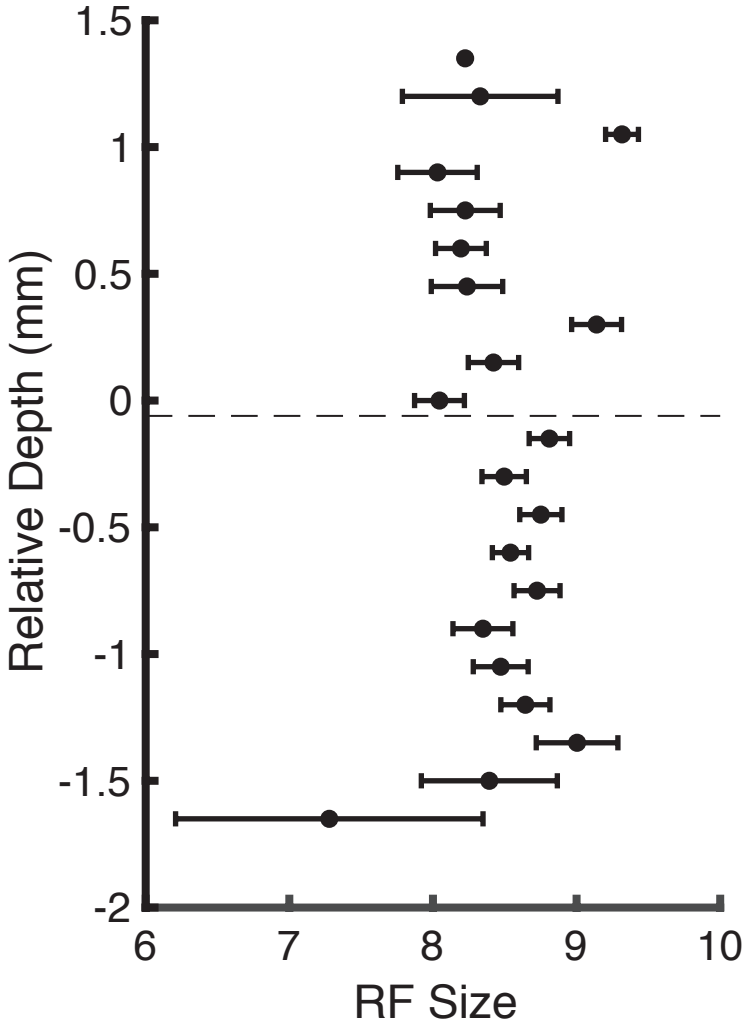



**Figure S1**
Pettine, Steinmetz and Moore

