2 + 4^2 = 30$ , and one needs to compute  $P(\sum_{i=1}^4 X_i^2 \leq 30)$  where  $X_1, X_2, X_3, X_4 \sim \mathcal{M}(10; 1/4, 1/4, 1/4, 1/4)$ . One straightforward way to do this is by simulation (a so-called Monte Carlo approximation): simulate these multinomially distributed random variables several times and count the proportion of times in which the sum of squares of the simulated random variables is less than 30 to approximate  $P(\sum_{i=1}^4 X_i^2 \leq 30)$ . This simple algorithm is expressed in Matlab code in Appendix A.

A careful reading of the derivation above indicates that the Poisson variability test covers a somewhat more general null hypothesis:  $m_1, m_2, \dots, m_n$  are *conditionally* independent and Poisson distributed, given the Poisson means  $\lambda_1, \lambda_2, \dots, \lambda_n$  (see Appendix B). Thus the null hypothesis includes: a) the standard model of inhomogeneous Poisson processes with a fixed rate function across trials (this is the case  $\lambda_1 = \lambda_2 = \dots = \lambda_n$ ), and b) trial-varying inhomogeneous Poisson processes (as stated above), but also c) processes in which a particular inhomogeneous rate function is chosen from an ensemble of possible rate functions for each stimulus presentation, randomly, *but not necessarily independently*.

## 2.2 Analysis of Cell Recordings

Our interest in the trial to trial variability of inferotemporal cells emerged as a consequence of the simple impression that, given an appropriate stimulus, neuronal responses at least *seemed* reliable, especially near the onset of the stimulus (Figure 3). Given the common view that cell firing patterns throughout cortex are invariably irregular, we set out to more rigorously explore this issue.

Figure 3 about here

In order to tease apart the effects on variability of time after stimulus onset, we divided the neural response into epochs, consisting of disjoint equal-length time intervals, beginning 100 ms after the onset of presentation of a stimulus. We used intervals of length 100 ms and 50 ms in the experiments described below. These choices were made for empirical reasons and are not due to limitations imposed by the test, which can be used for time periods of any length. The pairing of a cell and a stimulus presented repeatedly to the cell we labelled a *cell-stimulus pair*. For a fixed epoch, associated with each particular cell-stimulus pair are the data points  $m_1, m_2, \dots, m_n$  of spike counts for each of the  $n$  presentations of the stimulus to that cell during the epoch. For each cell-stimulus pair,  $m_1, m_2, \dots, m_n$ , (in particular the statistics  $\sum_{i=1}^n m_i$  and  $\sum_{i=1}^n m_i^2$ ), determine the  $p$ -value, the minimal value of  $\alpha$  that the Poisson variability test will reject, which we compute as per above.

We analyzed 328 cell-stimulus pairs, drawn from recordings from a total of 27 cells (18 from the first monkey and 9 from the second). These cells were selected on the basis of their clear visual response to at least one visual stimulus and based on the quality of the single unit isolation, which was as high as possible to prevent the uncertainty introduced by a noisy signal from contaminating our data. The number of presentations  $n$  varied with each cell-stimulus pair, ranging from  $n = 2$  to  $n = 14$ , with a median value of 7.

Using 100 ms epochs, 47 of the 328 pairs were rejected by the Poisson variability test at a significance level of 5% in the 100 to 200 ms period following stimulus onset. In contrast, the number of rejected cell-stimulus pairs decreased rapidly and progressively at subsequent epochs, further away from onset: 19 rejected pairs between 200 to 300 ms, 11 rejected pairs between 300 and 400 ms, and fewer further on. Similar trends are evident for 50 ms partitions, as illustrated in Figure 4.

#### *Pooling of results*

It would not be unexpected for a drug that has no effect to nevertheless show a “significant benefit” in three out of one hundred separate trials, if, say,  $\alpha = .05$ . After all, even if the null hypothesis is true, we can expect to reject it an average of 5/100 times when  $\alpha = .05$ . How surprising, then, are the *number* of rejections in each of the tested epochs? One way to assess the significance of the pooled data, within an epoch, is to assume the validity of the null hypothesis and the independence of the tests: the probability of any given number of rejections is then governed by the binomial distribution with parameter  $p = \alpha$ . But in the case of the Poisson variability test as applied here, this is too conservative. In fact, because the statistic  $\sum_{i=1}^n m_i^2$  is an integer, and because of the small numbers of samples involved in many of these tests, the actual probability of rejecting the null, when the null is true, can be substantially less than  $\alpha$  (where  $\alpha$  is used to determine the critical value  $f$ ), and in some cases even zero.

One possible approach to this problem is to include only data with a high number of samples. For example, repeating the analysis described above but including only the 45 cell-stimuli pairs with at least 10 presentations (i.e.,  $n \geq 10$ ), we obtain 15 rejections out of 45 cell-stimulus pairs in the 100 to 200 ms epoch, 2 rejected pairs in the 200 to 300 ms epoch, 4 rejected pairs in the 300 to 400 ms epoch, and 0 rejected pairs for all subsequent epochs. This approach is conservative, too, however, as any selection of a minimal threshold for the number of samples will be arbitrary: spike counts can still exhibit a surprising amount of regularity even with a few samples.

