Supporting Information

Tripathy et al. 10.1073/pnas.1221214110

SI Materials and Methods

Neuron Recordings. Whole-cell patch clamp recordings of mitral cells were obtained in vitro from mouse olfactory bulb slices using methods described previously (1). Mitral cells were identified under infrared differential interference contrast optics on the basis of their laminar position in the olfactory bulb and their morphology. All experiments were performed at 35 °C in standard Ringer's solution with excitatory (25 μM 2-amino-5-phosphonopentanoic acid and 10 μM 6-cyano-7-nitroquinoxaline-2,3-dione) and inhibitory (10 μM bicuculline) synaptic activity blocked.

Current-clamp recordings were performed while injecting neurons with a filtered white-noise current stimulus. Noise traces were generated by convolving a 2.5-s white-noise current with an alpha function of the form $t*exp(-t/\tau)$, where $\tau = 3$ ms. We chose this spectral structure as it generates reliable spiking in these neurons and corresponds to the timescale of fast synapses afferent to MCs (2). Each neuron received one of a small number of stimuli generated via this method (most neurons received one of three stimulus templates) and was presented ∼40 stimulus repeats. The amplitude (variance) of the noise used was between 5% and 40% of the direct current (100–800 pA, $\sigma = 20$–80 pA) offset for each cell, with the majority of cells receiving 10–20% of the dc offset. The variance of the noise was selected as previously described (2), to induce reliable firing without large input fluctuations. For all recordings, a 25 or 50 pA hyperpolarizing pulse was injected before stimuli were delivered to measure input resistance and membrane time constant, allowing us to track the stability of recordings over multiple trials. Only neurons whose firing patterns were stable across trials and fired a sufficient number of spikes in each trial (>5 Hz) were used in this study. Upon stimulation most neurons usually underwent a brief nonspiking adaptation period (111 ± 14 ms), which was assessed visually and excluded from the analysis.

Model Fitting. GLM models were fit and simulated using code provided by Jonathan Pillow (Departments of Psychology and Neurobiology, University of Texas at Austin, Austin, TX) (3). Models consisted of a temporal stimulus filter $\boldsymbol{k}$, a postspike history filter $\boldsymbol{h}$,and a constant bias term b. Stimulus and history filters were each represented using 10 spline-like cosine basis functions spaced logarithmically in time. The conditional intensity function of each neuron was modeled as $\lambda(t) = \exp(\boldsymbol{k}\cdot\boldsymbol{x} + \boldsymbol{h}\cdot\boldsymbol{r} + b)$, where $\boldsymbol{x}$ denotes the stimulus and $\boldsymbol{r}$ is the recorded spike response of the neuron. Before fitting, stimuli were downsampled to 1 KHz and standardized by subtracting the steady-state component and dividing by the amplitude of the stimulus noise. LNP models were fit using the spike-triggered average stimulus as the linear filter and estimating the spike-rate nonlinearity using 60 independent histogram bins.

Models were trained using all of the trials from the first 90% of the stimulus presentation and validated using the remaining 10%. In specific, we validated the fit of our models by comparing real and model peri-stimulus time histograms (PSTHs) computed from the test stimulus set (i.e., stimuli not used in the training of the model). We simulated model spike trains using the GLM to probabilistically generate spikes elicited by the test stimulus. PSTHs were computed by summing spikes across trials and smoothing with a Gaussian filter of width $\sigma = 2$ ms. The similarity between real and model PSTHs was reported using Pearson's correlation coefficient. For visualization, MC rasters were randomized across trials.

To assess whether the GLM fitting procedure could also fit neuron responses to multiple stimulus types, we performed an additional set of experiments on mitral cells ($n = 5$ neurons) where each neuron was stimulated with both a high- and low-frequency stimulus (white noise convolved with an alpha function with $\tau = 3$ ms and $\tau = 10$ ms, respectively). We found that the GLM modeling procedure could sufficiently fit neuron responses to each of these stimulus types, indicating that the fitting procedure is not specific to the particular stimulus type used to generate stimulus evoked responses in this study (Fig. S2).

