
**Large-scale two-photon imaging revealed super-sparse
population codes in V1 superficial layer of awake monkeys**

Shiming Tang^{1,2,3*}, Yimeng Zhang⁴, Zhihao Li⁴, Ming Li^{1,2,3}, Fang Liu^{1,2,3},
Hongfei Jiang^{1,2,3}, Tai Sing Lee^{4*}

¹Peking University School of Life Sciences and Peking-Tsinghua Center for
Life Sciences, China; ²IDG/McGovern Institute for Brain Research at Peking
University, China; ³Key Laboratory of Machine Perception (Ministry of
Education), Peking University, Beijing 100871, P. R. China; ⁴Center for the
Neural Basis of Cognition and Computer Science Department, Carnegie
Mellon University, Pittsburgh, Pennsylvania, United States

* To whom correspondence should be addressed (tangshm@pku.edu.cn and
tai@cnbc.cmu.edu)

Abstract

One general principle of sensory information processing is that the brain must optimize efficiency by reducing the number of neurons processing the same information. The sparseness of the sensory representations in a population of neurons reflects the efficiency of the neural code. Here we employ large-scale two-photon calcium imaging to examine the responses of a large population of neurons with single-cell resolution, within the superficial layers of area V1, while simultaneously presenting a large set of natural visual stimuli, to provide the first direct measure of the population sparseness in awake primates. The results show that only 0.5% of neurons respond strongly to any given natural image—indicating a tenfold increase in the inferred sparseness over previous measurements. These population activities are nevertheless necessary and sufficient to discriminate visual stimuli with high accuracy, suggesting that the neural code in the primary visual cortex is both super-sparse and highly efficient.

Introduction

The efficient coding hypothesis is an important organizing principle of any sensory system (*Barlow, 1981; Olshausen and Field, 1996*). It predicts that neuronal population responses should be sparse, though the optimal level of sparseness depends on many factors. Most of the experimental evidence of sparse coding comes from individual neurons that were exposed to a large set of natural image stimuli, measured using single-unit recording techniques (*Haider et al., 2010; Hromadka et al., 2008; Rust and DiCarlo, 2012; Vinje and Gallant, 2000*). The sparseness of a neuron's response to a large set of stimuli in these studies was used to infer the sparseness of the population responses. The first direct measurement of population response sparseness has been done with two-photon (2P) GCaMP6 signal imaging in rodents (*Froudarakis et al., 2014*). One potential confound, however, is that GCaMP6 responses are subject to saturation at neuronal firing rates above 60-80 Hz (*Chen et al., 2013; Froudarakis et al., 2014*), which would under-estimate the sparseness measures that can capture the peakedness or sharpness of the population response distributions. Thus, direct and accurate measurement of population sparseness of neuronal response, particularly in non-human primates, remains lacking.

In this study, we provided the first direct measurement of population sparseness from V1 of awake macaques. We performed 2P imaging on a large

population of neurons using the genetically encoded calcium indicator GCaMP5 (*Akerboom et al., 2012*), delivered with adeno-associated viruses (AAVs). We showed previously that GCaMP5 exhibits linear non-saturating responses across a wide range of firing rates (10-150 Hz) (*Li et al., 2017*), which allows us to accurately measure response sparseness of almost all neurons in layer 2 in V1, within a 850 μm x 850 μm field of view—the spatial scale of about one hypercolumn.

Results and discussion

Our 2P imaging of GCaMP5 recorded neuronal population calcium (Ca^{++}) responses to 2,250 natural images from V1 layer 2 in two awake macaques (about 1000 neurons each). The monkey performed a fixation task while stimuli were presented to the appropriate retinotopic position in the visual field. Each trial sequence consisted of: a blank screen presented for one second, followed by a visual stimulus for one second. Each activated cell's region-of-interest (ROI) was defined as the compact region (>25 pixels) in which brightness exceeded 3 standard deviations (stds) above baseline, for each individual differential image. The standard ratio of fluorescence change ($\Delta F/F_0$) of each of these ROIs during stimulus presentation was calculated as the neuron's response (*see Methods*).

The receptive fields (RFs) of the neurons were characterized using

oriented gratings and bars presented in various positions. The RF centers of the imaged neurons were located between 3° and 5° in eccentricity. In each trial, a stimulus of 4° x 4° in size was presented, randomly drawn from a set of 2,250 natural image stimuli (**Figure 1, Figure 1-Figure supplements 1c**). The entire set of stimuli was repeated three times. These natural images evoked robust visual responses in the imaged neurons (**Figure 1a,b, Figure 1-Figure supplements 1c**).

We recorded from 1,225 neurons in monkey A and 982 in monkey B, and found that only a few neurons in each monkey responded strongly to any given image (**Figure 1a,b**), though most of the neurons responded strongly to at least one of the images in the set (**Figure 1c**). The rank-ordered distributions of the population responses were always sharply peaked (**Figure 1d,e**). On average, the percentage of neurons that responded above each's half-maximum were 0.49% (6.0/1,225) and 0.42% (4.1/982) for monkey A and B respectively (**Figure 1f,g**). This is a measure of population sparseness that can capture the peakedness of the population response distribution (**see Methods**). In other words, only about 0.5% of the cells responded *strongly* to each image, indicating a very high degree of sparseness in the strong population responses or a very sharp population response distribution.

We also examined each neuron's stimulus specificity or life-time sparseness (**Willmore et al., 2011**). Interestingly, we found most cells responded strongly only to a small number of images in the whole stimulus set

(**Figure 2a,b**). The preferred images for individual neurons often shared common features. For example, neuron 653 of monkey A was most excited when its receptive field (0.8° in diameter) covered the lower rim of the cat's eye (indicated by the red dashed line in the inset in **Figure 2b**). The neuron's preference for the specific curve feature was further confirmed by checking its selectivity to a variety of artificial patterns (**Figure 2b**). Similarly, neuron 949 of Monkey A was selective to a different specific curvature embedded within the natural stimuli (**Figure 2d**). A more systematic characterization of these cells using artificial stimulus patterns was reported previously (**Tang et al., 2018**), in which we showed that many of these neurons are highly selective to specific complex patterns. To measure the sharpness of each neuron's stimulus tuning curve—its life-time response sparseness—we computed the percentage of the stimuli that excited each neuron $> 50\%$ of the peak response found from the entire stimulus set. This population average was 0.49% ($11/2,250$) in Monkey A and 0.41% ($9.3/2,250$) in Monkey B (**Figure 2e,f**) respectively. This suggests that a high degree of stimulus specificity or life-time sparseness thus goes hand-in-hand with a heightened population sparseness.

To understand how much information was carried in the sparse ensemble of population activities, we evaluated how well the sparse neural responses allow a decoder to discriminate the 2,250 stimuli (**Quiroga RQ and Panzeri, 2009; Froudarakis et al., 2014**). The entire population's decoding accuracy was 54% for Monkey A and 38% for Monkey B, whereas the chance accuracy

was 0.04% (1/2,250) (**Figure 3a,b**, horizontal dashed lines). These population decoding accuracies estimate the *highest achievable decoding performance* from the full population activities that we recorded, and serve as the upper limit of accuracy performance that can be achieved by decoders that are made from any subset of neurons in the population (**Fig. 3a,b**, horizontal dashed lines).

Now that we know the highest achievable decoding performance of the entire population, we can build on that idea to assess the contribution of strong sparse responses to the overall decoding accuracy. We zeroed all responses below the top 0.5% response threshold and built a new decoding model to discriminate the 2,250 stimuli. Here, we assume only the strong signals above the threshold would be conveyed to downstream neurons successfully. Remarkably, a decoding accuracy of 28% could be achieved for monkey A and 21% for monkey B, with only the top 0.5% of the strongest signals included (**Figure 3a,b** vertical gray lines). This means, transmitting the top 0.5% of the strong responses for each image was sufficient for realizing 50% of achievable decoding performance for both monkeys.

Conversely, we can assess the necessity of the top 0.5% of the responses by repeating the decoding experiment after removing the strongest signals by setting them to 0, and keeping the remaining 99.5% of the signals intact. The blue curves in Figure 3a and 3b show the decoding performance with signals above a range of percentage threshold “removed”. The intersections of the blue curves with the gray lines indicate the decoding performance dropped by

50% when the strongest 0.5% of the signals are removed. Thus, we showed that these strong and sparse signals contain the necessary signals for realizing half of the decoding performance.