www.pnas.org/cgi/doi/10.1073/pnas.1819398116

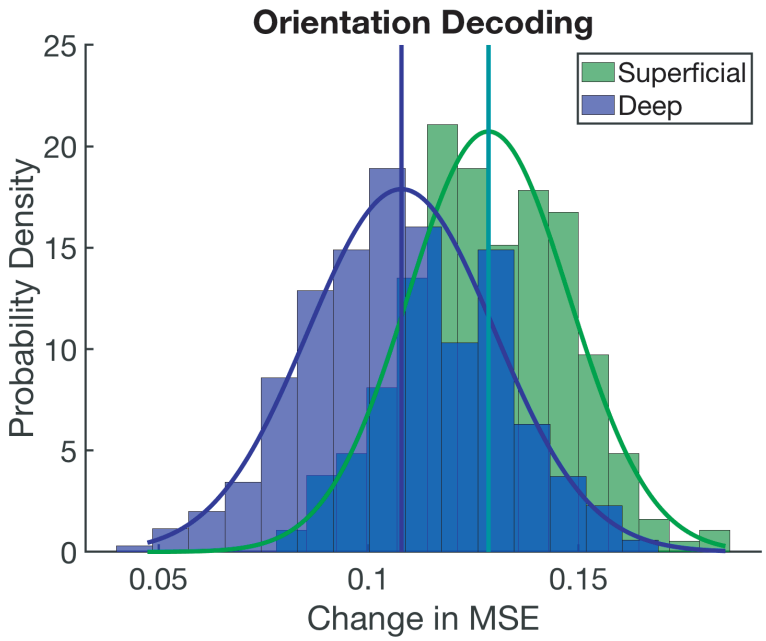



**Figure S2**

Pettine, Steinmetz