# Statistics of visual responses in primate inferotemporal cortex to object stimuli

### Sidney R. Lehky,<sup>1,2</sup> Roozbeh Kiani,<sup>3,4</sup> Hossein Esteky,<sup>3,5</sup> and Keiji Tanaka<sup>1</sup>

<sup>1</sup>Cognitive Brain Mapping Laboratory, RIKEN Brain Science Institute, Wako, Saitama, Japan; <sup>2</sup>Computational Neuroscience Laboratory, The Salk Institute, La Jolla, California; <sup>3</sup>Research Group for Brain and Cognitive Sciences, School of Medicine, Shaheed Beheshti University, Tehran, Iran; <sup>4</sup>Howard Hughes Medical Institute and Department of Neurobiology, Stanford University Medical School, Palo Alto, California; and <sup>5</sup>School of Cognitive Sciences, Institute for Research in Fundamental Sciences, Niavaran, Tehran, Iran

Submitted 12 November 2010; accepted in final form 2 May 2011

Lehky SR, Kiani R, Esteky H, Tanaka K. Statistics of visual responses in primate inferotemporal cortex to object stimuli. J Neurophysiol 106: 1097-1117, 2011. First published May 11, 2011; doi:10.1152/jn.00990.2010.-We have characterized selectivity and sparseness in anterior inferotemporal cortex, using a large data set. Responses were collected from 674 monkey inferotemporal cells, each stimulated by 806 object photographs. This 806  $\times$  674 matrix was examined in two ways: columnwise, looking at responses of a single neuron to all images (single-neuron selectivity), and rowwise, looking at the responses of all neurons caused by a single image (population sparseness). Selectivity and sparseness were measured as kurtosis of probability distributions. Population sparseness exceeded single-neuron selectivity, with specific values dependent on the size of the data sample. This difference was principally caused by inclusion, within the population, of neurons with a variety of dynamic ranges (standard deviations of responses over all images). Statistics of large responses were examined by quantifying how quickly the upper tail of the probability distribution decreased (tail heaviness). This analysis demonstrated that population responses had heavier tails than singleneuron responses, consistent with the difference between sparseness and selectivity measurements. Population responses with spontaneous activity subtracted had the heaviest tails, following a power law. The very light tails of single-neuron responses indicate that the critical feature for each neuron is simple enough to have a high probability of occurring within a limited stimulus set. Heavy tails of population responses indicate that there are a large number of different critical features to which different neurons are tuned. These results are inconsistent with some structural models of object recognition that posit that objects are decomposed into a small number of standard features.

monkey; extrastriate; object recognition; population coding; selectivity; sparseness; heavy tails; power law; Pareto distribution; stable distribution

INFEROTEMPORAL CORTEX is a high-level visual cortical area believed to be involved in object recognition (Desimone et al. 1984; Gross 2008; Logothetis and Sheinberg 1996; Rolls 2000; Tanaka 1996). Activation of inferotemporal neurons requires complex stimulus features compared with earlier areas (Kobatake and Tanaka 1994). Given that complexity, it is difficult to know if the most effective stimulus found for a neuron during experimental testing is likely to be close to the best possible stimulus, as only a limited number of images can be practically examined. The sample of images may give an unrepresentative view of what the neuron responds to. Similarly, only a limited sample of neurons in a population can be examined. How complex are the critical features of inferotemporal cells? Are they simple enough that something close to optimal is likely to be found within a limited set of randomly chosen object stimuli? And how many different critical features exist in a population? Are there just a few, as postulated by some structural models of object recognition, or are there a vast number of them, which may suggest other approaches to object recognition? We propose to examine such issues through a probabilistic analysis of responses using a large data set.

Responses of neurons that have been tested with a fixed set of visual stimuli can be measured along two dimensions (Fig. 1). First, we can measure the responses of a single neuron to different stimulus images ("single-neuron response"). From single-neuron responses, neural selectivity can be calculated. Second, we can measure the response of a population of neurons to a single image ("population response" or "image response"), from which population sparseness can be determined. As the equations for neural selectivity and population sparseness take the same form, it is convenient to refer to both as selectivity in a generic sense.

Although population sparseness has often been the parameter of theoretical interest, earlier studies tended to emphasize neural selectivity. Neural selectivity and population sparseness appeared to be implicitly treated as if they were interchangeable, both in the experimental literature (Rolls and Tovée 1995; Treves et al. 1999) and the modeling literature (Olshausen and Field 1997). However, Willmore and Tolhurst (2001) pointed out the importance of distinguishing the two concepts, and more recent work has incorporated consideration of that distinction in V1 (Lehky et al. 2005; Tolhurst et al. 2009) and inferotemporal cortex (Franco et al. 2007). The statistics of neural selectivity and population sparseness may or may not be identical. If their statistics are identical, then the system is called ergodic (Lehky et al. 2005). Franco et al. (2007) concluded that neural selectivity and population sparseness were identical in monkey inferotemporal cortex. Here we report, using a much larger data sample, that they are different. Selectivity, sparseness, and ergodicity are all issues we shall examine in detail.

The shape of the response probability distribution determines selectivity (with selectivity in the generic sense here). Example high-selectivity and low-selectivity probability density functions (pdfs) are shown in Fig. 2. High-selectivity responses have a greater probability of occurring at the ex-

Address for reprint requests and other correspondence: S. R. Lehky, Cognitive Brain Mapping Laboratory, RIKEN Brain Science Inst., Hirosawa 2-1, Wako-shi, Saitama 351-0198, Japan (e-mail: sidney@brain.riken.go.jp).



Fig. 1. Two ways of viewing neural response data. Statistical properties can be examined in terms of the response of a single neuron to a set of stimulus images or in terms of the response of a population of neurons to a single stimulus image.

tremes, very small or very large (i.e., heavier tails). Lowselectivity responses are more evenly distributed among intermediate levels.

When analyzing data we shall pay particular attention to the shape of the upper tail of the response probability distribution, as tail heaviness is a fundamental factor determining selectivity. Measuring tail heaviness allows an estimate of how likely we would be to obtain stronger neural responses if we were able to expand the stimulus set size, or the number of neurons recorded from, beyond what is experimentally practical. Tail heaviness has never been measured for visual neurons. Our approach will be to use a statistical technique called Pareto tail estimation. Because we are using a larger data set than those in previous studies, we are able to quantify tail characteristics of response probability distributions in a manner not feasible before.

Response statistics of neurons are of interest to theories of sparse representations (Bell and Sejnowski 1997; Field 1994; Hyvärinen et al. 2009; Olshausen and Field 1997; Simoncelli and Olshausen 2001). Characterizations of inferotemporal neurons presented here are relevant to those theories. Statistical properties of neural responses are also important to theories of population decoding (Jazayeri and Movshon 2006; Lehky and Sejnowski 1988; Pouget et al. 2000; Quian Quiroga and Panzeri 2009; Sanger 2003; Zhang et al. 1998). Within the field of object recognition, statistical characterization of inferotemporal neurons may constrain the manner in which object recognition theories are constructed. Previous examinations of inferotemporal selectivity or other probabilistic aspects of inferotemporal responses include those of Rolls and Tovée (1995), Baddeley et al. (1997), Treves et al. (1999), Krieman et al. (2006), Waydo et al. (2006), Franco et al. (2007), Lehky and Sereno (2007), Lehky and Tanaka (2007), and Zoccolan et al. (2007).

### METHODS

### Recording

Extracellular single cell recordings were collected from two macaque monkeys (*Macaca mulatta*) with tungsten microelectrodes. Recording site positions were estimated from the location of the guide tube, the electrode advancement indicated by micromanipulator readings, the estimated distribution and location of gray matter, and the depth of the ventral brain surface as indicated by changes in the characteristic noise when the electrode tip reached the ventral cortical surface. Cells were assigned to the following three cortical areas based on the criteria of Saleem and Tanaka (1996): superior temporal sulcus (STS), anterior dorsal TE (TEad), and anterior ventral TE (TEav). The STS region extended along the lower floor of the superior temporal sulcus until the lip of the sulcus. TEad extended from the lip of STS to the lateral lip of the anterior medial temporal sulcus (AMTS). Finally, TEav extended across the entire AMTS, including its lateral bank, and continued along the lateral half of the inferior temporal gyrus.

Penetration positions were evenly distributed over anterior 15–20 mm (*monkey 1*, right side) and anterior 13–20 mm (*monkey 2*, left side). Each penetration provided data from 3.6 cells on average, with 1.8 cells per recorded site. These values were similar for both STS and TE. Action potentials were isolated in real time with a template matching algorithm implemented in hardware (Wörgötter et al. 1986). Changing the template allowed data from more than one neuron to be recorded at a single site. A conventional spike window discriminator was also in place as an additional measure to ensure reliable isolation.

The selection of cells was not biased by response properties. All cells that remained reliably isolated throughout the stimulus presentation period were included in the data set (n = 674), regardless of selectivity.

This procedure would have missed cells that had zero spontaneous activity and that also had no response to any of the images presented as the electrode advanced (i.e., would have missed extreme highselectivity neurons). These "silent cells" on the upper tail of the selectivity probability distribution would be expected to be small in number, as our data showed a unimodal probability distribution of selectivities. There was no upswing in the distribution curve at high selectivities that would indicate the start of a second peak beyond our measurement range.

Recording procedures were in accord with National Institutes of Health guidelines as well as those of the Iranian Physiological Society. Further details of the recording methods have been described previously (Kiani et al. 2005, 2007).

The experimental procedures were submitted to the Committee for Care and Use of Experimental Animals, Iranian Society for Physiology and Pharmacology (CCUEA-ISPP) and were approved by this committee.

### Stimuli and Task

The stimulus set consisted of color photographs of natural and artificial objects, isolated on a gray background, each containing  $125 \times 125$ 



Fig. 2. Schematic probability density functions (pdfs) for high-selectivity and low-selectivity responses. Selectivity here can refer to either single-neuron selectivity or population selectivity (population sparseness). The high-selectivity pdf has a heavier tail than the low-selectivity pdf. These are generic examples intended to qualitatively illustrate the differences.

pixels. The larger dimension of each object extended  $\sim 7^{\circ}$ . Each neuron was tested with a mean of 1,271 images. All image presentations were repeated 9 ± 2 times (median: 10) to the neuron. Not all those images were the same for every neuron. To fill a complete response matrix (as in Fig. 1), we only used data from the 806 images that were presented to all 674 neurons. They included human, monkey, and nonprimate faces, human and animal bodies, fishes, reptiles, fruits, vegetables, trees, and various kinds of artifacts (examples are shown in Fig. 3 and the full set in Supplemental Fig. S1).<sup>1</sup>

The only task of the monkey was to fixate within  $2^{\circ}$  of a  $0.5^{\circ}$  fixation spot presented at the center of the screen. Eye position was monitored by an infrared eye-tracker.

At the start of a trial the monkey maintained fixation for 300 ms. After that, a series of images was presented with a rapid serial visual presentation (RSVP) paradigm (Földiák et al. 2004; Keysers et al. 2001). Each image was presented for 105 ms, followed immediately by the next image without gap. The order of images was chosen pseudorandomly. A trial lasted for a series of 60 images, or until the monkey broke fixation. The monkey received a juice reward every 1.5–2.0 s while maintaining fixation.

Our RSVP method enabled us to accurately measure neural responses to a large number of stimuli. It has been shown that cells in the monkey inferotemporal cortex retain stimulus selectivity during rapid serial presentations (De Baene et al. 2007; Edwards et al. 2003; Földiák et al. 2004; Keysers et al. 2001), down to as fast as 14–28 ms/stimulus. Also, backward masking has a minimal effect on the initial part of neuronal responses when the stimulus onset asynchrony is greater than 80 ms (Kovács et al. 1995; Rolls and Tovée 1994; Rolls et al. 1999).