Fig. 3a and 3b also provide a more complete picture of the decoding performance as a function of percentage threshold. While the top 0.5% of the signals are both necessary and sufficient for realizing 50% of the performance, 99% of decoding performance is not reached until the top 40% (for monkey A) and top 30% (for monkey B) of the responses are included (saturation of the red curves). However, without the top 5% responses, the decoding performance drops practically to zero (blue curve). Thus, having the strongest 5% of the responses are necessary for achieving full performance, but insufficient by themselves. The presence of other weaker signals in the population is required to achieve the full performance. The decoding results revealed the significant information contents are indeed carried by the superb sparse strong responses.

We note that, in the above analyses, the selection of responses to keep or remove are based on the absolute response magnitude of the cells, rather than on a normalized response magnitude of each cell as determined by its peak response. We made this decision because we assume that downstream neurons may know where the signals come from, but it's unclear whether they can know (or remember from historical responses) the peak response magnitude of the neuron that is providing the signals. For better comparison

with the population sparseness measure, we repeated the decoding experiment where the percentage threshold chosen was relative to the peak response of each individual neuron. We find that the results are qualitatively similar, and that strong sparse responses still carry a disproportionate amount of information (**Figure 3-Figure supplements 1**). Quantitatively, the top 1.5% of the responses are now required to achieve 50% of the decoder's highest achievable performance. This suggests that absolute response strengths may potentially convey more discriminable information to downstream neurons. The decrease in performance based on the top 0.5% responses selected with a relative threshold (Figure 3-Figure supplements 1, red curve) is understandable because the relative threshold will include some useless contribution from neurons with weak peak responses and excluding the more useful contribution of some neurons with high peak responses, amplifying the effect of noises particularly when a small number of neurons were selected.

In conclusion, this study provides the first simultaneous recording of a large dense population in V1 at single cell resolution in response to a large set of natural stimuli using 2P imaging in awake macaque. Whereas earlier studies provided life-time sparseness measurements in rodents (**Hromadka et al. 2008; Haider et al. 2010**), non-human primates (**Rolls and Tovee, 1995; Vinje and Gallant, 2000; rust and DiCarlo 2012**) and humans (**Quiroga et al. 2005**)—and also population sparseness measurement in rodents (**Froudarakis et al. 2014**)—our study provides the first direct measurement of

sparseness of large-scale neuronal population responses in awake macaque monkeys, made possible by the large-scale 2P imaging techniques. We found that a very small ensemble of neurons from V1's superficial layer would be active at a time in response to any given natural image. Using decoding analysis, we showed that these small ensembles of neural responses provide a surprisingly large amount of information for downstream neurons for discriminating complex image patterns in natural scenes.

Earlier studies inferred population sparseness based on measurements of life-time sparseness. We show here for the first time through direct measurements of population sparseness was indeed comparable to life-time sparseness measure. However, the level of sparseness we observed (0.5% at half maximal response) was considerably higher than earlier life-time sparseness estimates, based on single unit recording in macaques (*Rust and DiCarlo, 2012*). Studies on rodents have moreover yielded a considerable range of estimates of sparseness that vary across measurement techniques (*Haider et al., 2010; Hromadka et al., 2008; Froudarakis et al., 2014*). A single-unit study that used the cell-attached patch technique might have been the most accurate to date (*Hromadka et al., 2008*), and it showed that neurons were mostly silent in the awake auditory cortex, inferring that less than 2% of the neuronal population showed “well-driven” responses (> 20 Hz response frequency) to natural sounds. From our imaging observations, neurons in the superficial layers of V1 were densely packed, with small cell

bodies. It may not be possible to obtain stable and well isolated single-unit signals over several hours using extracellular recording methods. Our study has reduced the biases inherent to previous extracellular recording studies—in neuronal sampling, stimulus sampling, and in single-cell isolation—by imaging virtually all the neurons in a single field of view and in a particular layer with a large set of natural stimuli.

The high degree of population sparseness we observed is consistent with two recent conjectures from theoretical neuroscience. First, based on the metabolic costs of spiking, one group posits that fewer than 1% of the neurons should be substantially active concurrently in any brain area (*Lennin, 2003*). Second, and more importantly, theoretical sparse coding studies have suggested that because the number of V1 neurons is at least 200 times more abundant than its thalamic input, V1 neurons could be quite specialized in their feature selectivity and thus highly sparse in their population responses (*Olshausen, 2013; Rehn and Sommer, 2007*). We have indeed observed this very finding using 2P imaging techniques in the V1 superficial layers (*Tang et al. 2018*). These findings are reminiscent of the highly specific codes exhibited by neurons in the human medial temporal lobes (*Quiroga et al., 2005*), suggesting that many V1 neurons might be akin to highly specific “grandmother neurons”, though they may encode information in the form of an extremely sparse population code. The observed high degree of population and life-time sparseness are also consistent with our earlier observation that

V1 neurons in this layer were tuned to complex patterns with a great degree of specificity (*Tang et al., 2018*, see also *Hedgé and Van Essen, 2007*). These findings reveal the complexity and specificity of feature selectivity and the super-sparse neural representation within a V1 hypercolumn, providing new understanding of the neural codes in the macaque primary visual cortex.

Materials and methods

Key Resources Table

Reagent type (species) or resource	Designation	Source or reference	Identifiers	Additional information
gene ()				
strain, strain background (Macaque)	Rhesus monkeys	Beijing Prima Biotech Inc	http://www.primasbio.com/cn/Default	http://www.primasbio.com/cn/Default
genetic reagent ()				
cell line ()				
transfected construct ()				
biological sample ()				
antibody				
recombinant DNA reagent	AAV1.hSyn.GCaMP5G	Penn Vector Core	V5072MI-R	
sequence-based reagent				
peptide, recombinant protein				
commercial assay or kit				
chemical compound, drug				
software, algorithm	Matab 7.12.0 (R2011a)	MathWorks	Matab 7.12.0 (R2011a)	https://www.mathworks.com
software, algorithm	Codes for the decoding analysis and image movement correction	This paper	Codes for the decoding analysis and image movement correction	https://github.com/lelabcnbc/sparse-coding-elif2018
other				

All experimental protocols were approved by the Peking University Animal Care and Use Committee, and approved by the Peking University Animal Care and Use Committee (LSC-TangSM-5).

Subjects

The study used two adult rhesus monkeys (A and B), 4 and 5 years of age and weighing 5 and 7 kg (Li et al., 2018). Two sequential surgeries were performed

on each animal under general anesthesia and strictly sterile conditions. In the first surgery, a 16-mm hole was drilled in the skull over V1. The dura was opened to expose the cortex, into which 50-100 nl AAV1.hSynap.GCaMP5G.WPRE.SV40 (AV-1-PV2478, titer 2.37e13 (GC/ml), Penn Vector Core) was pressure-injected at a depth of ~500 μ m. After AAV injection, the dura was sutured, the skull cap was placed back, and the scalp was sutured. Then the animal was returned to its cage for recovery. Antibiotic (Ceftriaxone sodium, Youcare Pharmaceutical Group Co. Ltd., China) was administered for one week. After 45 days, a second surgery was performed, in which three head-posts were implanted on each animal's skull, two on the forehead and one on the back of the head. A T-shaped steel frame was connected to these head-posts for head stabilization during imaging. The skull and dura were later opened again to explore the cortex. A glass cover-slip (diameter 8 mm and thickness 0.17 mm) was glued to a titanium ring and gently pressed onto the cortical surface. A ring-shape GORE membrane (20 mm in outer diameter) was inserted under the dura. The titanium ring was glued to the dura and skull with dental acrylic to form an imaging chamber. The whole chamber (formed by thick dental acrylic) was covered by a steel shell to prevent breakage of the cover-slip when the animal was returned to the home cage.

Behavioral task

During imaging, each monkey sat in a standard primate chair with head restraint and performed a fixation task, which involved fixating on a small white spot (0.1°) within a window of 1° for over 2 seconds to obtain a juice reward. Eye position was monitored with an infrared eye-tracking system (ISCAN, Inc.) at 120 Hz.