and Moore

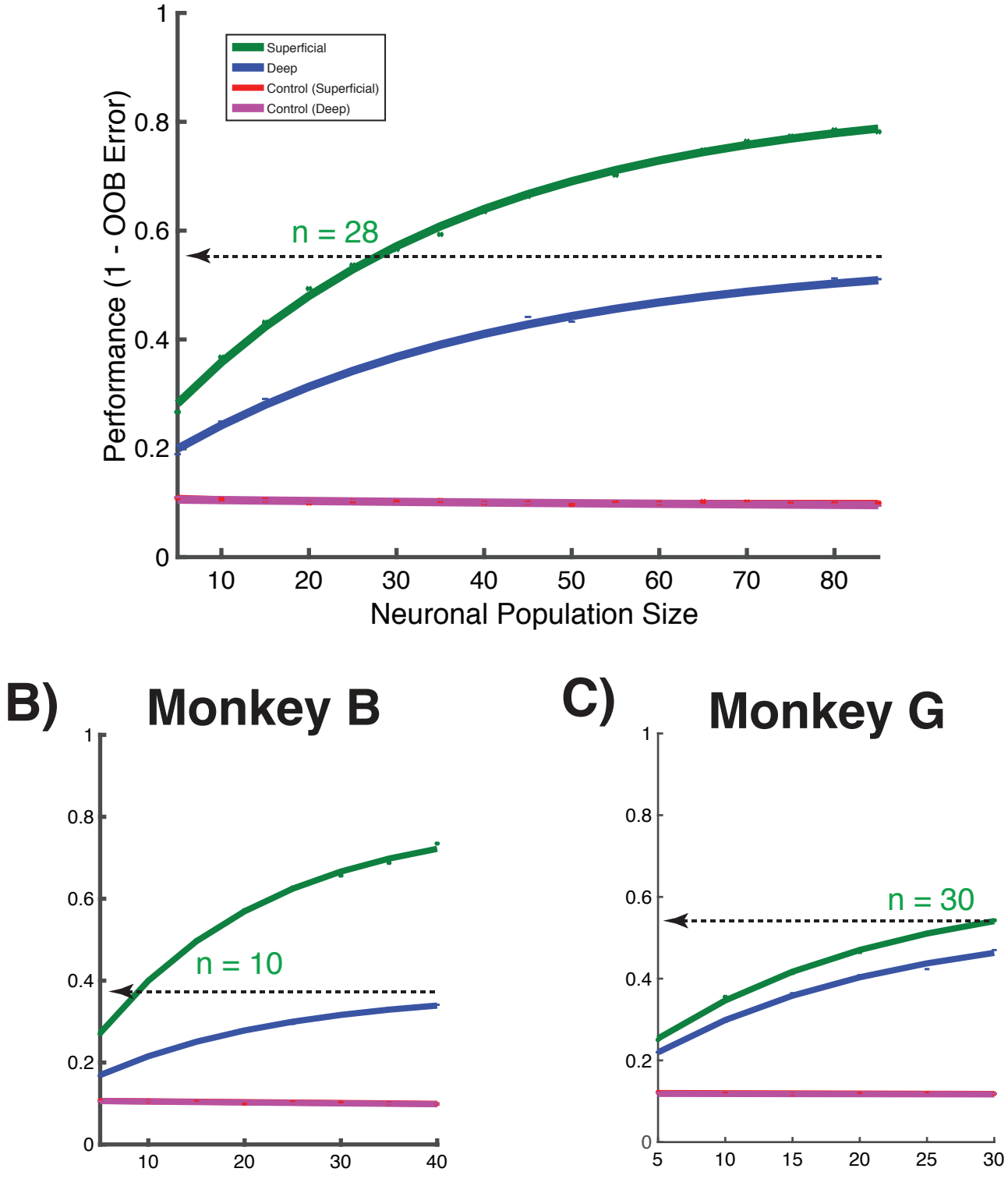
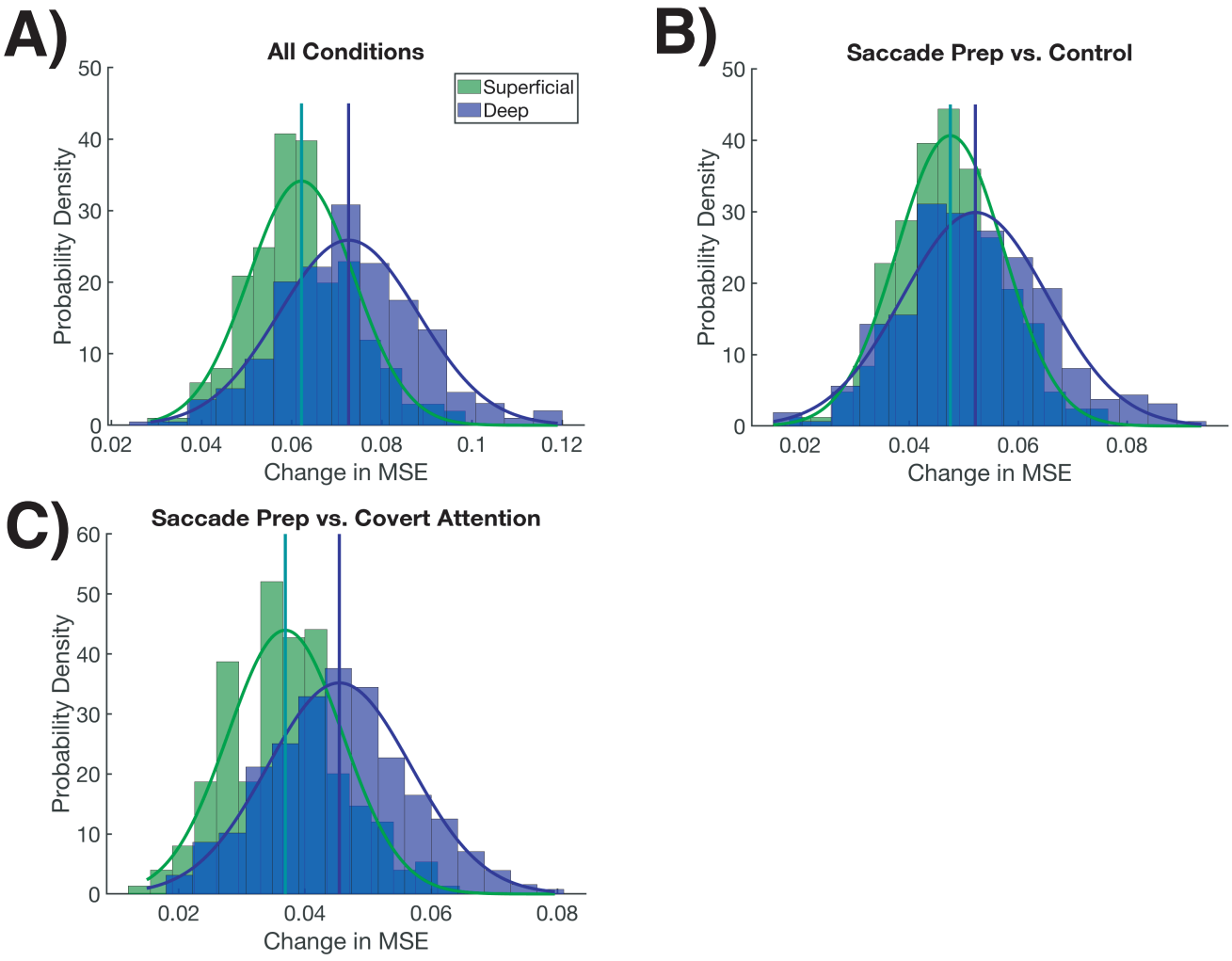



