

The Effects of Population Tuning and Trial-by-Trial Variability on Information Encoding and Behavior

Ramon Nogueira,^{1,2} Nicole E. Peltier,³ Akiyuki Anzai,³ Gregory C. DeAngelis,³ Julio Martínez-Trujillo,⁴ and Rubén Moreno-Bote^{1,5}

¹Center for Brain and Cognition & Department of Information and Communication Technologies, Universitat Pompeu Fabra, Barcelona, 08002, Spain,

²Center for Theoretical Neuroscience, Mortimer B. Zuckerman Mind Brain Behavior Institute, Columbia University, New York, New York 10027,

³Department of Brain and Cognitive Sciences, Center for Visual Science, University of Rochester, Rochester, New York 14627, ⁴Robarts Research Institute and Brain and Mind Institute, Schulich School of Medicine and Dentistry, Departments of Psychiatry, Physiology and Pharmacology, Western University, London, Ontario N6A 3K7, Canada, and ⁵Serra Hunter Fellow Programme, Universitat Pompeu Fabra, Barcelona, 08002, Spain

Identifying the features of population responses that are relevant to the amount of information encoded by neuronal populations is a crucial step toward understanding population coding. Statistical features, such as tuning properties, individual and shared response variability, and global activity modulations, could all affect the amount of information encoded and modulate behavioral performance. We show that two features in particular affect information: the modulation of population responses across conditions (population signal) and the inverse population covariability along the modulation axis (projected precision). We demonstrate that fluctuations of these two quantities are correlated with fluctuations of behavioral performance in various tasks and brain regions consistently across 4 monkeys (1 female and 1 male *Macaca mulatta*; and 2 male *Macaca fascicularis*). In contrast, fluctuations in mean correlations among neurons and global activity have negligible or inconsistent effects on the amount of information encoded and behavioral performance. We also show that differential correlations reduce the amount of information encoded in finite populations by reducing projected precision. Our results are consistent with predictions of a model that optimally decodes population responses to produce behavior.

Key words: attention; global activity; MT/V5; noise correlations; PFC; sensory processing

Significance Statement

The last two or three decades of research have seen hot debates about what features of population tuning and trial-by-trial variability influence the information carried by a population of neurons, with some camps arguing, for instance, that mean pairwise correlations or global fluctuations are important while other camps report opposite results. In this study, we identify the most important features of neural population responses that determine the amount of encoded information and behavioral performance by combining analytic calculations with a novel nonparametric method that allows us to isolate the effects of different statistical features. We tested our hypothesis on 4 macaques, three decision-making tasks, and two brain areas. The predictions of our theory were in agreement with the experimental data.

Introduction

Identifying the statistical features of neuronal population responses that affect the amount of encoded information and be-

havioral performance is critical for understanding neuronal population coding (Arandia-Romero et al., 2017; Panzeri et al., 2017). Changes in network states, such as global modulations of activity (Harris and Thiele, 2011; Luczak et al., 2013; Gutnisky et al., 2017), as well as changes in correlated noise among neurons, have been shown to constrain the amount of information encoded by neuronal populations (Zohary et al., 1994; Ecker et al., 2014; Lin et al., 2015; Schölvinck et al., 2015). Indeed, it has been suggested that changes in neuronal tuning, global activity modulations, and noise correlations affect behavioral performance in cer-

Received April 15, 2019; revised Oct. 4, 2019; accepted Oct. 11, 2019.

Author contributions: R.N., G.C.D., J.M.-T., and R.M.-B. designed research; R.N., N.E.P., A.A., and J.M.-T. performed research; R.N. analyzed data; R.N. and R.M.-B. wrote the first draft of the paper; R.N., A.A., G.C.D., J.M.-T., and R.M.-B. edited the paper; R.N., A.A., G.C.D., J.M.-T., and R.M.-B. wrote the paper.

R.N. was supported by National Science Foundation's NeuroNex Program DBI-1707398 and Gatsby Charitable Foundation GAT3419. R.M.-B. was supported by Ministry of Economy (MINECO) (Spain) BFU2017-85936-P and FLAGERA-PCIN-2015-162-C02-02 and Howard Hughes Medical Institute 55008742. G.C.D., N.E.P., and A.A., as well as experiments in the G.C.D. laboratory, were supported by National Eye Institute (NEI) Grants EY013644 and EY016178. J.M.-T. was supported by the Canadian Institutes of Health Research (CIHR) and Natural Sciences and Engineering Research Council (NSERC). We thank K.H. Britten, J.A. Movshon, W.T. Newsome, M.N. Shadlen, and E. Zohary for making data available in the Neural Signal Archive (<http://www.neuralsignal.org/>); W. Bair for maintaining this database; and Matthew Leavitt for useful comments on the manuscript.

The authors declare no competing financial interests.

Correspondence should be addressed to Ramon Nogueira at rn2446@columbia.edu.

<https://doi.org/10.1523/JNEUROSCI.0859-19.2019>

Copyright © 2020 the authors

tain conditions (Cohen and Newsome, 2008; Cohen and Maunsell, 2009; Mitchell et al., 2009; Gu et al., 2011; Verhoef and Maunsell, 2017; Ni et al., 2018). Additionally, theoretical studies have determined the exact pattern of noise correlations that limit the amount of encoded information for very large neuronal populations, so-called differential correlations (Moreno-Bote et al., 2014; Kanitscheider et al., 2015). However, the aspects of neuronal responses that most directly affect the amount of encoded information for finite neuronal populations are not clear because experimental designs often do not allow control over other statistical features that could potentially be involved. Furthermore, it is unknown whether the same features of population responses that affect the amount of encoded information also impact behavioral performance (Arandia-Romero et al., 2017; Panzeri et al., 2017).

In a population of N neurons, it is possible to define N mean firing rates, $N(N + 1)/2$ independent covariances, as well as features based on combinations of these quantities. What statistical features matter the most for encoding information? Do these same features also affect behavioral performance? To address these questions, we characterized the amount of encoded information and behavioral performance in three different tasks based on responses of multiple neurons recorded simultaneously in macaque monkeys. We examined neurons recorded in two different brain areas: the middle temporal area (MT), and area 8a in the lateral prefrontal cortex (LPFC). We developed a conditioned bootstrapping approach that allowed us to determine the features of neuronal population responses that influence the amount of information encoded and behavioral performance by generating fluctuations of one feature while keeping the other features constant. Using this approach, we found that the amount of information encoded in neuronal ensembles was primarily determined by two features: (1) the length of the vector joining the mean population responses in different experimental conditions (referred to here as population signal [PS]); and (2) the inverse population covariability projected onto the direction of the PS vector (projected precision [PP]). Contrary to previous suggestions (Zohary et al., 1994; Kanitscheider et al., 2015; Lin et al., 2015; Ecker et al., 2016; Gutnisky et al., 2017), other statistical features, such as mean pairwise correlations (MPC) and global activity (GA) modulations, did not affect the amount of information encoded when PS and PP were kept constant. Strikingly, we also found that PS and PP are predictive of behavioral performance, whereas MPC and GA are not.

We show that MPC and GA are correlated with, and thus confounded with, PS and PP. This can explain why previous studies report that changes in MPC (Zohary et al., 1994; Mitchell et al., 2009; Ecker et al., 2010; Renart et al., 2010; Gu et al., 2011; Ni et al., 2018) or GA (Kanitscheider et al., 2015; Lin et al., 2015; Ecker et al., 2016; Gutnisky et al., 2017) modulate the amount of information encoded and behavioral performance, as those effects may have been due to concomitant changes in PS and PP, not MPC or GA per se. Finally, we link our results with previous theoretical work by showing that PP is reduced when differential correlations are added into the system. Our results are broadly consistent with predictions of a model that optimally decodes population responses to produce behavior.

Materials and Methods

Theoretical expression for the amount of encoded information

Theoretical decoding performance (DP) for an arbitrary linear classifier. If we assume that the activity of a neuronal population $\mathbf{r}$ (N neurons)

follows a multivariate Gaussian distribution, the covariance matrix for $\mathbf{r}$ is stimulus-independent, the probability of presenting condition 1 is same as presenting condition 2 ($p(C_1) = p(C_2) = 0.5$), and the classification is based on a linear projection of $\mathbf{r}$ onto a scalar variable $z = \boldsymbol{\omega}^T \mathbf{r} + \omega_0$ (linear classifier) (see Fisher, 1936), then the performance of the linear classifier can be expressed as follows:

$$DP = \Phi\left(\frac{1}{2} \frac{\boldsymbol{\omega}^T \Delta \mathbf{f}}{\sqrt{\boldsymbol{\omega}^T \Sigma \boldsymbol{\omega}}}\right) \quad (2)$$

where $\Delta \mathbf{f} \equiv \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$, $\boldsymbol{\mu}_1 = \mathbb{E}[\mathbf{r} | C_1]$, $\boldsymbol{\mu}_2 = \mathbb{E}[\mathbf{r} | C_2]$, $\Sigma = \mathbb{E}[(\mathbf{r} - \boldsymbol{\mu}_1)(\mathbf{r} - \boldsymbol{\mu}_1)^T | C_1] = \mathbb{E}[(\mathbf{r} - \boldsymbol{\mu}_2)(\mathbf{r} - \boldsymbol{\mu}_2)^T | C_2]$, $\Phi(\cdot)$ is the cumulative Gaussian function and DP is the decoding performance. This expression gives the percentage of correct classifications (i.e., DP) that would be achieved by a linear classifier that reads out from a neuronal ensemble using an arbitrary set of weights, $\boldsymbol{\omega}$.

Theoretical DP for the optimal linear classifier. By optimizing Equation 2 with respect to $\boldsymbol{\omega}$, we find that $\boldsymbol{\omega}_{opt} \propto \Sigma^{-1} \Delta \mathbf{f}$, which corresponds to the solution for the linear discriminant analysis (LDA) (Fisher, 1936). We can substitute $\boldsymbol{\omega}$ for $\boldsymbol{\omega}_{opt}$ in Equation 2 to find the DP for the optimal classifier:

$$DP = \Phi\left(\frac{1}{2} \sqrt{\Delta \mathbf{f}^T \Sigma^{-1} \Delta \mathbf{f}}\right). \quad (3)$$

The term inside $\Phi(\cdot)$ is known as $d' = \sqrt{\Delta \mathbf{f}^T \Sigma^{-1} \Delta \mathbf{f}}$ (Averbeck and Lee, 2006; Chen et al., 2006). It is a scalar quantity; therefore, it remains invariant under unitary rotations of the reference frame. By rotating the original neuron-based orthogonal basis so that Σ (and thus Σ^{-1}) becomes diagonal, we can express Equation 3 as follows:

$$DP = \Phi\left(\frac{1}{2} |\Delta \mathbf{f}| \sqrt{\sum_{i=1}^N \frac{\cos^2 \hat{\theta}_i}{\sigma_i^2}}\right). \quad (4)$$

For a similar derivation in the case of linear Fisher information, see Moreno-Bote et al. (2014). The first term, $|\Delta \mathbf{f}|$, which we will refer to as population signal (PS), is the norm of the stimulus tuning vector $\Delta \mathbf{f}$. It measures the overall modulation of the activity of the neuronal population as a function of the stimulus conditions. The second term,

$\sqrt{\sum_{i=1}^N \frac{\cos^2 \hat{\theta}_i}{\sigma_i^2}}$, which we will call projected precision (PP), is a function of $\hat{\theta}_i$, the angle between the i -th eigenvector of the covariance matrix Σ

and the direction of the stimulus tuning vector $\mathbf{u}_{\Delta \mathbf{f}} \equiv \frac{\Delta \mathbf{f}}{|\Delta \mathbf{f}|}$, and σ_i^2 , the i -th eigenvalue of the covariance matrix (Fig. 1A,B). Thus, PP is the square root of the sum of squares of the eigenvalues of the precision matrix (inverse of the noise covariance matrix, Σ) projected onto the direction $\mathbf{u}_{\Delta \mathbf{f}}$. This rotation allows us to dissect the independent contributions of population tuning (first-order statistics) and trial-by-trial variability (second-order statistics) to the amount of information encoded by a neural population, which is not possible with the standard factorization of d' into signal and noise components (see Discussion).

Theoretical DP for shuffled neuronal recording data. Trial shuffling is a commonly used technique to destroy the trial-by-trial shared fluctuations among neurons while preserving their mean firing rate to different experimental conditions. To understand the effects of noise correlations on DP, it is useful to derive theoretical expressions of DP for shuffled data. Shuffling the activity of single neurons across trials for a fixed stimulus condition transforms the covariance matrix Σ into Σ_{sh} (Averbeck et al., 2006), which is only approximately diagonal due to finite data size effects. The theoretical expression of DP for the shuffled data is given by:

$$DP = \Phi\left(\frac{1}{2} \frac{\boldsymbol{\omega}^T \Delta \mathbf{f}}{\sqrt{\boldsymbol{\omega}^T \Sigma_{sh} \boldsymbol{\omega}}}\right). \quad (5)$$

The optimal classifier is therefore $\boldsymbol{\omega}_{opt} \propto \Sigma_{sh}^{-1} \Delta \mathbf{f}$, and Equation 4 becomes:

$$DP = \Phi\left(\frac{1}{2}|\Delta\mathbf{f}|\sqrt{\sum_{i=1}^N \frac{\cos^2\theta_i}{\sigma_i^2}}\right). \quad (6)$$

In this expression, PS remains the same as in Equation 4, whereas the PP for the shuffled data becomes $\sqrt{\sum_{i=1}^N \frac{\cos^2\theta_i}{\sigma_i^2}}$, where σ_i^2 is the variance of the response of each neuron and θ_i is the angle between the stimulus tuning axis $\mathbf{u}_{\Delta\mathbf{f}}$ and each vector of the neuron-based orthogonal basis defining the original N -dimensional space of the neuronal activity. The above equation can also be expressed as follows:

$$DP = \Phi\left(\frac{1}{2}\sqrt{\sum_{i=1}^N \frac{(\Delta\mathbf{f}_i)^2}{\sigma_i^2}}\right), \quad (7)$$

where $\sqrt{\sum_{i=1}^N \frac{(\Delta\mathbf{f}_i)^2}{\sigma_i^2}}$ can be understood as the square root of the sum of the signal-to-noise ratio of all neurons in the ensemble (Seung and Sompolinsky, 1993).

Theoretical DP under suboptimal readouts. Here, we derive theoretical expressions of DP for two suboptimal classifiers: one blind to response variability and another blind to pairwise correlations (Pitkow et al., 2015).

The variability-blind classifier takes into account the neuronal tuning, but it is blind to any elements of the covariance matrix (i.e., it considers the covariance matrix to be proportional to the identity matrix). Thus, the readout weights for this classifier are given by $\omega \propto \Delta\mathbf{f}$. Introducing this expression into Equation 2, we find the following:

$$DP = \Phi\left(\frac{1}{2}\sqrt{\sum_{i=1}^N \frac{|\Delta\mathbf{f}|^2}{\sigma_i^2 \cos^2\theta_i}}\right), \quad (8)$$

where σ_i^2 and θ_i are as defined previously. The variability-blind classifier is directly related to the “axis of discrimination” framework and associated d' measure previously proposed (Cohen and Maunsell, 2009). In this framework, the axis of discrimination becomes precisely the difference in means vector $\Delta\mathbf{f}$, which corresponds to a suboptimal readout direction that ignores both the shared and individual trial-by-trial variability of the population. The variability-blind PP is defined as follows:

$$\left(\sum_{i=1}^N \sigma_i^2 \cos^2\theta_i\right)^{-\frac{1}{2}}$$

which is a lower bound on the real population's PP. As shown in Fig. 1-1 (available at <https://doi.org/10.1523/JNEUROSCI.0859-19.2019.f1-1>), the analytical expression for encoded information based on the variability-blind classifier performs more poorly than the theoretical DP obtained using the full covariance matrix. Moreover, the degree of suboptimality of the variability-blind classifier grows with neural ensemble size.