Visual stimuli

Visual stimuli were generated using the ViSaGe system (Cambridge Research Systems) and displayed on a 17" LCD monitor (Acer V173, 80Hz refresh rate) positioned 45 cm from the animal's eyes. Each stimulus was presented for 1 second after a 1 second blank within a fixation period of 2 seconds. We estimated the RF sizes and positions of the imaged neurons with small drifting gratings and bars presented at different locations. The RFs were estimated to be 0.2° to 0.8° in size with RF locations between 3 to 5 degrees in eccentricity for both monkeys.

Drifting and oriented gratings were tested to examine the visual responses of imaged neurons (*Li et al., 2017*). Small patches (0.8° in diameter) of gratings with 100% contrast square waves were presented to the center of RFs of imaged cells, with 2 spatial frequencies (4.0 and 8.0 cyc/deg) at 2 temporal frequencies (1 and 2 Hz), 6 orientations, and 2 directions (30° apart).

A natural stimulus set (NS) of 2,250 $4^\circ \times 4^\circ$ stimulus patches extracted from different natural scene photos was used to examine the neuronal

responses to natural stimuli. The order of the stimuli was randomized in each session. These stimuli were tested on monkeys A and B, each with at least three repetitions.

Eye movement control

We analyzed the distribution of eye-positions during stimulus ON periods. The monkeys' fixation during stimulus presentation (from 1 to 2 second in the graph) was stable and accurate. The distribution of eye positions during stimulus presentation, with standard deviations less than 0.05° , which was significantly smaller than the typical receptive field sizes (ranging from 0.2 to 0.8 degrees) of neurons at 3-5 degree eccentricities. To examine whether the eye movement had significant contribution to the distribution of neuronal population responses, we compared the standard deviations (stds) of eye position in different population response classes of neurons: (1) weak responses ($\Delta F/F_0 < 0.5$), (2) sparse strong responses (one or two cells responded), (3) dense responses (more than ten cells responded). We found no statistically significant differences in the distribution of eye position data in these three classes (*Tang et al., 2018*), indicating that the observed effects were not caused by movement differences. The ROC and decoding analysis (*Figure 1-Figure supplement 2*), demonstrating the reliability of the neural responses across trials confirm that the sparse population responses were evoked by stimuli repeatedly, not by random eye-movement jitters.

Two-photon imaging

After a recovery period of 10 days from the second surgery, the animals were trained to maintain eye-fixation. Two-photon imaging was performed using a Prairie Ultima IV (In Vivo) 2P microscope (Bruker Nano, Inc., FMBU, formerly Prairie Technologies) powered by a Ti: Sapphire laser (Mai Tai eHP, Spectra Physics). The wavelength of the laser was set at 1000 nm. With a 16× objective (0.8 - N.A., Nikon), an area of 850µm × 850µm was imaged. A standard slow galvonometer scanner was used to obtain static images of cells with high resolution (1024 × 1024). The fast and resonant scan (up to 32 frames per second) was used to obtain images of neuron activity. The images were recorded at 8 frames per second by averaging each 4 frames. Infected cells of up to 700 µm in depth were imaged. We primarily focused on cells 160 µm to 180 µm deep, which contained a high density of infected cells.

Imaging data analysis

All data analyses were performed using customized Matlab software (The MathWorks, Natick, MA). The images from each session were first realigned to a template image (the average image of 1000 frames) using a normalized cross-correlation-based translation algorithm, to correct the X-Y offset of images caused by the motion between the objective and the cortex (*Li et al., 2017*).

The cell density was high in superficial V1, and many cell bodies were quite dim at rest. It was difficult to identify these cells directly by eye or by algorithm based on the morphology from their static images. We therefore identified ROIs for cell bodies based on their responses. The differential images (which were averaged frames of the ON stimulus period, from which we then subtracted the average of the stimulus OFF period, for each stimulus condition) were first filtered using low-pass and high-pass Gaussian filters (5 pixels and 50 pixels). Notably, these two filters were used solely for ROI identifications. In all further analyses, we used the raw data without any filtering. Connected subsets of pixels (>25 pixels) with average pixel value greater than 3 stds in these differential images were identified as active neuronal ROIs. Note that these 3 stds' empirical value was used only for deciding the ROIs of the activated cells, and was not used as a cutoff threshold for measuring neuronal responses (*Figure 1-Figure supplements 3*). The ratio of fluorescence change ($\Delta F/F_0$) of these ROIs was calculated for each activated cell. $\Delta F = F - F_0$, where F_0 is the baseline activity during the blank screen prior stimulus onset in each trial and F is fluorescence activity in the ROI during stimulus presentation in the trial. A neuropil-correction was performed with an index of 0.7 (*Chen et al., 2013*).

Sparseness measure

The sparseness measure is used to quantify the peakedness of the

response distribution. There are several different definitions of sparseness and corresponding sparseness measures (*Willmore et al., 2011*). One intuitive one for sparse codes, described by Willmore et al. (2011), is that “the population response distribution that is elicited by each stimulus is peaked (exhibits population sparseness). A peaked distribution is one that contains many small (approximately zero) magnitude values and only a small number of large values. Thus, a neural code will have high population sparseness if only a small proportion of the neurons are strongly active at any given time.” A measure consistent with this intuition is the percentage of neurons that responded strongly, above a certain threshold relative to their peak response. This measure has been used in other studies (*Rust and DiCarlo, 2012*), typically with a half-peak response threshold.

The sparseness measures based on the calculation suggested by Rolls and Tovee (*Rolls and Tovee, 1995; Vinje and Gallant, 2000*) are popular for quantifying sparseness of spiking data, but they are very sensitive to measurement noise and uncertain baselines due to nonlinearities and missing responses in the low spiking rate range (< 10 Hz) in calcium imaging (*Li et al., 2017*). The sparseness measure we used, which are based on the percentage of the cells or stimuli above the half-maximum of each neuron, is much less sensitive to low level activities (iceberg effect) or baseline fluctuations in the calcium signal in this study.

Stability and reliability of the neuronal measurements

For each single neuron, we examined whether the sparse strong responses ($\Delta F/F_0 > 50\%$ max) observed across the 2,250 stimuli were reliable across trials by performing the following ROC analysis (*Quiroga et al., 2005*): we set all the stimuli that produced mean responses greater than 50% of the observed maximum mean peak of the cell to be in the ON class, and all other stimuli to be in the OFF class. We computed the ROC for classifying the ON class against the OFF class based on the response of each single trial. If the responses above the half-maximum are stable across all trials, then the AUC (the area under the ROC curve) will be close to 1.0, because the ON and OFF classes are readily discriminable. The null hypothesis is that sparse strong responses will be spurious single trial epileptic responses, and not repeatable across trials. To test this hypothesis, we shuffled all the responses against the stimulus labels, and recomputed the mean responses for all the stimuli across the trials. We performed 1000 shuffles. We found that most of the shuffled cases have much lower average peak responses because of the mismatch of the rigorous sparse responses across trials, suggesting the reliability of the sparse responses in the original data. To make an even stricter and more fair comparison with the original data on ROC terms, for each shuffle, we recomputed the maximum responses, and used the half of this mean maximum as threshold to sort the stimuli into ON and OFF classes and repeated the ROC analysis to obtain the AUC for this shuffle. The probability of

the null hypothesis is the percentage of the time that the AUCs of the 1000 shuffles reach the AUC of the original data. With this ROC analysis, we found > 96% neurons with the probability of the null hypothesis $p < 0.01$ (**Figure 1-Figure supplements 2**).

Decoding Analysis

We used a nearest centroid classifier to discriminate the 2,250 images based on the population responses in each trial. Since each image was tested 3 times, the nearest centroid classifier was trained based on two trials for all images and tested on the hold-out trials. We repeated the procedure for each trial, performing 3-fold cross-validations.

For each monkey, we constructed neural response matrices (with dimension 2250×1225 for monkey A, and 2250×982 for monkey B) for three trials $X^{(1)}$, $X^{(2)}$, and $X^{(3)}$, that store the neural responses to all images in each trial as rows in that matrix. We trained and tested nearest-centroid classifiers via a three-fold cross-validation procedure across trials in a 2250-way image decoding task. Specifically, for trial t , during training, we computed the centroids of the other two trials $C^{(t)}$ (if $t=1$, $C^{(1)} = (X^{(2)} + X^{(3)})/2$, if $t=2$, $C^{(2)} = (X^{(1)} + X^{(3)})/2$, etc.) and stored $C^{(t)}$ in the classifier; during testing, given some row k of $X^{(t)}$, which is the population neural response vector to image k in trial t , the (trained) classifier computed the Euclidean distances between row k of $X^{(t)}$ and every row of $C^{(t)}$. The model outputted the index $(1, 2, \dots, 2249, 2250)$ of the row

in $C^{(t)}$ that gives the smallest distance. The correct output is k and all other outputs are incorrect. The average decoding accuracy for this trial is defined as the percentage of correct outputs over all rows of $X^{(t)}$. We repeated the above procedure for each trial and reported the average of three (average) decoding accuracies.