The 105-ms stimulus duration we used fell within a behaviorally valid range, as indicated by fixation durations in humans during free-view scanning of images (Henderson and Hollingworth 1998). The probability distribution of fixation durations between saccades in their data had a mode of 230 ms and an interquartile range of 186 ms (estimated from Fig. 2 of Henderson and Hollingworth 1998). Chimpanzees have a shorter mean fixation duration than humans during free-viewing of images (Kano and Tomonaga 2009). Under natural viewing conditions, therefore, the visual system is exposed to a rapid sequence of short-duration stimuli. As the eye moves between fixa-

<sup>1</sup> Supplemental Material for this article is available online at the Journal website.



Fig. 3. Example stimulus images, selected from complete set of 806 images. Photos courtesy of Hemera Photo-Objects/Jupiterimages.

tions, saccadic suppression blocks response cross talk between responses to successive images (Bremmer et al. 2009).

### Spike Train Analysis

We measured neural activity for each stimulus presentation during a 140-ms window, offset by the earliest significant response within the inferotemporal population (70 ms) (Tamura and Tanaka 2001). The time window for data analysis thus ran from 71 to 210 ms after stimulus onset. All analysis procedures described below were based on the mean response to each image over that time period. Responses to the last two stimuli in each series were not included because the monkey did not maintain fixation for the whole period used for the measurement of the response. To minimize contamination of neural activity measurement by responses to the previous stimuli, we excluded presentations with large activity (exceeding the spontaneous activity by  $2 \times SD$ ) in the 50-ms period immediately after the stimulus onset. This resulted in exclusion of 15% of presentations. Spontaneous activity was measured in a 200-ms window at the start of each trial, preceding the series of stimulus presentations.

### Describing the Response Probability Distributions

We were interested in determining which standard parametric probability distribution best fit the data, both for single-neuron responses and population responses. This was done by calculating the difference between the empirical cumulative distribution function (cdf) and the best-fit cdf of various parametric distributions. The empirical cdf was determined for each neuron (single-neuron response) or each stimulus image (population response) with the Kaplan-Meier estimator.

*Distributions tested.* Seven standard probability distributions were fit to each single-neuron or population response. The first six of these were:

Weibull cdf: 
$$F(x|a, b) = 1 - e^{-\left(\frac{x}{a}\right)^b}$$
 (1)

log-logistic cdf:  $F(x|\mu, \sigma) = \frac{1}{1 + e^{\frac{-(\ln(x) - \mu)}{\sigma}}}$  (2)

log-normal cdf: 
$$F(x|\mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \int_{0}^{x} \frac{e^{\frac{-(\ln(t) - \mu)^{2}}{2\sigma^{2}}}}{t} dt$$
 (3)

gamma cdf: 
$$F(x|a, b) = \frac{1}{b^a \Gamma(a)} \int_0^x t^{a-1} e^{\frac{t}{b}} dt$$
 (4)

exponential cdf: 
$$F(x|a) = 1 - e^{\frac{-x}{a}}$$
 (5)

normal cdf: 
$$F(x|\mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{\frac{\sigma}{2\sigma^2}} dt$$
 (6)

The seventh distribution, the  $\alpha$ -stable distribution, is more complicated. As the  $\alpha$ -stable distribution has not previously been used in experimental neurophysiology to our knowledge, it is described in more detail below.

 $\alpha$ -Stable distribution. The  $\alpha$ -stable distribution is a class of probability distributions useful for describing heavy-tailed random variables, including those with skewed distributions (Nolan 2011b; Samorodnitsky and Taqqu 1994; Zolotarev 1986). First formulated by Lévy (1925), it found initial application in finance (Mandelbrot 1963) and is now used in a variety of disciplines (Adler et al. 1998; Ghahfarokhi and Ghahfarokhi 2009; Mantegna and Stanley 2000; Nikias and Shao 1995).

Stable distributions are defined by their characteristic function (Fourier transform of the pdf):

### STATISTICS OF INFEROTEMPORAL RESPONSES

$$\varphi_X(u) = \int_{-\infty}^{\infty} e^{iux} f_X(x) dx$$
  
=  $\exp\left\{-\gamma^a \left|u\right|^a \left[1 - i\beta \operatorname{sign}(u) \tan \frac{\pi \alpha}{2}\right] + i\delta u\right\}$  if  $\alpha \neq 1$  (7)  
=  $\exp\left\{-\gamma \left|u\right| \left[1 + i\beta \operatorname{sign}(u) \frac{2}{\pi} \ln\left(|u|\right)\right] + i\delta u\right\}$  if  $\alpha = 1$ 

where  $f_x(x)$  is the pdf of random variable X. Stable distributions are parameterized in a variety of ways. We used the 1-parameterization here. No general equations for the pdf or cdf of stable distributions are known to exist in closed form, which limited the use of these distributions until sufficient computational power became widely available. There are four parameters associated with the distribution:  $S(\alpha,\beta,\gamma,\delta)$ . The first is  $\alpha$ , the index of stability or characteristic exponent, which determines tail heaviness. It may take values  $0 < \alpha \le 2$ . Smaller values of  $\alpha$  produce heavier tails, and all stable distributions with  $\alpha < 2$  have power law tails. The second parameter,  $\beta$ , specifies skewness and may take values  $-1 \leq \beta \leq$ 1. Positive values are right skewed, and negative values are left skewed. The parameters  $\gamma$  and  $\delta$  are scale and location parameters. Three special cases of stable distributions do have closed-form expressions for their pdf: the normal distribution, corresponding to  $S(2,\beta,\sigma/\sqrt{2},\mu)$  where  $\beta$  has no effect; the Cauchy distribution, corresponding to  $S(1,0,\gamma,\delta)$ ; and the Lévy distribution, corresponding to  $S(0.5,1,\gamma,\delta)$ .

Stable distributions have two unique properties that set them apart from other distributions, both familiar from the normal distribution special case. First, summing stable independent random variables with the same distribution shape (same index  $\alpha$ ) will produce a new distribution that retains that shape (shape is stable, hence the name of the distribution). A corollary of this is that a stable distribution convolved with itself does not change shape. It has been suggested that this property has significance for neural information processing (Holden 1975). Second, stable distributions are attractor probability distributions, as a generalized Central Limit Theorem applies to them (Gnedenko and Kolmogorov 1968). In the same way that sums of independent and identically distributed (i.i.d.) random variables with finite variance will converge to a normal distribution, sums of i.i.d. random variables with infinite variance (power law tails) will converge to other stable distributions.

Fitting distributions to data. All parameter fits for the first six distributions were maximum likelihood, except for the  $\sigma$  parameter of the log-normal distribution. That parameter was estimated by the square root of the unbiased estimate of the variance of the log of the data. Fittings for these distributions were done with the Matlab Statistics Toolbox. Parameter fits for the  $\alpha$ -stable distribution were done with the methods of Koutrouvelis (1980, 1981), with Matlab code downloaded from Veillette (2011). (Other  $\alpha$ -stable software is available from Nolan 2011a.)

After each of the seven distributions was fitted to the data, goodness of fit was compared. The error measure between the empirical cdf and a parametric cdf used was the Kolmogorov-Smirnov distance:

$$D = \max_{1 \le i \le n} \left| \mathbf{F}_{\text{emp}}(x_i) - \mathbf{F}(x_i) \right| \tag{8}$$

where F indicates cdf and n is the number of data points in the sample (number of neurons or number of images). Errors for each singleneuron response were then averaged to give a mean error for all neurons recorded from, and, similarly, errors for each population response were averaged to give a mean error for all images used. The best-fit parametric distribution had the smallest mean error (smallest mean Kolmogorov-Smirnov distance).

For purposes of fitting probability distributions, only neural firing rates greater than zero were included. Firing rates equal to zero were trimmed from the data. This was done because the discrete nature of spikes made it impossible to assign accurate spike rates at very low activity levels. For example, with a 140-ms trial duration, a single spike occurring over 10 trials translates to an average of 0.7 spikes/s. Mean responses below 0.7 spikes/s cannot be resolved. All firing rates below 0.7 would be mapped to zero in this example, leading to an artifactual excess of zeros in the data, with no way of knowing to what extent zero measured firing rate was actually 0.1 or 0.4 spikes/s. The shape of the lower tail of the response distribution is therefore ambiguous. As firing rate increases or trial duration increases, approximating a discrete process by a continuous probability distribution becomes increasingly accurate, and all this becomes less of an issue.

In addition to probability distributions of neural responses, we were interested in the distributions of response-modulations. Responsemodulations are responses with spontaneous activity subtracted. Unlike responses, which are always positive, response-modulations can be either positive or negative. However, the first five probability distributions above allow only positive values. For those distributions, the data was shifted upward to nonnegative values when fitting parameters. As we were interested in estimating the shapes of distributions and not their locations, that procedure was sufficient for our purposes.

#### Calculating Neural Selectivity and Population Sparseness

The selectivity/sparseness measure we used was kurtosis, which is a measure of the "peakedness" of a probability distribution. Kurtosis depends entirely on the shape of the distribution, independent of its position (e.g., mean) or scale (e.g., variance). The equation for the kurtosis selectivity index is:

$$SI_{K} = \frac{\sum_{i=1}^{n} (r_{i} - \bar{r})^{4}}{ns^{4}} - 3$$

$$\frac{\frac{1}{n}\sum_{i=1}^{n} \left(r_{i} - \sum_{i=1}^{n} \frac{r_{i}}{n}\right)^{4}}{\left[\frac{1}{n}\sum_{i=1}^{n} \left(r_{i} - \sum_{i=1}^{n} \frac{r_{i}}{n}\right)^{2}\right]^{2}} - 3$$
(9)

For single-neuron responses,  $r_i$  refers to the response of the neuron to the *i*th image and *n* refers to the number of images. For population responses,  $r_i$  is the response of the *i*th neuron in the population to a particular image and *n* refers to the number of neurons in the population. Mean response is denoted by  $\bar{r}$  and the standard deviation of the responses by *s*. Subtracting 3 rescales kurtosis so that a normal distribution has a kurtosis of 0. This rescaling procedure leads to what is called excess kurtosis (kurtosis in excess of that for a normal distribution), but here we shall refer to it simply as kurtosis.

Higher kurtosis means that more of the variance is due to infrequent large events on the tails, rather than frequent small events. However, as kurtosis is a global measure of the shape of the entire probability distribution and not just the tail, in some cases measures of kurtosis and measures of tail heaviness can become dissociated.

Kurtosis has been a popular measure of selectivity in the theoretical literature (e.g., Bell and Sejnowski 1997; Olshausen and Field 1996; Simoncelli and Olshausen 2001). Use of kurtosis in the experimental literature include Lehky et al. (2005), Lehky and Sereno (2007), as well as Tolhurst et al. (2009).