The correlation-blind classifier (Pitkow et al., 2015) only takes into account the signal-to-noise ratio of each neuron and ignores pairwise correlations among the neuronal ensemble (i.e., the off-diagonal terms in the covariance matrix are 0). The set of weights for this classifier, then, is given by $\omega_i \propto \frac{\Delta\mathbf{f}_i}{\sigma_i^2}$, which can also be written as

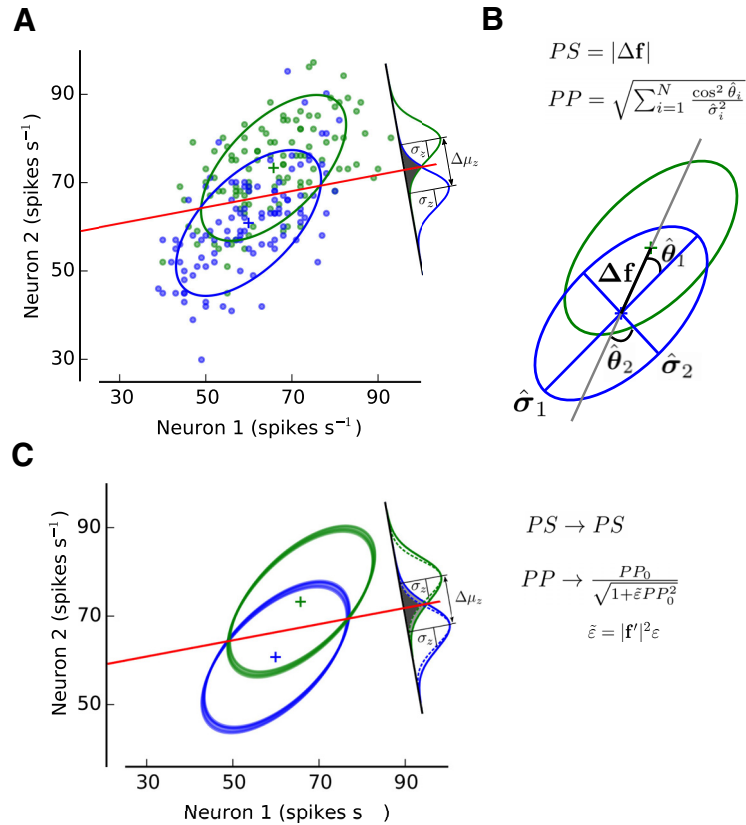


Figure 1. The information encoded by a neuronal ensemble is fully characterized by population signal (PS) and projected precision (PP). **A**, The trial-by-trial joint activity of a network consisting of N neurons can be characterized by an ellipsoid embedded in an N -dimensional space (a representative population of two MT neurons from Monkey 4 is shown here; see Fig. 2; see Materials and Methods). The covariance matrix and mean activity of the population determine the shape and location of the ellipsoids corresponding to stimulus 1 (s_1 ; green) and stimulus 2 (s_2 ; blue). The linear classifier is characterized by a linear boundary that separates the two clouds of points as well as possible (red line). The difference in mean value for the decision variable z and its SD are represented by $\Delta\mu_z$ and σ_z , respectively (signal-to-noise ratio of d'). **B**, Graphical depiction of PS and PP. PS is defined as the norm of the tuning vector, $\Delta\mathbf{f}$, which corresponds to the distance between the mean responses associated with s_1 and s_2 (distance between the centers of the green and blue ellipsoids). PP is calculated from the angles (θ_i) between each eigenvector of the covariance matrix and the tuning vector ($\Delta\mathbf{f}$) (orientation of each axis of the ellipsoid with respect to $\Delta\mathbf{f}$) as well as from their eigenvalues (σ_i^2 ; length of each axis). Greater information is encoded when the longest axis of the ellipsoid becomes orthogonal to the direction of $\Delta\mathbf{f}$. Mean pairwise correlations (MPC) is defined as the mean across all neuronal pairs of the trial-by-trial covariability (Pearson correlation) for a given set of trials. Global activity (GA) is defined as the mean activity across trials neurons and neurons for a set of trials. **C**, When information-limiting correlations are added into the system, the amount of encoded information is harmed due to a reduction of PP. Because PS accounts for the contribution of the population tuning on encoded information, it is not affected by the presence of information-limiting correlations. The magnitude of the reduction in PP will depend on the strength of information-limiting correlations (ε), the sensitivity of the population tuning to stimulus changes ($|\mathbf{f}'|^2$), and PP without information-limiting correlations (PP_0) (see Materials and Methods). See also Figure 1-1 (available at <https://doi.org/10.1523/JNEUROSCI.0859-19.2019.f1-1>), Figure 1-2 (available at <https://doi.org/10.1523/JNEUROSCI.0859-19.2019.f1-2>), and Figure 1-3 (available at <https://doi.org/10.1523/JNEUROSCI.0859-19.2019.f1-3>).

$\omega \propto \Sigma_{sh}^{-1} \Delta\mathbf{f}$. By substituting this expression into Equation 2, we obtain the following:

$$DP = \Phi\left(\frac{1}{2}|\Delta\mathbf{f}|\sqrt{\sum_{i=1}^N \frac{\cos^2\theta_i}{\sigma_i^2}}\right), \quad (9)$$

where σ_i^2 and θ_i are as defined earlier and σ_i^2 and θ_i correspond to the i -th eigenvalue and the angle between $\mathbf{u}_{\Delta\mathbf{f}}$ and the i -th eigenvector of the matrix $\Sigma_{sh}^{-1} \Sigma_{sh}^{-1}$, respectively. The correlation blind classifier, though suboptimal for correlation-intact data, is optimal for shuffled data from which pairwise correlations have been removed. As in the variability-blind case, PS is equivalent to that for the optimal case, while the

correlation-blind PP becomes $\frac{\sum_{i=1}^N \cos^2 \theta_i}{\sigma_i^2}$, which is also a lower bound on the true PP of a population of neurons.

Relationship to information-limiting correlations (differential correlations). There is a well-known characterization of the type of correlations that limit information in large neuronal populations, the so-called differential correlations (Moreno-Bote et al., 2014; Kanitscheider et al., 2015). The fundamental difference between the formulation of differential correlations and our approach is that the expression for differential correlations applies to correlations that limit information in very large neuronal populations ($N \rightarrow \infty$), while the statistical features PS and PP are guaranteed, under some assumptions, to be the only features that affect information for finite neuronal populations ($N < \infty$). However, as information-limiting correlations can also have an impact on the amount of encoded information for finite N , we examine in detail the relationship between these two approaches.

As described above, the task of our binary classifier is to assign one of two possible labels to a multidimensional pattern of activity. In most experimental situations, the two discrete labels are just particular instances of a continuous variable. For example, the direction of motion is a continuous variable, but the subject may be asked to discriminate between two particular directions (e.g., left vs right) in a motion discrimination task. Although we may measure neuronal activity only for particular values of a stimulus variable s (e.g., s_1 and s_2 ; $\Delta s \equiv (s_1 - s_2)$), the population tuning curve, $\mathbf{f}$, is a continuously varying function of s , $\mathbf{f}(s)$. Differential correlations (Moreno-Bote et al., 2014) are defined as a rank-one perturbation of the covariance matrix along the $\mathbf{f}'(s)$ direction:

$$\Sigma = \Sigma_0 + \varepsilon \mathbf{f}' \mathbf{f}'^T \quad (10)$$

where ε is the strength of the differential correlations, $\frac{d\mathbf{f}(s)}{ds} = \mathbf{f}'(s)$ is the derivative of the tuning curves, and Σ_0 is the covariance matrix without differential correlations. As we are interested in discrimination tasks for which the stimulus categories are not contiguous, we take the approximation $\mathbf{f}'(s) \approx \frac{\Delta \mathbf{f}}{\Delta s}$ for both stimulus conditions (Fig. 1C).

We can assess the impact of information-limiting correlations on the amount of encoded information in a finite population of neurons by directly inserting Equation 10 into Equation 2 to obtain the following:

$$DP = \Phi \left(\frac{1}{2} \frac{\omega^T \Delta \mathbf{f}}{\sqrt{\omega^T \Sigma_0 \omega + \varepsilon (\omega^T \mathbf{f}')^2}} \right). \quad (11)$$

In general, the optimal classifier takes the form $\omega_{opt} \propto \Sigma^{-1} \Delta \mathbf{f}$ (see previous section). Under $\mathbf{f}'(s) \approx \frac{\Delta \mathbf{f}}{\Delta s}$, this optimal readout direction is parallel to the optimal readout direction in the absence of differential correlations, $\omega_{0,opt} = \Sigma_0^{-1} \Delta \mathbf{f}$ (Moreno-Bote et al., 2014). However, as we will see now, DP is strongly affected by the presence of differential correlations. To see why, note that the amount of encoded information when there are no differential correlations ($\varepsilon = 0$) is $DP = \Phi \left(\frac{1}{2} d'_0 \right)$, where $d'_0 = \sqrt{\Delta \mathbf{f}^T \Sigma_0^{-1} \Delta \mathbf{f}}$. In contrast, with differential correlations, the amount of encoded information is given by the following:

$$DP = \Phi \left(\frac{1}{2} \frac{d'_0}{\sqrt{1 + \frac{\varepsilon}{\Delta s^2} d_0'^2}} \right). \quad (12)$$

Therefore, when differential correlations are introduced into the system, the original d'_0 is reduced by a factor $\sqrt{1 + \frac{\varepsilon}{\Delta s^2} d_0'^2}$, and thus encoded information is reduced. When the neuronal ensemble is very large ($N \rightarrow \infty$), Equation 12 becomes the following:

$$DP = \Phi \left(\frac{1}{2} \frac{\Delta s}{\sqrt{\varepsilon}} \right) \quad (13)$$

assuming that d'_0 grows monotonically with N (this is the case if Σ_0 does not contain differential correlations). Therefore, no matter how large the neuronal population is, information is bounded by $\Phi \left(\frac{1}{2} \frac{\Delta s}{\sqrt{\varepsilon}} \right)$.

To make more explicit the contribution of information-limiting correlations, PS, and PP, we can rewrite Equation 12 as follows:

$$DP = \Phi \left(\frac{1}{2} PS \times \frac{PP_0}{\sqrt{1 + \bar{\varepsilon} \times PP_0^2}} \right) \quad (14)$$

where PP_0 is the PP evaluated using Σ_0 instead of Σ and $\bar{\varepsilon} = \varepsilon |\mathbf{f}'|^2$. The larger the strength of information-limiting correlations (ε), the larger the decrease in encoded information in the neuronal population. In other words, when information-limiting correlations are introduced into the system, the statistical feature PS is not affected, whereas PP is reduced by a factor $\sqrt{1 + \bar{\varepsilon} \times PP_0^2}$. The framework presented in this study is complementary to the framework described by Moreno-Bote et al. (2014) by showing that, in finite neuronal populations, the contribution of trial-by-trial variability to encoded information is not necessarily dominated by differential correlations.

Comparison between theoretical and cross-validated DP on experimental data

We evaluated how well DP of the theoretical expression (DP_{th}) agrees with the DP of classifiers trained and tested on experimental data (DP_{cv}). We plotted DP_{th} against DP_{cv} and performed Type II regression (orthogonal linear regression) on the data. We assessed the similarity between the two measures of DP by computing the percentage of the variance in DP_{cv} that can be explained by DP_{th} as $\lambda_1 / (\lambda_1 + \lambda_2) \times 100$, where λ_1 and λ_2 correspond to the variance captured by the first and second principal components of the DP_{th} versus DP_{cv} plot, respectively. To account for possible idiosyncrasies from choosing an LDA when evaluating DP_{cv} and possible overfitting on DP_{cv} due to the limited number of trials in our datasets, we also computed DP_{cv} using logistic regression (LR) on both the test and the training sets (Fig. 1-2, available at <https://doi.org/10.1523/JNEUROSCI.0859-19.2019.f1-2>). In addition to the percentage of explained variance as the goodness-of-fit metric for the linear fit (% explained variance), we also considered the slope and intercept parameters as complementary goodness-of-fit metrics (Fig. 1-2, available at <https://doi.org/10.1523/JNEUROSCI.0859-19.2019.f1-2>) and obtained very similar overall results.

We compared how well the optimal expression (Eq. 4) approximated DP_{cv} with respect to the suboptimal expressions for DP_{th} (Eqs. 8, 9) by plotting the ratio between the percentages of explained variance (Fig. 1-1A, available at <https://doi.org/10.1523/JNEUROSCI.0859-19.2019.f1-1>). We performed the same analysis after shuffling across trials the activity of each individual neuron for a given stimulus condition (Fig. 1-1B, available at <https://doi.org/10.1523/JNEUROSCI.0859-19.2019.f1-1>). Finally, we compared an LDA, a quadratic discriminant analysis (QDA), and an LR. Since LDA performed the best among the three (Fig. 1-3, available at <https://doi.org/10.1523/JNEUROSCI.0859-19.2019.f1-3>), we chose LDA for computing DP_{cv} in all analyses presented here. For a detailed description of how these fits were performed for each dataset, see Experimental design and statistical analysis.

Conditioned bootstrapping method

We performed an analysis based on bootstrapping to test the effects of fluctuations in PS, PP, and other statistical quantities of neuronal responses on the amount of encoded information (DP_{cv}) and behavioral performance. The conditioned bootstrapping method involves two steps: (1) generating fluctuations in statistical quantities through bootstrapping, and (2) conditioning by selecting bootstrap samples that produce fluctuations in certain statistical features but not in others.

For a particular set of trials (subdataset, see below), bootstrap samples were generated by randomly selecting M trials (M being the total number

of trials for this particular subdataset) with replacement. For each bootstrap sample, we calculated the following quantities:

- Behavioral performance of the animal, denoted as B (see below for the definition of behavioral performance in each task).
- Theoretical DP, DP_{th} (Eq. 4).
- DP of a cross-validated linear classifier (LDA), DP_{cv} .
- PS and PP.
- MPC, defined as the average of all pairwise response correlations among neuron pairs for a fixed stimulus condition.
- GA of the neuronal population, defined as the mean neuronal activity across all neurons and trials.

We generated 2×10^4 bootstrap samples when analyzing behavior (see Fig. 9) and 10^3 when analyzing DP_{cv} (see Figs. 4, 6, and 8), as fitting the classifier is computationally expensive. From this procedure, we obtained a distribution of each quantity listed above. We will denote the obtained value of the statistical feature i for the bootstrap iteration j as x_{ij} , and the fluctuation of x_{ij} with respect to the median value of the distribution across bootstrap iterations as $\delta x_{ij} = x_{ij} - \bar{x}_i$, where $\bar{x}_i$ denotes the median of the distribution for that particular feature. We used the median rather than the mean so that the method works just as robustly for skewed distributions as for symmetrical distributions, although strongly skewed distributions were typically not observed (see Fig. 3). It is important to note that bootstrap samples could include nonunique trials. When evaluating statistical features computed using the covariance matrix, if the number of unique trials is small, singular matrices could be generated. Nevertheless, the number of unique trials in all datasets is typically 10 times larger than the number of neurons used in the analysis (see next section). Therefore, the rank of the bootstrapped covariance matrix is always bound by the number of neurons rather than by the number of unique trials subsampled for each bootstrap iteration. In other words, the probability of generating covariance matrices of full rank (number of neurons) in 2×10^4 (or 10^3) bootstrap iterations is effectively 1.

To assess what statistical features affect the amount of encoded information and behavioral performance the most, we evaluated the dependency of δDP_{cv} and δB on δx_i , where, as above, x_i represents a particular statistical feature of the neuronal responses. Here, dependency is measured by various quantities, including correlation. A dependency across bootstrap iterations between δDP_{cv} and δx_i (referred here as $p(\delta DP_{cv}, \delta x_i)$), and between δB and δx_i ($p(\delta B, \delta x_i)$) could result from dependencies mediated by a third quantity δx_k . For instance, Equation 4 predicts that DP_{cv} depends exclusively on PS and PP, but since PP decreases as the ensemble's noise (both individual and pairwise) increases, an inverse relationship between δMPC and δPP is expected. Therefore, we may find a correlation between δDP_{cv} and δMPC as a result of their dependencies on δPP . To estimate the strengths of the dependencies $p(\delta DP_{cv}, \delta x_i)$ and $p(\delta B, \delta x_i)$ that are not confounded by other variables, we developed a method for minimizing the correlation due to dependencies of δDP_{cv} (or δB) and δx_i on a third variable δx_k . Among all generated bootstrap samples, we selected those that yielded $\delta x_k \approx 0$ (i.e., x_k values within the ± 15 th percentile of its median value). Based on the selected samples, we computed $p(\delta DP_{cv}, \delta x_i | \delta x_k \approx 0)$ (or $p(\delta B, \delta x_i | \delta x_k \approx 0)$), that is, the dependency between δDP_{cv} (or δB) and δx_i conditioned on δx_k . This method can easily be extended to multiple conditioning variables ($\delta x_b, \delta x_r$, etc.) as long as there are enough bootstrap samples for the analysis that satisfy the conditions. Using this technique, we computed the following conditional dependencies:

- $p(\delta DP_{cv}, \delta PS | \delta PP \approx 0, \delta MPC \approx 0, \delta GA \approx 0)$ and $p(\delta B, \delta PS | \delta PP \approx 0, \delta MPC \approx 0, \delta GA \approx 0)$
- $p(\delta DP_{cv}, \delta PP | \delta PS \approx 0, \delta MPC \approx 0, \delta GA \approx 0)$ and $p(\delta B, \delta PP | \delta PS \approx 0, \delta MPC \approx 0, \delta GA \approx 0)$
- $p(\delta DP_{cv}, \delta MPC | \delta PS \approx 0, \delta PP \approx 0, \delta GA \approx 0)$ and $p(\delta B, \delta MPC | \delta PS \approx 0, \delta PP \approx 0, \delta GA \approx 0)$
- $p(\delta DP_{cv}, \delta GA | \delta PS \approx 0, \delta PP \approx 0, \delta MPC \approx 0)$ and $p(\delta B, \delta GA | \delta PS \approx 0, \delta PP \approx 0, \delta MPC \approx 0)$

Bootstrap samples were selected based on conditioned features, not on tested features, and the feature under study played no role on the conditioning procedure. For instance, to determine the real dependency between δDP_{cv} and δPS , we used only samples for which $\delta PP \approx 0, \delta MPC \approx 0, \delta GA \approx 0$, but δPS was unconstrained. To evaluate $p(\delta DP_{cv}, \delta x_i | \delta x_k \approx 0, \delta x_l \approx 0, \delta x_r \approx 0)$ (see Figs. 6, 10B) and $p(\delta B, \delta x_i | \delta x_k \approx 0, \delta x_l \approx 0, \delta x_r \approx 0)$ (see Figs. 9, 10C; Fig. 9-1, available at <https://doi.org/10.1523/JNEUROSCI.0859-19.2019.f9-1>), we used 10^3 and 2×10^4 bootstrap iterations, respectively. With the conditioning criterion of ± 15 th percentile around the median, $\sim 2.7\%$ of the total number of bootstrap iterations (27 for dependencies on DP_{cv} and 540 for B) satisfied the conditioning on all three x_k, x_b , and x_r from which to compute dependencies.