A better approach is to account for the actual significance values associated with individual tests as a function of sample size. Code for computing the actual significance levels is available at <http://www.dam.brown.edu/ptg/REPORTS/amarasingham/pvt.html>. Taking this into account, we can compute exactly the distribution on the number of rejections as a sum of Bernoulli (reject/accept) variables with rejection probabilities given by the actual significance levels (see also Amarasingham, 2004). This gives the exact significance of the observed number of rejections. When applied to the data, this pooling method reveals a consistent trend: the significance of the *number of rejections* is very high for the 100 ms to 200 ms period (significance  $1 \cdot 10^{-33}$ ), and decreases rapidly at subsequent epochs. This and a similar trend for the 50 ms partitions are also illustrated in Figure 4.

Figure 4 about here

### *Refractory period*

One of the most natural biophysical objections to Poisson models of neural activity is the refractory period (Berry and Meister, 1998), which, among other things, introduces strong short-term dependencies in spike train structure (“absolute refractory period”). A statistical test which is sensitive to reliability in the spike counts may merely reflect such local structures (or even not-so-local structures, as in the “relative refractory period”), particularly among cell-stimuli pairs with relatively high firing rates, since the effect of the refractory period on reliability will increase with the firing rate. Indeed, Kara et al. (2000) argue that, in retina, LGN, and V1, low spike count variability relative to the Poisson process is largely explained by absolute and relative refractory periods. One way to examine this interpretation for our data is to compare the results of the variability test across epochs, among cell-stimuli pairs that have the same mean spike counts. If the observed (super-Poisson) reliability is largely due to the refractory period or another interactive effect imposed on an “otherwise” Poisson process (e.g. a Poisson-refractory process, perhaps of the type proposed by Kass and Ventura (2001) ), then all other things being equal one would expect the effect to be independent of the time of occurrence relative to stimulus onset. Figure 5 provides a scatter plot of firing rate versus  $p$ -value, for cell-stimulus pairs separated by epoch. While there is a potential refractory period effect manifested by high firing rates near stimulus onset (the proportion of significant pairs grows along the  $x$ -axis in the first epoch), that is not enough to explain all the trends in the data. For a given window of firing rates (e.g. 40 Hz to 60 Hz), the fraction of significant cell-stimulus pairs is highest in the first epoch (100 ms to 200 ms), than in later ones. There is evidently a systematic effect of epoch on spike count reliability, which is independent of firing rate, and which cannot be explained by refractory period alone.

Figure 5 about here

### *Effect of eye position*

In an attempt to more carefully characterize the variable response observed in primary visual cortical cells, Gur et al. (1997) found that the use of moving stimuli coupled with precise control of stimulus position on the retina lowered the variance to mean ratio compared to previously reported values. In the present experiment, we did not reposition the stimulus in real time, but we did record eye position throughout each trial. Figure 6 shows one measure of the variability

due to eye movements during our task, by plotting the standard deviation of the monkeys' horizontal and vertical eye position as a function of time. In general, the variability at each time point was quite low (less than 0.25 deg), but there is a notable increase in this variability starting approximately 300 ms after the stimulus appeared. Even during periods of controlled fixation, eye movements are not completely abolished, and we cannot, therefore, rule out the possibility that increased positional uncertainty contributed to higher variability in later temporal epochs.

Figure 6 about here

### 3 Discussion

#### *Interpreting reliability and variability in neuronal responses*

There is an asymmetry in the implications of reliability versus variability in neural responses. Reliability is the more revealing of the two phenomena. For example, suppose that a neuron's outputs encode information by a temporally reliable spiking process, which would in turn imply reliability in the spike counts. Then, *provided* that the variable being encoded is kept constant or under control in repeated experimental trials, we would expect to observe spike count reliability in observations of the output of that neuron. If it is not held under control, however, then we might observe variability in the spike counts through the variation of the unknown encoding variable alone. Thus, under a temporally reliable coding hypothesis, an observation of spike count reliability provides the experimentalist with an additional clue that the neuron is actually encoding something which is being held under experimental control. The clue of reliability can usefully operate in addition to, or even separately from, that of elevated firing rates, which is the chief means by which hypotheses concerning encoded variables are evaluated in neurophysiology. Observations of variability, on the other hand, are subject to multiple explanations: either a reliable coding hypothesis is invalid, or the encoded variable is not being held under experimental control. These latter two alternatives, in the case of observed variability, will be quite difficult to distinguish.

This has implications for how we interpret observations of reliability and variability in the neural records. At one extreme, Poisson-like or super-Poisson variability in spike counts is simply accepted as a general principle, and models have been developed to explain this variability. For example, Shadlen and Newsome (1998) postulate mechanisms of synaptic integration designed specifically to preserve the response variability of neurons from one processing stage to the next, motivating this as one of the key features to be explained in cortical physiology.

Based on our data, we suggest that all cortical neurons should not be characterized as a static, homogeneous population with regard to variability of spike discharge. The rejection of the minimal Poisson hypothesis reported here does not imply that *all* cortical cells exhibit high reliability – only that some apparently do under some, particular, conditions. Many authors have remarked on the potential utility of reliable responses, particularly in the context of neural coding, and so identifying the conditions which generate them suggests itself as an important research topic. One interesting, possible, scenario is that the reliability of a cell’s discharge is state dependent. Levels of attention or motivation, for example, which are known to affect a cortical cell’s firing rate (see, Reynolds and Chelazzi, 2004, for a review) and to alter levels of neuronal synchronization (Fries et al., 2001), may also influence the variability of neural response.