Although kurtosis will be the main selectivity metric used here, a number of other metrics exist in the literature, which we will briefly examine. One of these is a selectivity index based on entropy of the pdf of neural responses, introduced by Lehky et al. (2005) and described more fully there:

Downloaded from http://jn.physiology.org/ at Stanford Univ Medical Center on January 18,

2013

 $SI_{E} = -\sum_{i=1}^{M} p(r_{i}) \log_{2}(p(r_{i})) \Delta r$ (10) Here  $p(r_{i})$  is the probability density of the *i*th bin in a discretized response pdf and  $\Delta r$  is the bin width. There are also two selectivity

response pdf and  $\Delta r$  is the bin width. There are also two selectivity indices based on activity fractions. One was introduced by Rolls and Tovée (1995) and slightly modified by Vinje and Gallant (2000):

$$SI_{R} = \frac{n}{n-1} \left( 1 - \frac{\left(\sum_{r=1}^{n} \frac{r_{i}}{n}\right)^{2}}{\sum_{r=1}^{n} \frac{r_{i}^{2}}{n}} \right)$$
(11)

This modified version reverses the scale of the original Rolls and Tovée (1995) formulation so that higher values of  $SI_R$  correspond to large selectivity. The other activity fraction measure was introduced by Zipser and colleagues (Moody et al. 1998):

$$SI_{Z} = \frac{n - \sum_{i=1}^{n} \frac{r_{i}}{r_{\max}}}{n - 1}$$
(12)

Another selectivity index is based on the Michaelson contrast of responses:

$$SI_{C} = (r_{max} - r_{min})/(r_{max} + r_{min})$$
(13)

This measure was not useful for our large data set because it saturated close to the ceiling value of 1.0 for every single-neuron response and every population response.

### Normalization

High population sparseness could arise as an artifact because neurons in the population might nonspecifically have different levels of activation. For example, if 99 neurons fired at 10 spikes/s to all stimuli and 1 neuron fired at 100 spikes/s to all stimuli, that would lead to calculation of high sparseness. However, that high sparseness would not be interesting from the perspective of neural coding. To control for this possibility, in addition to analyzing raw data we also performed selectivity/sparseness calculations and other statistical analyses using data in which the activity of each neuron was normalized with respect to mean firing rate across all stimuli:

$$r_i^{\text{norm}} = \frac{r_i}{\frac{1}{n}\sum_{i=1}^n r_i}$$
(14)

In principle, normalization has no effect on single-neuron selectivity. That is because kurtosis depends on the shape of the response pdf, and dividing all responses by the same constant changes the scale of the pdf curve but has no effect on its shape. On the other hand, normalization does affect population sparseness. For population sparseness, each neuron in the population undergoes normalization by a different number. Therefore, overall, there is a nonlinear transform in the shape of the pdf of population responses, leading to a change in kurtosis.

Although we normalized with respect to mean firing rate, there are different definitions of the normalization factor that could have been used. One general family of normalizations is based on the p-norm or Minkowski norm of the neural response vector  $(r_1, r_2, r_3, ..., r_n)$  where *n* is number of neurons in the population. The p-norm is defined as:

$$\operatorname{norm}_{p} = \left(\sum_{i=1}^{n} r_{i}^{p}\right)^{\frac{1}{p}}$$
(15)

For p = 1 the norm is the sum of responses in the population, for p = 2 the norm is the conventional Euclidean vector norm, and for  $p = \infty$  the norm is the maximum response in the population.

Simulation results from a model indicated that the value of normalized population sparseness strongly depended on the value of pused in the definition of normalization. As p increases, normalized sparseness becomes smaller. Depending on the value of p chosen, normalized sparseness could be either smaller or larger than unnormalized sparseness.

By normalizing with respect to mean response of a neuron, we effectively chose p = 1. That is because (sum of responses) and (mean response) are equivalent when normalizing population sparseness, as they differ by the same multiplicative factor (the population size) for every neuron in the population.

Choosing p = 1 provided the most direct way of addressing concerns that inhomogeneities with respect to mean responsiveness of different neurons artificially inflated the calculated population sparseness. However, we repeated all analyses using p = 2 (not presented), and the results were qualitatively the same as reported here.

After a preliminary analysis of the data, we decided that it would be interesting to examine normalization not only with respect to response magnitude but also with respect to the dispersion of responses. This was done by normalizing each neuron by the standard deviation of responses over all stimulus images:

$$r_i^{\text{norm}} = \frac{r_i}{\left(\frac{1}{n-1}\sum_{i=1}^n (r_i - \bar{r})^2\right)^{\frac{1}{2}}}$$
(16)

### Effect of Data Set Size

The effect of data set size on calculated statistics was examined by repeating the analyses using resampled subsets of the complete data set. This was carried out both for subsets of stimulus images (for single-neuron responses) and for subsets of neurons (for population responses). A random subset of a particular size was selected without replacement. The selectivity index was calculated for the subset, and the largest firing rate (extremal response) was determined for the subset as well. This was repeated 10,000 times for each subset size, each time using a different random sampling, and the results were averaged. Subset size ranged from 4 to 650 neurons for population responses.

### Pareto Tail Index

Large responses occurring on the upper tails of response pdfs were analyzed separately from the main body of data. Tail data were fit with a generalized Pareto distribution using maximum likelihood. This was done individually for the probability distribution for each singleneuron response and for each population response. Fitting was performed with the Matlab Statistics Toolbox.

Generalized Pareto tail analysis was introduced by Pickands (1975) and has subsequently found wide use (Coles 2001), particularly by the insurance industry for modeling rare large events on the upper tail of a sparse distribution (earthquakes, floods, etc.) (Embrechts et al. 1997). We are not aware of any previous application to neurophysiological data.

The pdf for the generalized Pareto distribution is a monotonically declining function defined by:

$$y = f(r|k, \sigma, \theta) = \frac{1}{\theta} \left( 1 + k \frac{(r-\theta)}{\sigma} \right)^{-1-\frac{1}{k}}$$
(17)

when  $k \neq 0$ . Within the  $k \neq 0$  case, when k > 0 the function is defined for  $\theta < r$  and when k < 0 it is defined for  $\theta < r < -\sigma/k$ . When k = 0, the generalized Pareto distribution becomes equal to the exponential distribution:

$$y = f(r|0, \sigma, \theta) = \frac{1}{\theta} e^{-\frac{(r-\theta)}{\sigma}}$$
(18)

defined for  $\theta < r$ . The parameter  $\sigma$  sets the scale. The parameter  $\theta$  defines the threshold for what is counted as a "large" response (threshold where the upper tail starts). In this study  $\theta$  was individually set for each neuron (single-neuron responses) or each image (for population responses) so that in each case the largest 10% of responses were included in the tail analysis. Finally, *k* is a shape parameter quantifying heaviness of the tail. It is also called the tail parameter, and is the parameter we are primarily interested in. The value of *k* determines qualitative properties of the tail. If k > 0 the tail follows a power law, if k = 0 the tail is exponential, and if k < 0 the tail is finite with zero probability of a response greater than  $-\sigma/k$ .

### RESULTS

Neurons in the data sample included all cells that had been stably isolated, so that the sample was not biased by response properties. Recording positions were evenly distributed across the broad cortical regions delineated in METHODS, which would have covered many feature columns (Fujita et al. 1992), and possibly many category domains if they exist in monkeys as in humans (Downing et al. 2006). Results for all three cortical areas (TEav, TEad, and STS) and for both monkeys appeared essentially the same, so their data were pooled for presentation here (Fig. 7 presents some unpooled results).

Although responses of inferotemporal neurons were predominantly excitatory, a few neurons produced mostly inhibitory responses. Of 659 neurons for which we have spontaneous activity data, 78 neurons (11.8%) produced large inhibitory responses more frequently than large excitatory responses, measured over the stimulus set of 806 images. A "large" response was defined as a deviation of >5 spikes/s from spontaneous activity.

### Selectivity and Sparseness

Single-neuron selectivities for 674 neurons are shown in Fig. 4*Ai*. Selectivity was defined as kurtosis (*Eq. 9*). A separate selectivity was calculated for each neuron based on its responses to 806 images. Figure 4*Aii* shows single-neuron selectivities for neural response-modulations (spontaneous activity subtracted) rather than neural responses. The sample size in this case was slightly smaller, 659 neurons, because spontaneous activity data were missing from 15 neurons.

Single-neuron selectivities for responses and for response-modulations were almost exactly the same. Mathematically they should be identical, but here the samples differed slightly. The reason they should be identical is that we are using a selectivity measure, kurtosis, that is based on the shape of the probability distribution. The shape of the probability distribution does not change if we subtract the same constant (spontaneous activity) from all responses of a single neuron.

Population sparseness based on neural responses is shown in Fig. 4*Bi*, while population sparseness based on response-modulations is shown in Fig. 4*Bii*. A separate sparseness value was calculated for each of 806 images, based on the probability distribution of neural activities in the population for each image.

Population sparseness was greater than single-neuron selectivity. This is shown by comparing average kurtosis values given in each panel of Fig. 4, A and B. The difference in kurtosis was significant at the P = 0.00001 level under a *t*-test (for kurtosis means) and a Wilcoxon rank sum test (for kurtosis medians). This was true for both responses and responsemodulations. This shows that measuring single-neuron selectivity is not equivalent to measuring population sparseness.

Population sparseness still was greater than single-neuron selectivity even after the response of each neuron in the population was normalized with respect to mean firing rate (Fig. 4, C and D). The difference between them remained significant at the P = 0.00001 level when comparing both means and medians.

Median selectivity for area STS was slightly lower than the other two inferotemporal areas, TEav and TEad, both for single-neuron selectivity and for population sparseness. For single-neuron responses, the kurtosis values were [TEav, TEad, STS] = [2.0, 2.1, 1.4]. For population responses the values were [TEav, TEad, STS] = [6.9, 9.3, 4.9]. Previous studies have reported that single-neuron selectivity in TEav and TEad were similar (Lehky and Tanaka 2007; Tamura and Tanaka 2001).

For comparison with the kurtosis selectivity index  $SI_K$ , values of the  $SI_E$ ,  $SI_R$ , and  $SI_Z$  measures (defined in *Eqs. 9–12*) are given in Table 1. These other metrics also show that population sparseness is greater than single-neuron selectivity.

### Effect of Neural Correlation on Population Sparseness

As recordings were conducted with single electrodes, calculations of population sparseness did not take into account correlations between neurons in the population. Such correlations would be expected to cause a decrease in estimated population sparseness. However, computer simulations based on correlation values in the literature indicated that inclusion of correlations would have had a very small effect on measured population sparseness.

The noise correlation coefficient between nearby units (within several hundred micrometers) is typically around 0.15– 0.20 (Constantinidis and Goldman-Rakic 2002; Gawne and Richmond 1993; Huang and Lisberger 2009; Reich et al. 2001; Smith and Kohn 2008; Zohary et al. 1994). Two studies using tetrodes report much lower correlations between nearby neurons (r = 0.005, Ecker et al. 2010; r = 0.02, Erickson et al. 2000). The correlation coefficient drops as the distance between neurons increases (Constantinidis and Goldman-Rakic 2002; Smith and Kohn 2008) and as stimulus duration decreases (Constantinidis and Goldman-Rakic 2002; Reich et al. 2001; Smith and Kohn 2008). For our recording conditions, with neurons in the population spread over a broad cortical area and with short stimulus durations (105 ms), a correlation coefficient of 0.05 would be a reasonable value based on the published data.

The effect of correlation on population sparseness was estimated through model simulations as follows. Mimicking the size of our data set, a population of 674 model neurons whose responses followed a gamma distribution was stimulated by 806 images. Noise correlations between neurons were set at chosen values with copula statistical methods. Thus we were able to compare sparseness for populations



Fig. 4. Selectivity and sparseness. A: single-neuron selectivity. For each neuron, selectivity was measured as kurtosis for the probability distribution of responses to 806 stimulus images. Ai: selectivity of responses for 674 neurons. Aii: selectivity of response-modulations for 659 neurons. Response-modulation is response with spontaneous activity subtracted. B: population sparseness. For each stimulus image, sparseness was measured as kurtosis for the probability distribution of responses across the population of neurons. Bi: population sparseness of responses for 806 images. Neural population size was 674 cells. Bii: population sparseness of response-modulations for 806 images. Neural population size was 659 cells. C and D: selectivity and sparseness for normalized data. Neural responses were normalized with respect to mean firing rate prior to calculations of singleneuron selectivity and population sparseness.

with zero and nonzero correlations. We examined two nonzero correlation coefficients, r = 0.2 as a worst-case value and r = 0.05 as a more likely value for our conditions. Compared with zero correlation, the results showed that a correlation of 0.2 caused population sparseness to decrease by 6.7% and a correlation of 0.05 caused sparseness to decrease by 1.8%.