The dependencies $p(\delta DP_{cv}, \delta x_i | \delta x_k \approx 0, \delta x_l \approx 0, \delta x_r \approx 0)$ and $p(\delta B, \delta x_i | \delta x_k \approx 0, \delta x_l \approx 0, \delta x_r \approx 0)$ were evaluated by using two metrics: (1) the Pearson correlation coefficient (Fig. 9-1B, available at <https://doi.org/10.1523/JNEUROSCI.0859-19.2019.f9-1>) and (2) the percent change (see Figs. 4, 6, 9, 10B, C; Fig. 9-1A, available at <https://doi.org/10.1523/JNEUROSCI.0859-19.2019.f9-1>). Percent change in DP_{cv} with respect to δx_i (% change DP_{cv}^i) was defined as follows:

$$\% \text{change } DP_{cv}^i = \frac{\langle DP_{cv}^{\delta x_i^+} \rangle - \langle DP_{cv}^{\delta x_i^-} \rangle}{\langle DP_{cv}^{\delta x_i^-} \rangle}, \quad (15)$$

where $\langle DP_{cv}^{\delta x_i^+} \rangle$ and $\langle DP_{cv}^{\delta x_i^-} \rangle$ represent the mean DP_{cv} values for the bootstrap iterations that produced δx_i above (δx_i^+) and below (δx_i^-) its median value $\bar{x}_i$, respectively. Similarly, percent change in B with respect to δx_i (% change B^i) was defined as follows:

$$\% \text{change } B^i = \frac{\langle B^{\delta x_i^+} \rangle - \langle B^{\delta x_i^-} \rangle}{\langle B^{\delta x_i^-} \rangle}, \quad (16)$$

where $\langle B^{\delta x_i^+} \rangle$ and $\langle B^{\delta x_i^-} \rangle$ represent the mean B values for the bootstrap iterations that produced δx_i above (δx_i^+) and below (δx_i^-) its median value $\bar{x}_i$, respectively. Because behavior (B) was quantified as the mean reaction time in the attentional task, the sign of $p(\delta B, \delta x_i | \delta x_k \approx 0, \delta x_l \approx 0, \delta x_r \approx 0)$ was inverted so that negative fluctuations in B correspond to positive fluctuations in performance (see Figs. 8, 9B, C; Fig. 9-1, available at <https://doi.org/10.1523/JNEUROSCI.0859-19.2019.f9-1>) for Monkeys 2 and 3 (attentional task with recordings in LPFC 8a). A negative correlation between the statistical feature δx_i and % change DP_{cv}^i and/or % change B^i signifies an anticorrelation between the particular statistical feature and the amount of encoded information and/or behavioral performance.

To confirm that our results still hold for different conditioning criteria, we also computed percent change B using bootstrap iterations that restricted the range of the conditioning variables x_k, x_b , and x_r to be within ± 10 th and ± 20 th percentile of their respective median values (Fig. 9-1, available at <https://doi.org/10.1523/JNEUROSCI.0859-19.2019.f9-1>). We observed no significant differences between the results from the different conditioning criteria.

In Figure 4, the dependencies $p(\delta DP_{cv}, \delta MPC)$ and $p(\delta DP_{cv}, \delta GA)$ were evaluated using percent change DP_{cv}^i (Eq. 15) without conditioning on the rest of statistical features of the neuronal responses. For each monkey and ensemble size, the reported dependencies corresponded to the median value across independent subdatasets (see detailed description of each experimental dataset in Experimental design and statistical analysis). Statistical significance was calculated with a two-sided Wilcoxon signed-rank test, with which we tested whether the median of the distribution of independently obtained values was significantly greater or less than zero. Likewise, in Figure 5, the dependencies $p(\delta PS, \delta MPC)$, $p(\delta PS, \delta GA)$, $p(\delta PP, \delta MPC)$, and $p(\delta PP, \delta GA)$ were evaluated with the Pearson correlation between the bootstrap fluctuations of these statistical features. For each ensemble size, the reported dependencies corresponded to the median values across independent subdatasets and monkeys, and statistical significance was calculated with a two-sided Wilcoxon signed-rank test as before. For Figure 8 and Fig. 10-1G (available at <https://doi.org/10.1523/JNEUROSCI.0859-19.2019.f10-1>), the dependency $p(\delta B, \delta DP_{cv})$ was eval-

uated by computing the Pearson correlation coefficient between δDP_{cv} and δB . For each monkey and ensemble size, the median value across independent subdatasets was reported, and statistical significance was calculated with a two-sided Wilcoxon signed-rank test as for Figures 4 and 5. Error bars indicate the 25th to 75th percentile of a distribution of medians obtained by sampling with replacement from the distribution of independent values (bootstrap error bars; 1000 iterations). Error bars are only provided for visual guidance.

Neural population model

We built a neural population model that captures the basic statistical properties of the neural responses that we observed experimentally. The model includes limited-range correlations, differential correlations, and realistic values for tuning sensitivity, firing rates, and Fano factors. The model also generated choices based on optimal read out of information. We evaluated the simulation results by repeating all of the analyses described in the previous sections, including describing the relationships between PS, PP, MPC, and GA with both encoded information and simulated behavioral performance.

Population activity model. Our neural population model consisted of $N = 1000$ neurons. Each neuron's firing rate was modeled as a function of stimulus parameter s (e.g., motion direction) and stimulus strength, ν (representing, e.g., motion coherence, which controls task difficulty). We considered two stimuli $s_1 = +1$ and $s_2 = -1$ (analogous to the two directions of motion around the discrimination boundary) at three stimulus strength levels, $\nu = \{0.16, 0.32, 0.48\}$. We defined mean firing rate μ of neuron k as a linear function of stimulus parameters (i.e., tuning curve) as follows:

$$\mu_k(s, \nu) = m_{1k} s \nu + m_{0k} \quad (17)$$

where m_{1k} and m_{0k} are the slope (sensitivity) and the baseline (spontaneous) firing rate of neuron k , respectively. The slope parameters m_{1k} were drawn randomly from a normal distribution centered at the origin with a SD of $\sigma_{m_1} = 1.3$, and m_{0k} was drawn from a gamma distribution with shape and scale parameters set at 20 and 1, respectively. The parameters of the distributions were chosen to approximate empirical distributions. In what follows, we will use $\mu_k(s, \nu)$ and $f_k(s, \nu)$ (defined earlier in Theoretical expression for the amount of encoded information) synonymously. We will use the terms spike count, firing rate, and neuronal activity during a stimulus presentation interchangeably as the stimulus duration was set to 1 s in our simulations.

Responses of neurons to identical stimuli vary from trial to trial, and the trial-by-trial variabilities are partially shared among neurons (noise correlation) (Cohen and Kohn, 2011; Kohn et al., 2016). We incorporated noise correlations into our model in the form of limited-range pairwise correlations between neurons k and l (Kanitscheider et al., 2015; Kohn et al., 2016) as follows:

$$\rho_{kl}^{gen} = \left(A \exp\left(-\frac{|m_{1k} - m_{1l}|}{\tau}\right) + C \right) (1 - \delta_{kl}) + \delta_{kl}, \quad (18)$$

where δ_{kl} is the Kronecker delta ($A = 0.1$, $C = 0$, and $\tau = 1$ are used in our model).

Then, we defined a generative covariance matrix as follows:

$$\Sigma_{kl}^{gen}(s, \nu) = \sigma_k(s, \nu) \sigma_l(s, \nu) \rho_{kl}^{gen}, \quad (19)$$

where $\sigma_k^2(s, \nu)$ represents the trial-by-trial variance of neuron k . We used $\sigma_k^2(s, \nu) = \eta \mu_k(s, \nu)$ ($\eta = 0.5$), so that Fano factors of the model neurons are within the range of values typically found in experiments (Shadlen and Newsome, 1998; Arandia-Romero et al., 2016; Nogueira et al., 2018). $\Sigma_{kl}^{gen}(s, \nu)$ is not yet the full model's covariance matrix, $\Sigma_{kl}(s, \nu)$, which includes differential correlations and other components as derived in the next section.

Based on $\mu(s, \nu)$ and $\Sigma_{kl}^{gen}(s, \nu)$, we generated neuronal activity as follows. First, we chose s and ν pseudo-randomly for each trial of 1 s duration such that there were 100 trials per stimulus condition (600 trials in total per simulated recording session). Then, we generated a preliminary response vector $\mathbf{r}'_j$ for each trial j as follows:

$$\mathbf{r}'_j = \mu(s_j, \nu_j) + M \mathbf{z}_j + \mu' \delta s_j, \quad (20)$$

where $\Sigma_{kl}^{gen} = MM^T$, $\mathbf{z}_j$ represents an N -dimensional vector whose elements are drawn from a zero-mean, unit-variance Gaussian distribution, μ' denotes the derivative of the population tuning curve ($\mathbf{f}$), which equals $\mathbf{m}_1$ in our model, and δs_j indicates a common sensory noise term drawn from a Gaussian distribution with zero mean and variance $\sigma_{\delta s}^2 = \epsilon$. The last term $\mu' \delta s_j$ introduces differential correlations to the population activity (Moreno-Bote et al., 2014; Kanitscheider et al., 2015). The generated population activity pattern $\mathbf{n}_j$ was calculated by applying multiplicative and additive global modulations to the vector $\mathbf{r}'_j$ before putting it through a Poisson process that generated spike counts for each model neuron, that is:

$$\mathbf{n}_j \sim \text{Poisson}(g_j \mathbf{r}'_j + g_j). \quad (21)$$

Here, g_j is the global modulation factor drawn from a gamma distribution with scale and shape parameter $\theta_g = 10^{-5}$ and $k_g = 10^3$, respectively (such that $\langle g_j \rangle = 1$ and $\sigma_{g_j}^2 = 10^{-5}$). The global modulation incorporates the short and long timescale fluctuations in population activity often seen *in vivo* recordings (Ecker et al., 2014; Goris et al., 2014; Lin et al., 2015; Arandia-Romero et al., 2016). We generated a total of 10 simulated recording sessions that reproduced experimentally realistic distributions of Fano factors and pairwise correlations (Fig. 10-1C,D, available at <https://doi.org/10.1523/JNEUROSCI.0859-19.2019.f10-1>).

Behavioral decision model. We also modeled trial-by-trial behavioral choices using an optimal linear classifier that reads out the activity of the simulated population. For a given set of trials, simulated behavioral performance was calculated as the percentage of correct choices, just like the behavioral performance in the task involved Monkeys 1 and 4. For modeling behavior, the optimal classifier was derived from an analytical expression (Eqs. 22, 23, below) rather than from data fitting since the simulated data had a limited number of trials (200 trials per stimulus intensity) and may result in overfitting. By introducing Equation 17 into the analytical expression for the optimal linear classifier derived in Theoretical expression for the amount of encoded information, we obtain the following:

$$\omega_{opt} = 2\Sigma^{-1} \nu \mathbf{m}_1, \quad (22)$$

where Σ is the covariance matrix for our population model, which is given by the following:

$$\Sigma_{kl} = (\sigma_g^2 + 1) \left(\Sigma_{kl}^{gen} + \epsilon \mu'_k \mu'_l \right) + \sigma_g^2 ((\mu_k + 1)(\mu_l + 1)) + (\mu_k + 1) \delta_{kl}. \quad (23)$$

For each trial j , a decision variable was chosen as follows:

$$d_j = \omega_{opt}^T \mathbf{n}_j + \omega_0, \quad (24)$$

where $\omega_0 = -2\nu \mathbf{m}_1^T \Sigma^{-1} \mathbf{m}_0$. The behavioral choice c for trial j is, then, given by the sign of the decision variable d , that is, $c_j = \text{sign}(d_j)$. Behavioral performance was evaluated by the number of correct classifications over the total number of trials.

Information encoded by the model. In this section, we derive a theoretical expression for DP (DP_{th}) of our model. In our model, $\mathbf{f}' \approx \frac{\Delta \mathbf{f}}{\Delta s}$, and therefore, d' is given by Equation 12. When $N \rightarrow \infty$:

$$d' = \frac{2\nu}{\sqrt{\epsilon}}, \quad (25)$$

and therefore

$$DP = \Phi\left(\frac{\nu}{\sqrt{\epsilon}}\right). \quad (26)$$

Equation 26 suggests that, when the input signal is noisy on a trial-by-trial basis, DP (the amount of information that can be extracted) of a very large neural ensemble does not saturate at perfect performance. In-

stead, it will approach an asymptote below 1 determined by the signal-to-noise ratio of the input signal $\frac{\nu}{\sqrt{\epsilon}}$ (see Fig. 10-1F, available at <https://doi.org/10.1523/JNEUROSCI.0859-19.2019.f10-1>).

Analysis of model simulation data. We performed the analyses on our simulated neuronal and behavioral data (see Comparison between theoretical and cross-validated DP on experimental data and Conditioned bootstrapping method). First, we examined how well our theoretical DP (DP_{th}) approximated the cross-validated DP (DP_{cv}) of a linear classifier (LDA) evaluated on the simulated data, in the same way as we did for the *in vivo* recording data. For a particular ensemble size, we randomly selected a neuronal ensemble from a recording session and stimulus intensity and computed DP_{th} and DP_{cv} . For each ensemble size, stimulus intensity and simulated recording session, we repeated this process 20 times. We then fitted a line to the relationship between DP_{th} and DP_{cv} , and computed the percentage of variance in DP_{cv} that was explained by DP_{th} , as described in Comparison between theoretical and cross-validated DP on experimental data (Fig. 10-1E, available at <https://doi.org/10.1523/JNEUROSCI.0859-19.2019.f10-1>).

We also conducted the conditioned bootstrapping analyses on the simulated data using the method described in Conditioned bootstrapping method. We examined the effects of bootstrap fluctuations in statistical features of the population activity on the amount of encoded information and behavioral performance. For each bootstrap iteration, we selected an ensemble of a given size (2, 4, 6, 8, and 10 units) and calculated the following quantities: theoretical DP (DP_{th}), cross-validated DP of a linear classifier (DP_{cv}), PS, PP, MPC, GA, and behavioral performance (B). Both DP_{th} and DP_{cv} were evaluated for inferring the stimulus (either s_1 or s_2) from the activity pattern of the population model. This process was repeated 20 times for each of the recording sessions and stimulus intensities. The dependencies of DP_{cv} and B on different features of the neural activity were quantified by percent change in DP_{cv} and B as defined in Equations 15 and 16, as well as by Pearson correlation coefficients. The results were averaged over repetitions. For each ensemble size, we thus obtained 30 independent samples (10 surrogate recording sessions $\times$ 3 stimulus intensities), and the median was reported. Statistical significance for the dependency on the variables DP_{cv} (or B) and δx_i in terms of $p(\delta DP_{cv}, \delta x_i | \delta x_k = 0, \delta x_l = 0, \delta x_r = 0)$ (or $p(\delta B, \delta x_i | \delta x_k = 0, \delta x_l = 0, \delta x_r = 0)$) for each ensemble size was calculated with a two-sided Wilcoxon signed-rank test, with which we tested whether the median of the distribution of 30 independently obtained values was significantly greater or less than zero. It is important to note that $p(\delta B, \delta x_i | \delta x_k = 0, \delta x_l = 0, \delta x_r = 0)$ as percent change of surrogate behavior (see Fig. 10C; Eq. 16) can be very small for relatively large surrogate ensemble sizes ($N \sim 100$) while it may exhibit strong Pearson correlation at the same time. In our analysis, however, we prefer using percent change of behavior because of its interpretability and robustness to outliers in the distribution of bootstrap fluctuations. The same bias could potentially arise when evaluating $p(\delta DP_{cv}, \delta x_i | \delta x_k = 0, \delta x_l = 0, \delta x_r = 0)$ as percent change of DP_{cv} . Nevertheless, in this case, this effect is typically masked by the strong dependency between the amount of encoded information and PS/PP, a consequence of using the same neuronal population to compute all the relevant quantities.