**Figure S3**

Pettine, Steinmetz and Moore




# Figure S4

Pettine, Steinmetz and Moore

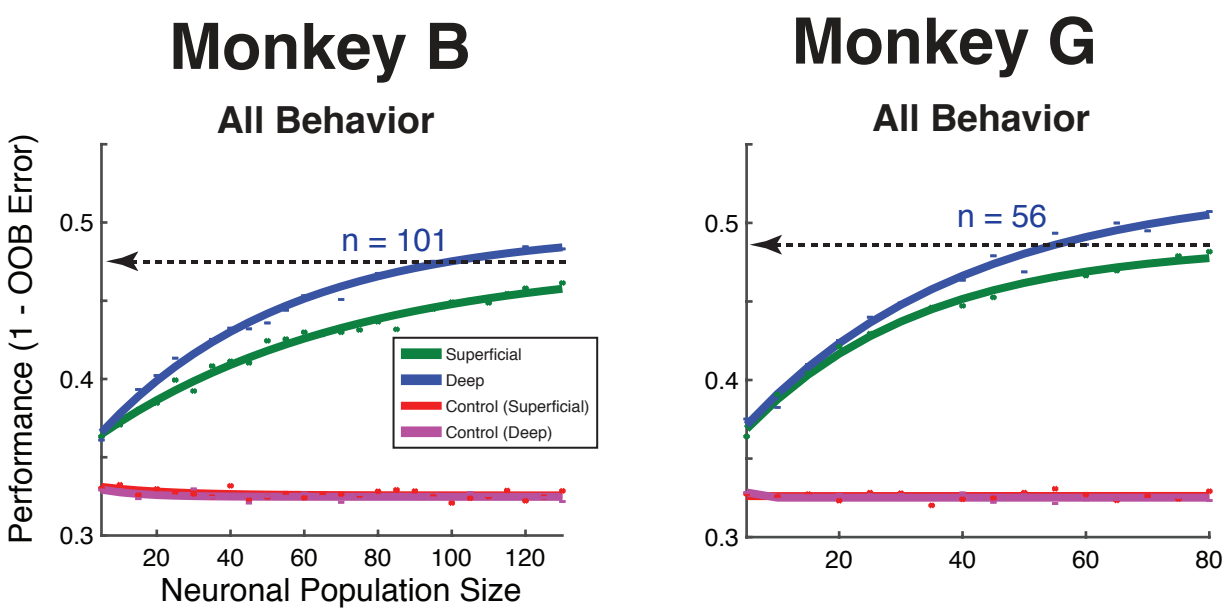



**Figure S5**

Pettine, Steinmetz and Moore

**Figure S1 RF diameters across depth.** The mean diameter and standard error of RFs measured across multiple electrode depths. The division between superficial and deep is indicated by the dotted line. The mean values for superficial units (8.4, $n$ = 104) and deep units (8.6, $n$ = 157) were statistically indistinguishable ($P$ = 0.3256, unpaired T-test).

**Figure S2 Feature importance histograms for orientation decoding.** Shown are separate histograms for superficial ($n$ = 257) and deep units ($n$ = 406) of the change in mean squared error (MSE) caused by permutation of information from a single unit. Also shown are probability density functions derived from the mean and standard deviation of each population's MSE values, along with vertical lines placed at their means. Note that the distributions are unimodal and normally distributed, with few outliers. We found that the mean level of MSE change for superficial units was significantly higher than that of deep units (superficial = 0.129, deep = 0.108, $P = 9.5 * 10^{-29}$ by Wilcoxon Rank Sum).

**Figure S3 Decoding orientation with single units**. Neuron dropping curves are shown, using only single units. A) Data from both monkeys combined. B) and C) data are separated for the monkeys B and G.

**Figure S4 Feature importance histograms for attentional decoding.** Same as figure S1, but for decoding attentional conditions. Decoding A) all three conditions, B) saccade preparation versus control, and C) covert attention versus control. All distributions are unimodal, normally distributed, and have few outliers. Using a Wilcoxon Rank Sum test, the distribution of deep units had a significantly higher mean than that of superficial units in all attentional conditions. A) Deep

= 0.073, superficial = 0.062, $P = 2.7 * 10^{-16}$. B) Deep = 0.052, superficial = 0.048, $P = 4.3*10^{-5}$. C) deep = 0.050, superficial = 0.044, $P = 3*10^{-7}$.

**Figure S5 Behavioral decoding neuron dropping curves, separated by monkey**. Decoding of behavioral conditions (covert, saccade and control) separated by monkeys B (left) and G (right).

## Supplementary Methods

### Cortical Column Laminar Recordings

#### *Electrode targeting: Use of MRI guidance to achieve perpendicularity*

We sought to achieve simultaneous recordings at sites located within a single cortical "column." In particular, the topographic organization of extrastriate visual cortex suggests that vertically separated neurons should have overlapping RFs, so we sought to record from a column principally by this definition. Since the cortical magnification factor (an estimate of how much cortical tissue corresponds to units of retinal space) is approximately 1 deg/mm (1), we could measure the approximate angle with the cortex by the distance between receptive fields measured on the deepest and most superficial recording contacts, and sought to keep this angle at 10 degrees or less, corresponding to a RF shift of ~0.5 degrees, given 2 mm thickness of cortex.

In order to achieve these perpendicular penetrations we employed an MRI targeting technique (2). We implanted the monkeys with custom built recording chambers made from PEEK-type plastic, rather than from titanium, to avoid "shadows" in the MRI images. While we did not employ ceramic skull screws, we took some care to ensure that the titanium skull screws and plates were not located close to the recording chamber and brain areas of interest. We filled a custom-made plastic cylinder with copper sulfate solution. We anesthetized the monkey and inserted this cylinder into the recording chamber, into which it fit snugly. We performed structural MRI imaging (1.5 Tesla; T-1 weighted image) to visualize the location and orientation of the recording chamber (visible due to the high-contrast copper sulfate solution within it) relative to the position of the prelunate gyrus within the brain. By manually identifying the contours of the prelunate gyrus, we could compute perpendicular vectors to it and project these back to the level of the electrode stage, thus identifying which penetration approach vectors were likely to yield perpendicular penetrations.