Those downward shifts in population sparseness/selectivity caused by including correlation are far too small to eliminate the nonequivalence between population sparseness and singleneuron selectivity observed here. Looking at results for normalized activity in Fig. 4, *C* and *D*, mean population selectivity was 396% greater than single-neuron selectivity for responses and 295% greater for response-modulations, both much larger than the correlation correction of 1.8%.

### Selectivity/Sparseness vs. Response Magnitude

We looked at the relationship between selectivity/sparseness and mean response. For single-neuron responses, we used the mean response of each neuron to all images. For population responses, we used the mean response of the

1103

Table 1.	Single-1	neuron sel	ectivity an	nd popula	ition s	parseness	for
neural rea	sponses, c	calculated	with four	metrics	in the	literature	

	Single	Neuron	Population		
	Selectivity	Normalized selectivity	Sparseness	Normalized sparseness	
SIK	1.88	1.88	9.61	7.98	
SIE	0.30	0.30	0.80	0.37	
SIR	0.29	0.29	0.56	0.36	
SIZ	0.77	0.77	0.89	0.85	

The four metrics are defined in *Eqs. 9–12*. Values represent medians over all neurons (single-neuron responses) or over all stimulus images (population responses).

population to each image. The same was done for responsemodulations.

Figure 5Ai shows single-neuron selectivity plotted as a function of the mean value of single-neuron responses. There was large scatter among different neurons for both their mean responses and their selectivities, with a moderately strong negative correlation between those two properties (r = -0.52). Highly selective neurons were associated with low mean response. For response-modulations (Fig. 5Aii), there was less scatter in activities, and they were weakly correlated with selectivity.

Population sparseness is plotted as a function of mean population response in Fig. 5Bi. The corresponding plot for response-modulations is in Fig. 5Bii. There was little change in the average level of activity in the population when stimulated by different images. This indicates that different images are encoded primarily by changes in the pattern of activity within the population, without large shifts in the overall level of population activity. Correlation between mean response level and population sparseness was low.

## Effect on Selectivity/Sparseness of Changing Stimulus Set Composition

Ideally the library of critical features embedded within the large and diverse image set we used would be an unbiased sampling of all features present in the monkeys' visual environment. However, because the characteristics of such a stimulus space are unknown, there was no way to design the stimulus set perfectly. Although our stimulus set had great diversity, there nevertheless may have been biases that affected neural response statistics. One possibility was that the stimulus set underrepresented high-effectivity stimuli. Another possibility was that the stimulus set underrepresented low-effectivity stimuli.

To examine whether selectivity measurements were highly sensitive to such potential biases, we resampled the existing data to create an increased emphasis toward either high-effectivity or low-effectivity stimuli. These were called augmented data sets. We formed the high-effectivity augmented data set ( $DS_{high}$ ) by taking data from a random sample of 100 images in the top quarter of the most effective stimuli and pooling them back into the base data set. The low-effectivity augmented data set ( $DS_{low}$ ) was formed in an analogous manner, resampling from the bottom quarter of effective stimuli. In each case the augmented data set size increased from 806 images.

These manipulations created a set of new single-neuron response vectors, each with 906 elements rather than 806 elements. The effects of single-neuron response selectivity were modest. Mean kurtosis of the base data set ( $DS_{base}$ ) was  $k_{base} = 3.50$ , while for  $DS_{high}$  and  $DS_{low}$  it was  $k_{high} = 2.47$  and  $k_{low} = 3.65$ , respectively. Normalizing single-neuron selectivities had no effect on these values.

Examining population selectivities for augmented data sets required an additional step. The resampling procedure itself

Fig. 5. Selectivity/sparseness vs. response magnitude. Correlation coefficients *r* are included. *A*: single-neuron selectivity as a function of the mean value of neural activities to all stimulus images. Each dot represents mean response of a different neuron. *Ai*: single-neuron responses; n = 674 neurons. *Aii*: single-neuron response-modulations; n = 659 neurons. *B*: population sparseness as a function of mean value of neural responses across the population to a single image. Each dot represents mean population response to a different image. *Bi*: population responses; n = 674 images. This plot omits 6 outliers extending to the right out to 45.6 spikes/s. *Bii*: population series series and the right correlations were significantly different from zero at the P = 0.001 level.



J Neurophysiol • VOL 106 • SEPTEMBER 2011 • www.jn.org

would not create new image response vectors but would merely insert repeats of existing ones into the augmented data set. To create new image response vectors we took the 100 repeated vectors and shuffled (permuted) their components before pooling them back into the original base data set. For example, to create the first component of the new vectors, the first components of all 100 repeat vectors were shuffled among themselves. This shuffling was repeated individually for each of the 674 components in an image response vector (equal to the 674 neurons in the data set). In this manner, responses to 100 new simulated random "images" with either low effectivity or high effectivity were created and pooled back into the original data set to create augmented data sets tilted toward those characteristics.

Changing the set of image response vectors in this way had little effect. For  $DS_{base}$  mean kurtosis was  $k_{base} = 12.51$  while for  $DS_{high}$  and  $DS_{low}$  it was  $k_{high} = 12.09$  and  $k_{low} = 12.73$ . After normalizing responses, the values were  $k_{base} = 5.31$ ,  $k_{high} = 5.34$ , and  $k_{low} = 5.17$ .

On the basis of these simulations, the observation that population sparseness was greater than single-neuron selectivity appeared to be robust in the face of moderate shifts in the characteristics of the stimulus set. The same held true when response-modulations were examined rather than responses.

### Effect of Data Set Size

The shape of the upper tail for the response probability distribution is a major factor determining selectivity and sparseness. Events occurring on the tail are rare. Small data samples have a relatively high likelihood of missing those rare large responses, thereby mischaracterizing tail properties. In particular, small data samples, by missing rare large events, underestimate the heaviness of the tail, and therefore underestimate selectivity and sparseness.

Here we examine two characteristics of responses expected to increase in magnitude as the size of the data set increases: *1*) selectivity/sparseness and *2*) maximum value of responses (extremal events). These characteristics were estimated as a function of data set size by resampling subsets of our complete data set (see METHODS). When dealing with single-neuron responses, data set size refers to the number of images tested on each neuron. When dealing with population responses, data set size refers to the number of neurons included within the population when coding each image.

The effect of data set size on single-neuron selectivity (red) and population sparseness (blue) is shown in Fig. 6A. Both neural selectivity and population selectivity (sparseness) increased as a function of data set size. However, population selectivity increased faster, behavior expected if the probability distribution of population responses had a heavier tail than the distribution of single-neuron responses. (see Fig. 16 for model simulations demonstrating tail heaviness effects). This held true for both responses and for response-modulations, using unnormalized and normalized data.

Maximum response also increased as a function of data set size, looking at either single-neuron responses or population responses (Fig. 6B). For single-neuron responses, maximum response refers to the largest response of a neuron across all stimulus images. For population responses, maximum response

refers to the activity of the most responsive neuron in the population when presented with one stimulus image.

Population responses were again more sensitive to data set size than single-neuron responses, indicating that probability distributions for population responses have heavier tails. With a heavier tail, as sample size increases there is greater probability of hitting an extremal event than with a thinner tail. Curves for normalized data were not included in Fig. 6*B* because a comparison of response magnitudes is not meaningful after both sets of data have been independently normalized with respect to mean response.

The greater value of population sparseness over singleneuron selectivity remained when we examined data for the individual *monkeys KH* (n = 322 neurons) and *SH* (n = 352neurons). For individual monkeys, the number of stimulus images was slightly larger than for the pooled data, being 1,016 images for *monkey KH* and 1,010 images for *monkey SH*. (The 806 images used in the pooled data represent the intersection of those two stimulus sets). Figure 7A shows results for the individual monkeys.

The difference between population sparseness and singleneuron selectivity also remained when we examined data for individual cortical areas rather than pooled data. Figure 7*B* shows results for area TEav (n = 271 neurons). Results were also similar for the other cortical areas examined, TEad (n =221 neurons) and STS (n = 152 neurons).

### Dynamic Range

We measured the dynamic range of each neuron, using as one measure the interquartile range of its responses across the image set (difference between the 25th and 75th percentiles of responses). The results for 674 neurons are shown in Fig. 8. The dynamic range of a neuron and its selectivity had low correlation (r = -0.27).

We were interested in seeing which class of inferotemporal neurons was more important for creating highly sparse population responses: high-selectivity neurons or high-dynamicrange neurons. Using population sparseness based on responsemodulations (so that spontaneous activity differences between neurons were already factored out), we "lesioned" the 10% of cells in the population with the highest selectivities or the 10% with the largest dynamic range. The results showed that removing high-selectivity neurons had little effect on population sparseness, while removing high-dynamic-range neurons caused a large drop (Fig. 9).

On the basis of these observations, we decided to measure sparseness when neural responses were normalized for dynamic range. Specifically, we normalized each neuron in the population with respect to the standard deviation of its responses over the stimulus image set (Eq. 16) (with standard deviation and interquartile range being conceptually similar measures of dispersion). This differs from the normalization presented above, which was normalization with respect to mean response. Normalizing for dynamic range caused population sparseness values to collapse, dropping lower than single-neuron selectivities (Fig. 10). The large drop in sparseness occurred even though the normalized neurons still retained a range of different mean activities and different single-neuron selectivities. This shows that a critical requirement for high population sparseness is a statistically inhomogeneous popula-

, 2013

Fig. 6. Effect of data set size on response statistics. A: selectivity/sparseness vs. data set size. For single-neuron responses, data set size was number of neurons. For population responses, data set size was number of stimulus images. Different data set sizes were created by bootstrap resampling subsets of the complete data. *Ai*: effects of data set size based on responses. *Aii*: effects of data set size based on response-modulations. *Aiii* and *Aiv*: same curves for normalized data. *B*: maximum response vs. data set size. *Bii*: effects of data set size based on responses. *Bii*: effects of data set size based on responses. *Bii*: effects of data set size based on responses. *Bii*: effects of data set size based on responses. *Bii*: effects of data set size based on response. *Bii*: effects of data set size based on response modulations. Selectivity and maximum response in these plots reflect mean values over 10,000 repeats of resampled data at each point.



tion with a diversity of dynamic ranges. Having different mean firing rates does not matter so much, nor are the single-neuron selectivities within the population of central importance. These experimental results are reproduced in the modeling shown in Fig. 15.

### Response Magnitude vs. Rank

Response magnitude versus response rank is plotted in Fig. 11 on a log-log scale. This is known as a Zipf plot (Newman 2005). Only the upper tails (largest 10% of responses) are shown in the plots. The single-neuron plots (red) were produced as follows. For each neuron, responses to all images were normalized relative to the largest response. These normalized responses were then averaged across all neurons. The population plots (blue) were produced as follows. For each image, responses of all neurons were normalized relative to the largest response of all neurons were normalized relative to the largest response of all neurons were normalized relative to the largest response of all neurons were normalized relative to the largest response in the population. The results were then averaged over all images.

For both responses and response-modulations, the population curve had a steeper slope than the single-neuron curve. This again indicates that the probability distribution for population responses has a heavier tail than that for single-neuron responses. The heavier tail of the population response increases the likelihood of a larger extremal response, which in turn pushes down the rest of the curve that is normalized relative to the extremal event.