Experimental design and statistical analysis

Our theory was tested on three different datasets obtained from 4 monkeys, each of which involved simultaneously recorded units (from 2 to ~ 50 units). One dataset involved pairs of MT neurons from a monkey (Monkey 1) performing a coarse direction discrimination task (Zohary et al., 1994). The second dataset involved LPFC (area 8a) neurons recorded from 2 monkeys (Monkeys 2 and 3) that were trained to perform a spatial attention task (Tremblay et al., 2015). The final dataset involved recordings from groups of MT neurons in a monkey (Monkey 4) trained to perform a fine direction discrimination task. The details for each experiment are described below.

Coarse direction discrimination task with recordings in area MT. This dataset has been previously described (Zohary et al., 1994), and is freely available at the Neural Signal Archive (<http://www.neuralsignal.org/nsa2004.2>). Data from 1 female adult macaque monkey (*Macaca mu-*

latta) are included in this dataset. The animal was trained to perform a coarse direction discrimination task (Britten et al., 1992), and pairs of single neurons were recorded in area MT. The monkey performed a coarse direction discrimination task, in which a noisy random-dot motion stimulus moved in one of two opposite directions. In each trial, a stimulus was presented for 2 s. The presented direction of motion was either the preferred or null direction of the recorded neurons. If the preferred directions of the two neurons differed substantially, the axis of discrimination was set to the preferred-null axis of the best-responding neuron. The null direction was defined as the direction opposite to the preferred direction. The strength of the motion signal was manipulated by controlling the fraction of dots that moved coherently from one frame to the next (motion coherence), whereas the remaining “noise” dots were randomly replotted on each video update. When motion coherence was 0%, all dots moved randomly, and there was no coherent direction in the net motion. When motion coherence was 100%, all dots moved coherently in either the preferred or null direction for the pair of neurons. A range of coherences between 0% and 100% was used to adjust task difficulty and measure neural and behavioral sensitivity. In each trial, the monkey was presented randomly with either the preferred or null direction of motion at a particular coherence. The monkey’s task was to report the direction of the net motion by making a saccade to one of two choice targets. The psychometric function averaged across sessions is shown in Figure 7A.

Each trial started with the appearance of a fixation point. After 2 s of stimulus presentation, the random dot pattern and the fixation point disappeared, and two choice targets appeared on the screen, corresponding to the two possible directions of motions (preferred or null) (see Fig. 2A). The monkey made a saccadic eye movement to one of the targets to report its perceived direction of motion. Data from this experiment (Zohary et al., 1994) consisted of 41 recording sessions in which pairs of single units were simultaneously recorded in area MT. Only sessions with random-dot motion stimuli generated from a variable seed were used.

Since stimulus strength was controlled by the motion coherence parameter, we subdivided each recording session according to the coherence presented in each trial. Trials belonging to the 0% coherence condition were discarded because stimulus identity and correct behavioral performance were not defined for this condition. To evaluate how accurate our analytical approximation was for the amount of information encoded by the neural population, we plotted DP_{th} against DP_{cv} for each recorded pair of neurons and coherence condition and fitted the data points with a Type II linear regression. The percentage of variance explained by the first principal component of the data represented the goodness of fit (see Comparison between theoretical and cross-validated DP on experimental data).

For the analysis involving behavior, we also discarded trials with coherences that elicited a mean behavioral performance $>98\%$ correct to avoid ceiling effects (25.6%, 51.2%, and 99.9% coherence were eliminated by this criterion). This was done because the bootstrap distributions for behavioral performance (B) would have very low variability for these high coherences; therefore, calculating $p(\delta B, \delta x_i | \delta x_k = 0, \delta x_l = 0, \delta x_r = 0)$ may be highly biased or even undefined. The criterion for discarding easy trials (98% percent correct) corresponds to eliminating conditions that are >2 SDs away from the mean of a cumulative Gaussian fit, since $\Phi(2\sigma) \approx 0.98$. After splitting recording sessions by coherence and discarding 0% and high coherences as described above, we obtained 187 independent subdatasets, each of which corresponded to a set of trials for a particular coherence level and recording session. The mean number of trials per subdataset was 75, ranging from 30 to 231 trials. For the analysis, we used a trial-by-trial population activity vector whose entries corresponded to the spike count from the entire 2 s stimulus duration for each neuron. All statistical features (PS, PP, GA, and MPC) as well as DP (and DP_{th}) were calculated using this time window.

From each subdataset, we generated bootstrap distributions for the following quantities: theoretical DP (DP_{th}), cross-validated (trained and tested on different sets of trials) DP of a linear classifier (DP_{cv}), PS, PP, MPC, GA, and B. Behavioral performance was defined as the fraction of correct choices of the monkey on each subdataset. Both DP_{th} and DP_{cv} were calculated by inferring the motion direction (preferred or null)

presented to the monkey on a trial-by-trial basis from the simultaneously recorded activity of a neuronal pair.

Significance for the dependency $p(\delta DP_{cv}, \delta x_i | \delta x_k = 0, \delta x_l = 0, \delta x_r = 0)$ (or $p(\delta B, \delta x_i | \delta x_k = 0, \delta x_l = 0, \delta x_r = 0)$) for the whole experiment (187 independent subdatasets) was calculated with a two-sided Wilcoxon signed-rank test, with which we tested whether the median of the distribution of 187 independently obtained values was significantly greater or less than zero.

Spatial attention task with recordings in LPFC (prearcuate area 8a). This dataset is described in detail by Tremblay et al. (2015). Two male monkeys (*Macaca fascicularis*), both 6 years old (Monkey F, 5.8 kg; Monkey JL, 7.5 kg), contributed to this dataset. In each monkey, a 96-channel “Utah” multielectrode array was chronically implanted in the left caudal LPFC. The multielectrode array was inserted on the prearcuate convexity posterior to the caudal end of the principal sulcus and anterior to the arcuate sulcus, a region cytoarchitecturally known as area 8a (Bullock et al., 2017). The monkeys were trained to covertly sustain attention to one of four Gabor stimuli (target) presented on a screen while ignoring the other three Gabor stimuli (distractors) (see Fig. 2B). At the beginning of each trial, a cue indicated which of the four Gabor stimuli was the target (cue period, 363 ms). After the cue period, all four Gabor stimuli appeared on the screen, which marked the start of the attentional period. The attentional period ended after a variable delay (585–1755 ms) when one or two Gabor stimuli changed orientation by 90°. Only correct “Target” trials were used in the analysis reported here (Tremblay et al., 2015).

The mean number of correct “Target” trials per recording session was 207 (range: 172–224) for Monkey JL and 221 trials (range: 198–246) for Monkey F. The mean number of simultaneously recorded units was 56 (range: 53–61) for Monkey JL and 54 (range: 44–66) for Monkey F. Neuronal recordings included both single units and multiunits. The mean percentage of single units over the total number of simultaneously recorded units was 44% and 40% for Monkey JL and Monkey F, respectively. Because units with very low firing rates precluded reliable statistical analysis, we excluded units with firing rates <1 Hz for all subsequent analyses. After this exclusion, the mean number of simultaneously recorded units was 51 (range: 50–53) for Monkey JL and 50 (range: 42–59) for Monkey F. The shortest attentional period used was 585 ms; therefore, we defined a fixed attentional time window of 585 ms starting at the end of the cue period. In this way, the firing rate of all units was calculated over the same time window. All statistical features (PS, PP, GA, and MPC) as well as DP (and DP_{th}) were calculated using this time window.

To maximize the statistical power of our analysis, we created a larger number of independent subdatasets as follows. For each recording session, we randomly selected 21, 10, 7, 5, and 4 nonoverlapping ensembles of size 2, 4, 6, 8, and 10, respectively, and repeated this process 5 times. Since the smallest number of simultaneously recorded units for both monkeys was 42 units, we chose values that maximized the number of nonoverlapping ensembles for each size.

To evaluate how accurate our analytical approximation was for the amount of information encoded by the neural population for a particular ensemble size, we plotted DP_{th} against DP_{cv} for each subdataset and fitted the data points with a Type II linear regression. The percentage of variance explained by the first principal component of the data represented the goodness of fit of our approximation (see Comparison between theoretical and cross-validated DP on experimental data).

For each subdataset and bootstrap iteration, we calculated the following quantities: DP_{th} , DP_{cv} , PS, PP, MPC, GA, and behavioral performance (B). Behavioral performance was quantified as the mean reaction time across trials (either the original subdataset or a bootstrap iteration) (for reaction time distributions from Monkeys 2 and 3, see Fig. 7B,C, respectively). Both DP_{th} and DP_{cv} were calculated by inferring the monkey's location of attention on a trial-by-trial basis from the simultaneously recorded activity of an ensemble. For decoding purposes, we considered two binary classification tasks. In the first task, the decoder classified the locus of attention as being on the right or the left side of the screen. In the second task, the decoder classified the locus of attention as being on the upper half or the lower half of the screen. The reported DP_{cv} (and DP_{th}) values are averages over the two decoding tasks.

For a given subdataset, the reported dependency relationship $p(\delta DP_{cv}, \delta x_i | \delta x_k = 0, \delta x_l = 0, \delta x_r = 0)$ (or $p(\delta B, \delta x_i | \delta x_k = 0, \delta x_l = 0, \delta x_r = 0)$) was evaluated as the mean across the five iterations and the two classification tasks (vertical and horizontal). For each ensemble size and monkey, the reported dependency $p(\delta DP_{cv}, \delta x_i | \delta x_k = 0, \delta x_l = 0, \delta x_r = 0)$ (or $p(\delta B, \delta x_i | \delta x_k = 0, \delta x_l = 0, \delta x_r = 0)$) was the median across recording sessions and nonoverlapping ensembles of units (independent subdatasets). We obtained 84, 40, 28, 20, and 16 (the number of nonoverlapping ensembles $\times$ 4 recording sessions per monkey) independent values for ensemble sizes of 2, 4, 6, 8, and 10 units, respectively, and assessed the median and tested significance. Significance was calculated with a two-sided Wilcoxon signed-rank test of whether the median of the distribution of independently obtained values for each ensemble size and monkey was significantly greater or less than zero.

Fine direction discrimination task with recordings in MT. This dataset was obtained specifically for this analysis, as part of a larger series of ongoing studies of how MT neurons represent local motion in the presence of background optic flow. One male adult macaque monkey (*M. mulatta*) was used in this experiment. The animal was surgically implanted with a circular head holding device, a scleral coil for measuring eye movements, and a recording grid (Gu et al., 2006, 2008). The animal was trained to perform a fine direction discrimination task (described below) with water as a reward for correct performance. Eye movements were measured and controlled at all times. Neuronal activity in MT/V5 was recorded with 24-channel linear electrode arrays (V-probes, Plexon). Spike waveforms were acquired by a Cerebus System (Blackrock). All experimental procedures conformed to National Institutes of Health guidelines and were approved by the University Committee on Animal Resources at the University of Rochester.

In each recording session, a V-probe was inserted into MT/V5 and was allowed to settle for ~ 30 min. We then performed standard tests (DeAngelis and Uka, 2003) to map receptive fields and to measure the direction tuning of the recorded units. These measurements were used to determine the location and size of the stimulus aperture such that it was larger than the receptive fields of the units under study. However, stimulus motion was not tailored to the preferred directions and speeds of the recorded neurons; a fixed set of stimulus velocities was used across sessions for the fine discrimination task. A total of 75 units were recorded across 3 sessions. The mean number of simultaneously recorded units was 25 (range: 24–27). Most recordings included multiunit activity as well as 2–4 well-isolated single units. Because units with very low firing rates precluded reliable statistical analysis, we excluded units with firing rates <1 Hz for all subsequent analyses. Despite the exclusion, the mean number of simultaneously recorded units remained at 25 (range: 24–27).

The monkey was trained to perform a fine direction discrimination task. A pattern of random dots, presented within a circular aperture, moved upward in the visual field with either a rightward or leftward component. The animal's task was to report whether the perceived motion was up-right or up-left by making a saccade to one of two choice targets. The motion stimulus was presented at 100% coherence in one visual hemifield and was localized and sized according to the receptive fields of the recorded units. Stimulus motion within the aperture followed a Gaussian velocity profile with an SD of 333 ms and a duration of 2 s. Once the monkey fixated for 200 ms, the motion stimulus appeared and began to move. The stimulus was presented stereoscopically at zero disparity, such that motion appeared in the plane of the display. The experimental protocol involved 7 directions of motion relative to vertical: -12° , -6° , -3° , 0° , 3° , 6° , and 12° , where negative and positive values correspond to leftward and rightward motion, respectively. After stimulus presentation was completed, two saccade targets appeared, 5° to the right and left of the fixation target, and the monkey reported his perceived direction of motion by making a saccade to one of the targets (see Fig. 2C). Because stimuli were presented at 100% coherence, task difficulty was controlled by the direction of motion with respect to vertical (see Fig. 7D). The mean number of trials per session was 742 (range: 735–756). In each recording session, the same number of trials were presented for each direction of motion.

Since direction of motion controlled the difficulty of the task, this parameter was considered to be the stimulus strength. To control for

stimulus strength, we divided data from each recording session according to motion direction. The task of our decoder is to correctly classify each trial as having rightward or leftward motion direction; therefore, trials from each recording session were split into 4 independent subdatasets ($\pm 12^\circ$, $\pm 6^\circ$, $\pm 3^\circ$, and 0° stimulus directions relative to vertical). As described above for the coarse discrimination task, we subsequently discarded stimulus conditions that were either ambiguous (0° motion direction; correct behavioral performance is not defined) or too easy ($\pm 12^\circ$ directions; mean behavioral performance ≥ 0.98). The analysis was thus performed on 6 independent subdatasets (3 recording sessions $\times$ 2 absolute motion directions). To quantify neuronal activity, we computed firing rates in a time window that included ± 1 SD around the peak of the Gaussian velocity profile of the stimulus (666 ms window width). All statistical features (PS, PP, GA, and MPC) as well as DP (and DP_{th}) were calculated using this time window.

To increase statistical power, we created a larger number of independent subdatasets by sampling randomly nonoverlapping ensembles of particular sizes. Namely, we considered ensemble sizes of 2, 4, 6, 8, and 10, which yielded 12, 6, 4, 3, and 2 nonoverlapping ensembles of units, respectively. We repeated this sampling procedure 20 times. Since the smallest number of simultaneously recorded units was 24, we chose values that maximized the number of nonoverlapping ensembles for each size.

We evaluated how accurate DP_{th} was with respect to DP_{cv} by following the same procedure as for Monkeys 2 and 3 (see previous section). For each subdataset and bootstrap iteration, we calculated the following quantities: DP_{th} , DP_{cv} , PS, PP, MPC, GA, and behavioral performance (B). For this task, behavioral performance was defined as the fraction of correct choices for that particular subdataset. Both DP_{th} and DP_{cv} were calculated by inferring whether motion direction was leftward or rightward of vertical for each trial based on the neuronal activity pattern. For a particular randomly constructed ensemble, the reported dependency relationship $p(\delta DP_{cv}, \delta x_i | \delta x_k \approx 0, \delta x_l \approx 0, \delta x_r \approx 0)$ (or $p(\delta B, \delta x_i | \delta x_k \approx 0, \delta x_l \approx 0, \delta x_r \approx 0)$) was evaluated as the mean across the 20 iterations. For each ensemble size, the reported dependency $p(\delta DP_{cv}, \delta x_i | \delta x_k \approx 0, \delta x_l \approx 0, \delta x_r \approx 0)$ (or $p(\delta B, \delta x_i | \delta x_k \approx 0, \delta x_l \approx 0, \delta x_r \approx 0)$) was the median across recording sessions, stimulus strengths, and nonoverlapping ensembles of units. We used 72, 36, 24, 18, and 12 (the number of nonoverlapping ensembles $\times$ 3 independent recording sessions $\times$ 2 stimulus strengths) independent values to assess the median and test its significance for ensemble sizes of 2, 4, 6, 8, and 10 units, respectively. Significance was calculated with a two-sided Wilcoxon signed-rank test for whether the median of the distribution of independent values for each ensemble size was significantly greater or less than zero.

Data and software availability

The datasets generated in this study and the code used for their analysis are available from the corresponding author upon reasonable request.

