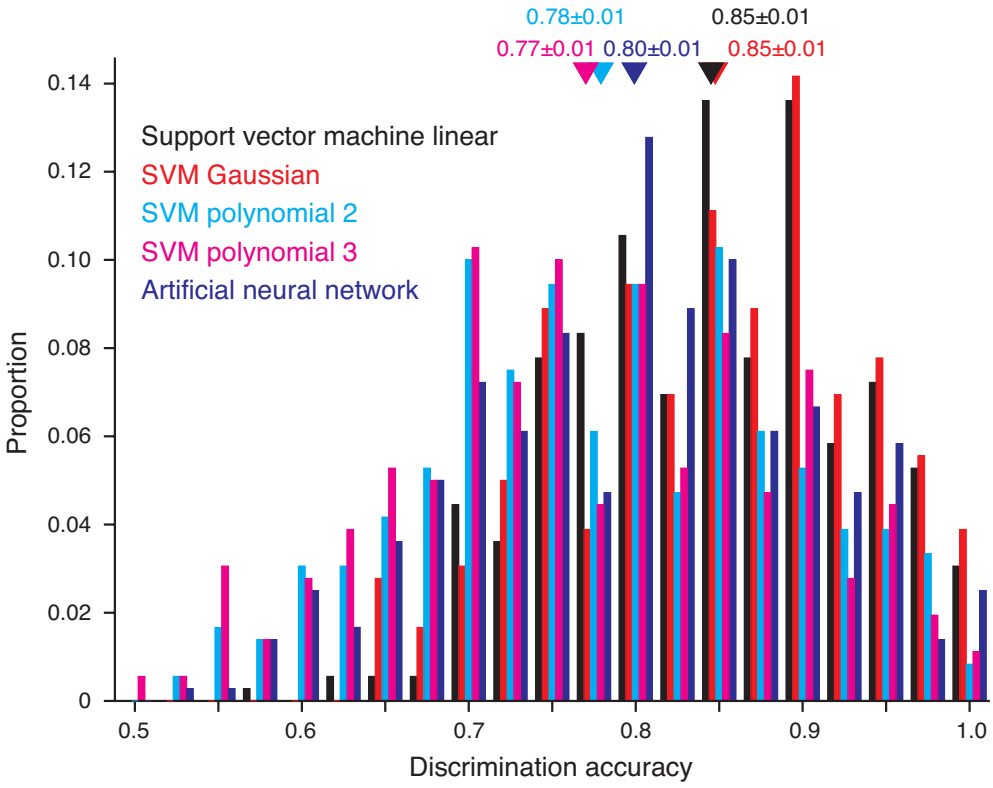
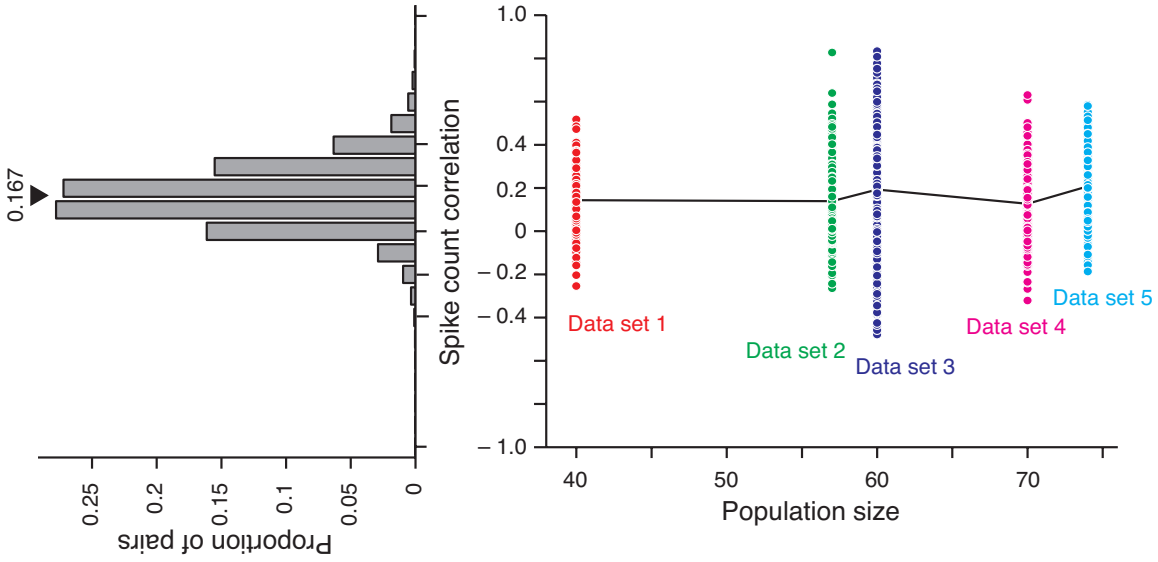
# Decoding the activity of neuronal populations in macaque primary visual cortex: Supplementary materials