It should be noted in the Zipf plots that the second largest response was almost the same size as the largest response. For single-neuron activities to different images, the second largest response was 0.893 relative to the largest. For population activities to a single image, the second largest response was 0.795 compared with the largest. Moreover, those fractions depended on data set size, increasing for larger data sets (not shown). This last point was confirmed by doing Zipf plots for different data set sizes (analogous to the analysis presented in Fig. 6). By extrapolation, for very large data sets beyond what we used, the second largest response would approach equality with the largest response.

The Zipf rank-order plots show that even for highly selective/ sparse representations, the best response does not strongly dominate other responses. With a large data set, the chances are high



Fig. 7. Results for individual monkeys and individual cortical area. A: comparison of population sparseness and single-neuron selectivity for *monkey KH* as well as for *monkey SH. B*: comparison of population sparseness and single-neuron selectivity in cortical area anterior ventral TE (TEav). Kurtosis is plotted as a function of data set size, where different data set sizes were formed by resampling subsets of the complete data, as in Fig. 6. For population sparseness, data set size refers to number of neurons in the population. For single-neuron selectivity, data set size refers to the number of stimulus images presented to each neuron.

for finding a second stimulus or a second neuron that is almost as good as the best one. This does not affect the underlying response probability distribution, which is what determines selectivity/ sparseness.



Fig. 8. Dynamic range of neurons. Dynamic range was measured here as interquartile range of responses. Dynamic ranges for responses and response-modulations of each neuron are the same, as subtracting a constant (spontaneous activity) has no effect upon this calculation.

### Probability Density Functions of Responses

We investigated which probability distribution provided the best overall description of single-neuron responses and population responses. Seven distributions were examined: Weibull, log-logistic, log-normal, gamma, exponential, normal, and  $\alpha$ -stable (*Eqs. 1–7*). The error measure used was the Kolmogorov-Smirnov distance (*Eq. 8*) between the empirical cdf of responses and the best-fit cdf for each of the seven parametric distributions. The Kolmogorov-Smirnov distance was averaged over all neurons (single-neuron responses) or all stimulus images (population responses).

For single-neuron responses, the gamma distribution gave the best fit by a small margin over other distributions (Fig. 12A). Fits for responses (Fig. 12Ai) and response-modulations (Fig. 12Aii) were very close (theoretically they should be identical). Although the gamma distribution provided the best fit, it was not significantly better than the Weibull, log-logistic, or log-normal distributions, based on standard errors of mean Kolmogorov-Smirnov distances over the set of all single-neuron responses.

1107



Fig. 9. Contributions of high-selectivity neurons and high-dynamic-range neurons to population sparseness. *A*: population sparseness based on response-modulations, including all neurons in the population (659 neurons). *B*: population sparseness after high-selectivity neurons were removed from the population. *C*: population sparseness after high-dynamic-range neurons were removed from the population. In each case, the top 10% of high-selectivity neurons or high-dynamic-range neurons were removed from the data set, leaving a population size of 593 neurons in both *B* and *C*.

The equation for the gamma pdf is:

$$\mathbf{f}(\mathbf{x}|a,b) = \frac{1}{b^a \Gamma(\mathbf{a})} x^{a-1} e^{-\frac{\lambda}{b}}$$
(19)

For best-fit gamma distributions, the median value of a was 3.0 with interquartile range 1.7. The median value of b was 2.0 with interquartile range 1.6. Correlation between the two parameters was -0.04.

Plots of pdfs based on best-fit gamma parameters are shown in Fig. 12*B*, *i* and *ii*. Three plots are shown in the figure, with shape parameter *a* taken at its 5th, 50th, and 95th percentile values over all neurons, while scale parameter *b* was kept at its 50th percentile value. The shapes of the response pdfs are quite different in each case, indicative of the diversity in the response statistics for different neurons. If we had included the spread of values for the scale parameter *b*, the diversity would have appeared even greater.

Best-fit probability distributions for population activities are shown in Fig. 13. Population probability distributions based on responses were dissimilar to those based on response-modulations. For responses, the log-normal distribution gave the best fit by a small but significant margin (Fig. 13*Ai*). For responsemodulations, the  $\alpha$ -stable distribution gave an extremely good fit with a much smaller Kolmogorov-Smirnov distance than any of the other tested distributions (Fig. 13*Aii*).

The pdf for the log-normal distribution is:

$$f(x|\mu,\sigma) = \frac{1}{x\sigma\sqrt{2\pi}}e^{\frac{-(\log(x) - \mu)^2}{2\sigma^2}}$$
(20)

For population pdfs based on responses, the median best-fit value of  $\mu$  was 1.7 and its interquartile range was 0.13. The median value of  $\sigma$  was 1.0 with interquartile range 0.03. The correlation between the two parameters was 0.3.

Plots of population response pdfs based on best-fit lognormal parameters are shown in Fig. 13*Bi*. Plots are shown with parameters  $\mu$  and  $\sigma$  each set to their 5th, 50th, and 95th percentile values from among neurons in the population. All three distributions are similar, demonstrating that the probability distribution of neural responses within a population remains much the same for different images.

The  $\alpha$ -stable distribution gave the best description of population response-modulations. The pdf of the  $\alpha$ -stable distribution is given by the inverse Fourier transform of Eq. 7, which cannot be expressed in closed form. For the two  $\alpha$ -stable parameters describing shape, the median value of  $\alpha$  was 1.31 with interquartile range 0.09, while the median value of  $\beta$  was 0.66 with interquartile range 0.25. The correlation between the two parameters was -0.29. This value of  $\alpha$  indicates a power law tail, with a tail heaviness between a normal distribution ( $\alpha = 1$ ) and a Cauchy distribution ( $\alpha = 2$ ). The positive  $\beta$  value indicates a positively skewed distribution.

Plots of population response-modulation pdfs based on bestfit  $\alpha$ -stable parameters are shown in Fig. 13*Bii*. For these plots,  $\alpha$  was set at its 5th, 50th, and 95th percentile values and  $\beta$  was set to its 95th, 50th, and 5th percentile values (reverse order because of the negative correlation with  $\alpha$ ). The scale and location parameters  $\gamma$  and  $\delta$  were set to their median values ( $\gamma = 2.32$ ;  $\delta = 2.91$ ). Again, there was little variation in the distribution of responses within the population to different images.

A previous study found that the probability distribution of single-neuron responses was described by a gamma distribution (Franco et al. 2007), in agreement with our findings. They also reported that population responses followed an exponential distribution, while we found a log-normal distribution gave the best fit. However, our fitting procedure only included responses greater than 0 spikes/s (see METHODS). Our short trial durations together with the discrete nature of action potentials left very low spikes rates poorly characterized, leading to an artifactual excess of zeros. It is possible that data in which the low end of the response distribution was better characterized would lead to an overall description different from what we found. Nevertheless, in our data set, for the range greater than 0 spikes/s, a log-normal distribution gave a better fit than an exponential distribution.



Fig. 10. Population sparseness when activities of each neuron in the population were normalized with respect to standard deviation of responses over all images, rather than normalized with respect to mean response. *A*: sparseness based on neural responses. *B*: sparseness based on neural response.

Although the histograms in Figs. 12A and Fig. 13A indicate which probability distribution gave relatively the best fit, they do not indicate in an absolute sense how good that best fit was. That question was examined by comparing *1*) Kolmogorov-Smirnov (K-S) distance between the data and the best-fit parametric distribution with 2) K-S distances between random numbers generated from the best-fit distribution and the best-fit distribution itself.

The idea was that the K-S distances based on the data and the K-S distances based on random numbers would be comparable in size if the fit between the data and the parametric probability distribution were good. On the other hand, distances based on the data would be much larger than distances based on random numbers if the fit were poor.

For each single-neuron response or each population response, the empirical cdf was determined from the data and the K-S distance (Eq. 8) between the empirical cdf and the best-fit parametric cdf was calculated. Next, random numbers were generated from the best-fit distribution, and their cdf was determined. The K-S distance between the random-number cdf and the best-fit parametric cdf was then measured. The size of the random number set was equal to the size of the data set. For each data vector (each single-neuron response or each population response), the random number procedure was repeated 1,000 times. Using those 1,000 replications, the fraction of times that the data K-S distance was larger than the randomnumber K-S distance was computed. Finally, an average fraction over all data vectors was calculated. If a best-fit parametric distribution perfectly described the data, that fraction would have a chance level of 0.5. For poorer fits the fraction becomes larger.

The results were as follows. For single-neuron data (both responses and response-modulations), K-S distance between the data and the best-fit gamma distribution was almost always larger than the K-S distance between gamma-distributed random numbers and the best-fit gamma distribution (0.91 fraction of the time). This indicates that the gamma distribution did not provide a very accurate description of the data, even though it was the best among the parametric distributions tested. A better description is likely to exist, although under a P = 0.05 criterion the hypothesis that the data are gamma distributed cannot be rejected.



Fig. 11. Zipf plots, showing response magnitude vs. response rank on log-log coordinates. Only upper tails are included (top 10% of response magnitudes). Also shown are the best-fit lines for the initial portion of the curves. *Ai*: plots based on responses. *Aii*: plots based on response. *Aii*: plots based on response-modulations. *Bi* and *Bii*: same plots as *A*, but based on normalized data.

1109

J Neurophysiol • VOL 106 • SEPTEMBER 2011 • www.jn.org

### STATISTICS OF INFEROTEMPORAL RESPONSES





For population responses, the K-S distance between the data and the best-fit log-normal distribution was larger than the distance between log-normally distributed random numbers and the best-fit log-normal distribution 0.80 fraction of the time. The log-normal distribution therefore provided only a modestly good approximation to population responses.

For population response-modulations, K-S distance between the data and best-fit  $\alpha$ -stable distribution was larger than the distance between  $\alpha$ -stable random numbers and best-fit  $\alpha$ -stable distribution 0.38 fraction of the time. Given this performance, the  $\alpha$ -stable distribution was a very good description of population response-modulations.

Moving from unnormalized to normalized population results, probability densities for normalized population activities are shown in Fig. 13, C and D. In this case the best-fit distribution was the log-logistic distribution, whose pdf is given by:

$$f(x|\mu,\sigma) = \frac{e^{\frac{-(\log(x)-\mu)}{\sigma}}}{\sigma\left(1+e^{\frac{-(\log(x)-\mu)}{\sigma}}\right)^2}$$
(21)

The median value of the scale parameter  $\sigma$  was 0.340 with interquartile range 0.022. The log-logistic distribution provided a good fit for these data, as the K-S distance between data and the best-fit log-logistic distribution was larger than the distance based on log-logistic random numbers about half the time (0.46 fraction).

For normalized population response-modulations, the best-fit distribution was the  $\alpha$ -stable distribution, although it was not significantly better than the log-logistic distribution. For the two  $\alpha$ -stable shape parameters, the median value of  $\alpha$  was 1.70 with interquartile range 0.08 while the median value of  $\beta$  was 0.83 with interquartile range 0.24. The

correlation between the two parameters was 0.17. The  $\alpha$ -stable distribution gave a good fit to these data, as the K-S distance between data and the best-fit  $\alpha$ -stable distribution was larger than distances based on  $\alpha$ -stable random numbers 0.28 fraction of the time.

Normalized single-neuron probability distributions are not presented, as the normalization process has no effect on them. Normalized single-neuron distributions would be essentially identical to those in Fig. 12.