In our experiments, we first set the $X^{(t)}$'s defined above to be the original recorded neural responses and computed the decoding accuracies for both monkeys. We refer to the accuracies obtained from original neural data as "achievable decoding accuracies". Later, to evaluate the amount of information in the strong sparse portions of the neural data, we set $X^{(t)}$'s to be thresholded versions of the original data. We tried two classes of thresholding methods: "top only" (red in Figure 3) and "top excluded" (blue in Figure 3). In "top only", we only kept the largest $p\%$ of the responses across images and trials in the thresholded version and made the smaller $(100 - p)\%$ of the responses to be zero. In "top excluded", which is complementary to "top only", we set the largest $p\%$ of the responses to be zero and kept the smaller $(100-p)\%$ of the responses. For both "top only" and "top excluded", we evaluated decoding accuracies at the following percentages (p 's) (crosses in Figure 3): 0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, and 99.

Given the population sparseness was computed based on half-maximum of each individual neuron's response, we also repeated the decoding experiment

using a percentage threshold that is relative to each neuron's peak response, rather than the absolute response threshold, to select the "top responding" neurons to be included or excluded and reset the responses to be excluded in the data matrix to 0 accordingly as before, for training and testing the decoder.

Software Available

The code used for the decoding analysis and image movement correction can be found in <https://github.com/leelabcnbc/sparse-coding-elif2018>

ACKNOWLEDGMENTS

We are grateful to many colleagues for their insightful discussion and generous help on this paper, and in particular, to Stephen L. Macknik, Martinez-Conde Susana and Shefali Umrana for editing the manuscript. We thank Wenbiao Gan for the early provision of AAV-GCaMP5; and Peking University Laboratory Animal Center for excellent animal care. We acknowledge the Janelia Farm program for providing the GCaMP5-G construct, specifically Loren L. Looger, Jasper Akerboom, Douglas S. Kim, and the Genetically Encoded Calcium Indicator (GECI) project at Janelia Farm Research Campus Howard Hughes Medical Institute.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

REFERENCES

- Akerboom J, Chen TW, Wardill TJ, Tian L, Marvin JS, and Mutlu S. 2012. Optimization of a GCaMP calcium indicator for neural activity imaging. *Journal of Neuroscience* **32(40)**: 13819-13840.
- Barlow HB. 1981. "The ferrier lecture, 1980. Critical limiting factors in the design of the eye and visual cortex," *Proceedings of the Royal Society of London Series B* **212**: 1-34.
- Chen TW, Wardill TJ, Sun Y, Pulver SR, Renninger SL, Baohan A, Schreiter ER, Kerr RA, Orger MB, Jayaraman V et al. 2013. Ultrasensitive fluorescent proteins for imaging neuronal activity. *Nature* **499**: 295–300.
- Denk W, Strickler JH and Webb WW. 1990. Two-photon laser scanning fluorescence microscopy. *Science* **248**: 73–76 (1990).
- Froudarakis E, Berens P, Ecker AS, Cotton RJ, Sinz FH, Yatsenko D, Saggau P, Bethge M and Tolias AS. 2014. Population code in mouse V1 facilitates read-out of natural scenes through increased sparseness. *Nature Neuroscience*. **17(6)**: 851-857.
- Haider B, Krause MR, Duque A, Yu Y, Touryan J, Mazer JA, and McCormick DA. 2010. Synaptic and network mechanisms of sparse and reliable visual cortical activity during nonclassical receptive field stimulation. *Neuron*. **65**: 107–121.
- Hedgé, J. & Van Essen, D. C. 2007. A comparative study of shape

497 representation in macaque visual areas V2 and V4. *Cereb. Cor.* **17**:
498 1100-1116.

499 Hromadka T, DeWeese MR, Zador AM. 2008. Sparse representation of sounds
500 in the unanesthetized auditory cortex. *PLoS Biology*. **6(1)**: 0124-0137.

501 Lennie P. 2003. The cost of cortical computation. *Current Biology* **13**: 493-497.

502 Li M, Liu F, Jiang H, Lee TS, & Tang S. 2017. Long-term two-photon imaging in
503 awake macaque monkey. *Neuron*, **93**: 1049-1057.

504 Olshausen BA and Field DJ. 1996. Emergence of simple-cell receptive field
505 properties by learning a sparse code for natural images.
506 *Nature* **381(6583)**: 607-609.

507 Olshausen BA. 2013. Highly overcomplete sparse coding. *Proceed of SPIE*
508 **8651**: 86510S-4

509 Olshausen BA and Field DJ. 2005. How close we are to understanding V1?
510 *Neural Computation* **17(8)**, 1665-1699.

511 Quiroga RQ, Reddy L, Kreiman G, Koch C & Fried I. 2005. Invariant visual
512 representation by single neurons in the human brain. *Nature* **435(7045)**:
513 1102.

514 Quiroga RQ, Panzeri S. 2009. Extracting information from neuronal
515 populations: information theory and decoding approaches. *Nature*
516 *Reviews Neuroscience* **10(3)**:173.

-
- Rehn M and Sommer ET. 2007. A network that uses few active neurons to code visual input predicts the diverse shapes of cortical receptive fields. *J. Comp Neuroscience* **22(2)**: 135-146.
- Rolls ET and Tovee MJ. 1995. Sparseness of the neuronal representation of stimuli in the primate temporal visual cortex. *Journal of Neurophysiology* **73.2**: 713-726.
- Rust NC and DiCarlo JJ. 2012. Balanced increases in selectivity and tolerance produce constant sparseness along the ventral visual stream. *The Journal of Neuroscience* **32 (30)**: 10170 –10182.
- Tang S, Lee TS, Li M, Zhang Y, Xu Y, Liu F, Teo B and Jiang H. 2018. Complex Pattern Selectivity in Macaque Primary Visual Cortex Revealed by Large-Scale Two-Photon Imaging. *Current Biology* 28, 38–48.
- Vinje WE and Gallant JL. 2000. Sparse Coding and Decorrelation in Primary Visual Cortex During Natural Vision. *Science* **287**: 1273-1276.
- Willmore B D, Mazer J A, Gallant J L. 2011. Sparse coding in striate and extrastriate visual cortex. *Journal of Neurophysiology*, **105(6)**: 2907.

FIGURE LEGENDS

536

537 **Figure 1** Population sparseness of neuronal responses of V1 layer 2 neurons
538 to natural scenes. (a and b) Ca images of the neuronal population response to
539 two different natural images as shown in the insets. Typically, only a few
540 neurons, among the nearly 1000 neurons measured (1,225 neurons for
541 Monkey A or 982 neurons for Monkey B), responded strongly to a single patch
542 of natural scenes. (c) The overall neuronal population responses to all 2,250
543 natural images. Each cell was color-coded according to the response intensity
544 to its optimal stimulus respectively. (d and e) The distributions of neuronal
545 population responses to the two natural images respectively. Abscissa
546 indicates the 1,225 neurons that showed significant response to natural
547 images, in ranked order according to their responses to each image. Ordinate
548 indicates $\Delta F/F_0$. (f and g) Frequency histograms showing the distributions of
549 the number of stimuli (out of 2,250) (y-axis) with different population
550 sparseness, measured by the number of neurons activated strongly (x-axis). It
551 shows that less than 0.5 % of the cells (6 cells out of 1225 for monkey A, and
552 4.1 cells out of 982 for monkey B) responded above half of their peak
553 responses for any given image on the average.

554

555 The following figure supplements are available for figure 1:

556 **Figure supplement 1.** Two-photon calcium imaging in awake macaque
557 monitoring the neuronal activity in V1 layer 2 evoked by natural stimuli.

Figure supplement 2. Two-photon image and neuronal responses in monkey

B.

Figure supplement 3. The ROIs of activated cells overlaid on a two-photon image.