Arnulf B A Graf, Adam Kohn, Mehrdad Jazayeri and J Anthony Movshon

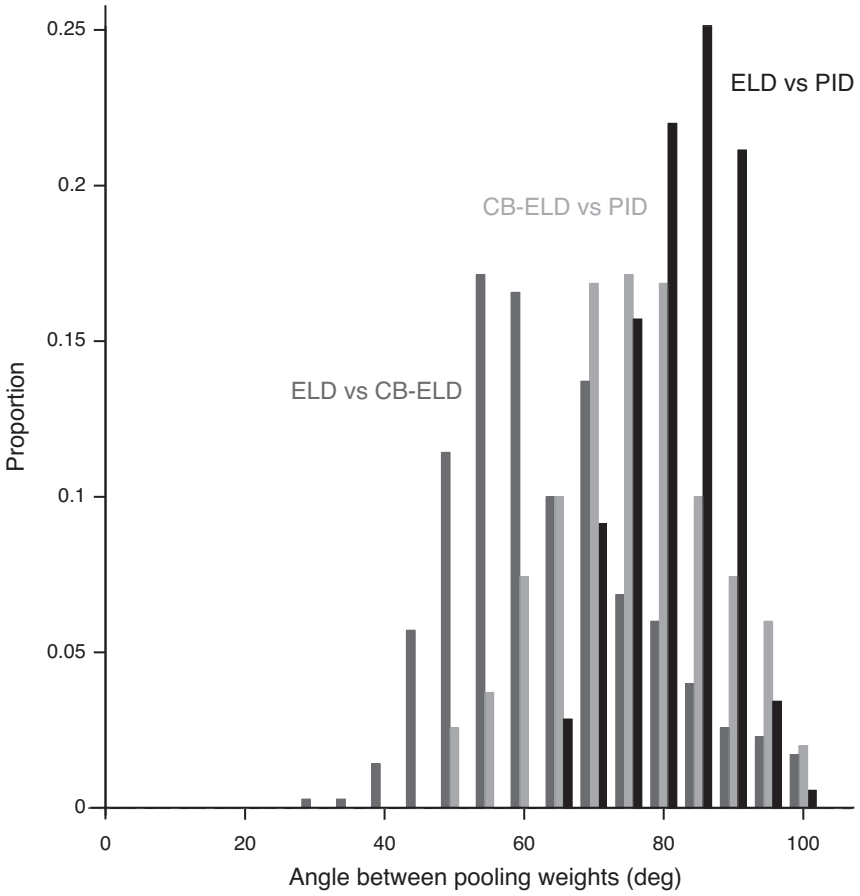



**Supplementary Figure 1.** Distribution of discrimination accuracies (10-fold cross-validated) between neuronal responses corresponding to neighboring orientations (separated by 5 deg) across 72 orientations and 5 data sets. The response of a neuronal population to a stimulus orientation is a point in the high-dimensional neuronal response space, each dimension representing the spike count of one neuron. Repetitions of the same stimulus (50 trials) create a cluster of dots, yielding a total of 72 clusters of 50 population responses. Discrimination of neuronal responses is equivalent to finding a