### Pareto Analysis of the Tails

In the previous section we looked at the response pdfs as a whole. Here, given the importance of the tails of the distribution for selectivity and sparseness, we focus on examining the upper tail in more detail. When fitting pdfs to the complete data, we may be generating models that fit well near the peak of the data distribution but poorly on the tails. As there is little data on the tails, they have a weak influence on the overall fit of the probability model.

Based on higher kurtosis in population responses than in single-neuron responses, we expect probability distributions for population responses to have heavier tails than singleneuron responses.

Heaviness of upper tails was quantified by fitting generalized Pareto distributions (GPDs) (*Eqs. 17* and *18*) to tail data. A separate GPD was fit for each neuron (single-neuron responses) or each image (population responses). Of particular interest was the value of the parameter k of the best-fit distribution. That parameter, known as the tail index, describes the shape of the probability density tail. Larger values of kindicate a heavier tail.

Histograms of the generalized Pareto tail index are shown in Fig. 14*Ai* for response data. The mean value of the tail index for population responses was not significantly different from zero ( $\bar{k} = -0.05$ ). The mean tail index for single-





neuron responses was much smaller, being strongly negative  $(\bar{k} = -0.43)$ . The difference in the mean tail indices confirms that population responses have heavier tails than single-neuron responses. The difference was significant under a one-tail *t*-test (P < 0.0001).

Population response-modulations also had heavier tails than single-neuron response-modulations (Fig. 14Aii). The mean

value of the tail index for population response-modulations ( $\bar{k} = 0.24$ ) was much larger than that of single-neuron response-modulations ( $\bar{k} = -0.44$ ).

After normalization, population responses and responsemodulations still had heavier tails than single-neuron data. The mean tail index for normalized population data was greater than that of the single-neuron data (Fig. 14B).

### STATISTICS OF INFEROTEMPORAL RESPONSES



Fig. 14. Upper tail heaviness for response probability density functions, as quantified by Pareto tail indices. *Ai*: Pareto tail indices calculated individually for the upper tail of each single-neuron response (red) and each population response (blue). Larger values of the tail index indicate heavier tails. *Aii*: Pareto tail indices calculated using response-modulations rather than responses. *Bi* and *Bii*: Pareto tail indices using normalized data.

Values of the tail parameter k correspond to qualitatively different types of tails. For single-neuron responses and response-modulations (k < 0), the Pareto analysis indicated that probability distributions had finite upper tails. For unnormalized population responses ( $k \approx 0$ ), the tails were exponential. Power law tails (k > 0) were found for normalized population responses. Power law tails also occurred for population response-modulations, both normalized and unnormalized.

### Modeling

In this section we describe simulations that qualitatively reproduced various aspects of the data.

Simulations reproducing greater population sparseness than single-neuron selectivity. This was done by creating a population of model neurons that were statistically inhomogeneous. Each neuron had a different probability distribution of response magnitudes when presented with a set of images. Postulating that the pdf shape was different for each neuron goes beyond saying that each neuron was most responsive to a different critical feature.

The responses of each model neuron to a set of stimulus images were defined by a gamma probability distribution (*Eq. 19*). To create an inhomogeneous population, the gamma distribution was different for each neuron. This was achieved by setting the two parameters in the gamma distribution *a* and *b* to random values. The values of the two parameters were themselves gamma distributed, so that for each neuron they were set by the Matlab gamma random number generator with the following characteristics: a = gamrnd(4, 0.5) and b = gamrnd(2, 0.5). The number of model neurons and the number of stimulus images were set to match the data.

This modeling resulted in single-neuron selectivity and population sparseness values presented in Fig. 15. It demonstrates that a neural population with inhomogeneous response distributions reproduces the observation that population sparseness is greater than single-neuron selectivity (comparing Fig. 15*A* with Fig. 15*B*). Normalizing neural responses with respect to mean response had little effect on population sparseness (Fig. 15*C*). However, normalizing with respect to the standard deviation of responses led to a sharp drop in sparseness, below single-neuron selectivities (Fig. 15*D*). This is the same pattern seen when the data were normalized relative to mean response (Fig. 4, *C* and *D*) or normalized relative to standard deviation (Fig. 10). Again in this model we see the critical importance of heterogeneity in the dynamic range of different neurons for creating high population sparseness.

We examined the effects of including Poisson noise in the model. This was done by replacing each value in the deterministic process by a Poisson-distributed random number having the same mean value as the deterministic process. Introducing this noise produced very minor effects on the model results shown in Fig. 15, regardless of the number of trials that each image was presented to a neuron.

We tried using a homogeneous neural population with Poisson noise, in which every neuron had an identical response probability distribution. This failed to replicate the observation that population sparseness is greater than neural selectivity. In our simulations under those conditions, population sparseness and single-neuron selectivity were within 2% of each other.

These simulations also produced heavier tails for the probability distribution of population responses than for singleneuron responses. Tail heaviness was measured in the same manner described above for the data, fitting a generalized Pareto distribution to the upper tail (largest 10% of responses). For single-neuron responses, the mean value of the tail index





was  $\bar{k} = -0.04$ . For population responses, it was  $\bar{k} = 0.12$ . Larger values of the tail index indicate heavier tails.

The tail index observed in the data for single-neuron responses ( $\bar{k} = -0.43$ ) was smaller than that produced by the modeling ( $\bar{k} = -0.04$ ). This indicates that the probability distributions for single-neuron responses in reality had upper tails that were much thinner than those of the gamma distribution used in the modeling.

### Simulations Varying Tail Heaviness

The Pareto tail analysis showed that population responses have heavier tails than single-neuron responses, which was reproduced by the modeling in the previous section. This section examines whether various other differences between population responses and single-neuron responses presented above might be explained as consequences of tail heaviness.

The approach here was to take random numbers distributed in accord with a GPD and subject them to the same analyses methods used for the data. Two sets of random numbers were generated that differed only in the heaviness of their tails. In one set, the tail index k was set to -0.44 (lighter tail), and in the other it was set to -0.05 (heavier tail), values determined by experimental results. The other two parameters of the GPD were held fixed and arbitrarily set to  $\sigma = 1$  and  $\theta = 0$  (so that no attempt was made in the random number sets to match mean firing rates in the data). These sets of random numbers only simulated the upper tails of the data, and not their full probability distributions.

Figure 16 gives the results when the two sets of random numbers were subjected to some of the same analyses used previously. Figure 16A shows kurtosis as a function of data set size using Pareto random numbers, Fig. 16B shows maximum response as a function of data set size, and Fig. 16C shows Zipf plots of the random numbers. These correspond to Fig. 6, A and B, and Fig. 11, respectively in the data analysis.

Fig. 15. Selectivity results based on model simulations using inhomogeneous neural populations. Neurons had gamma-distributed responses over the set of stimulus images. The probability distribution for each neuron was different, making the population inhomogeneous. *A*: single-neuron selectivity for 674 model neurons. *B*: population sparseness for 806 images. *C*: population sparseness when activity of each neuron was normalized with respect to mean response over all images. *D*: population sparseness when activity of each neuron was normalized with respect to standard deviation of responses over all images. Normalized single-neuron selectivities are not shown because normalization has no effect on them.

In each case, the differences between the lighter tail random numbers (red) and heavier tail random numbers (blue) qualitatively captured the differences between single-neuron responses and population responses. There are obvious quantitative differences, perhaps due to the fact that we only simulated tail behavior and not the entire distribution and only looked at average characteristics without including the heterogeneity found in the data. Nevertheless, these simple simulation results do point out the importance of tail characteristics in determining the selectivity/sparseness properties of neural responses as well as other statistical properties we have been concerned with.

### DISCUSSION

These results quantify high sparseness/selectivity both in single-neuron activities (Fig. 4, A and C) and in population activities (Fig. 4, B and D). Inferotemporal single-neuron selectivity and population sparseness were both substantially higher than previously found in striate cortex with the same kurtosis measure (Lehky et al. 2005), even when the inferotemporal stimulus set size was adjusted (using Fig. 6A) to match the smaller set used by Lehky et al. (2005).

This high selectivity or sparseness in inferotemporal cortex need not be interpreted in terms of efficient coding for information transmission. Although sparse coding theories applied to V1 and other early visual structures have emphasized information theoretic notions of efficient coding, sparseness may contribute to other aspects of visual information processing such as speed of learning or memory capacity (Földiák 2002) that may be more relevant for the high-level visual representations in inferotemporal cortex (see also discussion by Lehky et al. 2005).

Population sparseness was greater than single-neuron selectivity. Although measurements of single-neuron selectivity have sometimes been used as a proxy for population sparse-



Fig. 16. Simulation results showing influence of tail heaviness. Random numbers were generated with a generalized Pareto distribution and were analyzed in the same manner as the data. Two values of the Pareto tail index were used, k = -0.05 (heavier tail) and k = -0.43 (lighter tail). A: kurtosis as a function of size of synthetic data set. B: maximum response as a function of size of synthetic data set. C: response magnitude vs. rank. In all cases, differences between the lighter-tailed distribution (red) and the heavier-tailed distribution (blue) qualitatively mimicked differences seen in the data between single-neuron responses and population responses.

ness, these data show that they are not the same. The observation that population sparseness is greater than single-neuron selectivity contradicts an earlier study that reported that they were the same (Franco et al. 2007). That different conclusion may have been the result of a much smaller data sample in that study (44 neurons, 20 stimulus images). As shown in Fig. 6, the difference between neural selectivity and population sparseness decreases for small samples. In addition, standard errors increase when the sample size is small. A combination of a smaller magnitude of the effect combined with larger standard errors may have obscured observation of such a difference in the earlier study.

Selectivity/sparseness was measured under a RSVP paradigm, with each stimulus image presented for a duration of 105 ms. Previous work has shown that neural selectivity is dependent on stimulus duration (Fig. 5 of Keysers et al. 2001), with selectivity decreasing for very short presentations. However, the data of Keyser et al. (2001) indicated that this decrease in selectivity was not a significant effect until stimulus durations fell below 42 ms, far below the speed we were using. Moreover, in those data there did not appear to be a change in selectivity going from 111 ms to 222 ms, suggesting that selectivity as a function of duration already plateaus at the stimulus duration we were using. The ability to identify or categorize rapidly presented images in psychophysical backward masking tasks plateaus for stimulus durations beyond ~50 ms (Bacon-Macé et al. 2005; Lehky 2000). This again suggests that our conditions were within the range where duration effects on visual processing cease to be a major factor.

Neurons were heterogeneous in their characteristics, differing in spontaneous activity, selectivity, dynamic range, tail heaviness, and their response probability distribution as a whole (among properties we were concerned with). Because of this heterogeneity, responses from neurons in a population therefore do not represent independent and identically distributed (i.i.d.) observations often assumed in statistical analyses.

If all neurons had identical response probability distributions, then population sparseness and single-neuron selectivity would necessarily be equal. The fact that sparseness and selectivity are not equal is a prima facie indication of statistical heterogeneity among neurons in the population. The question arises of what is the specific nature of the heterogeneity underlying high population sparseness. Normalizing neural responses with respect to mean response was not sufficient to destroy high population sparseness (compare Fig. 4, B and D). On the other hand, when we normalized with respect to the dynamic range of neurons (standard deviation of their responses), then high population sparseness disappeared (compare Fig. 4B and Fig. 10). These observations are replicated in the modeling (Fig. 15). Furthermore, removing neurons with high single-neuron selectivities from the population did not have a great effect on population sparseness (Fig. 9). It therefore appears that a critical requirement for high population sparseness is to have an inhomogeneous population with diversity in the dynamic ranges of the component neurons. High population sparseness is not dependent on having high singleneuron selectivities. This last observation serves to dissociate the concepts of "population sparseness" and "single-neuron selectivity." Possibly there are biophysical limits on how high single-neuron selectivities can go, and dynamic range diversity among neurons in a population is a mechanism for achieving higher population sparseness that transcends those limits.