#### *Behavioral relevance*

The present data were collected under conditions where the recorded neural signal was in no obvious way related to the animals’ behavior. However, numerous previous studies have demonstrated the importance of inferotemporal cortex in successful execution of complex pattern recognition tasks (Logothetis and Sheinberg, 1996). The speed with which monkeys can perform these tasks is of special interest, because they give some hints as to when, and for how long, neuronal signals emanating from IT cortex may be integrated to drive recognition behaviors. Using stimuli similar to those employed in this study, for example, Vogels (Vogels, 1999a) found that monkeys could accurately categorize tree and non-tree stimuli in under 300 ms. More recently, Sheinberg et al. (unpublished observations) found that highly similar complex visual images could be discriminated by button response in approximately 425 ms, but that clear evidence for stimulus identity was evident in the monkeys oculomotor behavior approximately 200 ms following stimulus onset. This upper bound of 200 ms is intriguing because it leaves little time for extensive averaging of highly variable signals. Indeed, the average onset latency of selective neuronal responses in IT cortex falls somewhere between 100 ms and 140 ms (Perrett, Rolls, and Caan, 1982; Vogels, 1999b; Sheinberg and Logothetis, 2001). Thus, the temporal epoch between initial IT cell activation and discriminatory motor response is quite short, on the order of 100 ms or less. Interestingly, it is precisely during this epoch where we find that individual neurons are most likely to respond reliably to particular visual stimuli.

#### *Models*

Identifying models which usefully capture the variability of a cortical spike train remains a central problem in the statistical modeling of neural data. A commonly-invoked model is the slow-varying inhomogeneous Poisson process, which follows in a natural way from a rate coding viewpoint (Shadlen and Newsome, 1998): the *signal* embedded in a neural spike train is an underlying *rate*,

varying on coarse time intervals on the order of tens to hundreds of milliseconds, and hence the precise placement of spikes is random and irrelevant. Poisson processes arguably underly a majority of the descriptive and analytic techniques still being employed in neural data analysis.

On the other hand, several studies speak to the inappropriateness of inhomogeneous Poisson models for spike trains (Kass, Ventura, and Brown, 2004). One explicit reason for this involves short-term dependencies between spikes, as imposed by biophysical phenomena such as the refractory periods and bursting. However, non-Poisson behavior on such fine timescales does not necessarily imply that the spike counts on coarser timescales are not well-approximated by a Poisson process. At issue are conclusions of the form exemplified in Koch (1999, p. 355; see also Dayan & Abbot, 2001, p. 32): “[These results] support the hypothesis that once the refractory period is accounted for, a Poisson hypothesis constitutes a first-order description of cortical spike trains.”. Similarly, Kara et al. (2000) argue that much of the super-Poisson reliability which they report, via a Fano factor analysis, might be accounted for by refractory effects which are differentially affected by varying firing rates. In contrast, our rejection of the minimal Poisson hypothesis is not easily explained by a simple (time-independent) refractory period combined with an otherwise Poisson process, at least in the absence of other forms of regularity in the firing rate, since the rejection occurs near stimulus onset and not further away even among populations of cell-stimulus pairs with similar average firing rates.

Another objection to standard inhomogeneous Poisson models arises from the implicit assumption of statistical stationarity across trials. The typical Fano factor analysis provides one example of such an assumption. The relation of the Fano factor to inhomogeneous Poisson processes, even as a heuristic, requires that there be no variability across trials in the parameters of the process being observed. Such an assumption is the norm in neural data analysis, yet in its absence the significance of many widely-employed descriptive statistics, such as the peri-stimulus time histogram, the coefficient of variation, and auto- and cross-correlograms, is unclear, at best.

Avoiding assumptions of trial-to-trial stationarity requires a more delicate approach. The Poisson variability test is one example. Another example arises naturally in *reconstruction* paradigms (Brown et al., 1998; Brockwell, Rojas, and Kass, 2004; Wu et al., 2004; Zhang et al., 1998; Shoham et al., 2005), in which the relationship between behavioral or external variables and the firing rate are explicitly encoded in a model. Once fitted to data, such models can be used to predict, or reconstruct, the behavioral variables from the spike times. However, evaluating the validity of Poisson spike count models by model-fitting (Brown et al., 1998; Brown et al., 2002; Shoham et al., 2005) requires that the postulated firing rate models are valid and that their parameters have accurately been estimated. Justi-

fying such an assumption will often be difficult. In particular, conclusions of excess variability (Brown et al., 1998; Fenton and Muller, 1998) can perhaps always be attributed to the effects of additional variables that have not been modeled.<sup>3</sup>

In this light, a rejection of the minimal Poisson null hypothesis is unusual because it does not only demonstrate that a single model is not a good fit to the data, but that an entire class of models is inappropriate.