#### *Achieving desired approach vectors*

We employed a custom-built targeting device to angle and rotate the electrode into any desired orientation and position in three dimensions. The device consisted of a “double-eccentric” mechanism for positioning the electrode in the x-y plane of the well, a tilting mechanism, and a rotating mechanism. All four coordinates could be set with sub- millimeter precision using notches engraved in the device. The V4 recording chambers on

both monkeys projected from the monkeys' heads at an angle such that there was a unique point on the chamber's perimeter at the lowest elevation. This point was identified computationally in the MRI images and was identified on the chamber itself by filling the chamber with saline solution until the liquid first contacted the lip of the chamber. With this point of alignment between the MRI images and the physical well, the exact X, Y, tilt, and rotation coordinates for an approach vector specified by the MRI images were geometrically determined.

***Electrode targeting: Assessing perpendicularity with RF overlap***

RF positions and extents were estimated by computing the number of multi-unit spikes recorded on each channel in the 200 ms period following stimulus onset for each of probe location in a RF-mapping task. During this task, subjects fixated a small (~0.3 d.v.a.) white dot against a medium gray background. On each trial the six flash positions were selected from one of the rows of the grid in random order. A horizontally oriented grating was flashed for 50ms at each position, with a 150-250 ms variable delay between flashes. The flashes occurred at a total of 36 locations on a 6x6 grid with 3 d.v.a. spacing (total coverage 15x15 d.v.a.). If the subject maintained fixation within a 1.8 d.v.a. square window until after the sixth flash, he received a juice reward.

The upper right position of the grid was at the fovea such that only the lower left visual field was covered by the mapping. This 6x6 matrix of response counts was cubic spline interpolated to produce the full "RF map" and a 75%-of-max contour was determined, defining the RF border. The center of mass of the portion of the RF map within the RF border was defined as the RF center. This analysis was performed after recording RF-mapping task responses but before the change-detection task, so that a stimulus position could be chosen at a location that fell within the RF borders for all channels. If such a position was found, the recording was included in further analyses.

***Electrode targeting: Depth alignment***

We lowered electrodes into the brain rapidly (~25 μm/sec) until one channel was in the cortex, based on visual examination of LFP and spiking activity being recorded concurrently. Then we advanced the electrode slowly (~5 μm/sec) until the uppermost electrode contact was near the point of entering the brain, being recorded during advancement. We withdrew the electrode 500μm to release compression of the brain caused by the

electrode. During this brief withdrawal, no apparent change in the LFP or spiking activity was observed, confirming that this served to relax the cortex rather than change the position of the electrode relative to the brain. This manipulation qualitatively improved stability and recording quality. After reaching this position, the full-field flash task was run to assess the depth.

During the full-field flash task, monkeys fixated a small (~0.3 d.v.a.) white dot against a black background. The monitor turned maximal white for one frame (~8ms) then back to black. The flash occurred six times per trial with variable delays in the range of 150-250ms. If the monkey maintained fixation within a 1.8 d.v.a. square window until after the sixth flash, he received a juice reward. Approximately 30 trials, or 180 flashes, were completed per day. We computed the current source density (CSD) response to the full-field flashes. The CSD reflects the spatial and temporal position of current sources and sinks (i.e. where current flows into and out of the extracellular space, respectively) along the length of the electrode, given certain assumptions likely to be true for our recordings (3). The CSD can be computed discretely as the second spatial derivative of the LFP for each point in time, that is:

$$D(z) = \frac{\emptyset(z+h) - 2\emptyset(z) + \emptyset(z-h)}{h^2}$$

where z is the position in depth, h is the distance between potential measurements (in our case, 150 μm), and $\Phi$ (z) is the potential measured as a function of depth. We also calculated the CSD according to the inverse estimation method (4), and display the results of this calculation, which produces smoother, higher resolution plots of CSD, in figures for clarity. However, results were qualitatively indistinguishable with both methods. Borders between current sinks of interest were manually identified and channel depths were computed, in mm, relative to these borders.

***Depth registration***

In all included recordings, a prominent current sink was identified near the middle of the electrode, approximately 40-50 ms after flash onset. This was often followed by another sink just below the first, peaking approximately 100 ms after flash onset. These two sinks appeared in every included recording, and we therefore aligned the recordings on these functional markers of cortical lamina. In many recordings, further

sinks were observed near the top of the probe at ~150 ms and near the bottom of the probe at ~50 ms. Because the widths of all four of these sinks, when present, were highly consistent from recording to recording, we assigned each channel a depth relative to this central feature.