Besides increased sparseness, another possible advantage for diversity in the dynamic range of inferotemporal neurons is increased accuracy in the representations they form. Neural modeling has shown that a mixture of different tuning curve widths increases the precision of representations compared with having uniform tuning curve widths (Wilke and Eurich 2002).

It is conceivable that a more diverse stimulus image set may have changed the results reported here. Quantifying what constitutes stimulus diversity for an image set or what constitutes an unbiased sampling of all possible visual stimuli in the environment is a difficult problem, particularly if the significant stimulus variable may not be an image as a whole but unknown critical features embedded within the image. What we can report, however, are strong and unambiguous results differentiating single-neuron selectivity and population sparseness using a stimulus set that appears to be far more diverse than in any previous study.

Training modifies the selectivity of inferotemporal neurons (Booth and Rolls 1998; Kobatake et al. 1998; Logothetis et al. 1995). Such changes can occur even for passive viewing of visual stimuli, where the images do not have any explicit significance for the monkeys in terms of task performance (Freedman et al. 2006). In this study, the monkeys repeatedly viewed the stimulus set over a period of months under a passive viewing task. Thus the process of measuring selectivity may have perturbed selectivity to some extent. However, a recent review of plasticity effects in inferotemporal cortex concluded that adult learning introduced only small modulations upon preexisting neural object representations (Op de Beeck and Baker 2010). We therefore believe that the essential selectivity properties we report are inherent characteristics of inferotemporal responses and not created by exposure to the stimuli during data collection.

Bowers (2009) has argued that the existing neurophysiological data are consistent with a localist or "grandmother cell" type of visual representation rather than sparse coding. In such a localist representation, the activity of a single cell is sufficient to uniquely identify an object. We believe that the data presented here do not support "grandmother cell" coding. The Zipf plots (Fig. 11) show relative response magnitudes as a function of rank order. On average, the second largest response was almost the same size as the largest response (89.3% for single-neuron responses, 79.5% for population responses). This is not a characteristic of "grandmother cell" coding. For larger stimulus set sizes, as occur under real-world conditions, the difference between largest and second largest responses would be even smaller.

"Grandmother cell" coding models generally include the assumption of a hierarchy of cells with increasing selectivity. The region we recorded from, anterior inferotemporal cortex (AIT), is the last unimodal visual area, at the top of ventral stream hierarchy. The primary projection of area TE in AIT is to perirhinal cortex (Saleem and Tanaka 1996; Suzuki and Amaral 1994), a polymodal area. Visual selectivity in perirhinal cortex is not greater than in TE (Lehky and Tanaka 2007; Naya et al. 2003). Thus it seems unlikely that there is some "higher area" where "grandmother cells" are commonplace. Nevertheless, it is still possible that grandmother cell encoding could occur for a small number of special objects to which the observer was highly exposed and which also had strong behavioral significance.

Tail heaviness can be viewed as the fundamental statistical characteristic differentiating population responses and singleneuron responses, with the observed difference in selectivity being secondary to that. The ordering of selectivity/sparseness in unnormalized data was, from lowest to highest: 1) singleneuron responses = single-neuron response-modulations; 2) population responses; 3) population response-modulations. The ordering of tail heaviness (tail index k in Fig. 14), from lightest to heaviest, was identical to the ordering for selectivity.

Measured selectivity was dependent on the size of the data set, increasing for larger data samples, either the number of images for single-neuron selectivity, or the number of neurons for population selectivity (Fig. 6, see also modeling in Fig. 16). This behavior, predicted by Lehky et al. (2005), occurs as it becomes more likely with larger data sets to hit a rare event on the upper tail.

The effect of data set size was stronger for responses that are more highly selective (or sparse). The heavy upper tail of selective distributions provides a nonnegligible probability of getting responses that are much larger than typical ones (those near the peak of the distribution). The outlying "large" responses, although infrequent, significantly alter the response vector as a whole. However, they can be missed if we only have a small sampling of responses. On the other hand, low selectivity responses with a thinner tail can be well characterized by the pattern of responses near the peak of the distribution, leading to a dense representation with less benefit from collecting a large data set.

The heaviness of tails was directly examined by fitting generalized Pareto distributions to tail data. Heavier tails were indicated by a larger tail index k. Under a Pareto analysis, tails fall into three qualitative categories. If k < 0 the tail is finite, if k = 0 the tail is exponential, and if k > 0 the tail follows a power law. For unnormalized results, the measured Pareto tail indices indicated finite tails for single-neuron data (both responses and response-modulations), exponential tails for population responses, and power law tails for population response-modulations. Observation of this power law relationship in the data connects the neurophysiology with a broad body of physical theory related to power laws (Newman 2005).

A finite tail for single-neuron response distributions means that, as we add more images to the stimulus set, at some point for each neuron we stop finding new images that produce a larger response than any previous image. This is an indication that the critical feature for each inferotemporal neuron is simple enough that it is likely to occur within a modest sampling of images.

[The statement that single-neuron responses stop increasing as more images are added does not contradict the failure of the maximum response vs. stimulus set size curve to asymptote for single-neuron responses in Fig. 6*B*. There is a scatter in the tail indices for different neurons (Fig. 14), with a small fraction of them having positive values, indicating that they do not have finite tails. The average maximum response over many neurons will be dominated by that small fraction with nonfinite tails.]

A heavier-tailed responses for population data (either exponential or power law tails) means that as we include more neurons in a population being stimulated by one particular image, we keep finding neurons that respond better to that image than any previous neuron (with no end in sight with our sample of on the order of  $10^3$  neurons). This suggests that there is an indefinitely large number of different critical features embedded in each image to which different neurons are tuned. By "indefinitely large" we do not mean infinite, as any physical system is obviously finite. Even though the number of critical features in each image is large, the total number of critical features scoled for in inferotemporal cortex must be far larger still, as most neurons do not respond strongly to each image.

Perhaps one way to see how a single image can generate a large number of critical features is the following. Think of the image as being randomly cut up into a jigsaw puzzle. Each piece is a feature. You can then go back and cut the same image into a different jigsaw puzzle, producing a different set of features, and repeat the same process indefinitely. At the end, all the pieces from all the different puzzles are pooled into

2013

a single pile, giving a vast array of different but overlapping features generated from a single image. These features may not be describable in a simple manner by a limited set of geometric parameters.

Having a large number of neurons tuned to different critical features would be at odds with structural models for object representation that posit a small number of discrete, standard shape primitives (Biederman 1987; Marr and Nishihara 1978). These analyses therefore offer a constraint on models of object representation.

Although the focus here has been on shape responses of inferotemporal neurons, in many cases the visual processing of objects requires spatial localization of stimuli ("where") as well as identification of "what" (Edelman 2002). Inferotemporal neurons, in addition to being shape selective, are also location selective (Op de Beeck and Vogels 2000), and location information can be extracted from their responses (Lehky and Sereno 2011; Sereno and Lehky 2011). It is therefore possible that "where" information for object processing might be represented internally within the ventral visual pathway rather than involving the dorsal pathway. Examining this issue would require further quantification of inferotemporal spatial responses and how they interact with shape responses.

Overall, we find that the statistics of single-neuron responses and population responses in inferotemporal cortex are very different. This difference ultimately depends on statistical heterogeneity in the responses of different neurons in the population, most importantly on the existence of diverse values for response dynamic range. On the basis of examination of tail heaviness of probability distributions, we conclude that the critical features for individual neurons are not very complex, but there is an indefinitely large number of them. We suggest that this finding constrains models of object representation.

### GRANTS

This research was supported by the Iran National Science Foundation (Tehran, Iran) and by the Japan Society for the Promotion of Science (JSPS) through its Funding Program for World-Leading Innovative R&D on Science and Technology (FIRST Program).

### DISCLOSURES

No conflicts of interest, financial or otherwise, are declared by the author(s).

### REFERENCES

- Adler R, Feldman R, Taqqu MS (Editors). A Practical Guide to Heavy Tails: Statistical Techniques and Applications. Boston, MA: Birkhäuser, 1998.
- Bacon-Macé N, Macé MJ, Fabre-Thorpe M, Thorpe SJ. The time course of visual processing: backward masking and natural scene categorisation. *Vision Res* 45: 1459–1469, 2005.
- Baddeley R, Abbott LF, Booth MC, Sengpiel F, Freeman T, Wakeman EA, Rolls ET. Responses of neurons in primary and inferior temporal visual cortices to natural scenes. *Proc R Soc Lond B Biol Sci* 264: 1775–1783, 1997.
- Bell AJ, Sejnowski TJ. The "independent components" of natural scenes are edge filters. *Vision Res* 37: 3327–3338, 1997.
- Biederman I. Recognition-by-components: a theory of human image understanding. *Psychol Rev* 94: 115–147, 1987.
- **Booth MCA, Rolls ET.** View-invariant representations of familiar objects by neurons in the inferior temporal visual cortex. *Cereb Cortex* 8: 510–523, 1998.
- **Bowers JS.** On the biological plausibility of grandmother cells: implications for neural network theories in psychology and neuroscience. *Psychol Rev* 116: 220–251, 2009.