Results

We start by showing that fluctuations of MPC and GA influence the amount of information encoded in population activity, con-

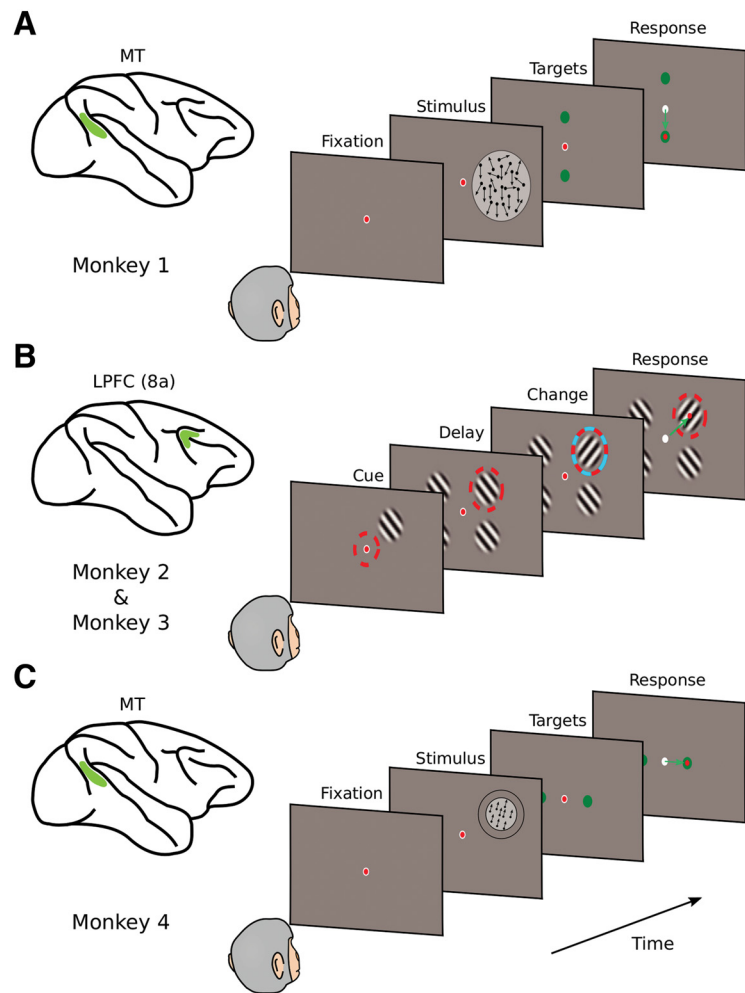


Figure 2. Three behavioral tasks used to test theoretical predictions in macaque monkeys. **A**, One monkey performed a coarse direction discrimination task (Monkey 1) while pairs of units were recorded in the MT area (Britten et al., 1992). After stimulus presentation (random dots moving toward the preferred or null direction of the neurons under study), the monkey reported the direction of motion by making a saccade to one of two targets. Difficulty was controlled by varying the percentage of coherently moving dots in the stimulus. **B**, Two monkeys performed an attentional task (Monkeys 2 and 3) while ~ 50 units were recorded simultaneously from the LPFC (area 8a) (Tremblay et al., 2015). Four Gabor patterns were presented on the screen, and the task was to make a saccade to the attended location after a change in orientation of the cued Gabor. **C**, One monkey performed a fine direction discrimination task (Monkey 4) while ~ 25 units were recorded simultaneously from area MT. After presentation of a fully coherent random dot stimulus, the monkey had to report whether dots moved leftward or rightward of vertical by making a saccade to one of two targets. Difficulty was controlled by making the left/right component of motion very small. For details, see Materials and Methods.

sistent with some previous studies. We then demonstrate that these effects of MPC and GA are eliminated when PS and PP are controlled for, whereas isolated fluctuations of PS and PP strongly predict the amount of information encoded in the population. Finally, we demonstrate that fluctuations of PS and PP are correlated with behavioral performance, and we compare this finding with predictions of an optimal decoding model.

MPC and GA correlate with the amount of encoded information

MPC and GA have been thought to modulate the amount of information encoded in neuronal populations (Zohary et al., 1994; Renart et al., 2010; Ecker et al., 2011, 2016; Kanitscheider et al., 2015; Lin et al., 2015; Gutnisky et al., 2017). We tested the correlation between these statistical features and the amount of information in three different datasets consisting of simultaneously recorded units (2 to ~ 50 single units or multiunits) in 4

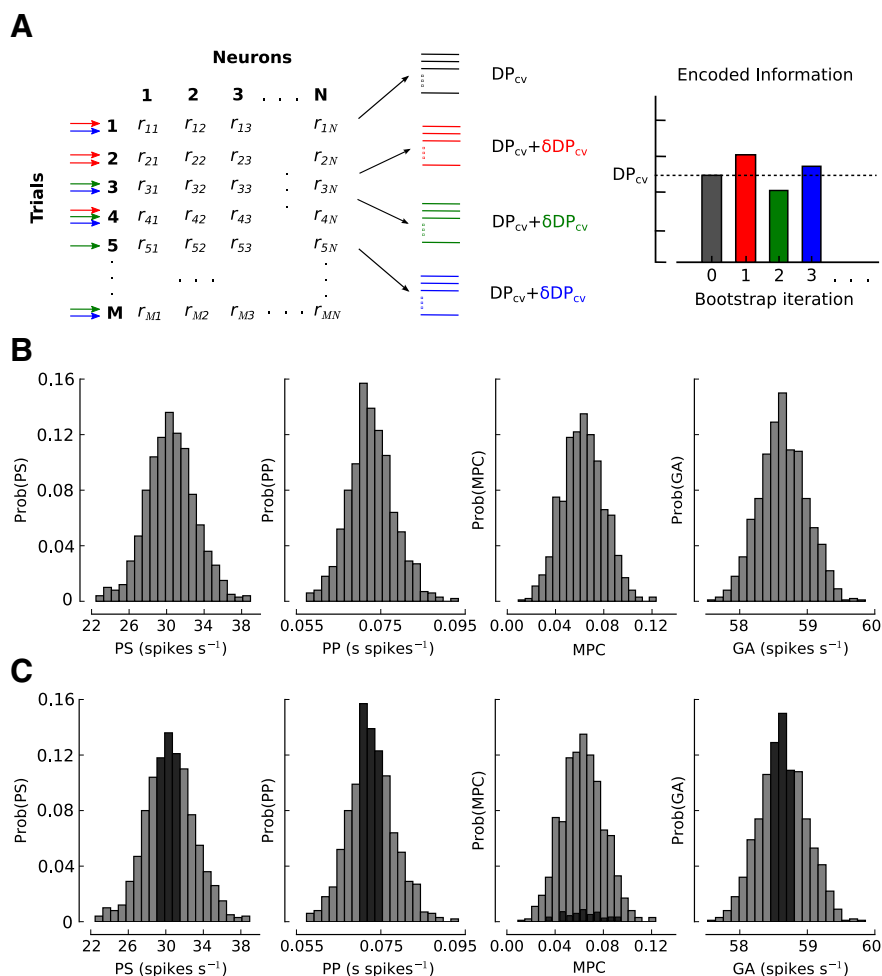


Figure 3. Dependencies between encoded information and statistical features of neuronal responses can be evaluated by the conditioned bootstrapping method. **A, B.** By subsampling trials with replacement (bootstrap) from the original dataset, distributions of the values of encoded information (DP_{cv}) (**A**) and statistical features of the neuronal activity (**B**) are generated. Example distributions of the values of PS, PP, MPCs, and GA for Monkey 4 (ensemble size = 10 units). **C.** A conditioning analysis is performed to determine the impact of fluctuations of each feature of the neuronal response on information by selecting subsets of bootstraps in which the other features are held close to their distributional medians. For instance, to study the effect of MPC on information, bootstraps are selected such that PS, PP, and GA are fixed near their median values (dark regions in the distributions). If bootstrap fluctuations in MPC (while fixing PS, PP, and GA) do not modulate DP_{cv}, we can conclude that MPC do not play a role in the amount of information encoded by a neural network. This approach is also used to study the effects of bootstrap fluctuations of neuronal activity on behavioral performance.

monkeys, two brain areas, and three tasks: (1) pairs of MT neurons recorded while the animal performed a coarse motion discrimination task (Zohary et al., 1994) (Monkey 1, Fig. 2A); (2) LPFC (area 8a) neurons recorded with a Utah array while 2 animals performed an attentional task (Tremblay et al., 2015) (Monkeys 2 and 3; Fig. 2B); and (3) MT neurons recorded with a 24-channel linear array while the animal performed a fine motion discrimination task (Monkey 4; Fig. 2C). Behavioral performance in these tasks was defined as the percentage of correctly reported directions of motion (Monkeys 1 and 4) or as the mean reaction time when correctly detecting a change in orientation of the attended Gabor patch (Monkeys 2 and 3). The amount of encoded information for each dataset was quantified as the cross-validated DP (DP_{cv}) of a linear classifier that reads out the activity of the recorded neuronal population to predict (1) which of two opposite directions of motion was presented in the coarse motion task (Monkey 1); (2) which Gabor patch was cued in the attention task (Monkeys 2 and 3); or (3) whether the stimulus motion was right or left of vertical in the fine direction discrimination task (Monkey 4). In each case, the

classifier was trained on the activity of neuronal ensembles recorded in area MT for the motion tasks or LPFC (area 8a) for the attention task (for details, see Materials and Methods).

To evaluate whether the amount of encoded information (DP_{cv}) was modulated by MPC and GA, we used a nonparametric method to produce fluctuations of MPC and GA generated by bootstrapping trials from each neural recording dataset (Fig. 3A, B). We then examined how fluctuations of MPC and GA affected DP_{cv} (Fig. 4; see Materials and Methods). We found that an increase in MPC tended to produce a decrease in DP_{cv}, therefore reducing the amount of encoded information. This was largely consistent across tasks and animals. For instance, for Monkey 1 (Fig. 4A), an increase of MPC significantly reduced the amount of encoded information by -0.05% (Wilcoxon signed-rank test, $p = 2.6 \times 10^{-3}$; see Materials and Methods). Qualitatively similar results were found for Monkeys 2–4 (Fig. 4B–D). In contrast, positive bootstrap fluctuations of GA tended to increase the amount of encoded information for some animals, particularly those from which we made PFC recordings. For Monkey 2 (Fig. 4B), bootstrap fluctuations of GA produced significant positive changes in DP_{cv} by 0.91% (ensemble size 2, Wilcoxon signed-rank test, $p = 6.9 \times 10^{-15}$; see Materials and Methods). For Monkey 3 (Fig. 4C), results were qualitatively similar to those obtained for Monkey 2, but for Monkeys 1 and 4, no significant effect was observed. Some of the size differences in the effects of MPC and GA on DP may be due to differences in behavioral tasks and/or brain areas (MT vs LPFC 8a). In particular, the population of LPFC 8a neurons was somewhat unbalanced with respect to spatial preference ($\sim 70\%$ contralateral preference), whereas direction selectivity in populations of MT neurons is generally quite balanced. This might explain the overall stronger effects of MPC and GA fluctuations on encoded information in the PFC animals (Monkeys 2 and 3) because those fluctuations have a stronger chance to spontaneously align with the individual, but unbalanced, sensory preferences.

Although the observed correlations between MPC, GA, and the amount of encoded information are generally consistent with those from previous experimental and theoretical studies (Zohary et al., 1994; Renart et al., 2010; Ecker et al., 2011, 2016; Kanitscheider et al., 2015; Lin et al., 2015; Gutnisky et al., 2017), the effects are rather small and somewhat inconsistent. It is important to note that the results reported in Figure 4 are obtained by correlating MPC and GA with DP_{cv} using all bootstrap samples. Thus, the correlation between MPC and GA fluctuations with DP_{cv} could be mostly due to other statistical features that co-fluctuate with MPC or GA. Indeed, as shown

what unbalanced with respect to spatial preference ($\sim 70\%$ contralateral preference), whereas direction selectivity in populations of MT neurons is generally quite balanced. This might explain the overall stronger effects of MPC and GA fluctuations on encoded information in the PFC animals (Monkeys 2 and 3) because those fluctuations have a stronger chance to spontaneously align with the individual, but unbalanced, sensory preferences.

below, MPC and GA are confounded with PP and PS, and do not exhibit significant correlations with DP_{cv} once PP and PS are held constant.

PS and PP determine the amount of information encoded in neuronal population responses

In this section, we identify the statistical features of population tuning and trial-by-trial variability that affect the amount of encoded information (as measured by DP_{cv}), and we test our predictions on the neural data described above.

Under some common assumptions (see Materials and Methods), an analytical expression for the theoretical DP (DP_{th}) of a linear classifier can be derived (Averbeck and Lee, 2006). This expression takes the form $DP_{th} = \Phi(d'/2)$, where $\Phi(x)$ is the cumulative normal function and $d' = \sqrt{\Delta \mathbf{f}^T \Sigma^{-1} \Delta \mathbf{f}}$ is the signal-to-noise ratio generalized for a population of neurons (Averbeck and Lee, 2006; Chen et al., 2006; Gutnisky et al., 2017). The term $\Delta \mathbf{f}$ is the vector joining the means of the population responses in the two stimulus conditions, and Σ is the stimulus-invariant noise covariance matrix of the neuronal population. To understand the roles of population tuning and trial-by-trial variability in determining the amount of encoded information, it is useful to rewrite this equation by rotating the original N -dimensional neural response space along the eigenvectors of the covariance matrix as follows:

$$DP_{th} = \Phi\left(\frac{1}{2}|\Delta \mathbf{f}| \sqrt{\sum_{i=1}^N \frac{\cos^2 \hat{\theta}_i}{\sigma_i^2}}\right), \quad (1)$$

where $\hat{\theta}_i$ represents the angle between the i -th eigenvector of the covariance matrix and the unitary direction defined by the stimulus vector $\Delta \mathbf{f}$, and σ_i^2 denotes the i -th eigenvalue of the covariance matrix (Fig. 1). Equation 1 reveals that the amount of information encoded by a neural population can be divided into two independent components: the first-order contribution from the population tuning (PS = $|\Delta \mathbf{f}|$) and the second-order contribution from the trial-by-trial variability ($PP = \sqrt{\sum_{i=1}^N \frac{\cos^2 \hat{\theta}_i}{\sigma_i^2}}$). It is important to remark that d' is often expressed as the signal-to-noise ratio of the linear classifier ($d' = \frac{\Delta \mu_z}{\sigma_z}$), where under optimality, the signal is given by $\Delta \mu_z = \Delta \mathbf{f}^T \Sigma^{-1} \Delta \mathbf{f}$, and the noise is given by $\sigma_z = \sqrt{\Delta \mathbf{f}^T \Sigma^{-1} \Delta \mathbf{f}}$. While this d' formulation aims to differentiate signal from noise with respect to the classifier's decision variable (Averbeck and Lee, 2006) (Fig. 1), it does not separate the contributions of firing rate modulation (PS) and trial-by-trial variability (PP) of the population to the amount of encoded information because $\Delta \mu_z$ and σ_z both contain terms associated with PS and PP.

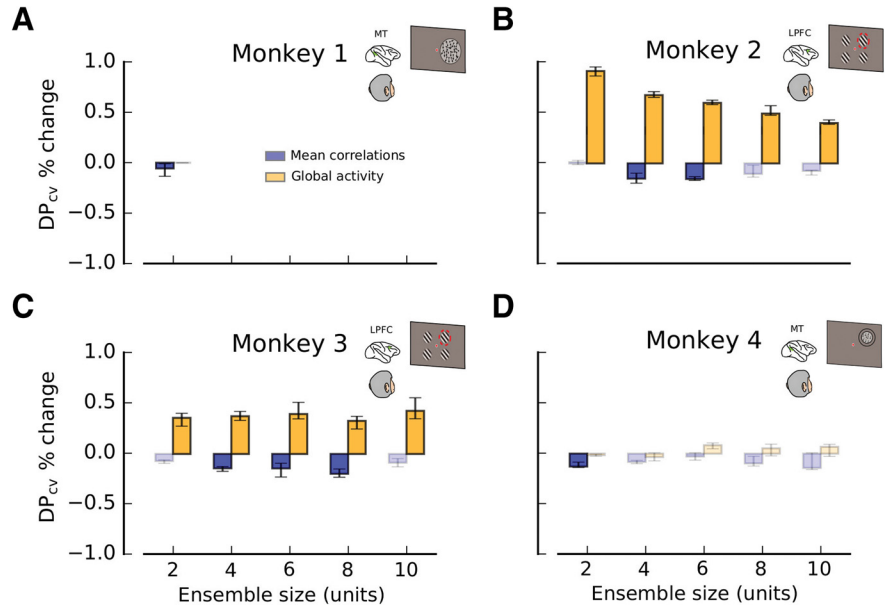


Figure 4. MPC and GA correlate with encoded information. **A**, Percent change in the amount of encoded information (DP_{cv}) related to fluctuations of MPC (blue) and GA (orange). Fluctuations of MPC and GA are produced through a bootstrap process in which virtual instances of the same experiment are generated by sampling trials with replacement from a particular dataset (Fig. 3B; see Materials and Methods; not to be confused with the conditioning analysis described in Fig. 3C). Smaller values of MPC produce significantly larger values of DP_{cv} ; thus, they are negatively correlated. Data are shown from pairs of neurons recorded in MT during a coarse direction discrimination task (Monkey 1; Fig. 2A). **B**, **C**, Percent change in DP_{cv} as a function of ensemble size (2, 4, 6, 8, and 10 units), for recordings from LPFC 8a during an attentional task (Monkeys 2 and 3; see Fig. 2B; see Materials and Methods). GA has a strong modulatory effect on encoded information on both monkeys. **D**, Analogous results for MT ensembles recorded during a fine direction discrimination task (Monkey 4; see Fig. 2C; see Materials and Methods). MPC has a consistent negative effect on encoded information. In all panels, error bars indicate the 25th to 75th percentile of the distribution of bootstrap medians. High-contrast colored bars represent significant deviations from 0 (Wilcoxon signed rank test, $p < 0.05$).