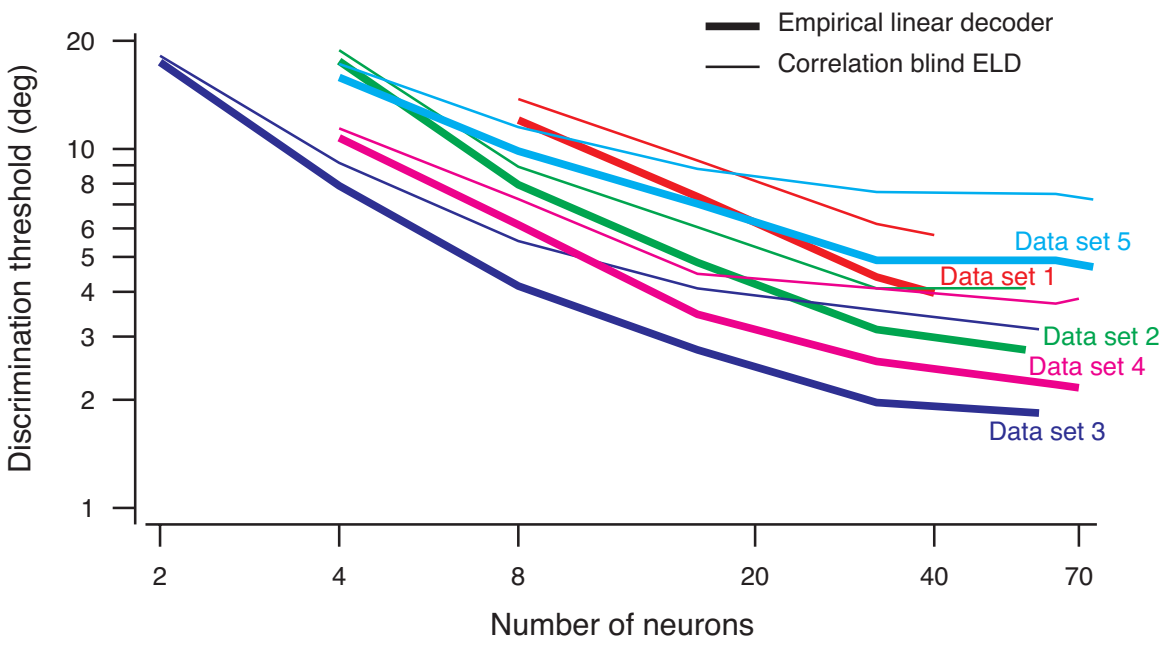
measure that best separates population responses corresponding to neighboring orientations. We used the Support Vector Machine (SVM, see Methods) because it allowed us to use the same algorithm for linear and nonlinear discriminations (Vapnik 2000; Schölkopf and Smola 2002): the SVM variants only differed by how they measured the interactions between neuronal responses. Comparing SVM variants tells us how well the data can be linearly discriminated, independently of the discrimination algorithm. The linear SVM used the Euclidean dot product as kernel function: $K(x,y) = x^T y$. For the nonlinear SVM, we chose polynomial kernel functions $K(x,y) = (1 + x^T y)^p$ of second and third degrees ( $p = 2 \; or \; 3$) because these kernels represent the next degrees of complexity above the linear one. We also evaluated a Gaussian kernel function $K(x,y) = \exp(-\gamma \|x - y\|^2)$ with $\gamma = \left[ mean(\|x - y\|) \right]^{-2}$ as this kernel preserves vicinities between neuronal responses. We found that the nonlinear SVM variants, despite their higher complexity, either performed similarly (Gaussian kernel) or significantly worse (polynomial kernels) than the linear SVM. For the polynomial kernels, a quadratic or cubic nonlinearity was forced on the data: the SVM could not choose to ignore the nonlinear terms imposed by the kernel function, akin to fitting a quadratic function through linear data. The low discrimination accuracy encountered for the polynomial kernels is a strong indicator of the linear separability of our data. Although there is an infinite family of nonlinear kernel functions, our results show that any improvement in decoding accuracy over linear discrimination will likely be modest at best. Finally, we tested a member from another family of nonlinear discriminators: the artificial neuronal network (ANN). The ANN was composed of an input layer (N units, where N is the neuronal population size), a hidden layer of 5 units and one output unit. The sign of the output unit defined which of 2 stimulus classes the neuronal response belonged to. The ANN consisted of sigmoidal units and was trained using batch backpropagation. We found that the linear SVM outperformed the ANN. In summary, our findings show that neuronal responses are best separated linearly, and that the assumption of linearity is neither restrictive nor inaccurate. As a consequence, discriminations between neuronal responses can be modeled by a log-likelihood ratio that is linear in the neuronal population response, yielding also a linear log-likelihood function. For the purpose of population decoding, neuronal responses can thus be considered as belonging to the exponential family with linear sufficient statistics (Ma, Beck et al. 2006; Beck, Ma et al. 2008).