From a pedagogical point of view, the rejection of a null hypothesis implies only that: the rejection of a hypothesis. It does not identify an alternative. However, the form of the statistical test, and in particular the distribution of its power among the space of alternative hypotheses, may shed light on the data-generating process. The Poisson variability test is based on a measure of the reliability in the spike counts, and hence the rejection of the minimal Poisson hypothesis suggests that the spike train near stimulus onset may be a finely structured temporal process with high precision on many or most of the spikes.

In any case, there is strong statistical evidence that neuronal responses near the onset of a stimulus cannot in general be well modeled by a simple Poisson process, even allowing for a rate function that varies from trial to trial or exhibits transients and other time-dependent activity.

### *Small Samples*

The statistical significance of the results is somewhat surprising in light of the small sample sizes, ranging from 2 to 14 per cell-stimulus pair. Indeed, as mentioned previously, some of the observations are so small that rejection of  $H_0$  at 5% significance is mathematically impossible. For example, even if *every* 100 ms post-stimulus epoch had *exactly* 2 spikes (corresponding to the *most* reliable outcome at 20 Hz), the Poisson variability test would only achieve 5% significance with 4 or more trials. In fact, to take one example, based on the numbers of trials and numbers of observed spikes in the analysis of the 100 ms to 200 ms epoch, for 50% of the cell-stimulus pairs it was *a priori impossible* to reject the null hypothesis.

### *Summary*

To summarize, we have derived a method for testing the Poisson nature of neural spike trains. Applying this method to data recorded from primate inferotemporal cortex shows that the neural response is highly regular near stimulus

---

<sup>3</sup>Taking such a possibility to its logical conclusion, it could be that the critical variables are in fact unknown, or hidden. Hidden variables effectively act randomly, and thus accounting for them involves making the rate function  $\lambda(t)$  itself random (see also Ventura et al. (2005)). A Poisson process equipped with a *random* rate function is a Cox process (Cox, 1955). Nevertheless, from our point of view the important feature of Cox processes is that the spike counts are Poisson noise, once the instantaneous firing rate has been specified. As a consequence, spike counts generated by a Cox process are included in the minimal Poisson hypothesis here (see §2.1).

onset of relevant visual stimuli and not later on. These results provide a simple and rigorous test of a Poisson hypothesis which avoids the assumption of identical conditions across trials that is commonly made in neuroscience.

## 4 Appendix A

The following MATLAB function computes the  $p$ -value of the Poisson Variability Test  $\alpha(n, N, S)$  by Monte Carlo approximation (here using 10,000 Monte Carlo samples). For example, given a vector of spike counts contained in `counts`, one could obtain an approximate  $p$ -value by calling `pvtalpha_mc( length(counts), sum(counts), sum(counts.*counts) )`. These approximations can typically be computed in seconds.

```
function [pvalue] = pvtalpha_mc(n,N,S)
% Returns the p-value of the Poisson variability test
% by Monte Carlo approximation (here using 10,000 samples).
% n is the number of trials, N is the total number of spikes (across trials)
% and S is the sum of squares of the spike counts across trials (see text)
mc_iter=10000; % mc_iter specifies number of Monte Carlo samples to use
h=hist( ceil( unifrnd(0,1,N,mc_iter)*n ), n ); % generate multinomial samples
i=sum(h.*h,1); % compute sum of squares
pvalue=sum(i<=S)/mc_iter; % output pvalue
return
```

## 5 Appendix B

Here we demonstrate that the Poisson variability test remains valid under the (more general) null hypothesis that  $m_1, m_2, \dots, m_n$  are *conditionally* independent and Poisson distributed, given the Poisson means  $\lambda_1, \lambda_2, \dots, \lambda_n$ .

Let us denote the critical region  $\mathcal{C} = \{\sum_{i=1}^n m_i^2 \leq f(n, \alpha, \hat{\mu})\}$ . We have shown that, for any fixed  $\lambda_1, \lambda_2, \dots, \lambda_n$ ,  $P(\mathcal{C}) \leq \alpha$  if  $m_1, m_2, \dots, m_n$  are independent. As a consequence,  $P(\mathcal{C}|\lambda_1, \lambda_2, \dots, \lambda_n) \leq \alpha$ , for all  $\lambda_1, \lambda_2, \dots, \lambda_n$ , provided  $m_1, m_2, \dots, m_n$  are *conditionally* independent given  $\lambda_1, \lambda_2, \dots, \lambda_n$ . Therefore, under this assumption,

$$P(\mathcal{C}) = \sum_{\lambda_1, \lambda_2, \dots, \lambda_n} P(\mathcal{C}|\lambda_1, \lambda_2, \dots, \lambda_n)P(\lambda_1, \lambda_2, \dots, \lambda_n) \leq \sum_{\lambda_1, \lambda_2, \dots, \lambda_n} \alpha P(\lambda_1, \lambda_2, \dots, \lambda_n) = \alpha, \quad (5)$$

*independently* of the joint distribution on  $(\lambda_1, \lambda_2, \dots, \lambda_n)$ , which thus may be of any form.