Figure 2 Life-time sparseness in neuronal responses of V1 layer 2 neurons to natural scenes. (a and b) The response of one example cell (cell 653) to the entire set of natural scene stimuli, exhibiting high stimulus specificity. (c and d) Another example cell (cell 949) also shows high stimulus specificity. (e and f) The distributions of stimulus specificity of neurons, in terms of half-height width of the stimulus tuning curves. Each cell would typically respond strongly to less than 0.5% of the natural images in our test set.

The following figure supplements are available for figure 2:

Figure supplement 1. Reliability analysis of neuronal responses.

Figure 3. Image decoding performance as a function of the percentage of only the strongest responses used (red curves) or removed (blue curve). Y axes show the cross-validated decoding accuracy on the 2,250-way image classification task. Dash lines are the referential “achievable decoding performance” in accuracy using the original entire neural population responses; red lines (“top only” in legends) show the decoding accuracies when different

percentages of top responses were kept and lower responses were removed (set to zero); blue lines (“top excluded” in legends) show the decoding accuracies when different percentages of top response were removed (set to zero) and lower responses were kept. X axes show the percentage of top responses included (red curves) and excluded (blue curves). Check **Decoding Analysis** for details. Gray vertical lines highlight the decoding accuracies including or excluding the top 0.5% responses. Since our classification task is a 2,250-way one, the chance accuracy is 1/2,250, or about 0.4%.

Figure 1 - Figure supplement 1. Two-photon calcium imaging in awake macaque monitoring the neuronal activity in V1 layer 2 evoked by natural stimuli. (a and b) Two-photon images of neurons expressing GCaMP5 at zoom 1X, 2X respectively. (c) Natural stimuli evoked robust neural activity probed by calcium indicator GCaMP5.

Figure 1 - Figure supplement 2. Two-photon images and neuronal responses in monkey B. (a, b) An average two-photon image from monkey B. (c, d) Differential images show sparse strong neuronal responses to natural stimuli.

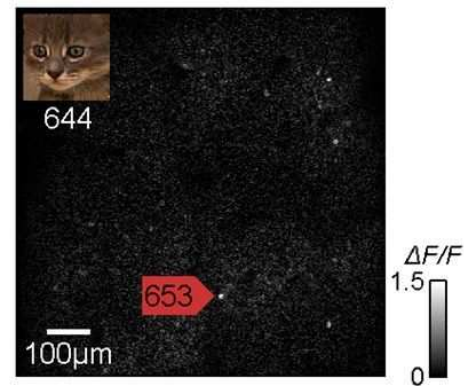
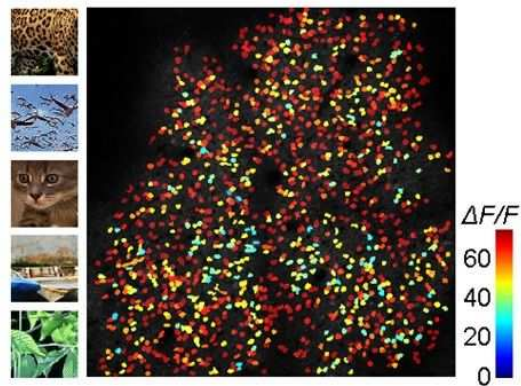
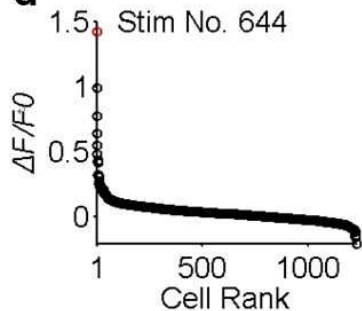
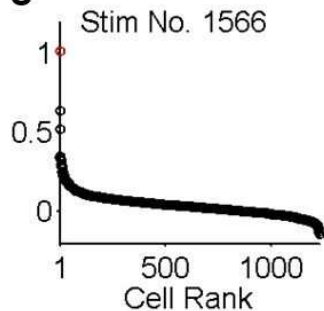
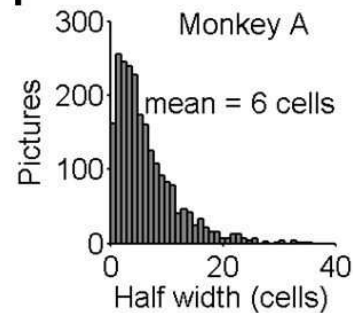
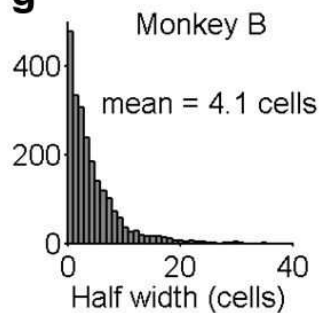
Figure 1 - Figure supplement 3. The ROIs are overlaid over an two-photon image of a 850 × 850 μm region under a 16X objective, showing the ROIs extracted based on activities were well matched to the cell bodies.

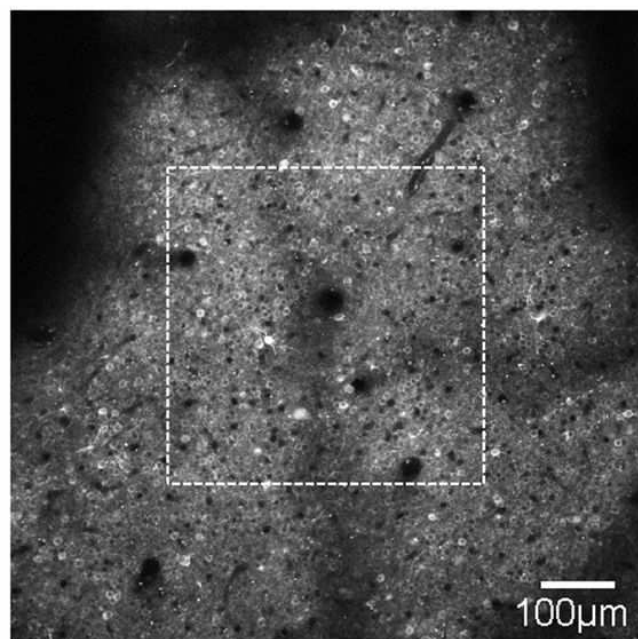
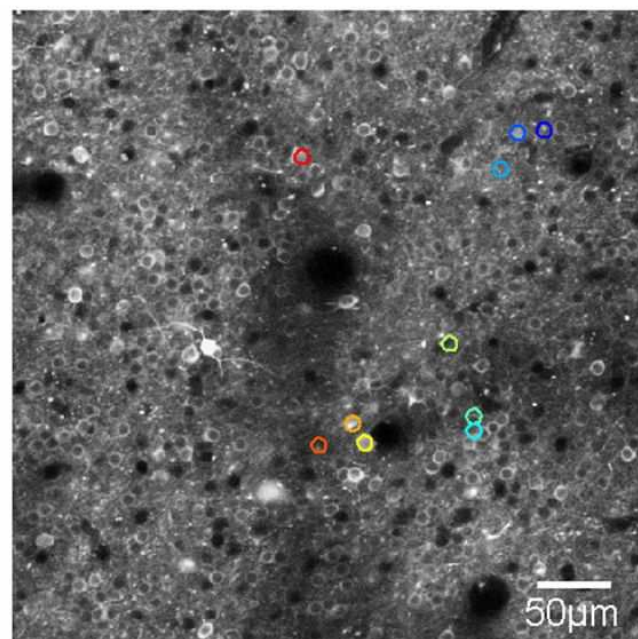
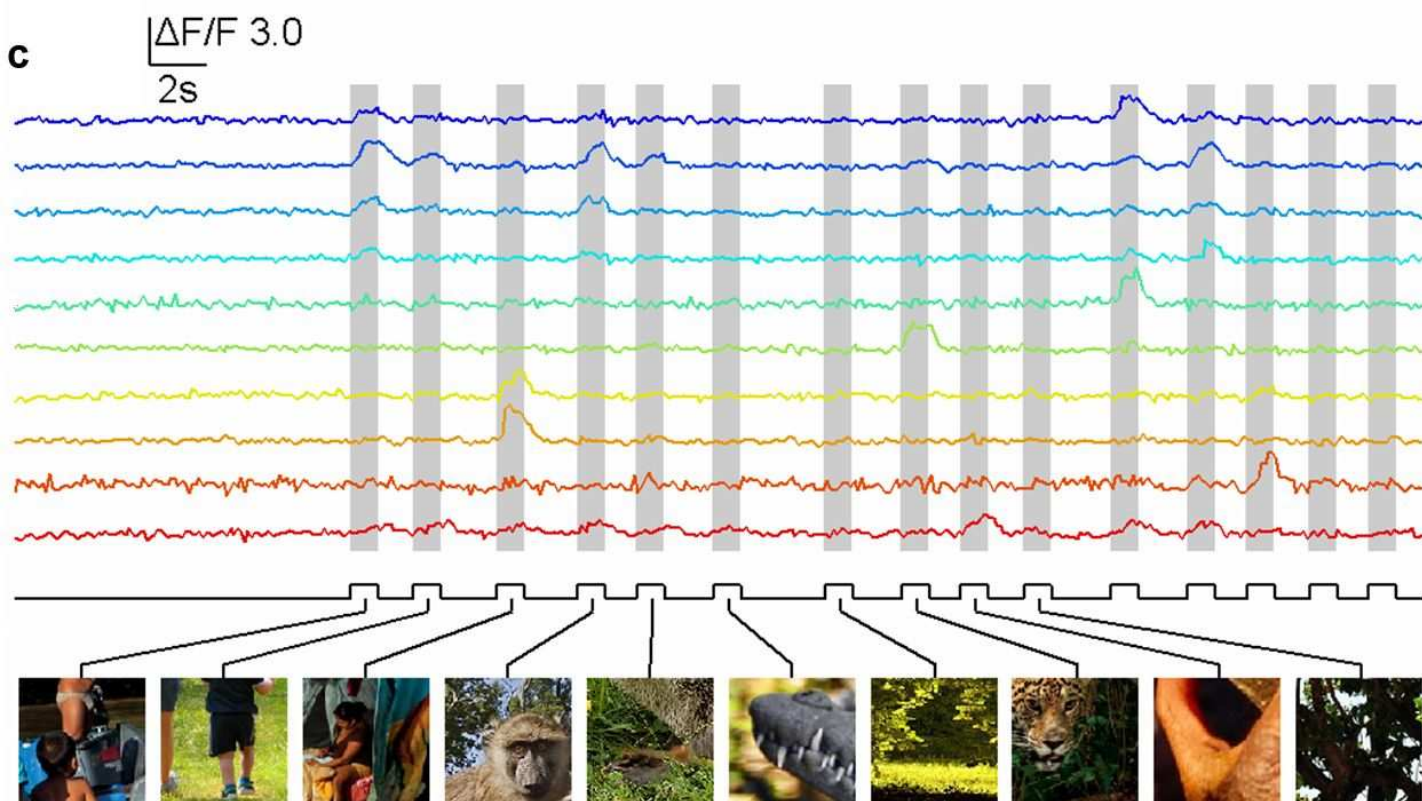
602