**Supplementary Figure 2.** Distribution of spike count correlation coefficients. The spike count correlation coefficient represents the trial-to-trial variability of the response of a neuronal pair across stimulus orientations and trials. The total density of pairwise spike count correlation coefficients, computed across all neuronal pairs in the 5 data sets, showed that most pairs of neurons had correlated trial-to-trial variability (left panel). Also, the spike count correlation coefficients did not differ significantly between the data sets (right panel, each dot represents the coefficient of neuronal pair and the line connects the mean coefficient of each data set).

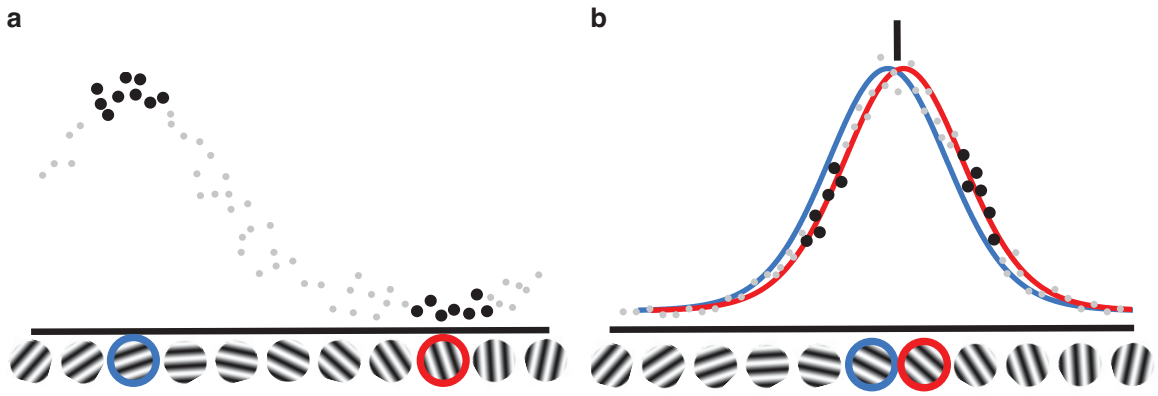



**Supplementary Figure 3.** Distribution of angles between the pooling weights (***W*** in eq. 1, with as many components as neurons in the population) of the Empirical Linear Decoder (ELD), the correlation-blind ELD (CB-ELD) and the Poisson Independent Decoder (PID) across 72 orientations and 5 data sets. The distributions were significantly different from zero, showing that all decoders relied on different internal pooling mechanisms. The weights of the ELD and CB-ELD were most similar, reflecting that both decoders were derived from the same computational framework applied to different data structures (data with or without correlations). The weights of both the ELD and CB-ELD differed strongly from those of the PID, illustrating that the PID and ELD had different internal neuronal pooling mechanisms.