## References

- Amarasingham, A (2004). Statistical methods for the assessment of temporal structure in the activity of the nervous system. Ph.D. diss., Brown University.
- Berry, M J and M Meister (1998). Refractoriness and neural precision. *Journal of Neuroscience* 18(6): 200–2211.
- Berry, M J, D K Warland, and M Meister (1997). The structure and precision of retinal spike trains. *Proc. Natl. Acad. Sci.* 94: 5411–5416.
- Brockwell, A E, A L Rojas, and R E Kass (2004). Recursive bayesian decoding of motor cortical signals by particle filtering. *Journal of Neurophysiology* 91: 1899–1907.
- Brown, E N, R Barbieri, V Ventura, R E Kass, and L M Frank (2002). The time-rescaling theorem and its application to neural spike data analysis. *Neural computation* 14(2): 325–346.
- Brown, E N, L M Frank, D Tang, M C Quirk, and M A Wilson (1998). A statistical paradigm for neural spike train decoding applied to position prediction from ensemble firing patterns of rat hippocampal place cells. *Journal of Neuroscience* 18(18): 7411–7425.
- Buracas, G T, A M Zador, M R DeWeese, and T D Albright (1998). Efficient discrimination of temporal patterns by motion-sensitive neurons in primate visual cortex. *Neuron* 20: 959–969.
- Cox, D R (1955). Some statistical methods connected with series of events. *Journal of the Royal Statistical Society, Series B* 17: 129–164.
- Dayan, P and L F Abbott (2001). *Theoretical Neuroscience*. MIT Press, Cambridge, Massachusetts.
- Fenton, A A and R U Muller (1998). Place cell discharge is extremely variable during individual passes of the rat through the firing field. *Proc Natl Acad Sci USA* 6: 3182–3187.
- Fries, P, J H Reynolds, A Rorie, and R Desimone (2001). Modulation of oscillatory neuronal synchronization by selective visual attention. *Science* 291: 1560–1563.
- Gabbiani, F and C Koch (1998). Principles of spike train analysis. In Koch, C and I Segev, editors, *Methods in neuronal modeling: from synapses to networks*, pp. 313–360. MIT Press.

- Gershon, E D, M C Wiener, P E Latham, and B J Richmond (1998). Coding strategies in monkey v1 and inferior temporal cortices. *Journal of Neurophysiology* 79(3): 1135–1144.
- Gur, M, A Beylin, and D M Snodderly (1997). Response variability in primary visual cortex (v1) of alert monkey. *Journal of Neuroscience* 17(8): 2914–2920.
- Kara, P, P Reinagel, and R C Reid (2000). Low response variability in simultaneously recorded retinal, thalamic, and cortical neurons. *Neuron* 27: 635–646.
- Kass, R E and V Ventura (2001). A spike-train probability model. *Neural Computation* 13: 1713–1720.
- Kass, R E, V Ventura, and E N Brown (2004). Statistical issues in the analysis of neuronal data. *Journal of Neurophysiology* 94: 8–25.
- Koch, C (1999). *Biophysics of Computation: Information Processing in Single Neurons*. Oxford University Press, New York, New York.
- Logothetis, N K and D L Sheinberg (1996). Visual object recognition. *Annual Review of Neuroscience* 19: 577–621.
- McAdams, C J and J H R Maunsell (1999). Effects of attention on the reliability of individual neurons in monkey visual cortex. *Neuron* 23: 765–773.
- Movshon, J A (2000). Reliability of neuronal responses. *Neuron* 27: 412–414.
- Muller, J R, A B Metha, J Krauskopf, and P Lennie (2001). Information conveyed by onset transients in responses of striate cortical neurons. *Journal of Neuroscience* 21(17): 6978–6990.
- Oram, M W, M C Wiener, R Lestienne, and B J Richmond (1999). Stochastic nature of precisely times spike patterns in visual system responses. *Journal of neurophysiology* 81: 3021–3033.
- Perrett, D. I., E. T. Rolls, and W. Caan (1982). Visual neurones responsive to faces in the monkey temporal cortex. *Exp Brain Res* 47(3): 329–42.
- Reynolds, J H and L Chelazzi (2004). Attentional modulation of visual processing. *Annual Review of Neuroscience* 27: 611–647.
- Shadlen, M N and W T Newsome (1998). The variable discharge of cortical neurons: implications for connectivity, computation, and information coding. *Journal of Neuroscience* 18(10): 3870–3896.

- Sheinberg, D L and N K Logothetis (1997). The role of temporal cortical areas in perceptual organization. *Proceedings of the National Academy of Sciences* 94: 3408–3413.
- Sheinberg, D L and N K Logothetis (2001). Noticing familiar objects in real world scenes: the role of temporal cortical neurons in natural vision. *Journal of Neuroscience* 21: 1340–1350.
- Shoham, S, L Paninski, M R Fellows, N G Hatsopoulos, J P Donoghue, and R A Normann (2005). Statistical encoding model for a primary motor cortical brain-machine interface. *IEEE Transactions on Biomedical Engineering* 52(7): 1312–1322.
- Softky, W R and C Koch (1993). The highly irregular firing of cortical cells is inconsistent with temporal integration of random epsps. *Journal of Neuroscience* 13: 334–350.
- Teich, M C, C Heneghan, S B Lowen, T Ozaki, and E Kaplan (1997). Fractal character of the neural spike train in the visual system of the cat. *Journal of the Optical Society of America A* 14(3): 529–546.
- Tolhurst, D J, J A Movshon, and A F Dean (1983). The statistical reliability of signals in single neurons in cat and monkey visual cortex. *Vision Research* 23: 775–785.
- Ventura, V, C Cai, and R E Kass (2005). Trial-to-trial variability and its effect on the time-varying dependence between two neurons. *Journal of Neurophysiology* 94: 2928–2939.
- Vogels, R. (1999a). Categorization of complex visual images by rhesus monkeys. part 1: behavioural study. *Eur J Neurosci* 11(4): 1223–38.
- Vogels, R. (1999b). Categorization of complex visual images by rhesus monkeys. part 2: single-cell study. *Eur J Neurosci* 11(4): 1239–55.
- Wu, W, M J Black, D Mumford, Y Gao, E Bienenstock, and J P Donoghue (2004). Modeling and decoding motor cortical activity using a switching kalman filter. *IEEE Transactions on Biomedical Engineering* 51(6): 933–942.
- Zhang, K, I Ginsburg, B L McNaughton, and T J Sejnowski (1998). Interpreting neuronal population activity by reconstruction: unified framework with application to hippocampal place cells. *Journal of Neurophysiology* 79: 1017–1044.