There are two salient points to make regarding the relationships between PS/PP and other measures of neuronal population information. First, it is important to emphasize that PP is not simply the projection of correlated noise (covariance matrix) onto the vector joining the means of the population responses in the two stimulus conditions, $\Delta \mathbf{f}$ (for additional details, see Materials and Methods). For instance, it is possible to have a finite projection of correlated noise onto $\Delta \mathbf{f}$, but PP can nevertheless be very large. This can happen if there is an eigenvector of the covariance matrix that is not orthogonal to $\Delta \mathbf{f}$ with a very small eigenvalue, indicating the presence of highly correlated sets of neurons. Second, previous studies have identified the exact pattern of noise correlations that limit information for very large neuronal populations, the so-called differential correlations (Moreno-Bote et al., 2014; Kanitscheider et al., 2015). Differential correlations add neuronal activity fluctuations along $\mathbf{f}'$ (parallel to the discrimination direction, $\Delta \mathbf{f}$), and they are known to be the only type of correlations that can limit information for very large neuronal populations ($N \rightarrow \infty$). We show (see Materials and Methods) that adding differential correlations into the system reduces PP but does not affect PS. The reduction of PP by differential correlations is $\sqrt{1 + \bar{\epsilon} \times PP_0^2}$, where $\bar{\epsilon} = \epsilon |\mathbf{f}'|^2$ and PP_0 is the PP evaluated on the original covariance matrix without information-limiting correlations. The variable ϵ directly measures the strength of differential correlations. Thus, differential correlations can have a strong effect on PP, but they are not the only factor in determining PP for finite neuronal populations.

We find that Equation 1 provides a very good approximation to the empirical measure of the amount of encoded information,

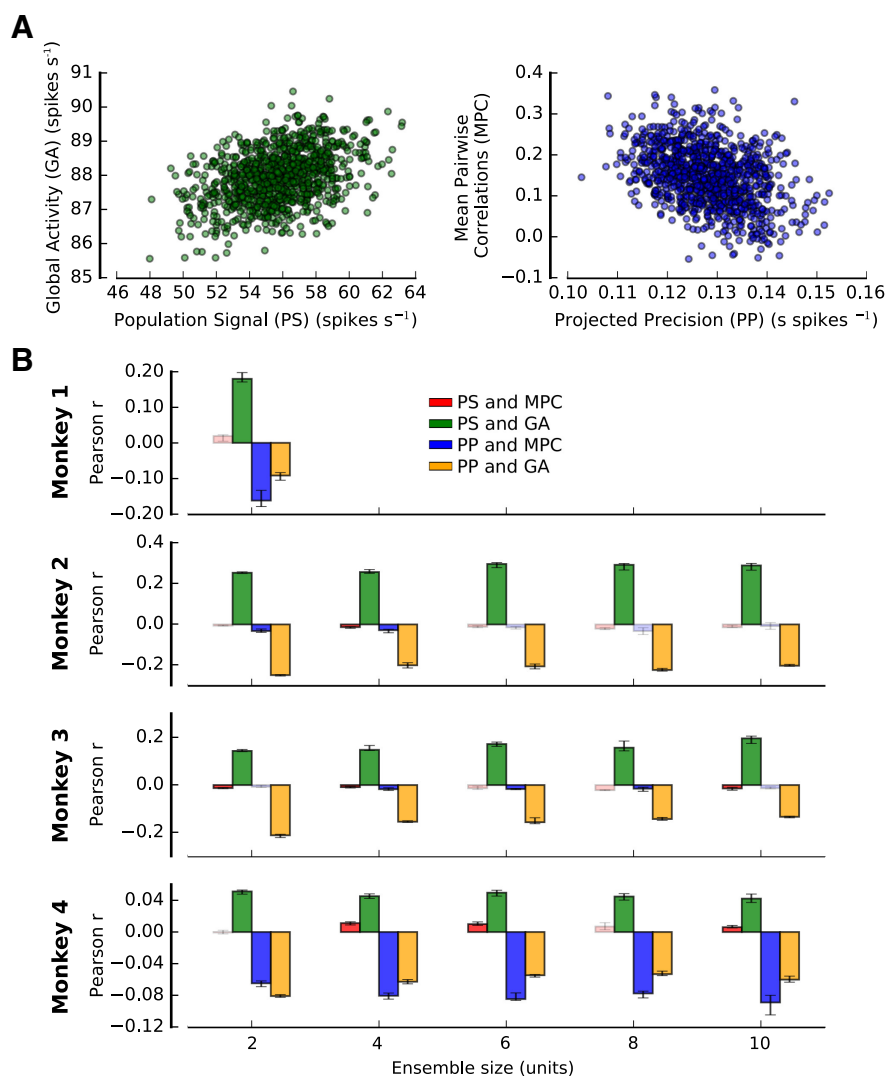


Figure 5. Bootstrap fluctuations in PS, PP, MPC, and GA are correlated. **A**, Examples of how bootstrap fluctuations (without conditioning) in GA are correlated with bootstrap fluctuations in PS (left), and fluctuations of MPC are correlated with PP (right) for pairs of neurons recorded in MT during a coarse direction discrimination task (Monkey 1; Fig. 2A). **B**, Median Pearson correlation between fluctuations in PS and MPC (red), PS and GA (green), PP and MPC (blue), and PP and GA (orange) for each monkey and ensemble size (see Materials and Methods). Error bars indicate the 25th to 75th percentile of the distribution of bootstrap medians. High-contrast colored bars represent significant deviations from 0 (Wilcoxon signed rank test, $p < 0.05$). Correlations between these different statistical features of the neuronal activity are likely to explain reported dependencies of encoded information on MPC and GA.

as it accounts for more than 96% of the variance in the cross-validated DP of a linear classifier (DP_{cv}) for all datasets (Fig. 1-2, available at <https://doi.org/10.1523/JNEUROSCI.0859-19.2019.f1-2>). In addition, Equation 1 is a better approximation to DP_{cv} than analytical expressions that simplify the covariance structure by removing the off-diagonal terms (Averbeck and Lee, 2006; Averbeck et al., 2006; Pitkow et al., 2015) or calculations that assume a covariance proportional to the identity matrix (“axis of discrimination” method; see Materials and Methods; Fig. 1-1, available at <https://doi.org/10.1523/JNEUROSCI.0859-19.2019.f1-1>) (Cohen and Maunsell, 2009). Finally, LDA produced better matches to DP_{cv} across different ensemble sizes and monkeys than LR and QDA did, justifying our use of LDA to evaluate the amount of information encoded by neural populations (DP_{cv} ; Fig. 1-3, available at <https://doi.org/10.1523/JNEUROSCI.0859-19.2019.f1-3>).

Having identified analytically the two features of population tuning and trial-by-trial variability that should determine the

amount of encoded information (i.e., PS and PP), we now test the central prediction that information depends exclusively on PS and PP and does not depend on other statistical features, such as MPC and GA, unless those features are correlated with PS and PP. Indeed, we found that bootstrap fluctuations of MPC, GA, PS, and PP showed substantial correlations among them (Fig. 5; see Materials and Methods). For instance, for Monkey 1, the correlations between fluctuations of PS and GA, PP and MPC, and PP and GA were all highly significant (ensemble size 2, $\rho_{PS,GA} = 0.18$, $p = 2.4 \times 10^{-24}$; $\rho_{PP,MPC} = -0.16$, $p = 1.1 \times 10^{-11}$; $\rho_{PP,GA} = -0.091$, $p = 1.1 \times 10^{-12}$), whereas the correlation between fluctuations of PS and MPC was not statistically significant (ensemble size 2, $\rho_{PS,MPC} = 0.019$, Wilcoxon signed-rank test, $p = 0.15$; see Materials and Methods). For Monkeys 2–4, results were qualitatively similar; although for Monkeys 2 and 3, the correlation $\rho_{PP,MPC}$ was weaker. Therefore, we suspected that the relationships between MPC (and GA) and the amount of encoded information (Fig. 4) would be reduced, if not eliminated, once the modulations in MPC (GA) were made independent of PS, PP, and GA (MPC) by selecting bootstrap samples with constant PS, PP, and GA (MPC) values.

We tested these predictions by using a conditioned bootstrapping method to evaluate the effect of fluctuations in one feature on the amount of encoded information while the values of other features are kept constant (Fig. 3C; see Materials and Methods). For example, to determine the effects of MPC on DP_{cv} independent of PS, PP, and GA, we generated bootstrap samples from the original dataset and selected those bootstrap iterations that produced PS, PP, and GA values within a narrow range around their medians. Then, for the selected data, we evaluated

the percent change in DP_{cv} introduced by the bootstrap fluctuations in MPC. Any dependence of DP_{cv} on MPC, then, would not be explainable by the dependencies of DP_{cv} on PS, PP, or GA. This method can be used to test the effects of any statistical feature on the amount of encoded information by fixing other features to their representative values, thus isolating the individual effects. There are other possible approaches to isolating the effects of different statistical features on the amount of encoded information (DP_{cv}), such as a model-dependent analysis based on GLMs. However, due to the linearity assumptions underlying GLMs and the potential collinearity between several statistical features, it is preferable to use a model-independent approach based on conditioned bootstrapping (see Materials and Methods).

Applying this conditioned bootstrapping approach to our datasets, we found that bootstrap fluctuations of PS and PP greatly affected the amount of encoded information, whereas fluctuations of MPC and GA produced negligible changes in DP_{cv} across

different tasks, animals, and neuronal ensemble sizes (Fig. 6). As anticipated, when fluctuations of MPC and GA were conditioned on PS and PP, the effects reported in Figure 4 largely vanished (Fig. 6, right column). For Monkey 1 (Fig. 6A), fluctuations of PS and PP produced significant positive changes in the amount of encoded information by 5.43% (Wilcoxon signed-rank test, $p = 3.2 \times 10^{-37}$; see Materials and Methods) and 2.75% ($p = 1.3 \times 10^{-35}$), respectively. In contrast, fluctuations of MPC and GA had no significant effects on DP_{cv} ($p = 0.33$ and $p = 0.16$, respectively). For Monkeys 2–4 (Fig. 6B–D), results were qualitatively similar to those obtained for Monkey 1 (Table 6-1, available at <https://doi.org/10.1523/JNEUROSCI.0859-19.2019.t6-1>). We also compared the difference in percent change in DP_{cv} for all pairs of statistical features and found again that PS and PP had the most significant effects on the amount of encoded information (Table 6-2, available at <https://doi.org/10.1523/JNEUROSCI.0859-19.2019.t6-2>). The observed dependency of DP_{cv} on PP is generally weaker than that on PS. This could be explained at least partially by the fact that PP is a second-order statistic and therefore is expected to be noisier than the first-order statistic, PS, when estimated from limited experimental data.

PS and PP are also the strongest predictors of behavioral performance

We have shown that PS and PP are the major statistical features of population activity that affect the amount of encoded information in neuronal populations. But do these features also have an impact on behavioral performance? We reasoned that, if downstream neurons that contribute to behavioral choices can extract most of the information encoded by a neuronal population, then behavioral performance should depend on PS and PP while it is largely independent of MPC and GA. However, since there could be many layers of nonlinear computations between the information encoding stage and the final behavioral choice, finding such a relationship is not guaranteed *a priori*.

Behavioral performance was measured as the percentage of correctly reported directions of motion (Monkeys 1 and 4) (Fig. 7A,D) or as the mean reaction time (Monkeys 2 and 3) (Fig. 7B,C). We first confirmed that modulations in the amount of encoded information (DP_{cv}) over different bootstrap samples were significantly correlated with the corresponding changes in behavioral performance (Fig. 8). Although the reported correlations are weak, they are nevertheless consistent with the predic-

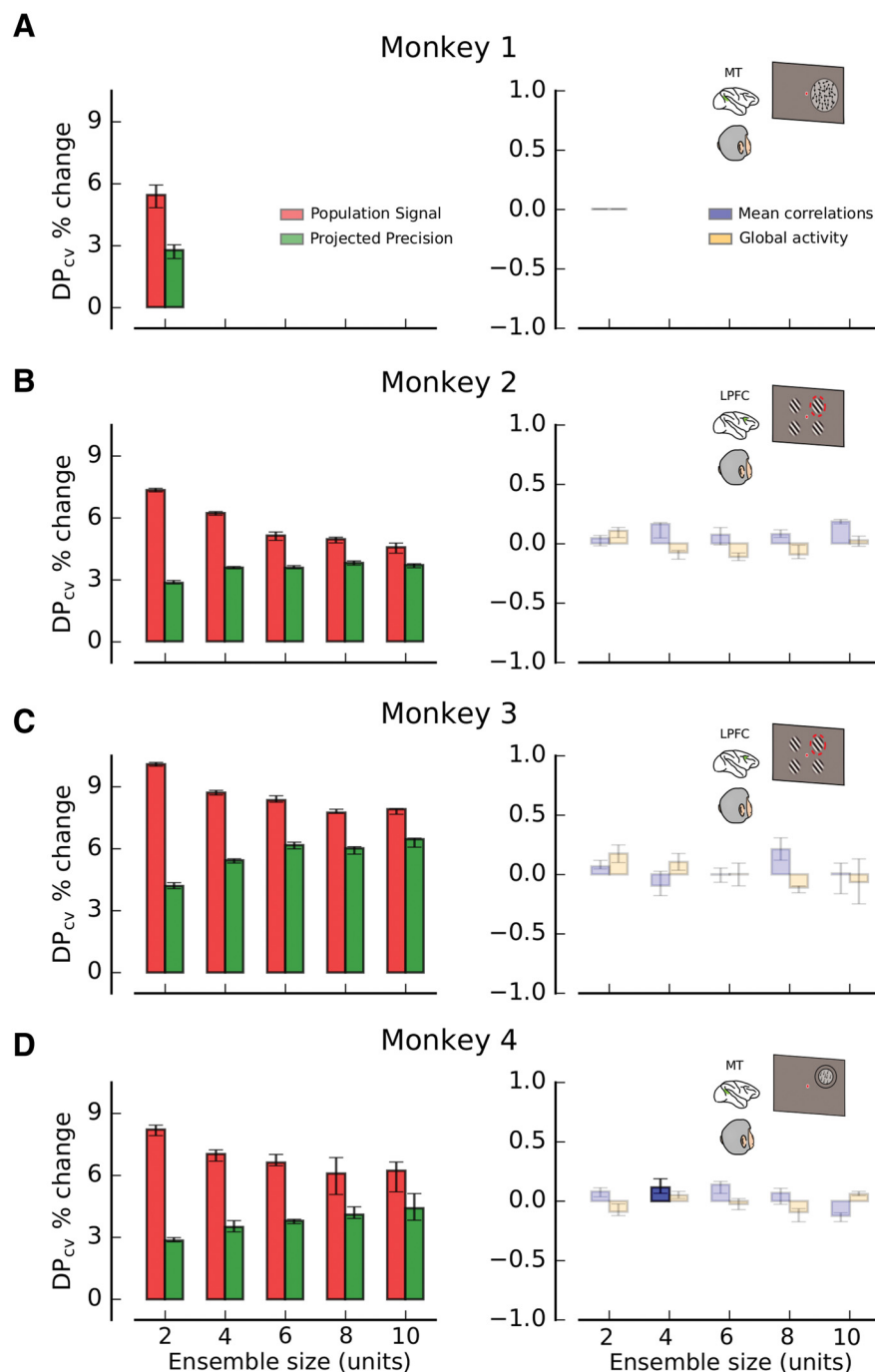


Figure 6. Encoded information depends mainly on PS and PP once other variables are controlled. **A**, Percent change in the amount of encoded information (DP_{cv}) when changes in one statistical feature of neuronal responses are isolated by the conditioned bootstrapping method described in Figure 3C. Red represents PS. Green represents PP. Blue represents MPC. Orange represents GA. Only PS and PP produce significant changes in DP_{cv} when the other three features are kept constant (e.g., PS is tested by using bootstrap samples for which the values of PP, MPC, and GA are all close to their respective median values). Data are shown from pairs of neurons recorded in MT during a coarse direction discrimination task (Monkey 1; see Fig. 2A; see Materials and Methods). **B**, **C**, Percent change in DP_{cv} as a function of ensemble size (2, 4, 6, 8, and 10 units), for recordings from LPFC 8a during an attentional task (Monkeys 2 and 3; see Fig. 2B; see Materials and Methods). **D**, Analogous results for MT ensembles recorded during a fine direction discrimination task (Monkey 4; see Fig. 2C; see Materials and Methods). In all panels, error bars indicate the 25th to 75th percentile of the distribution of bootstrap medians. High-contrast colored bars represent significant deviations from 0 (Wilcoxon signed rank test, $p < 0.05$). See also Table 6-1 (available at <https://doi.org/10.1523/JNEUROSCI.0859-19.2019.t6-1>) and Table 6-2 (available at <https://doi.org/10.1523/JNEUROSCI.0859-19.2019.t6-2>).