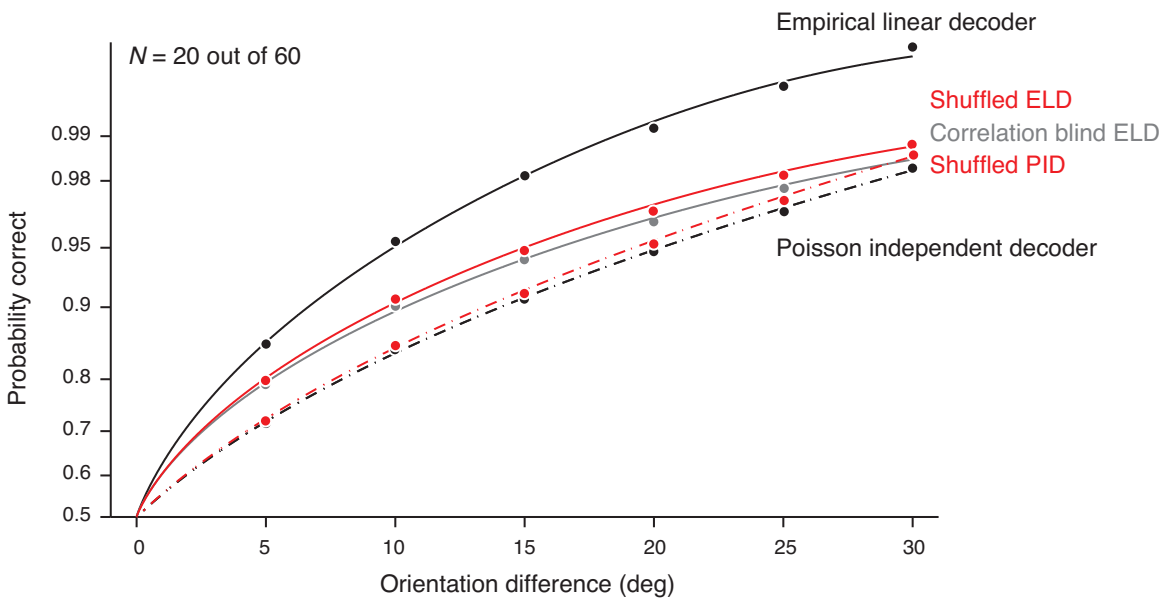



**Supplementary Figure 4.** Orientation discrimination thresholds of the Empirical Linear Decoder (ELD) and the correlation-blind ELD (CB-ELD) with respect to the neuronal population size. The discrimination threshold to reach an accuracy of 0.75 was computed from the population neurometric function based on discrimination accuracies averaged across random neuronal samples of a given size (covering number $n_c = 2$, see Methods). The thresholds of the CB-ELD were always higher than those of the ELD, suggesting that ignoring correlations hurts an empirical linear decoding scheme. Furthermore, the ELD mainly used pairwise correlations (covariance of the responses, or second order moments) because the impact of correlations on the discrimination thresholds was similar across population sizes. Also, the discrimination thresholds reached an asymptotic regime well before reaching the full population size, suggesting that small populations of $\sim 10$ neurons yielded a decoding accuracy relatively close to its asymptotic value for a given data set. This finding shows that the decoding analyses presented in the main document, which were either based on entire populations or large subsets ($N \geq 10$), reflect to a full extent the decoding accuracy that can be obtained from our neuronal samples. Finally, for a given population sample size, the discrimination

thresholds varied significantly across the data sets. This reflects the heterogeneity in neuronal tuning properties (amplitude, bandwidth, preferred orientation, baseline) and response variability. In particular, we suppose that the difference between the data sets is dominated by varying degrees of coverage in orientation of the underlying individual tuning curves, e.g. a small population of neurons with preferred orientations tiling (0,360) will yield a better decoding accuracy across all orientations than a larger population with neurons concentrated in a small range of orientations (see also Discussion).