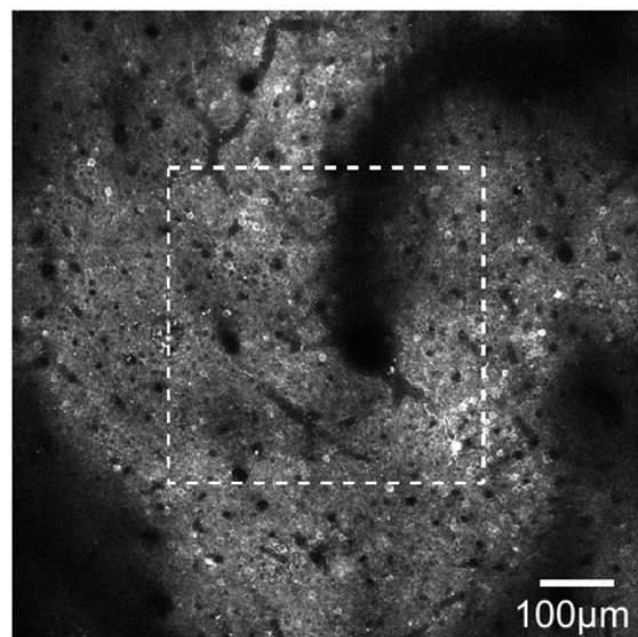
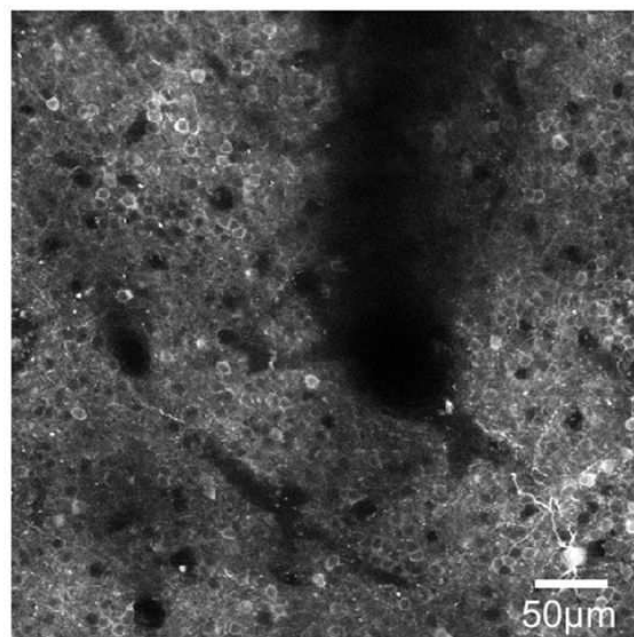
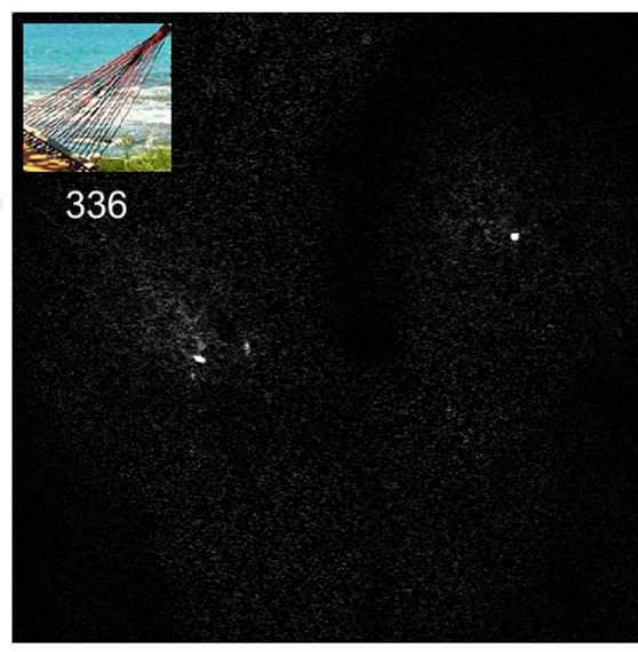
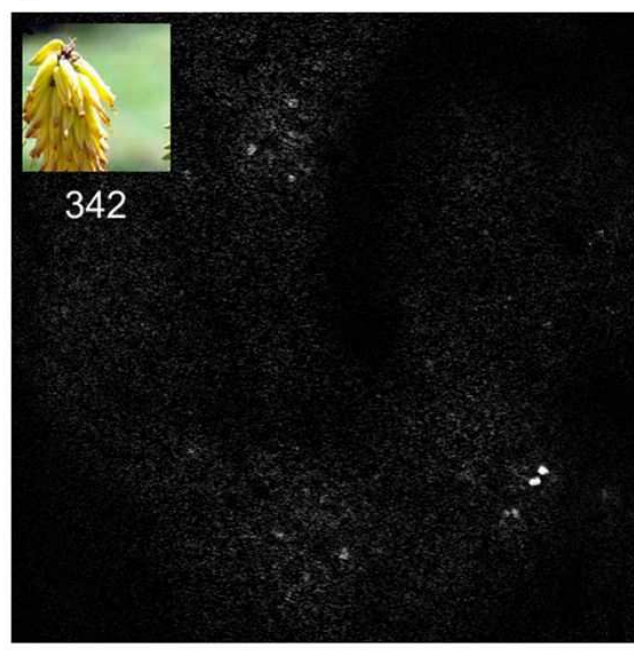
603 **Figure 2 - Figure supplement 1.** Reliability analysis of neuronal responses. (a)
604 The tuning curve of neuron #653's responses (averaged across trials) to the
605 2250 stimuli. (b) The tuning curve computed from one random shuffle across
606 all trials. (c) The ROC of original trials (red curve) against those of 99 shuffled
607 trials (gray). The AUC of the original data is 0.999. The AUC of the example in
608 (b), colored in blue, is 0.68. (d) The distribution of the AUC's of all 1000
609 shuffled cases. The probability, that the shuffled AUC can reach the raw data's
610 AUC, is less than 0.001 ($p < 0.001$) for neuron #653.