- Bremmer F, Kubischik M, Hoffmann KP, Krekelberg B. Neural dynamics of saccadic suppression. J Neurosci 29: 12374–12383, 2009.
- **Coles S.** An Introduction to Statistical Modeling of Extreme Values. London: Springer, 2001.
- **Constantinidis C, Goldman-Rakic PS.** Correlated discharges among putative pyramidal neurons and interneurons in the primate prefrontal cortex. *J Neurophysiol* 88: 3487–3497, 2002.
- **De Baene W, Premereur E, Vogels R.** Properties of shape tuning of macaque inferior temporal neurons examined using rapid serial visual presentation. *J Neurophysiol* 97: 2900–2916, 2007.
- **Desimone R, Albright TD, Gross CG, Bruce CJ.** Stimulus-selective properties of inferior temporal neurons in the macaque. *J Neurosci* 4: 2051–2062, 1984.
- Downing PE, Chan AW, Peelen MV, Dodds CM, Kanwisher N. Domain specificity in visual cortex. *Cereb Cortex* 16: 1453–1461, 2006.
- Ecker AS, Berens P, Keliris GA, Bethge M, Logothetis NK, Tolias AS. Decorrelated neuronal firing in cortical microcircuits. *Science* 327: 584–587, 2010.
- Edelman S. Constraining the neural representation of the visual world. *Trends* Cogn Sci 6: 125–131, 2002.
- Edwards R, Xiao D, Keysers C, Földiák P, Perrett D. Color sensitivity of cells responsive to complex stimuli in the temporal cortex. *J Neurophysiol* 90: 1245–1256, 2003.
- Embrechts P, Klüppelberg C, Mikosch T. Modelling Extremal Events for Insurance and Finance. Berlin: Springer, 1997.
- Erickson CA, Jagadeesh B, Desimone R. Clustering of perirhinal neurons with similar properties following visual experience in adult monkeys. *Nat Neurosci* 3: 1143–1148, 2000.
- Field DJ. What is the goal of sensory coding? *Neural Comput* 6: 559–601, 1994.
- Földiák P. Sparse coding in the primate cortex. In: Handbook of Brain Theory and Neural Networks (2nd ed.), edited by Arbib MA. Cambridge, MA: MIT Press, 2002, p. 1064–1068.
- Földiák P, Xiao D, Keysers C, Edwards R, Perrett DI. Rapid serial visual presentation for the determination of neural selectivity in area STSa. *Prog Brain Res* 144: 107–116, 2004.
- **Franco L, Rolls ET, Aggelopoulos NC, Jerez JM.** Neuronal selectivity, population sparseness, and ergodicity in the inferior temporal visual cortex. *Biol Cybern* 96: 547–560, 2007.
- Freedman DJ, Riesenhuber M, Poggio T, Miller EK. Experience-dependent sharpening of visual shape selectivity in inferior temporal cortex. *Cereb Cortex* 16: 1631–1644, 2006.
- Fujita I, Tanaka K, Ito M, Cheng K. Columns for visual features of objects in monkey inferotemporal cortex. *Nature* 360: 343–346, 1992.
- Gawne TJ, Richmond BJ. How independent are the messages carried by adjacent inferior temporal cortical neurons? *J Neurosci* 13: 2758–2771, 1993.
- **Ghahfarokhi MA, Ghahfarokhi PB.** Applications of stable distributions in time series analysis, computer sciences and financial markets. *Proc World Acad Sci Eng Technol* 49: 1027–1031, 2009.
- Gnedenko BV, Kolmogorov AN. Limit Distributions for Sums of Independent Random Variables. Cambridge, MA: Addison-Wesley, 1968.
- Gross CG. Single neuron studies of inferior temporal cortex. *Neuropsychologia* 46: 841–852, 2008.
- Henderson JM, Hollingworth A. Eye movements during scene viewing: an overview. In: *Eye Guidance in Reading and Scene Perception*, edited by Underwood G. Oxford, UK: Elsevier, 1998, p. 269–293.
- Holden AV. A note on convolution and stable distributions in the nervous system. *Biol Cybern* 20: 171–173, 1975.
- Huang X, Lisberger SG. Noise correlations in cortical area MT and their potential impact on trial-by-trial variation in the direction and speed of smooth-pursuit eye movements. J Neurophysiol 101: 3012–3030, 2009.
- Hyvärinen A, Hurri J, Hoyer PO. Natural Image Statistics: a Probabilistic Approach to Early Computational Vision. London: Springer, 2009.
- Jazayeri M, Movshon JA. Optimal representation of sensory information by neural populations. *Nat Neurosci* 9: 690–696, 2006.
- Kano F, Tomonaga M. How chimpanzees look at pictures: a comparative eye-tracking study. *Proc R Soc Lond B Biol Sci* 276: 1949–1955, 2009.
- Keysers C, Xiao DK, Földiák P, Perrett DI. The speed of sight. J Cogn Neurosci 13: 90–101, 2001.
- Kiani R, Esteky H, Mirpour K, Tanaka K. Object category structure in response patterns of neuronal population in monkey inferior temporal cortex. J Neurophysiol 97: 4296–4309, 2007.

- Kiani R, Esteky H, Tanaka K. Differences in onset latency of macaque inferotemporal neural responses to primate and non-primate faces. J Neurophysiol 94: 1587–1596, 2005.
- Kobatake E, Tanaka K. Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. *J Neurophysiol* 71: 856–867, 1994.
- Kobatake E, Wang G, Tanaka K. Effects of shape-discrimination training on the selectivity of inferotemporal cells in adult monkeys. *J Neurophysiol* 80: 324–330, 1998.
- Koutrouvelis IA. An iterative procedure for the estimation of the parameters of stable laws. *Commun Stat Simul Comput* 10: 17–28, 1981.
- Koutrouvelis IA. Regression-type estimation of the parameters of stable laws. J Am Stat Soc 75: 918–928, 1980.
- Kovács G, Vogels R, Orban GA. Cortical correlate of pattern backward masking. Proc Natl Acad Sci USA 92: 5587–5591, 1995.
- Krieman G, Hung CP, Kraskov A, Quian Quiroga R, Poggio T, DiCarlo JJ. Object selectivity of local field potentials and spikes in the macaque inferior temporal cortex. *Neuron* 49: 433–445, 2006.
- Lehky SR. Fine discrimination of faces can be performed rapidly. J Cogn Neurosci 12: 848–855, 2000.
- Lehky SR, Sejnowski TJ. Network model of shape-from-shading: neural function arises from both receptive and projective fields. *Nature* 333: 452–454, 1988.
- Lehky SR, Sejnowski TJ, Desimone R. Selectivity and sparseness in the responses of striate complex cells. *Vision Res* 45: 57–73, 2005.
- Lehky SR, Sereno AB. Comparison of shape encoding in primate dorsal and ventral visual pathways. J Neurophysiol 97: 307–319, 2007.
- Lehky SR, Sereno AB. Population coding of visual space: modeling. Front Comput Neurosci 4: 155, 2011.
- Lehky SR, Tanaka K. Enhancement of object representations in primate perirhinal cortex during a visual working-memory task. J Neurophysiol 97: 1298–1310, 2007.
- Lévy P. Calcul des Probabilités. Paris: Gauthier-Villars, 1925.
- Logothetis NK, Pauls J, Poggio T. Shape representation in the inferior temporal cortex of monkeys. *Curr Biol* 5: 552–563, 1995.
- Logothetis NK, Sheinberg DL. Visual object recognition. *Annu Rev Neurosci* 19: 577–621, 1996.
- Mandelbrot B. The variation of certain speculative prices. J Business 36: 394–419, 1963.
- Mantegna RN, Stanley HE. Introduction to Econophysics: Correlations and Complexity in Finance. Cambridge, UK: Cambridge Univ. Press, 2000.
- Marr D, Nishihara NK. Representation and recognition of the spatial organization of three-dimensional shapes. *Proc R Soc Lond B Biol Sci* 200: 269–294, 1978.
- **Moody SL, Wise SP, di Pellegrino G, Zipser D.** A model that accounts for activity in primate frontal cortex during a delayed matching-to-sample task. *J Neurosci* 18: 399–410, 1998.
- Naya Y, Yoshida M, Miyashita Y. Forward processing of long-term associative memory in monkey inferotemporal cortex. J Neurosci 23: 2861– 2871, 2003.
- Newman WE. Power laws, Pareto distributions and Zipf's law. *Contemp Phys* 46: 323–351, 2005.
- Nikias CL, Shao M. Signal Processing with Alpha-Stable Distributions and Applications. New York: Wiley, 1995.
- Nolan JP. Information on Stable Distributions. academic2.american.edu/ ~jpnolan/stable/stable.html [accessed May 1, 2011a].
- Nolan JP. Stable Distributions—Models for Heavy Tailed Data. academic2.american.edu/~jpnolan/stable/chap1.pdf [accessed May 1, 2011b].
- **Olshausen BA, Field DJ.** Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature* 381: 607–609, 1996.
- **Olshausen BA, Field DJ.** Sparse coding with an overomplete basis set: a strategy employed by V1? *Vision Res* 37: 3311–3325, 1997.
- **Op de Beeck H, Vogels R.** Spatial sensitivity of macaque inferior temporal neurons. *J Comp Neurol* 426: 505–518, 2000.
- **Op de Beeck HP, Baker CI.** The neural basis of visual object learning. *Trends Cogn Sci* 14: 22–30, 2010.
- **Pickands J.** Statistical inference using extreme order statistics. *Ann Stat* 3: 119–131, 1975.

- Pouget A, Dayan P, Zemel R. Information processing with population codes. Nat Rev Neurosci 1: 125–132, 2000.
- Quian Quiroga R, Panzeri S. Extracting information from neuronal populations: information theory and decoding approaches. *Nat Rev Neurosci* 10: 173–185, 2009.
- Reich DS, Mechler F, Victor JD. Independent and redundant information in nearby cortical neurons. *Science* 294: 2566–2568, 2001.
- **Rolls ET.** Functions of the primate temporal lobe cortical visual areas in invariant visual object and face recognition. *Neuron* 27: 205–218, 2000.
- **Rolls ET, Tovée MJ.** Processing speed in the cerebral cortex and the neurophysiology of visual masking. *Proc R Soc Lond B Biol Sci* 257: 9–15, 1994.
- **Rolls ET, Tovée MJ.** Sparseness of the neuronal representation of stimuli in the primate temporal visual cortex. *J Neurophysiol* 73: 713–726, 1995.
- Rolls ET, Tovée MJ, Panzeri S. The neurophysiology of backward visual masking: information analysis. *J Cogn Neurosci* 11: 300–311, 1999.
- Saleem KS, Tanaka K. Divergent projections from the anterior inferotemporal area TE to the perirhinal and entorhinal cortices in the macaque monkey. *J Neurosci* 16: 4757–4775, 1996.
- Samorodnitsky G, Taqqu MS. Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance. Boca Raton, FL: CRC, 1994.
- Sanger TD. Neural population codes. Curr Opin Neurobiol 13: 238–249, 2003.
- Sereno AB, Lehky SR. Population coding of visual space: comparison of spatial representations in dorsal and ventral pathways. *Front Comput Neurosci* 4: 159, 2011.
- Simoncelli EP, Olshausen BA. Natural image statistics and neural representation. Annu Rev Neurosci 24: 1193–1216, 2001.
- Smith MA, Kohn A. Spatial and temporal scales of neuronal correlation in primary visual cortex. J Neurosci 28: 12591–12603, 2008.
- Suzuki WA, Amaral DG. Perirhinal and parahippocampal cortices of the macaque monkey: cortical afferents. J Comp Neurol 350: 497–533, 1994.
- Tamura H, Tanaka K. Visual response properties of cells in the ventral and dorsal parts of the macaque inferotemporal cortex. *Cereb Cortex* 11: 384–399, 2001.
- Tanaka K. Inferotemporal cortex and object vision. *Annu Rev Neurosci* 19: 109–139, 1996.
- Tolhurst DJ, Smyth D, Thompson ID. The sparseness of neuronal responses in ferret primary visual cortex. J Neurosci 29: 2355–2370, 2009.
- Treves A, Panzeri S, Rolls ET, Booth MCA, Wakeman EA. Firing rate distributions and efficiency of information transmission of inferior temporal cortex neurons to natural visual stimuli. *Neural Comput* 11: 601–632, 1999.
- Veillette M. Alpha-Stable Distributions Matlab Code. math.bu.edu/people/ mveillet/research.html [Accessed May 1, 2011].
- Vinje WE, Gallant JL. Sparse coding and decorrelation in primary visual cortex during natural vision. *Science* 287: 1273–1276, 2000.
- Waydo S, Kraskov A, Quian Quiroga R, Fried I, Koch C. Sparse representation in the human medial temporal lobe. *J Neurosci* 26: 10232–10234, 2006.
- Wilke SD, Eurich CW. Representational accuracy of stochastic neural populations. *Neural Comput* 14: 155–189, 2002.
- Willmore B, Tolhurst DJ. Characterizing the sparseness of neural codes. *Network* 12: 255–270, 2001.
- Wörgötter F, Daunicht WJ, Eckmiller R. An on-line spike form discriminator for extracellular recordings based on an analog correlation technique. *J Neurosci Methods* 17: 141–151, 1986.
- Zhang K, Ginzburg I, McNaughton BL, Sejnowski TJ. Interpreting neuronal population activity by reconstruction: unified framework with application to hippocampal place cells. J Neurophysiol 79: 1017–1044, 1998.
- Zoccolan D, Kouh M, Poggio T, DiCarlo JJ. Trade-off between object selectivity and tolerance in monkey inferotemporal cortex. *J Neurosci* 27: 12292–12307, 2007.
- Zohary E, Shadlen MN, Newsome WT. Correlated neuronal discharge rate and its implications for psychophysical performance. *Nature* 370: 140–143, 1994.
- Zolotarev VM. One-Dimensional Stable Distributions. Providence, RI: Am. Mathematical Soc., 1986.

Downloaded from http://jn.physiology.org/

'at Stanford Univ Medical Center

9 9

January 18

201

ω