Depths were measured in millimeters, and positive depths indicate channels superficial relative to the CSD feature. In some sessions, further CSD recordings at deeper locations revealed that no further current sources or sinks of comparable magnitude could be identified below these CSD features, assuring us that our electrode covered the depth of cortex. Two alignments of these functionally defined layers with anatomical cortical layers seem possible. The uppermost sink could correspond to layer 2/3 (together), and the larger sink to layer 4 (Figure 1*C*). Alternately, the four visible sinks could correspond to layer 2, 3, 4, and 5 in order from superficial to deep. On the one hand, the first assignment seems reasonable as the thickness of the layers known histologically matches the thickness of these CSD features reasonably well, and our expectation from primary sensory areas is that layers 4 and 6 will have the earliest responses (5–7). However, the cortex may well be compressed around the electrode as it is inserted thus skewing the measured layer thicknesses. Layer 2 and 3 are well-differentiated cytoarchitecturally in V4 unlike in V1, suggesting they may not appear as a single sink. Furthermore, the earliest driving visual inputs into V4 are probably not from the ventral stream (8), which project into layer 4 (9), and may instead arrive from the pulvinar nucleus of the thalamus (10, 11), which synapses into deep layer 3 (12). This would indicate that the lower sink may correspond with the N95 marker used in previous studies to identify the granular layer (6, 13–15).

## Data Analysis

### *Tuning and Modulation Indices*

To determine the tuning of each single neuron, we calculated the firing rate on each trial during a 300 ms block, from 50 ms to 350 ms relative to stimulus onset. We then labeled the trials by stimulus orientation, and used a Kruskal-Wallis test to compare orientation distributions. If $P < 0.005$ we categorized the neuron as tuned. We then used a Chi-squared test to compare the proportion tuned in superficial versus deep layers. We also fit a Gaussian tuning function to the each neurons normalized (to the neuron's maximum activity) average

firing rate for the eight stimuli using parameters for amplitude (a), preferred orientation (b), width (c) and baseline (d). The formula was given by:

$$r(\theta) = a \times e^{-\left((\theta - b)/c\right)^2} + d$$

To obtain the parameters and goodness of fit measures, we used the Matlab fit function with nonlinear least squares, and constraints of 0 for the lower bound of all variables, and an upper bound of π for b and 8 for c. To determine if the neuron was well fit by the function, we used an adjusted $R^2$ cutoff of 0.70. For each neuron, the averaged firing rates were rotated around π until the optimal fit was achieved. We then compared the function parameters of superficial and deep layer neurons. As the sample sizes of superficial and deep neurons were unequal, we used bootstrapping without replacement to match the sample sizes, and repeated each test 1000 times. The reported p-values are the mean of those produced by a Wilcox signed-rank test.

***Attention Modulation***

For each neuronal unit, we calculated the mean firing rate during the cue epoch from -500 ms to 0ms relative to the blank period. For each unit, we then calculated the attention modulation indices for eye movement preparation and covert attention relative to the orthogonal control, using the standard formula:

$$Modulation\ Index = \frac{Attention - Orthogonal}{Attention + Orthogonal}$$

We then used a mixed effects model with fixed effects for neural depth, attention condition and an interaction term (implemented with the R package nlme (16)). To make layer comparisons within this omnibus model, we used three orthogonal contrasts: superficial behavioral conditions, deep behavioral conditions and superficial neuronal units versus deep neuronal units. In all tests, we included a random intercept for each neuronal unit, to control for repeat measures.

**Stimulus and Attention Classification**

***Feature Matrix***

We assembled a dataset composed of neuronal firing rates recorded across the columnar arrays and across multiple experimental sessions (23 sessions from Monkey G; 20 sessions from Monkey B; 86 superficial neurons; 181 deep neurons) for all units for which we recorded a minimum number of trials per orientation ($N$ = 20), or attention condition ($N$ = 200). Each column of the feature matrix was a specific neuron's firing rate, and each row of that column was the neuron's firing rate on a specific trial. The rows of each column were aligned, so that they shared the same label for orientation or attention condition (depending on the epoch). The number of rows associated with each orientation or attention condition were matched, so that chance level was 12.5% for the orientation epoch and 33% for the cue epoch. Each neuronal unit had multiple columns in the feature matrix, corresponding to the number of bins in which firing rates were calculated. Trials within a tested condition were shuffled for each unit in order to destroy any possibility of correlation among simultaneously recorded units. The firing rates for the orientation epoch were calculated in two 150 ms time bins, from 50 ms following stimulus onset to 350 ms following stimulus onset. For attentional decoding during the cue epoch, firing rates were calculated in four 150 ms time bins, from 50 ms following stimulus onset to 650 ms following stimulus onset. This provided a gross temporal pattern which was noted to improve performance in Nandy et al. (2016) (17). When building feature matrices with variable population sizes, we randomly sampled a population that size from all available units. This process was repeated 100 times, generating a unique of feature matrix for each run of the decoder.