Computation of Neuronal Statistics using GLM Models. We were interested in computing neuronal statistics like average firing rates and trial-to-trial reliability from the fitted GLM models. We computed these by simulating long experiments (∼2 min) of continuous stimulation and computing desired statistics based on these responses. We computed neuron reliability by stimulating each model neuron with multiple trials ($n = 50$) of the same stimulus and calculating reliability as the average zero-lag correlation across trials using a bin size of 5 ms.

To calculate to what extent neurons were driven by intrinsic (history plus bias) versus stimulus components (Fig. S3*B*), we used the model to simulate spike trains while storing the stimulus and intrinsic "currents," which generated the spike trains. Here the stimulus-driven component consists of the convolution of neuron's stimulus filter with the input stimulus; whereas, the intrinsic component is defined as the bias term plus convolution of neuron's spike train with its postspike filter. We calculated the ratio of intrinsic to stimulus inputs as $\langle[stim]_+\rangle / \langle[intrinsic]_+\rangle$ where $[\cdot]_+$ indicates selection of positive values of the currents and $\langle\cdot\rangle$ indicates the mean.

Stimuli Generation for Simulations. We generated zero-mean Gaussian stimuli $\boldsymbol{x}$ with a defined temporal correlation structure and length n by first generating a signal autocorrelogram with the desired spectral structure. This autocorrelogram was used to define a Toeplitz $n*n$ covariance matrix C where the elements of C indicate the pairwise correlations between points of $\boldsymbol{x}$. Correlated stimuli were then generated using the Cholesky decomposition to find a matrix L such that $C = L*L^T$, then multiplying L with a series of uncorrelated normal random variables of length n.

Here we chose eight broadly different stimuli statistics: three stimuli generated via convolving white noise with an alpha function defined as $t*exp(-t/\tau)$, where $\tau = 3$, 5, and 10 ms; three Ornstein–Uhlenbeck processes with $\tau = 10$, 20, and 40 ms, which have flat followed by $1/f^2$-like frequency profiles; a pure white-noise stimulus, with cutoff at 500 Hz; lastly, a "naturalistic" stimulus generated by combining an 8-Hz oscillatory stimulus with an Ornstein–Uhlenbeck process with $\tau = 10$ ms (displayed in Fig. S5).

Decoding. We decoded the population spiking responses using the maximum a posteriori (MAP) estimator (4), which finds the most probable stimulus given a particular population spike response. Stimuli $\boldsymbol{x}$ (typically of length ∼0.5 s, with sampling rate 1 KHz) were decoded from simulated spike responses $\boldsymbol{r}$ by computing the mode of the posterior distribution, $\overset{argmax}{x}\ p(\boldsymbol{x}|\boldsymbol{r})$, where $p(\boldsymbol{x}|\boldsymbol{r}) \propto p(\boldsymbol{r}|\boldsymbol{x})p(\boldsymbol{x})$ via Bayes' rule. Here $p(\boldsymbol{r}|\boldsymbol{x})$ is the likelihood of a response given a stimulus and is given by the set of uncoupled neuron encoding models and $p(\boldsymbol{x})$ is a multivariate Gaussian prior specifying the specific stimulus autocorrelation structure (with covariance matrix, C, used to generate stimuli). In specific, stimuli were decoded using a recently described method (4) that takes advantage of a convenient Gaussian approximation on the posterior distribution $p(\boldsymbol{x}|\boldsymbol{r})$ and its log-concavity to exactly compute the maximum (i.e., the mode) of the posterior

distribution via numerical optimization techniques. This method also provides an estimate of the uncertainty of the stimulus representation (Fig. S4 *F* and *G*). Matlab code for decoding and all other methods related to the simulation and analysis of spike trains generated from GLM models (detailed below) can be found at https://github.com/stripathy/mitral_cell_diversity.