## 6 Figure Legends

Figure 1: Basic fixation task performed by the monkeys. In a single observation period, between three and five individual visual stimuli were flashed on and off as the monkey fixated a spot in front of the images. At the end of the trial, the monkey received juice reward for reacquiring the spot as it jumped from the center of the screen to one of four randomly selected peripheral locations.

Figure 2: Analog voltage trace of neuronal activity from a single trial from one of the recordings used in the analysis. The high signal-to-noise ratio here is typical of the units employed in this study.

Figure 3: Reliable response from a single cell, indicating that particular stimuli can elicit highly regular trial-to-trial spike discharges from a neuron. **A.** Response to three of the stimuli used during testing of this cell. Each plot includes the spike rasters aligned to the onset of the stimulus (time indicated by the left vertical green line). At the bottom of each plot is an estimate of the instantaneous firing rate. **B.** Enlarged view of the aligned spike times for all responses to the most effective stimulus shown in **A** (taken from two separate blocks). **C.** A temporally expanded view of the spiking activity near stimulus onset, with the number of spikes occurring in the 50 ms shaded area indicated to the right of each trial. Note that the mean of the counts far exceeds the variance of the counts (ratio: 7.6 times).

Figure 4: Summary of the rejections from the Poisson variability test across a total of 328 cell-stimulus pairs for each epoch. The 800 milliseconds following stimulus onset were partitioned into disjoint intervals of equal length called epochs. Epochs are labelled along the x-axis by the starting time of the interval with which they are associated. The first row (a and b) summarizes the results from the 100 ms epoch partition, and the second row (c and d) from the 50 ms epoch partition. The first column (a and c) illustrates the total number of rejections vs. epoch. The second column (b and d) graphs the epoch-by-epoch significance of the number of rejections of the Poisson hypothesis, towards excess regularity, under the assumption that the results of the individual cell-stimulus pair tests are independent. For the first three 100 ms epochs (100 ms, 200 ms, and 300 ms; first 3 bars, 1st row, 2nd column) the significances are  $10^{-33}$ ,  $10^{-9}$ , and  $10^{-3}$ ,

respectively. For the first four 50 ms epochs (100 ms, 150 ms, 200 ms, and 250 ms; first 4 bars, 2nd row, 2nd column) the significances are  $10^{-15}$ ,  $10^{-23}$ ,  $10^{-10}$ , and 0.0008.

Figure 5: A scatter plot of  $\log_{10} p$ -value vs. firing rate in hertz, across all cell-stimulus pairs, separated in 100 millisecond epochs. The horizontal dashed line is the line of 5% significance ( $\log_{10}(.05) \approx -1.301$ ). The firing rate dimension was partitioned into subintervals of 20 Hz range (depicted by the vertical dotted lines), and the proportion of cell-stimulus pairs for which the minimal Poisson hypothesis was rejected (rejection proportion, r.p.) is provided. The data is plotted separately for (a) the 100 to 200 ms epoch, (b) the 300 to 400 ms epoch, and (c) the 500 to 600 ms epoch, with respect to stimulus onset. (Note that no particular meaning is attached here to the choice of 20 Hz partitions.)

Figure 6: Eye position variability. The average standard deviation of both horizontal and vertical eye position for the 800ms period following stimulus onset indicates that although there was relatively little positional uncertainty, the variability did increase later in the trial, and this may have contributed to variability later in the neuronal response.

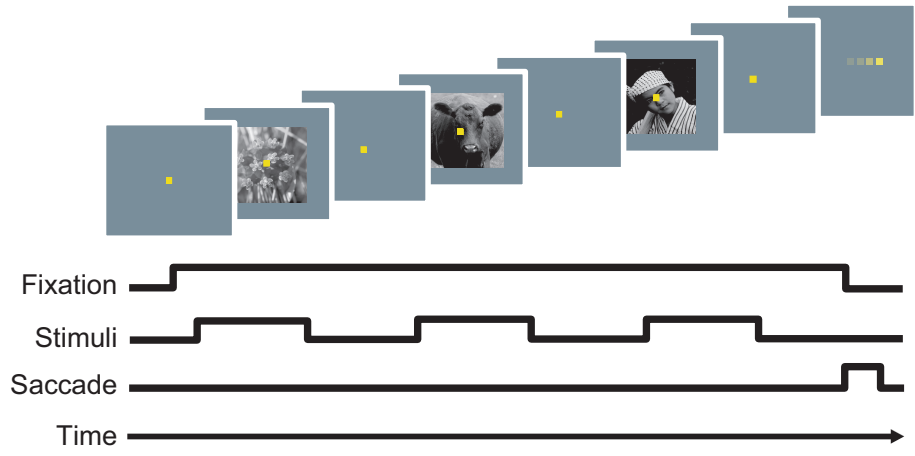


Figure 1: See Figure Legends above.