tions of a model that reads out neuronal population activity optimally to produce behavioral choices (see next section). We then performed the conditioned bootstrap analysis on behavioral performance separately for each stimulus strength or task difficulty

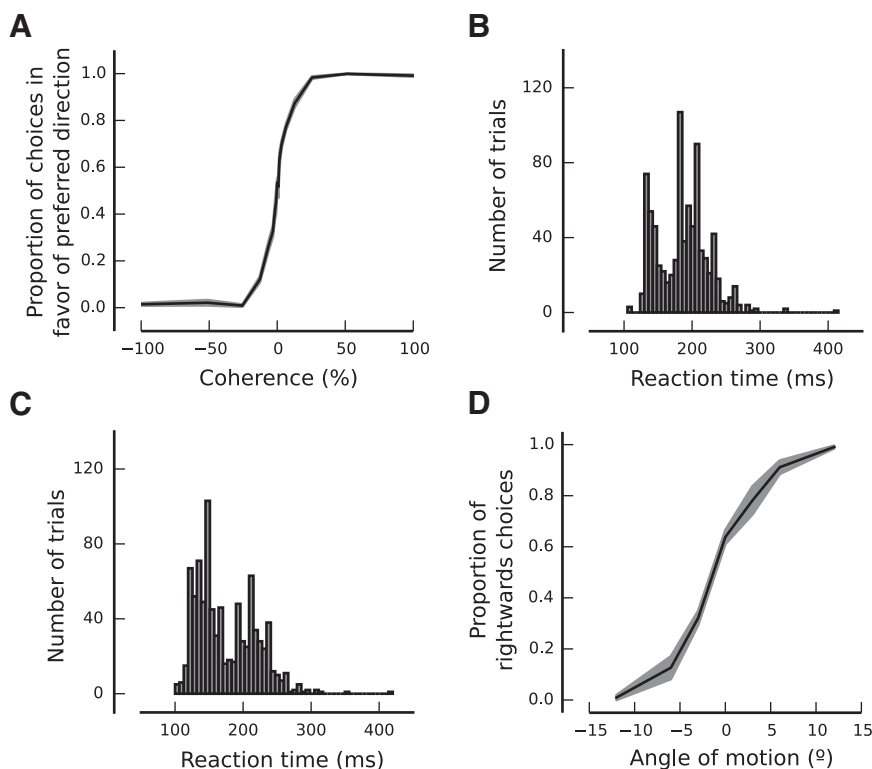


Figure 7. Definitions of behavioral performance for each task. **A**, Psychometric curve averaged across recording sessions for Monkey 1. Proportion of choices in favor of the preferred direction as a function of motion coherence (positive values of coherence correspond to the preferred direction of motion). Gray region represents SEM. **B**, **C**, Distributions of reaction times for Monkeys 2 and 3. For the attentional task, reaction time is defined as mean time (across trials in a particular dataset) from the change in stimulus orientation until the saccade to the cued Gabor pattern. **D**, Psychometric curve averaged across recording sessions for Monkey 4. Proportion of rightward choices as a function of the direction of motion with respect to vertical (positive and negative values denote motions with a rightward and leftward component, respectively). Gray region represents SEM.

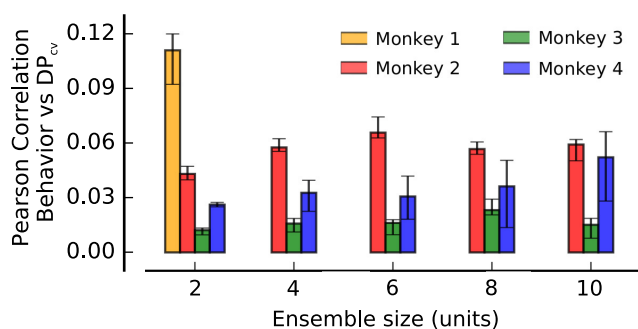


Figure 8. Amount of encoded information correlates with behavioral performance. Pearson correlation between fluctuations in encoded information (DP_{cv}) and fluctuations in monkeys' performance (see Materials and Methods). Across all datasets and ensemble sizes, trials associated with larger encoded information are also significantly associated with better task performance. Error bars indicate the 25th to 75th percentile of the distribution of bootstrap medians, and significant deviations from 0 (high-contrast colored bars) are calculated by a Wilcoxon signed rank test (not significant if $p > 0.05$).

to control for trivial dependencies (see Materials and Methods). We found that bootstrap fluctuations of PS and PP predicted significant modulations of behavioral performance across different datasets and ensemble sizes, whereas changes in behavioral performance produced by fluctuations in MPC and GA were either weak or inconsistent across different animals and ensemble sizes (Fig. 9). For example, bootstrap fluctuations of PS or PP, while keeping all other statistical features fixed, predicted significant changes in behavioral

performance for Monkey 1 (Fig. 9A; behavioral change predicted by PS alone: 1.59%, Wilcoxon signed-rank test, $p = 3.6 \times 10^{-14}$; PP alone: 0.53%, $p = 1.4 \times 10^{-3}$). However, the average change in behavioral performance predicted by bootstrap fluctuations of either MPC or GA alone were very small and not statistically significant for this animal (MPC: 0.06%, Wilcoxon signed-rank test, $p = 0.33$; GA: -0.03% , $p = 0.44$). We obtained qualitatively similar results for other monkeys and tasks (Fig. 9B–D; Table 9-2, available at <https://doi.org/10.1523/JNEUROSCI.0859-19.2019.t9-2>), except that bootstrap fluctuations of GA produced significant fluctuations in behavioral performance for Monkey 2, but not Monkey 3, in the attentional task. We also analyzed the data using Pearson's correlation coefficient in place of percent change in DP_{cv} and obtained qualitatively similar results (Fig. 9-1, available at <https://doi.org/10.1523/JNEUROSCI.0859-19.2019.f9-1>). These results were also robust across different conditioning thresholds (Fig. 9-1, available at <https://doi.org/10.1523/JNEUROSCI.0859-19.2019.f9-1>; see Materials and Methods) and for all pairs of statistical features (Table 9-3, available at <https://doi.org/10.1523/JNEUROSCI.0859-19.2019.t9-3>).

In summary, we find that PS and PP are the primary statistical features of neural activity that exhibit modulatory effects on behavioral performance, whereas MPC and GA show little or no consistent effects. These results are in line with those described earlier for the amount of encoded information (Fig. 6).

An experimentally constrained model accounts for the findings

To determine whether our experimental findings (Fig. 9) are consistent with an optimal readout of a neuronal population much larger than the observed ensembles, we developed a model constrained by the observed statistical properties of the neural responses (Fig. 10A; Fig. 10-1A, available at <https://doi.org/10.1523/JNEUROSCI.0859-19.2019.f10-1>; see Materials and Methods). The model simulates a number of trials (M trials in total) in which a stimulus ($s = \{+1, -1\}$) drives a large population consisting of $N = 1000$ neurons. Each neuron has a linear tuning curve with a slope ($\mathbf{m}_1$) drawn from a normal distribution. Neuronal correlations combined limited-range dependencies (Kanitscheider et al., 2015; Kohn et al., 2016), differential correlations (Moreno-Bote et al., 2014), and multiplicative and additive gains (Arandia-Romero et al., 2016). These correlations were generated through a sequence of steps as follows. For each trial j , a vector $\mathbf{r}'_j$ (size $N = 1000$ neurons) was drawn from a multivariate Gaussian with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}^{\text{gen}}$. The response vector $\mathbf{r}'_j$ was then corrupted with sensory noise, δs_j , which introduces differential correlations and limits the amount of information encoded in large populations of neurons (Moreno-Bote et al., 2014). Multiplicative and additive gains were then applied

(Arandia-Romero et al., 2016) (g_j), and a final Poisson step converted each response rate into an observed spike count n_j (Kanitscheider et al., 2015) (see Materials and Methods). Fano factors that approximately matched typical values found experimentally were used (Shadlen and Newsome, 1998; Arandia-Romero et al., 2016; Nogueira et al., 2018). Finally, based on the neuronal activity generated by the network, we generated surrogate choices (c_j) by optimally decoding the whole population of 1000 simulated neurons ($c_j = \text{sign}(\omega_{opt}^T n_j + \omega_0)$) on a trial-by-trial basis. The presence of differential correlations ensures that the amount of encoded information could not scale indefinitely with the number of neurons.

To compare with our experimental data, we randomly selected small ensembles (up to 10 model neurons) out of the full population (consisting of 1000 neurons) to test how fluctuations in PS and PP of the small ensembles correlate with encoded information and simulated behavior of the full network (Fig. 10-1, available at <https://doi.org/10.1523/JNEUROSCI.0859-19.2019.f10-1>). We repeated the analyses performed on the experimental data (Figs. 6, 9). Consistent with the experimental data, PS and PP computed from these small ensembles of model neurons are the only features that correlated significantly with DP_{cv} and surrogate behavioral performance, whereas MPC and GA did not (Fig. 10B,C). We also found that changes in the model's simulated behavioral performance associated with bootstrap fluctuations of PS and PP were quite small (Fig. 10C; ensemble size 2; surrogate behavioral change predicted by PS alone: 0.32%, Wilcoxon signed-rank test, $p = 2.12 \times 10^{-6}$; PP alone: 0.028%, $p = 0.049$) and approximately similar in magnitude to the experimental results (Fig. 9). As expected, when neuronal ensembles of up to 100 neurons were sampled from the full population, fluctuations of PS and PP had a larger effect on surrogate behavioral performance (ensemble size 100; surrogate behavioral change predicted by PS alone: 0.90%, Wilcoxon signed-rank test, $p = 1.73 \times 10^{-6}$; PP alone: 0.21%, $p = 4.38 \times 10^{-3}$). Finally, correlations between fluctuations of DP_{cv} and simulated behavioral performance were weak (Fig. 10-1G, available at <https://doi.org/10.1523/JNEUROSCI.0859-19.2019.f10-1>) and comparable with those observed in the experimental data (Fig. 7). In summary, the weak correlations that we observed experimentally between PS/PP and behavioral performance are broadly consistent with the weak dependencies between neuronal activity and choice observed using single-neuron analyses (Britten et al., 1996; Uka and DeAngelis, 2004; Nienborg and Cumming, 2009; Haefner et al., 2013), and

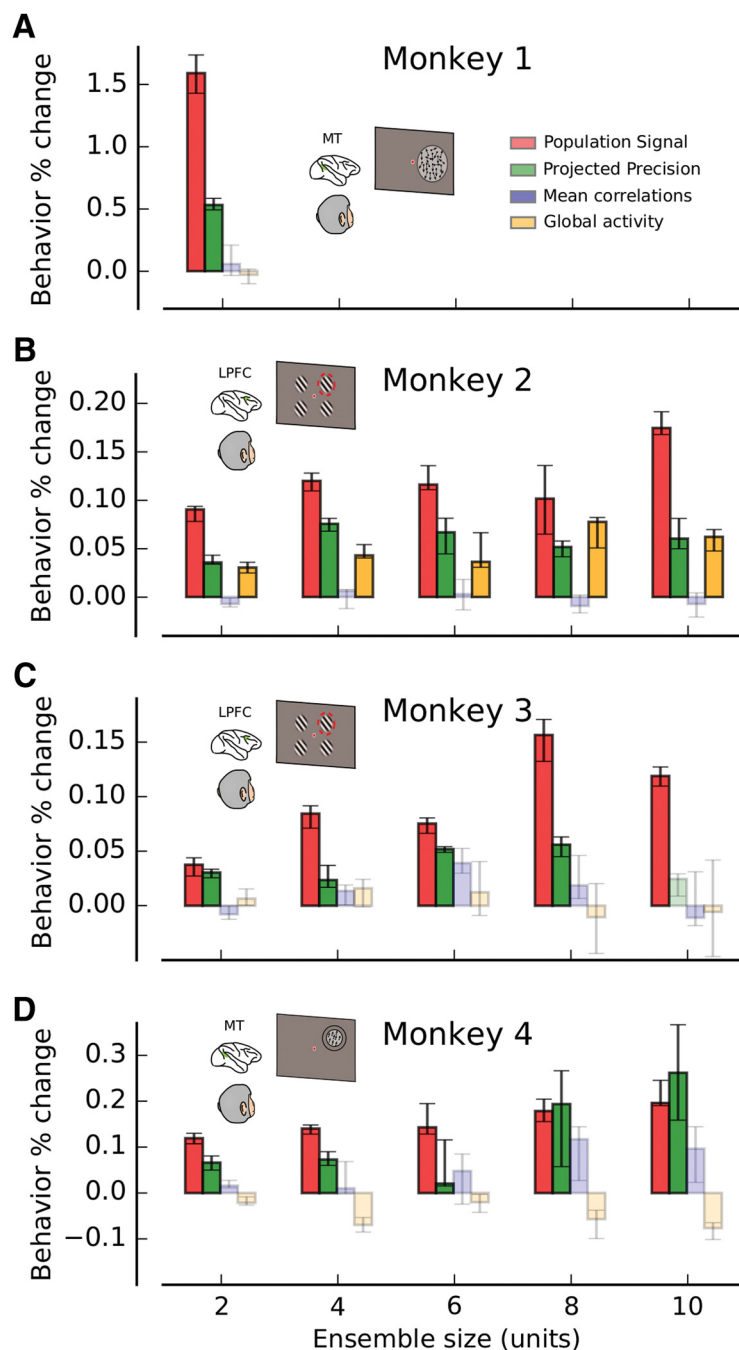


Figure 9. PS and PP best predict behavioral performance. **A**, Monkey 1. **B**, Monkey 2. **C**, Monkey 3. **D**, Monkey 4. Percent change in behavioral performance, as a function of ensemble size, is shown for each behavioral task when the conditioned bootstrapping approach is used to isolate fluctuations in PS (red), PP (green), MPC (blue), and GA (orange). See also Figure 9-1 (available at <https://doi.org/10.1523/JNEUROSCI.0859-19.2019.f9-1>), Table 9-2 (available at <https://doi.org/10.1523/JNEUROSCI.0859-19.2019.t9-2>), and Table 9-3 (available at <https://doi.org/10.1523/JNEUROSCI.0859-19.2019.t9-3>).

are also consistent with our hypothesis that behavior is generated by an optimal readout of a much larger population of neurons.