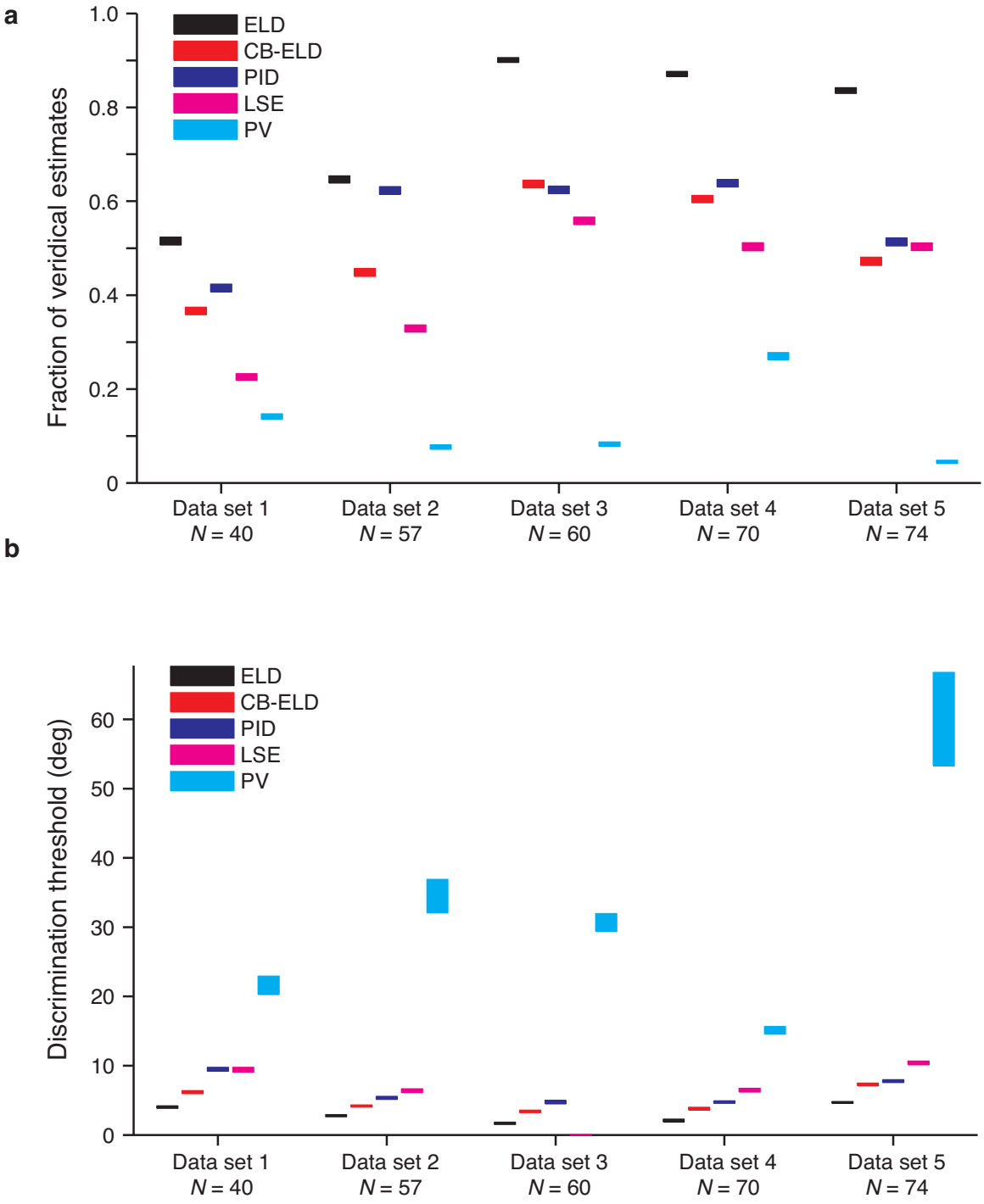



**Supplementary Figure 5.** Illustration of a coarse (a) and fine (b) discrimination using the neuronal population activity. The neurons are aligned with respect to their preferred orientation (gratings). The dots denote the population activity for a given stimulus orientation. We here illustrate discrimination between two orientations (indicated by the colored circles) on the basis of the population activity they elicit. (a) When the “blue” orientation is presented to a neuronal population, neurons with preferred orientations close to the “blue” orientation have the strongest response, and neurons at the “red” orientation respond much less vigorously. The opposite is true the “red” orientation is presented (not shown). Accurate discrimination between distant orientations can then be based on the location of the peaks of the population activity corresponding to the two orientations. The neurons with preferred orientations at the discriminanda are thus most informative. (b) For fine discriminations the peaks corresponding to the population activity elicited by each orientation are virtually superimposed, as shown by the average population activity (colored bell-shaped colored curves). In this case the peaks of the population activity are less informative for discrimination. However, the “flanks” of the population activity allow us to discriminate best between the two orientations: the population activity indicated by the dots here corresponds to the “red” orientation. The neurons with preferred orientations further apart from the discriminanda (“red” and “blue” orientations) are thus most useful for discriminating between nearby orientations.