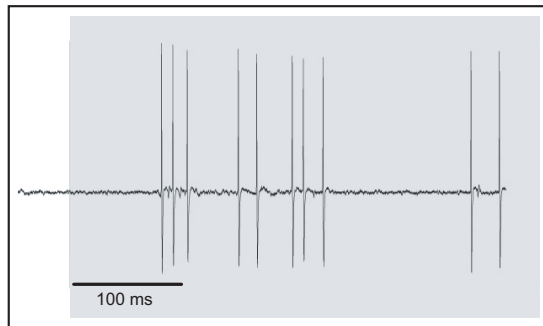


Figure 2: See Figure Legends above.

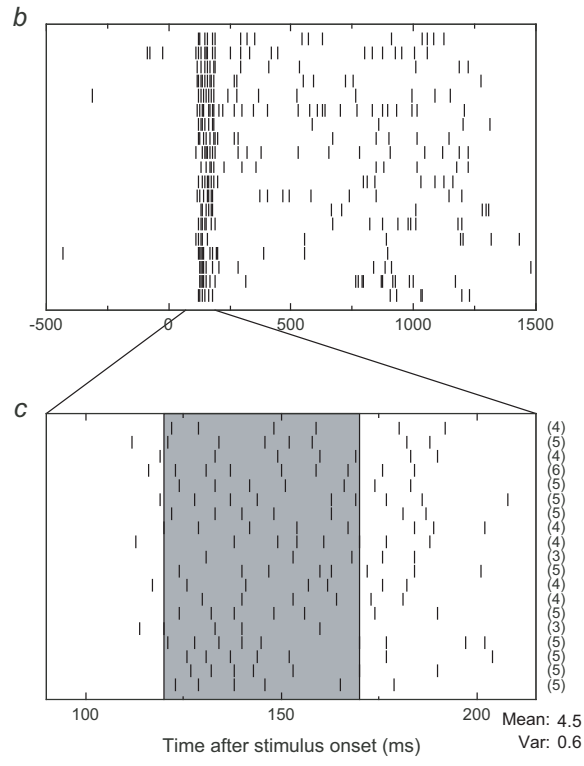
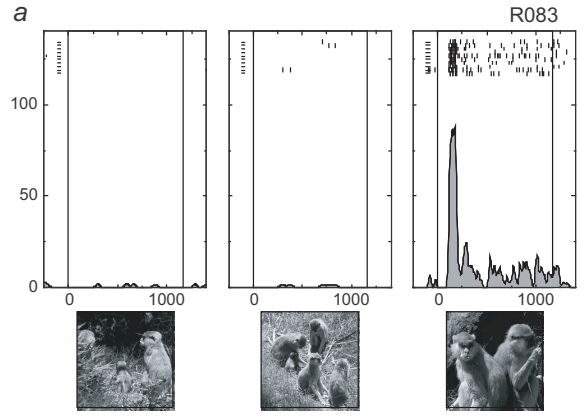


Figure 3: See Figure Legends above.

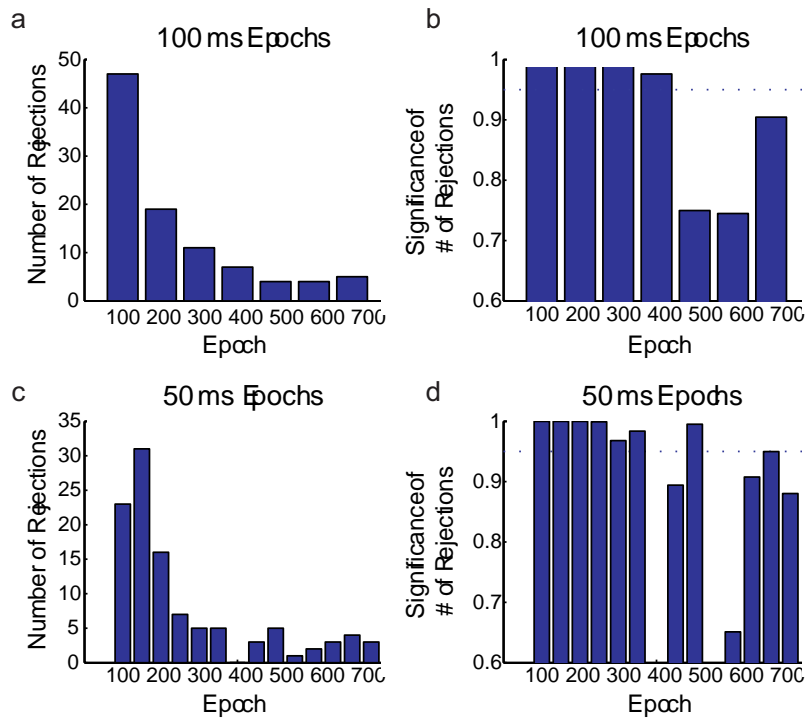


Figure 4: See Figure Legends above.