***Random Forest Classification***

We used a Random Forest decoder, similar to that used in Nandy et al. (2016) (17), as implemented by Matlabs (Mathworks TM) treebagger function. In addition to decoding based on firing rate, Random Forest can decode based on differences in firing rate variability, even when mean firing rates are equal (18, 19)[,]. Furthermore, rather than comparing each orientation to the others in turn, the decoder simultaneously considers all orientations. The decoder's decision trees were trained on bags of trials (matrix rows), selected through bootstrapping with replacement, and tested each decision tree on trials not included in the training bag. This out-of-bag (OOB) error was used as the performance measure. It is significantly more conservative than cross validation, but has the advantage of using all available data when training the decoder.

Furthermore, the bootstrapped sampling method has the traditional advantages associated with bootstrapping, such as revealing the true underlying distribution from the available training data, and reducing the impact of outlier trials (19). The decoder then used a boosting method to create decision trees. At each branch point, a random subset of the features (square root of the total number of features) was chosen to calculate potential decision boundaries. Each of the features in the subset was used as a linear threshold for linearly partitioning the population of trials. The Gini impurity (GI) of the original sample, as well as of the two partitions was calculated using the formula:

$$GI = 1 - \sum_{i=1}^{J} p_i$$

Where J is the number of classes, and $p_i$ is the probability of choosing stimulus class *i* at random from the sample. The GI of the two partitions was averaged, and subtracted from the GI of the parent sample. The feature with the greatest decrease in GI was used at the decision boundary at that branch point. The use of a random subset of features reduces the influence of outlier features, allowing one to be less careful about the neurons selected for use in decoding. Stopping criteria for the decision trees was when either all the trials at a branch point had the same label ($GI = 0$), or there were only 5 trials at the branch point. We set the number of trees to 200. To determine these hyperparameters, we decoded orientation using all neurons from both monkeys, while varying the number of trees and the number of leaves. Performance was best with a stopping rule of 5 trials in a leaf, and began to asymptote with 200 trees. Thus, we fixed the hyperparameters to those values for all subsequent analyses. The decoder was trained and tested using each of the 100 feature matrices, producing a distribution of decoder performance.

***Neuron-dropping Curves***

We used neuron-dropping curves to assess the performance of the decoder. Also known as learning-curves, these are a standard tool in the machine learning to assess whether performance limitations are due to the decoder, or to the quantity of data. When computing these functions, the quantity of data used for decoding is varied and an error rate (or performance level) is plotted as a function of that quantity. The presence of an asymptote indicates that the decoder has reached maximal performance, whereas the absence of an

asymptote indicates more data is needed. We then fit a saturating function and compared both the rate of rise, and the asymptotic value between populations.

We created pseudo-populations, starting with 5 units, and then incrementing by 5 until the maximal number of available units was reached. For each population size, we randomly sampled the requisite number from the larger population with replacement, repeating this process 100 times to bootstrap a representative distribution. To this range of performance levels, we fit the saturating function,

$$f(s) = a \times e^{(-b \times s)} + c,$$

where *s* is the size of the population, *a* controls the y intercept, *b* the slope and *c* specifies the function asymptote. This was implemented using the Matlab fit function with the method non-linear least squares. A confidence interval of 95% was derived from the fitting process.

***Timecourse Decoding***

To assess the decoder performance prior to an event of interest (stimulus onset or cue onset), we created a feature matrix using a single 150 bins of firing rate from each neuron for each trial. We then recalculated that bin using time steps of 25 ms, moving from a period of time prior to event of interest to a period subsequent. Behavior was decoded separately using each feature matrix and performance was measured using OOB error. The time value of each performance measure was assigned using the leading edge of the bin. This meant that if the bin ended, for example, at stimulus onset, the 150 ms prior to stimulus onset was used as the performance measure at time 0. We chose this as it is a much more conservative estimate of latency than choosing the center of the bin, or the tailing edge.

For both stimulus decoding, we swept the time period from 100 ms prior appearance of stimulus to 350 ms after it. For behavioral decoding over time, we swept from 100 ms prior to the cue to 650 ms subsequent to the cue.

***Feature Importance***

We used a heavily validated permutation test to assess feature importance (18). In this method, the decoder is trained on a full set of valid features. The decoder is then tested by sequentially permuting the features, and measuring the change in mean-squared-error (MSE) caused by permuting that specific feature. This is

repeated on several separately trained versions of the decoder until a distribution of MSE impacts is created for each feature. The measure of each feature's importance is then taken as the mean of their MSE distribution.