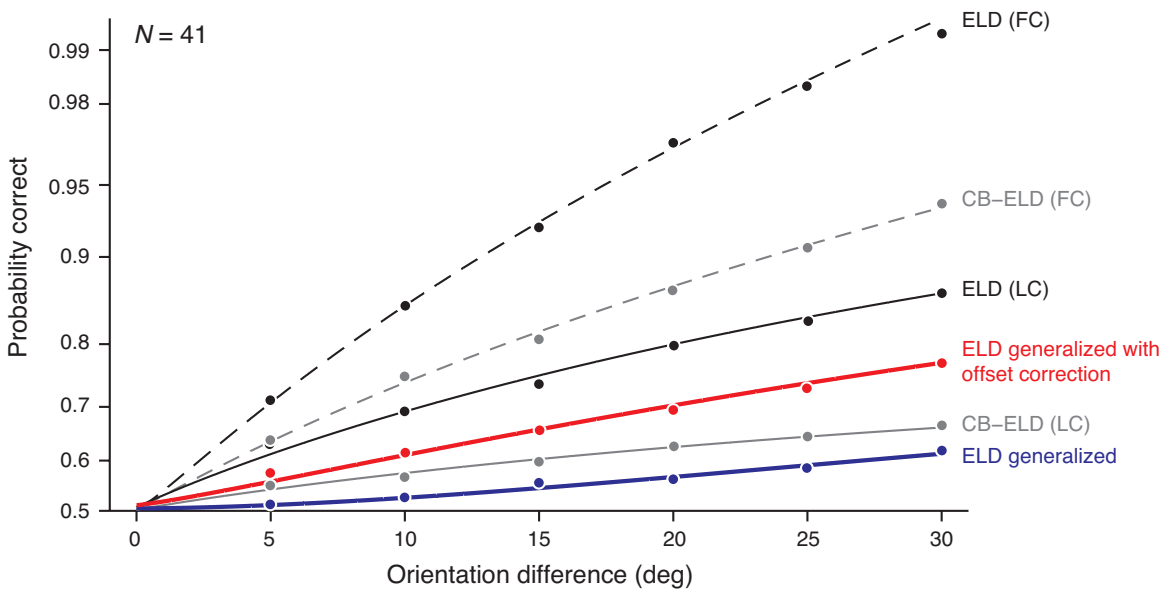

**Supplementary Figure 6.** Population neurometric functions of the Empirical Linear Decoder (ELD), the correlation-blind ELD (CB-ELD), the shuffled ELD, the Poisson Independent Decoder (PID), and the shuffled PID for the neuronal data of Fig. 4a. The CB-ELD is trained on shuffled data and tested on unshuffled (raw) data. Comparing the CB-ELD to the full ELD addresses how correlations affect the computations for decoding (Averbeck, Latham et al. 2006) by quantifying how much information is lost when (raw) neuronal responses are decoded using an algorithm that ignores correlations. This so-called diagonal case illustrates whether a downstream neuron has to know about correlations to increase the information available for decoding. Because neuronal responses are correlated (raw data), the diagonal case assesses how much decoding accuracy is lost by ignoring correlations. The CB-ELD performed worse than the ELD, showing that ignoring correlations hurts the decoding accuracy for an empirical read-out rule. In the shuffled scenario, the ELD is trained *and* tested on the shuffled data (Averbeck, Latham et al. 2006). This is an ideal case for the PID because the PID assumes neuronal independence. As such, the shuffled PID yielded more accurate decoding than the raw PID. On the other hand, the shuffled ELD shows how correlations affect the total amount of information in population codes by investigating how well a decoder generalizes to a different data structure. The shuffled ELD outperfomed the CB-

ELD because the shuffled ELD is optimized for the testing data as it is trained *and* tested on the same data structure (this is not the case in the diagonal case). Furthermore, the shuffled ELD yielded worse decoding accuracy than the raw ELD. This suggests that correlated data is more easily separable than uncorrelated data. Correlations can thus help an empirical decoder to increase its accuracy. The shuffled case is, however, less relevant when studying the computations in the brain because the neuronal data are correlated, and hence the data to be decoded (testing data) always contains the correlated variability between neurons.




**Supplementary Figure 7.** Comparison of the performance of decoders belonging to the linear log-likelihood framework and point estimators. The estimation and discrimination accuracies of the decoders belonging to the linear log-likelihood framework, which were

investigated in the main document, are shown in panels (a) and (b) respectively. We compared their decoding accuracies to two point estimators that directly inferred single estimates from the neuronal population response by linearly weighting them, without using an intermediate representation such as the log-likelihood function. The point estimators establish a direct correspondence between the neuronal response and the outcome of the estimation. In the population vector (PV) (Georgopoulos, Schwartz et al. 1986), each neuron votes for its preferred direction proportionaly to its response. The orientation estimate corresponding to a given population activity is derived from the sum of the votes corresponding to each neuron. We computed the discrimination accuracy using the position of this estimate with respect to the discrimination boundary (average of the angle between both alternatives). The poor decoding accuracy of the PV can be attributed to two factors: the PV is known (Pouget, Zhang et al. 1998) to be a suboptimal decoder when the neuronal response distributions are not Gaussian (we showed that our responses were Poisson-like, see Supplementary Fig. 1) and when the orientation tuning curves are not cosine functions (see tuning curves in Fig. 1). While the PV used the neuronal responses to scale a set of basis vectors, the Least Squares Estimator (LSE) (Salinas and Abbott 1994; Pouget, Zhang et al. 1998) improves the PV by optimizing these weights to minimize the squared error of the estimates. Although the LSE outperformed the simple PV, both decoders yielded decoding accuracies significantly below those obtained from the decoders belonging to the linear log-likelihood framework. In summary, the linear log-likelihood framework yielded more accurate estimates than the point estimators. This finding makes a strong argument in favor of decoding using a re-encoded version of the neuronal responses, e.g. the log-likelihood function, instead of directly decoding neuronal responses.