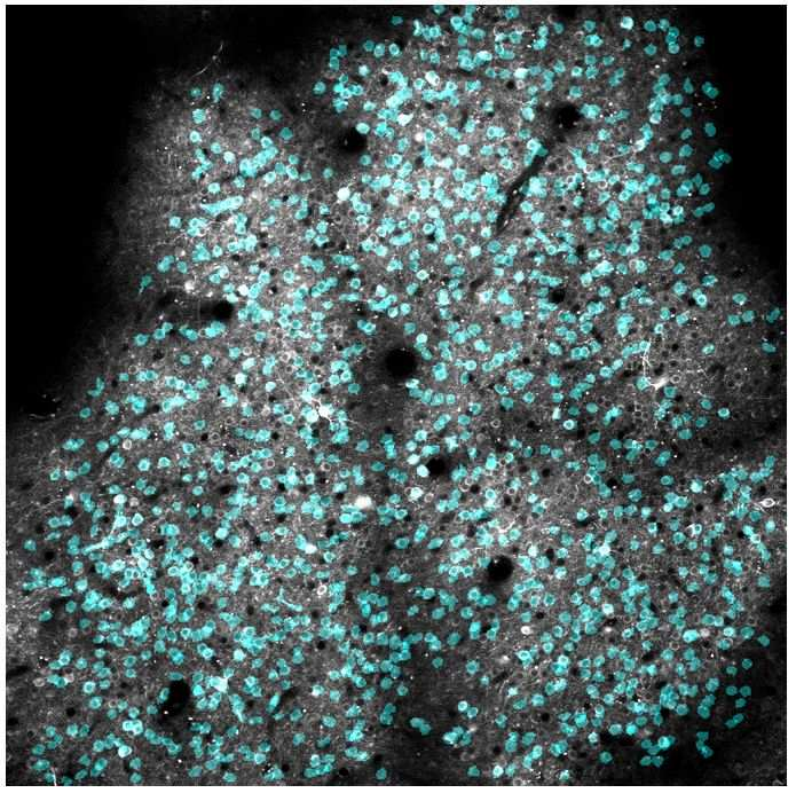
611

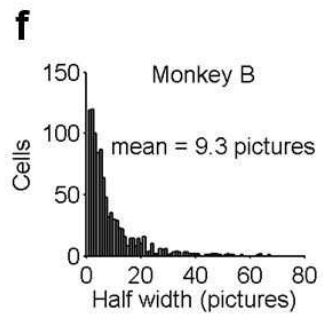
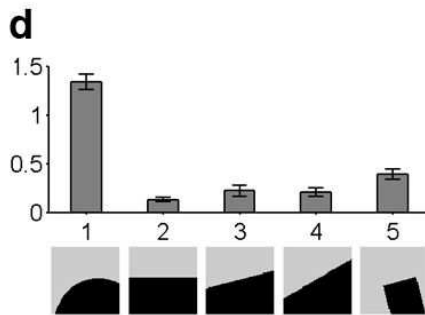
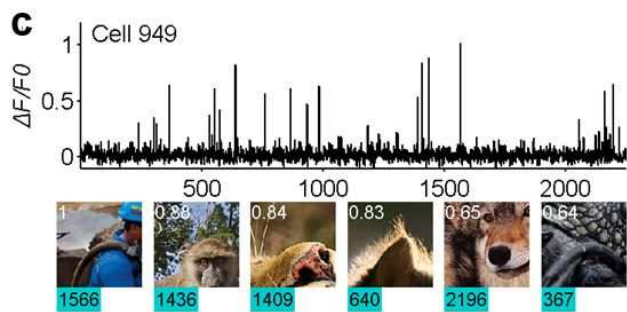
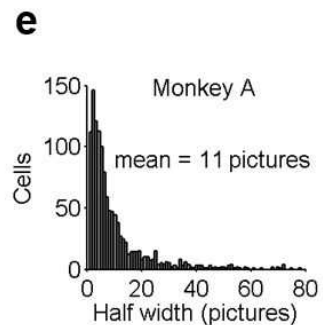
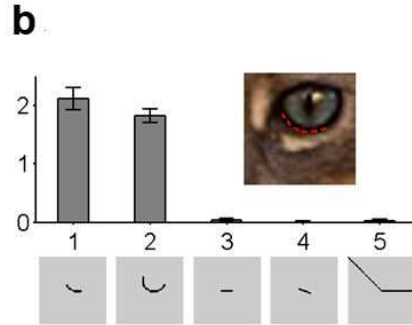
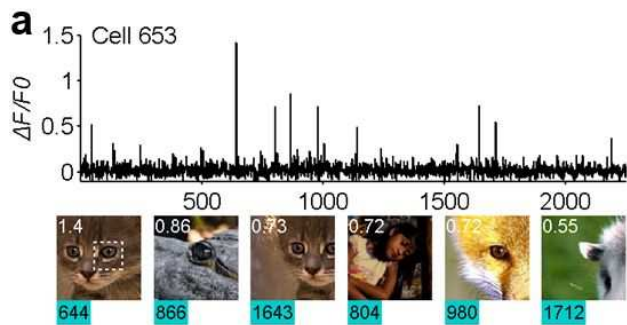
612 **Figure 3 - Figure supplement 1.** Image decoding performance as a function
613 of the percentage of top responding neurons selected to be included or
614 excluded using a threshold relative to the peak response of each neuron. See
615 Methods for details. Results were qualitatively similar to Figure 3.

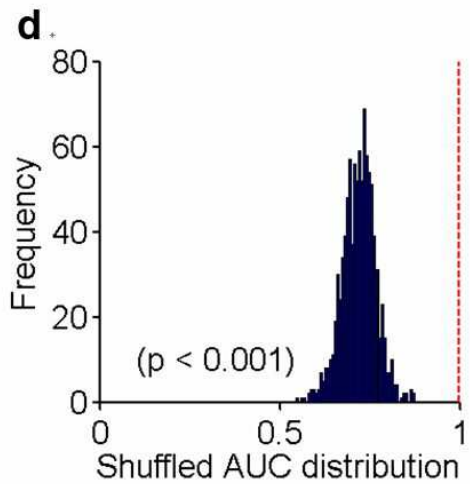
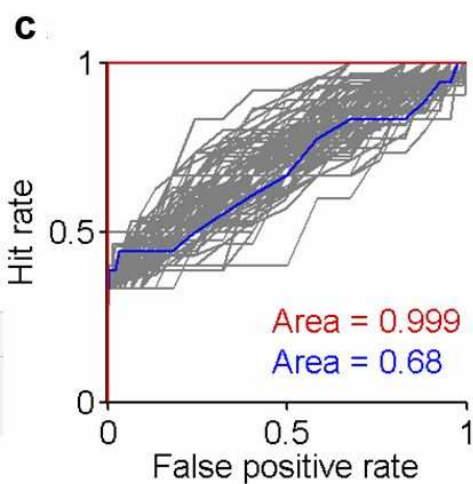
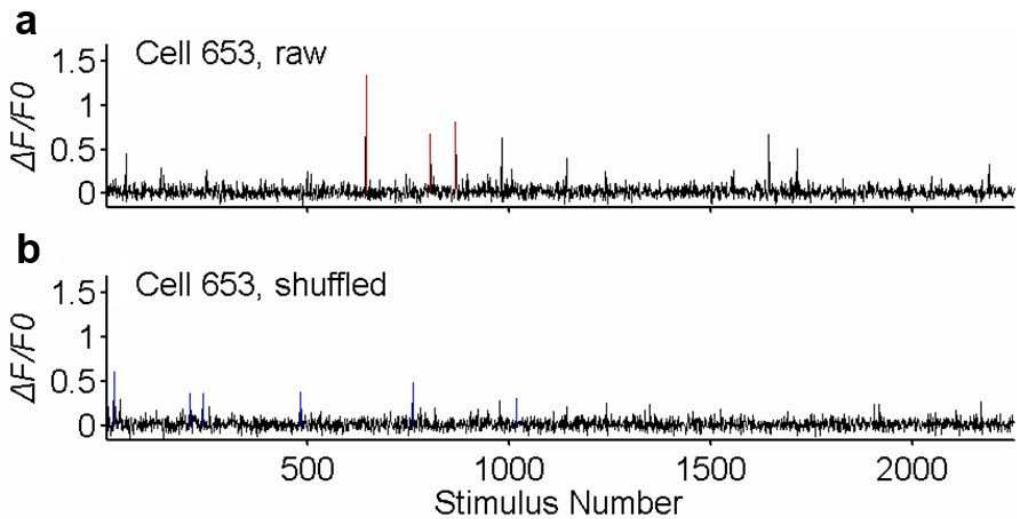
a**b****c****d****e****f****g**