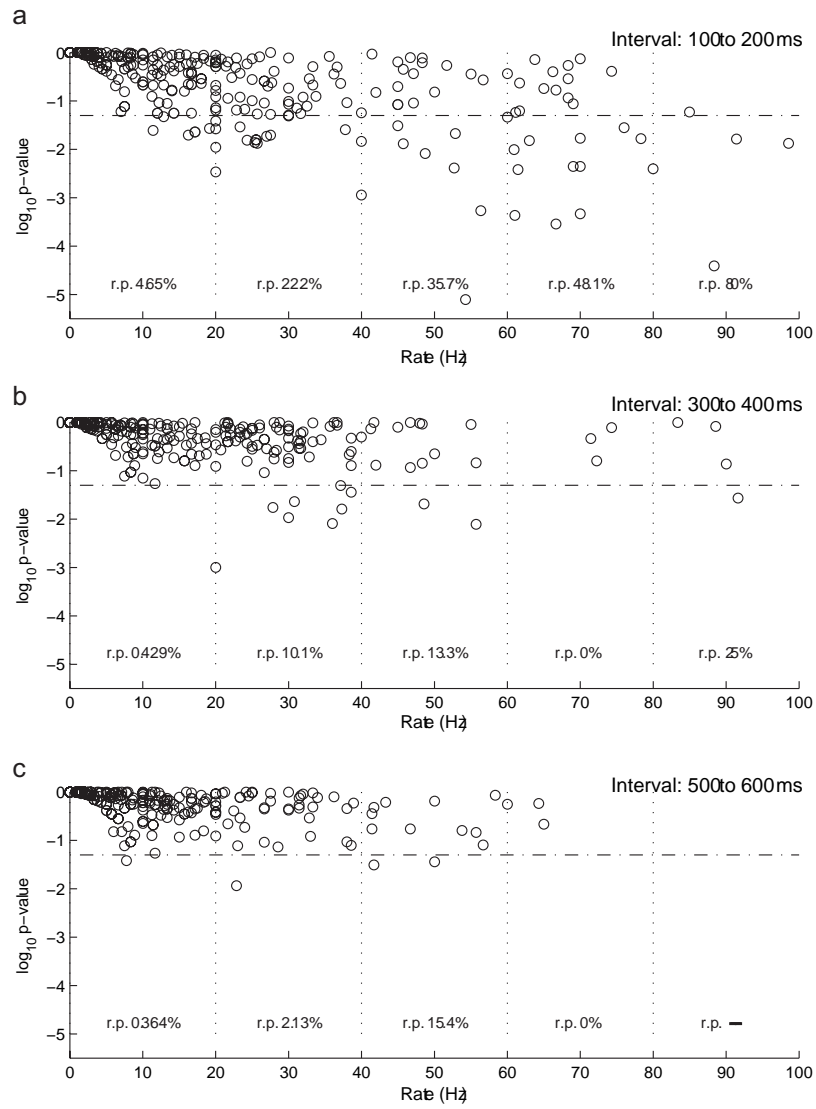


Figure 5: See Figure Legends above.

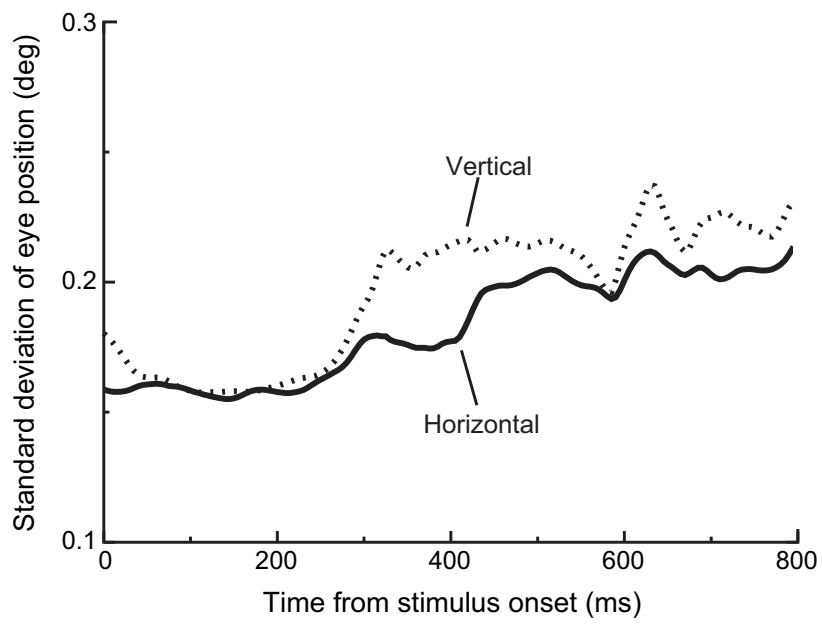


Figure 6: See Figure Legends above.