**Supplementary Figure 8.** Population neurometric functions of the Empirical Linear Decoder (ELD) and the correlation-blind ELD (CB-ELD) for a data set of 41 neurons recorded at full contrast (FC) and at a lower contrast of 0.1 (LC). First, we computed the discrimination accuracy of the ELD and CB-ELD on the full contrast data (black and gray dotted lines) and on the low contrast data (black and gray full lines). We found that neuronal population responses were more accurately discriminated when they were elicited by full contrast stimuli rather than by low contrast stimuli (dotted versus full lines). This suggests that for the low contrast data, the neuronal response variability (e.g. spike count variance) occluded some of the response difference (e.g. between the mean spike count), thus hampering discrimination. Furthermore, we found that the spike count correlations were higher on the low contrast data ( $r_{sc} = 0.53$ ) than on the full contrast data ( $r_{sc} = 0.23$ ), in agreement with another study (Kohn and Smith 2005). The impact of correlations on the discrimination accuracy was also stronger on the low contrast data. The cost of ignoring interneuronal correlations for decoding thus scaled with their magnitude. Second, we addressed the question of contrast invariance by using the parameters (discrimination weights and offset) computed at full contrast to decode responses elicited by low contrast stimuli. We found that the discrimination accuracy of

this generalized ELD decreased dramatically (black versus blue lines), reflecting that the linear log-likelihood framework is not contrast invariant *per se*. The linear term (the combination of the pooling weights with the neuronal responses) can scale with lower responses. The stimulus-dependent offset remains, however, unchanged and is thus out of scale for the low contrast data. In order to overcome this problem, we adjusted the offset to minimize the discrimination error on the low contrast data (red line). This allowed us to take the offset out of the decoding problem and analyze how well the linear term scales with contrast. The offset-corrected generalized ELD yielded significantly better discrimination accuracy on low contrast data than its non-corrected variant (red versus blue lines). However, the discrimination performance of the ELD trained on the low contrast data was still superior to the offset-corrected generalized ELD (black versus red lines). We suppose that this difference is attributed to a subset of neurons pooled by the ELD: these neurons were vigorously driven by high contrast stimuli and were not responsive to low contrast stimuli, thus hampering discrimination accuracy. In summary, our results suggest that the linear log-likelihood framework is close to contrast invariant when the offset term is corrected for each stimulus contrast.

References


Averbeck, B. B., P. E. Latham, et al. (2006). "Neural correlations, population coding and computation." Nat Rev Neurosci **7**(5): 358-66.

Beck, J. M., W. J. Ma, et al. (2008). "Probabilistic population codes for Bayesian decision making." Neuron **60**(6): 1142-52.

Georgopoulos, A. P., A. B. Schwartz, et al. (1986). "Neuronal population coding of movement direction." Science **233**(4771): 1416-9.

Kohn, A. and M. A. Smith (2005). "Stimulus dependence of neuronal correlation in primary visual cortex of the macaque." J Neurosci **25**(14): 3661-73.

Ma, W. J., J. M. Beck, et al. (2006). "Bayesian inference with probabilistic population codes." Nat Neurosci **9**(11): 1432-8.

Pouget, A., K. Zhang, et al. (1998). "Statistically efficient estimation using population coding." Neural Comput **10**(2): 373-401.

Salinas, E. and L. F. Abbott (1994). "Vector reconstruction from firing rates." J Comput Neurosci **1**(1-2): 89-107.

Schölkopf, B. and A. Smola (2002). Learning with Kernels, MIT Press.

Vapnik, V. (2000). The Nature of Statistical Learning Theory, Springer.