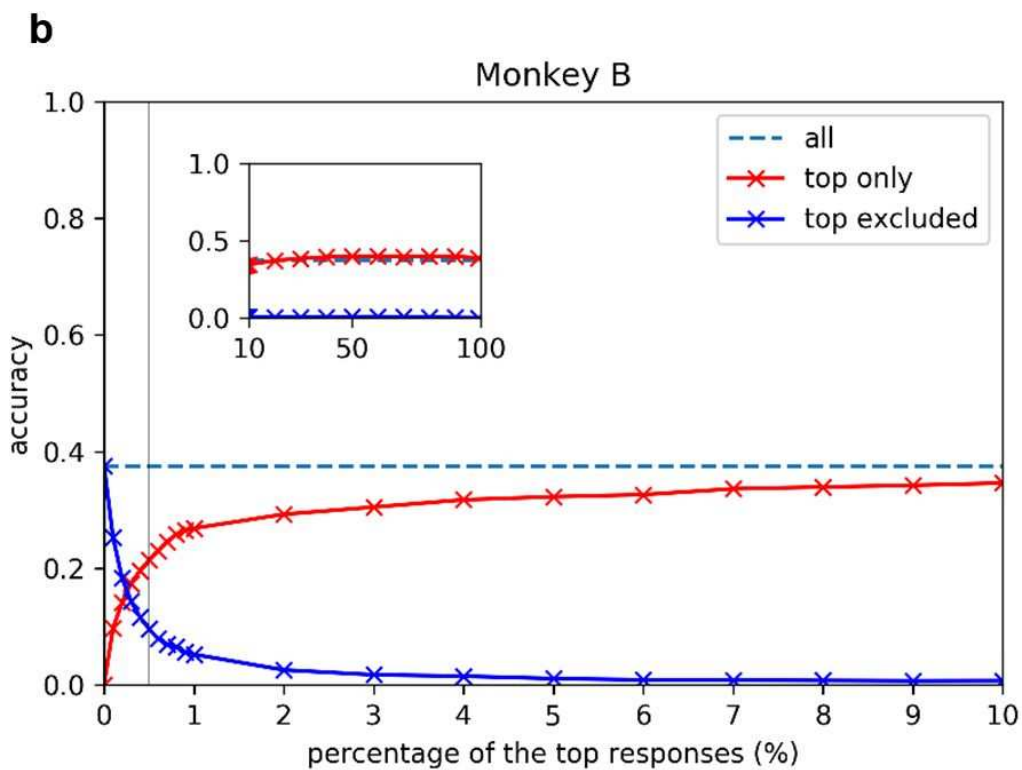
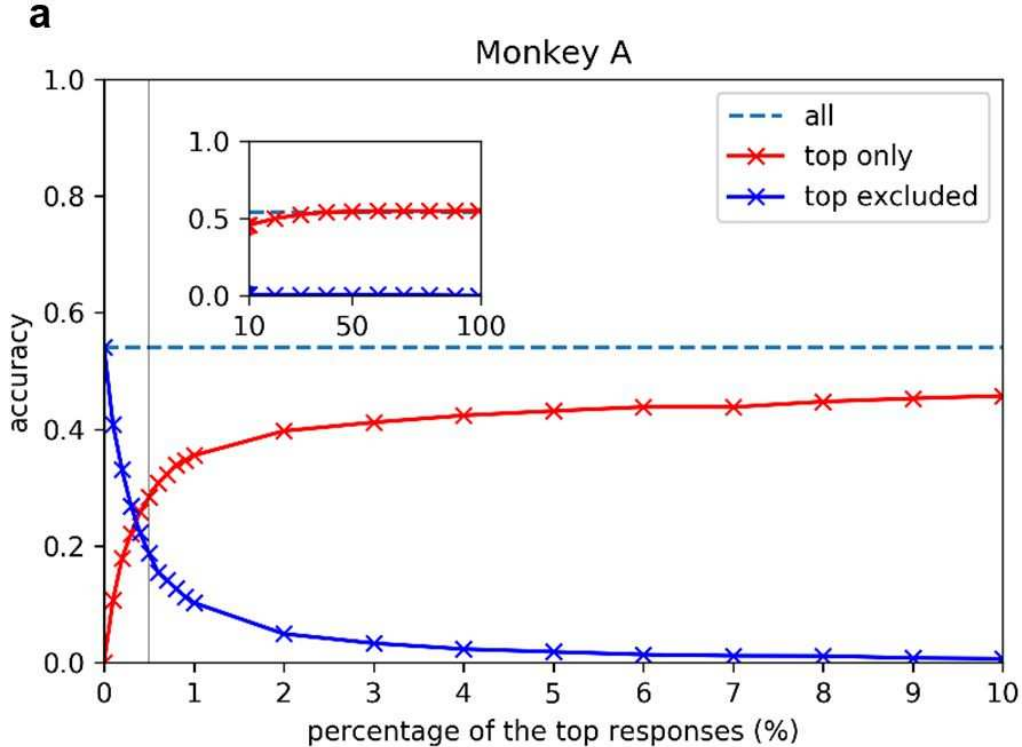
a**b****c**

a**b****c****d**



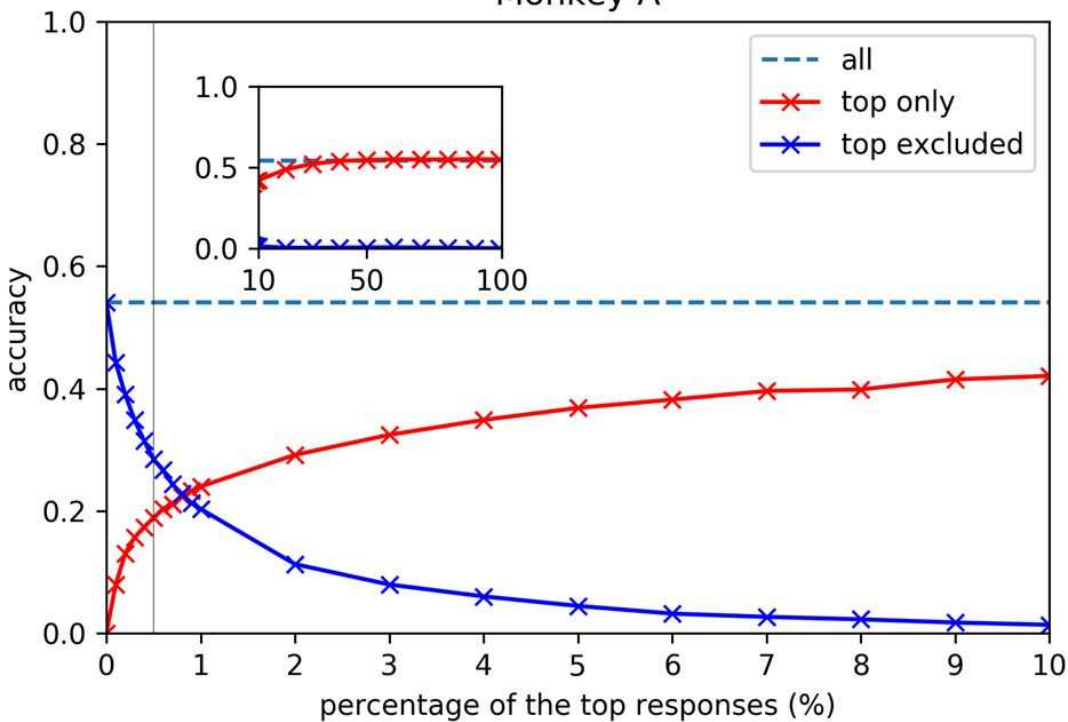






a

Monkey A

**b**

Monkey B