To implement this method, we simplified the feature matrix so that a single feature was associated with each unit. This was done by creating a single bin for each unit that was 300ms in duration for orientation decoding, and 600ms in duration for attentional decoding. Given the unequal number of recorded units in superficial and deep populations and the need for matched population sizes to compare MSE, we capped population sizes at 257 units for orientation decoding and 215 units for attentional decoding. In the case of superficial units, this encompassed the entire population. In the case of deep units, we randomly subsampled from the full population on each trial. MSE distribution estimates for each unit were taken from 100 cycles of the decoder.

***Citations***


1. Gattass R, Sousa AP, Gross CG (1988) Visuotopic organization and extent of V3 and V4 of the macaque. *J Neurosci* 8(6):1831–1845.
2. Kalwani RM, Bloy L, Elliott MA, Gold JI (2009) A method for localizing microelectrode trajectories in the macaque brain using MRI. *Journal of Neuroscience Methods* 176(2):104–111.
3. Mitzdorf U (1985) Current source-density method and application in cat cerebral cortex: investigation of evoked potentials and EEG phenomena. *Physiological Reviews* 65(1):37–100.
4. Pettersen KH, Devor A, Ulbert I, Dale AM, Einevoll GT (2006) Current-source density estimation based on inversion of electrostatic forward solution: Effects of finite extent of neuronal activity and conductivity discontinuities. *Journal of Neuroscience Methods* 154(1):116–133.
5. Hansen BJ, Dragoi V (2011) Adaptation-induced synchronization in laminar cortical circuits. *PNAS* 108(26):10720–10725.
6. Schroeder CE, Mehta AD, Givre SJ (1998) A spatiotemporal profile of visual system activation revealed by current source density analysis in the awake macaque. *Cereb Cortex* 8(7):575–592.
7. Swadlow HA, Gusev AG, Bezdudnaya T (2002) Activation of a Cortical Column by a Thalamocortical Impulse. *J Neurosci* 22(17):7766–7773.
8. Chen C-M, et al. (2007) Functional Anatomy and Interaction of Fast and Slow Visual Pathways in Macaque Monkeys. *Cereb Cortex* 17(7):1561–1569.
9. Ungerleider LG, Galkin TW, Desimone R, Gattass R (2008) Cortical Connections of Area V4 in the Macaque. *Cereb Cortex* 18(3):477–499.
10. Guillery RW, Sherman SM (2002) Thalamic Relay Functions and Their Role in Corticocortical Communication: Generalizations from the Visual System. *Neuron* 33(2):163–175.
11. Shipp S (2003) The functional logic of cortico–pulvinar connections. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 358(1438):1605–1624.
12. Edward G. Jones (1985) *The thalamus* (Springer Science+Business Media, LLC, Plenum Press, New York). Limited fascim. ed. Available at: http://proxy.library.nyu.edu/login?url=https://ebookcentral.proquest.com/lib/nyulibrary-ebooks/detail.action?docID=1296504 [Accessed July 10, 2018].
13. Bollimunta A, Chen Y, Schroeder CE, Ding M (2008) Neuronal Mechanisms of Cortical Alpha Oscillations in Awake-Behaving Macaques. *J Neurosci* 28(40):9976–9988.
14. Givre SJ, Schroeder CE, Arezzo JC (1994) Contribution of extrastriate area V4 to the surface-recorded flash VEP in the awake macaque. *Vision Research* 34(4):415–428.

15. Trongnetrpunya A, et al. (2016) Assessing Granger Causality in Electrophysiological Data: Removing the Adverse Effects of Common Signals via Bipolar Derivations. *Front Syst Neurosci* 9. doi:10.3389/fnsys.2015.00189.
16. Pinherio J, Bates D, DebRoy S, Sarkar D (2018) *nlme: Linear and Nonlinear Mixed Effects Models* Available at: https://CRAN.R-project.org/package=nlme [Accessed April 10, 2018].
17. Nandy AS, Mitchell JF, Jadi MP, Reynolds JH (2016) Neurons in Macaque Area V4 Are Tuned for Complex Spatio-Temporal Patterns. *Neuron* 91(4):920–930.
18. Strobl C, Malley J, Tutz G (2009) An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychological methods* 14(4):323.
19. Ho TK (2002) A Data Complexity Analysis of Comparative Advantages of Decision Forest Constructors. *Pattern Anal Appl* 5(2):102–112.