Discussion

Identifying the statistical features of neural population responses that determine the amount of encoded information and predict behavioral performance is essential for understanding the link between neuronal activity and behavior. Based on data collected from two brain areas and three behavioral tasks, we identified two critical features: the length of the vector joining the mean re-

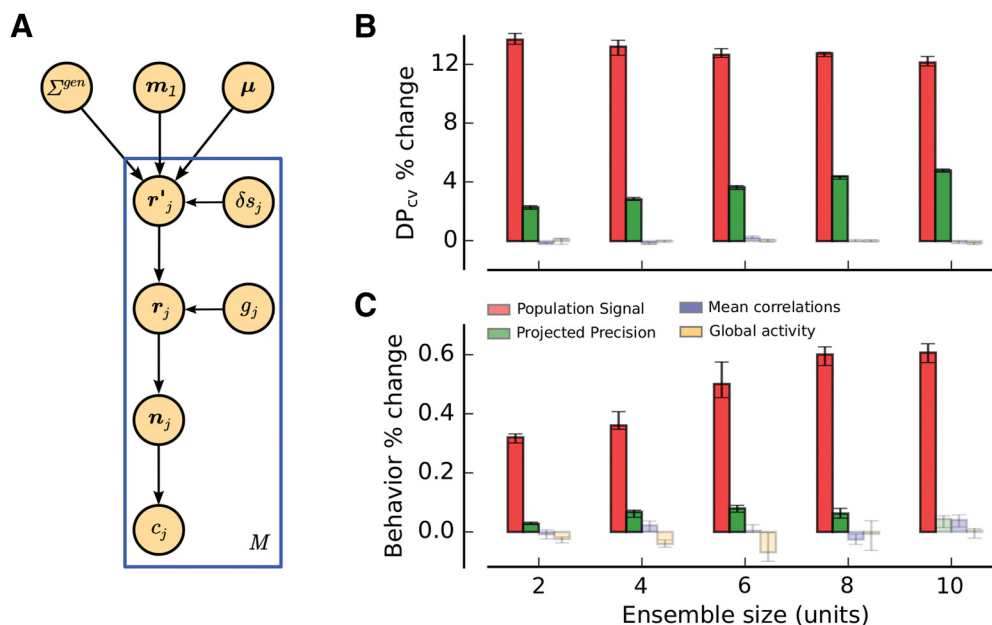


Figure 10. An experimentally constrained neural population model accounts for the empirical findings. **A**, Generative model for the simulated neuronal responses. A population of N model neurons was characterized by linear tuning curves with slope m_1 . An intermediate activity pattern (r'_j) was obtained by drawing an N -dimensional sample from a multivariate Gaussian distribution (mean μ and covariance Σ^{gen}) and then corrupting it with sensory noise (δs_j) on every trial j (r_j) (M trials in total). A homogeneous response gain modulation (g_j) and a Poisson step were applied to produce the final population spike count (n_j). The choice of the virtual agent (c_j) was obtained by an optimal readout of the population activity pattern on each trial (see Materials and Methods). **B**, Percent change in the amount of information encoded by the model (DP_{cv}), as a function of the ensemble size, when bootstrap fluctuations of different features of the neural activity are isolated. Only isolated bootstrap fluctuations in PS and PP influence the amount of information encoded by the population, consistent with the experimental observations (Fig. 6). **C**, Percent change in behavioral performance predicted by fluctuations in different statistical features of model population activity. PS and PP are the most important factors influencing behavioral performance. The magnitude of changes in behavioral performance is similar to that observed experimentally (Fig. 9). See also Figure 10-1 (available at <https://doi.org/10.1523/JNEUROSCI.0859-19.2019.f10-1>).

sponses of the population of neurons across two conditions to be distinguished (PS), and the inverse of the population covariability projected onto the direction of that vector (PP). We found that changes in PS and PP are significantly correlated with the amount of encoded information and behavioral performance, but MPC and GA are not, once PS and PP are held fixed. Our experimental results are consistent with predictions of a neuronal population model with realistic neuronal tuning, variability, and correlations that is readout optimally to generate behavioral choices.

Although previous studies have examined how input signals are represented in the average activity of neurons (Hubel and Wiesel, 1959; Mountcastle et al., 1967; Renart and van Rossum, 2012), described the types of neuronal correlations that can benefit or harm encoded information (Zohary et al., 1994; Abbott and Dayan, 1999; Averbeck et al., 2006; Ecker et al., 2011; Moreno-Bote et al., 2014; Verhoef and Maunsell, 2017), and cautioned about directly linking MPC with information (Abbott and Dayan, 1999; Averbeck et al., 2006; Moreno-Bote et al., 2014), they have not clearly separated the primary features of neuronal population activity that constrain the amount of encoded information and affect behavioral performance. As shown above, the traditional interpretation of neuronal population sensitivity, d' , considers the signal and noise components of the classifier's decision variable. However, contrary to this classic interpretation, our framework separates the contributions of population tuning (first-order statistic) and trial-by-trial variability (second-order statistic) of the neuronal responses to the amount of encoded information. Separating these two contributions of d' is important for several reasons. First, as we have shown, the effects of statistical features, such as MPC and GA, on the amount of encoded information and behavioral performance can be under-

stood in terms of the contributions by the first-order (PS) and second-order (PP) statistical components. This suggests that PS and PP are fundamental statistical features that underlie many other statistical features that affect the amount of information encoded and behavioral performance. Second, when addressing how learning affects neuronal coding, for example, it is important to understand not just how d' changes during learning but also how learning alters the tuning curves of individual neurons (i.e., PS) and the coordination among neurons along relevant task axis (i.e., PP). Therefore, distinguishing the first- and second-order contributions may help us understand circuit-level models of learning.

Previous studies have also identified the exact pattern of noise correlations that limit information for very large neuronal populations (Moreno-Bote et al., 2014; Kanitscheider et al., 2015), but as we have shown, the precision of the neuronal code for finite neuronal populations depends only partially on differential correlations, which act by reducing PP. Thus, one important contribution of our study is identifying PS and PP as the critical features of neuronal activity that modulate the amount of encoded information and predict behavioral performance for finite populations sizes.

Another contribution of this study involves demonstrating that other features of neuronal activity, namely, MPC and GA, have little or no impact on the amount of encoded information and behavioral performance once PS and PP are controlled for (Zohary et al., 1994; Shadlen and Newsome, 1998; McAdams and Maunsell, 1999; Kanitscheider et al., 2015; Lin et al., 2015; Arandia-Romero et al., 2016; Ecker et al., 2016; Gutnisky et al., 2017; Verhoef and Maunsell, 2017). We developed a novel approach for studying correlations among multiple variables while eliminating the effects of other variables (conditioned bootstrapping method). This method uses bootstrapping to generate con-

tinuous distributions of multiple statistical features so that fluctuations of some features can be computed for a subset of bootstrap samples that yielded no fluctuations in other features. This method allowed us to isolate the effects of a particular feature on both the amount of encoded information and behavioral performance. Thus, our approach clearly differs from the traditional trial shuffling method that destroys all dependencies in the data to examine the effects of correlations (Romo et al., 2003; Kohn and Smith, 2005; Tremblay et al., 2015; Leavitt et al., 2017). It also differs from maximum entropy models that generate surrogate data while fixing the first and second moments of the generated distribution to desired values (Elsayed and Cunningham, 2017).

Previous studies have suggested that MPC and GA affect behavioral performance (Cohen and Newsome, 2008; Cohen and Maunsell, 2009; Mitchell et al., 2009; Gu et al., 2011; Ni et al., 2018). On the surface, these studies appear to be at odds with our main finding that only PS and PP should affect the amount of encoded information or behavioral performance. Indeed, we find that bootstrap fluctuations of MPC and GA have little effect on the amount of encoded information or behavioral performance when PS and PP are kept constant. Our results suggest that the previous studies found effects of MPC and GA on behavioral performance because these variables are themselves correlated with PS and PP (Fig. 5). Indeed, there may be many other statistical features of neuronal responses that seemingly influence the amount of encoded information and behavioral performance, but such relationships could be explained as a byproduct of correlations of these statistical features with PS and PP.

Although theoretical research has proposed that fluctuations of GA can modulate the amount of encoded information (Kanitcheider et al., 2015; Lin et al., 2015; Ecker et al., 2016), a recent study based on large neuronal populations in monkey primary visual cortex found that the amount of information does not vary with large fluctuations in GA (Arandia-Romero et al., 2016). Another recent study found a modest, but significant, increase in DP when population activity decreased (Gutnisky et al., 2017), which appears to be inconsistent with our theoretical prediction. Again, however, some of the previous reports of positive effects of GA may have arisen from correlations between fluctuations in GA and PS. Indeed, we have found a highly significant correlation between PS and GA. Therefore, an important question is whether there is a residual effect of GA on the amount of encoded information in the previous studies after the contributions of PS and PP are eliminated. Our theory predicts that there should not be, and the conditioned bootstrapping approach that we have developed would be useful for teasing apart the statistical features (e.g., PS and PP) that fundamentally affect the amount of encoded information and behavioral performance from heuristic parameters, such as GA, which may have only secondary effects on information or behavior.

In conclusion, based on information metrics that are readily applicable to neuronal data, we developed a theory-driven analysis that has identified the statistical features of neural population responses that modulate the amount of information encoded by cortical neuronal populations. We found an excellent agreement between the theory and the experimental results, which indicates that the assumptions of our approach are reasonable. A critical finding is that PS and PP are the primary features that correlate with behavioral performance and that one must be careful when interpreting effects of other statistical features that may covary with PS or PP. Finally, we showed that our results are consistent

with predictions of a model that optimally decodes population responses to produce behavior.

References

- Abbott LF, Dayan P (1999) The effect of correlated variability on the accuracy of a population code. *Neural Comput* 11:91–101.
- Arandia-Romero I, Tanabe S, Drugowitsch J, Kohn A, Moreno-Bote R (2016) Multiplicative and additive modulation of neuronal tuning with population activity affects encoded information. *Neuron* 89:1305–1316.
- Arandia-Romero I, Nogueira R, Mochol G, Moreno-Bote R (2017) What can neuronal populations tell us about cognition? *Curr Opin Neurobiol* 46:48–57.
- Averbeck BB, Lee D (2006) Effects of noise correlations on information encoding and decoding. *J Neurophysiol* 95:3633–3644.
- Averbeck BB, Latham PE, Pouget A (2006) Neural correlations, population coding and computation. *Nat Rev Neurosci* 7:358–366.
- Britten KH, Shadlen MN, Newsome WT, Movshon JA (1992) The analysis of visual motion: a comparison of neuronal and psychophysical performance. *J Neurosci* 12:4745–4765.
- Britten KH, Newsome WT, Shadlen MN, Celebrini S, Movshon JA (1996) A relationship between behavioral choice and the visual responses of neurons in macaque MT. *Vis Neurosci* 13:87–100.
- Bullock KR, Pieper F, Sachs AJ, Martinez-Trujillo JC (2017) Visual and pre-saccadic activity in area 8A of the macaque monkey lateral prefrontal cortex. *J Neurophysiol* 118:15–28.
- Chen Y, Geisler WS, Seidemann E (2006) Optimal decoding of correlated neural population responses in the primate visual cortex. *Nat Neurosci* 9:1412–1420.
- Cohen MR, Kohn A (2011) Measuring and interpreting neuronal correlations. *Nat Neurosci* 14:811–819.
- Cohen MR, Maunsell JH (2009) Attention improves performance primarily by reducing interneuronal correlations. *Nat Neurosci* 12:1594–1600.
- Cohen MR, Newsome WT (2008) Context-dependent changes in functional circuitry in visual area MT. *Neuron* 60:162–173.
- DeAngelis GC, Uka T (2003) Coding of horizontal disparity and velocity by MT neurons in the alert macaque. *J Neurophysiol* 89:1094–1111.
- Ecker AS, Berens P, Keliris GA, Bethge M, Logothetis NK, Tolias AS (2010) Decorrelated neuronal firing in cortical microcircuits. *Science* 327:584–587.
- Ecker AS, Berens P, Tolias AS, Bethge M (2011) The effect of noise correlations in populations of diversely tuned neurons. *J Neurosci* 31:14272–14283.
- Ecker AS, Berens P, Cotton RJ, Subramaniam M, Denfield GH, Cadwell CR, Smirnakis SM, Bethge M, Tolias AS (2014) State dependence of noise correlations in macaque primary visual cortex. *Neuron* 82:235–248.
- Ecker AS, Denfield GH, Bethge M, Tolias AS (2016) On the structure of neuronal population activity under fluctuations in attentional state. *J Neurosci* 36:1775–1789.
- Elsayed GF, Cunningham JP (2017) Structure in neural population recordings: an expected byproduct of simpler phenomena? *Nat Neurosci* 20:1310–1318.
- Fisher RA (1936) The use of multiple measurements in taxonomic problems. *Ann Hum Genet* 7:179–188.
- Goris RL, Movshon JA, Simoncelli EP (2014) Partitioning neuronal variability. *Nat Neurosci* 17:858–865.
- Gu Y, Watkins PV, Angelaki DE, DeAngelis GC (2006) Visual and nonvisual contributions to three-dimensional heading selectivity in the medial superior temporal area. *J Neurosci* 26:73–85.
- Gu Y, Angelaki DE, DeAngelis GC (2008) Neural correlates of multisensory cue integration in macaque MSTd. *Nat Neurosci* 11:1201–1210.
- Gu Y, Liu S, Fetsch CR, Yang Y, Fok S, Sunkara A, DeAngelis GC, Angelaki DE (2011) Perceptual learning reduces interneuronal correlations in macaque visual cortex. *Neuron* 71:750–761.
- Gutnisky DA, Beaman C, Lew SE, Dragoi V (2017) Cortical response states for enhanced sensory discrimination. *Elife* 6:e29226.
- Haefner RM, Gerwinn S, Macke JH, Bethge M (2013) Inferring decoding strategies from choice probabilities in the presence of correlated variability. *Nat Neurosci* 16:235–242.
- Harris KD, Thiele A (2011) Cortical state and attention. *Nat Rev Neurosci* 12:509–523.
- Hubel DH, Wiesel TN (1959) Receptive fields of single neurones in the cat's striate cortex. *J Physiol* 148:574–591.
- Kanitcheider I, Coen-Cagli R, Pouget A (2015) Origin of information-limiting noise correlations. *Proc Natl Acad Sci U S A* 112:E6973–E6982.

- Kohn A, Smith MA (2005) Stimulus dependence of neuronal correlation in primary visual cortex of the macaque. *J Neurosci* 25:3661–3673.
- Kohn A, Coen-Cagli R, Kanitscheider I, Pouget A (2016) Correlations and neuronal population information. *Annu Rev Neurosci* 39:237–256.
- Leavitt ML, Pieper F, Sachs AJ, Martinez-Trujillo JC (2017) Correlated variability modifies working memory fidelity in primate prefrontal neuronal ensembles. *Proc Natl Acad Sci U S A* 114:E2494–E2503.
- Lin IC, Okun M, Carandini M, Harris KD (2015) The nature of shared cortical variability. *Neuron* 87:644–656.
- Luczak A, Bartho P, Harris KD (2013) Gating of sensory input by spontaneous cortical activity. *J Neurosci* 33:1684–1695.
- McAdams CJ, Maunsell JH (1999) Effects of attention on orientation-tuning functions of single neurons in macaque cortical area V4. *J Neurosci* 19:431–441.
- Mitchell JF, Sundberg KA, Reynolds JH (2009) Spatial attention decorrelates intrinsic activity fluctuations in macaque area V4. *Neuron* 63:879–888.
- Moreno-Bote R, Beck J, Kanitscheider I, Pitkow X, Latham P, Pouget A (2014) Information-limiting correlations. *Nat Neurosci* 17:1410–1417.
- Mountcastle VB, Talbot WH, Darian-Smith I, Kornhuber HH (1967) Neural basis of the sense of flutter-vibration. *Science* 155:597–600.
- Ni AM, Ruff DA, Alberts JJ, Symmonds J, Cohen MR (2018) Learning and attention reveal a general relationship between population activity and behavior. *Science* 359:463–465.
- Nienborg H, Cumming BG (2009) Decision-related activity in sensory neurons reflects more than a neurons causal effect. *Nature* 459:89–92.
- Nogueira R, Lawrie S, Moreno-Bote R (2018) Neuronal variability as a proxy for network state. *Trends Neurosci* 41:170–173.
- Panzeri S, Harvey CD, Piasini E, Latham PE, Fellin T (2017) Cracking the neural code for sensory perception by combining statistics, intervention, and behavior. *Neuron* 93:491–507.
- Pitkow X, Liu S, Angelaki DE, DeAngelis GC, Pouget A (2015) How can single sensory neurons predict behavior? *Neuron* 87:411–423.
- Renart A, van Rossum MC (2012) Transmission of population-coded information. *Neural Comput* 24:391–407.
- Renart A, de La Rocha J, Bartho P, Hollender L, Parga N, Reyes A, Harris KD (2010) The asynchronous state in cortical circuits. *Science* 327:587–590.
- Romo R, Hernández A, Zainos A, Salinas E (2003) Correlated neuronal discharges that increase coding efficiency during perceptual discrimination. *Neuron* 38:649–657.
- Schölvinck ML, Saleem AB, Benucci A, Harris KD, Carandini M (2015) Cortical state determines global variability and correlations in visual cortex. *J Neurosci* 35:170–178.
- Seung HS, Sompolinsky H (1993) Simple models for reading neuronal population codes. *Proc Natl Acad Sci U S A* 90:10749–10753.
- Shadlen MN, Newsome WT (1998) The variable discharge of cortical neurons: implications for connectivity, computation, and information coding. *J Neurosci* 18:3870–3896.
- Tremblay S, Pieper F, Sachs A, Martinez-Trujillo J (2015) Attentional filtering of visual information by neuronal ensembles in the primate lateral prefrontal cortex. *Neuron* 85:202–215.
- Uka T, DeAngelis GC (2004) Contribution of area MT to stereoscopic depth perception: choice-related response modulations reflect task strategy. *Neuron* 42:297–310.
- Verhoef BE, Maunsell JH (2017) Attention-related changes in correlated neuronal activity arise from normalization mechanisms. *Nat Neurosci* 20:969–977.
- Zohary E, Shadlen MN, Newsome WT (1994) Correlated neuronal discharge rate and its implications for psychophysical performance. *Nature* 370:140–143.