Mutual Information Calculation. We calculated the mutual information (4) of the population response $\boldsymbol{r}$ about the stimulus $\boldsymbol{x}$ as $I(\boldsymbol{x};\boldsymbol{r}) = H(\boldsymbol{x}) - H(\boldsymbol{x}|\boldsymbol{r})$. $H(\boldsymbol{x})$ denotes the entropy of the stimulus and is defined by the multivariate Gaussian stimulus prior $p(\boldsymbol{x})$ and $H(\boldsymbol{x}|\boldsymbol{r})$ denotes the conditional entropy of the stimulus given the response and is estimated by approximating the posterior distribution $p(\boldsymbol{x}|\boldsymbol{r})$ as a multivariate Gaussian $N(\boldsymbol{x_{map}}, C)$, where the covariance matrix, C, is computed as a byproduct of our decoding method. Here we use the fact that the entropy of a Gaussian with covariance matrix C is $\ln\sqrt{(2\pi e)^n|C|}$, where $|\cdot|$ denotes matrix determinant and n is the dimension of the stimulus. Estimates of $I(\boldsymbol{x};\boldsymbol{r})$ were obtained by averaging $H(\boldsymbol{x}|\boldsymbol{r})$ over responses elicited to multiple stimuli realizations ($n = 50$). It is important to note that because this method estimates the entropy of the posterior distribution, it generally provides a better estimate of the mutual information than the commonly used lower-bound estimate of $I(\boldsymbol{x};\boldsymbol{r})$ obtained via the optimal linear estimator (5), especially when the neurons are nonlinear and not well described by an LNP model.

We computed a normalized measure of the redundancy or synergy (6) of a pair of neurons $\boldsymbol{a};\boldsymbol{b}$ relative to each of the neurons independently as $\frac{I(\boldsymbol{x};\boldsymbol{a}) + I(\boldsymbol{x};\boldsymbol{b}) - I(\boldsymbol{x};\boldsymbol{a},\boldsymbol{b})}{I(\boldsymbol{x};\boldsymbol{a},\boldsymbol{b})}$. Positive values indicate informational redundancy and negative values indicate synergy.

To elaborate on our finding of synergistic pairs of neurons (Fig. 2*F*), we note that due to computational constraints we can only decode stimuli of relatively short lengths (0.5 s). Therefore, we will tend to underestimate the information rates of neurons that fire at low firing rates. For example, when performing stimulus decoding to calculate the information rate of a single neuron with a very low firing rate, it may fire zero spikes during the time interval and thus encodes no stimulus information. However, when considering two such neurons, the two will be much more likely to fire at least one spike between them, and thus encode some nonzero stimulus information. In this example, the case of a two-neuron pair would appear synergistic relative to a single neuron alone. Therefore, if we could simulate arbitrarily long stimulus presentations we would expect this apparent synergy effect to disappear.

Calculating Population Stimulus Generalization. To calculate how well heterogeneous and homogeneous populations generalized across stimuli of differing types, we computed the generalizability for each population type. Here generalizability is defined as $corr(ranks_{stim1}, ranks_{stim2})$, or the correlation between population ranks on pairs of stimulus types.

GLM Dimensionality Reduction. We chose to reduce the dimensionality of the space defined by neuronal GLM parameters using principal component analysis (Fig. S6) for visualization and further analysis. Principle components (PCs) were generated by first concatenating waveforms of stimulus, postspike, and bias components across all neurons and standardizing before performing PC analysis. Postspike and bias terms were transformed to units of log(Gain) before concatenating. The first 10 ms of postspike filters were removed and not included in analysis.

Computing Population Diversity. We calculated population diversity as the mean Euclidean distance of GLM parameters computed between all pairs of neurons in a population. We excluded the first 10 ms of the postspike filters across neurons as most neurons were refractory during this period. The average diversity of heterogeneous populations was computed by averaging over 50,000 randomly sampled populations. When reporting the uncertainty in the diversity of randomly sampled populations (Fig. 4*D*), we chose to show a measure of the population variance (interquartile range) as opposed to SEs.

We sampled populations that varied greatly in their amount of diversity (from low to high; Fig. 4*F* and Figs. S9 and S10) through implementing stratified sampling where we first sampled 2 million five-neuron populations and then further subsampled this set to pick populations that varied uniformly in their diversity.

Eliminating Diversity in a Single GLM Dimension. We constructed populations which had diversity eliminated along a single GLM dimension (stimulus, postspike, or bias) by modifying the neuron model parameters from the ones based on the recorded neurons (Fig. S11). For example, to sample neurons where diversity in the stimulus filter had been eliminated, we set the stimulus filter for all neurons that of the mean stimulus filter computed over all neurons. We further ensured that mean of the firing rates across neurons were similar between the original and diversity-reduced populations.

1. Padmanabhan K, Urban NN (2010) Intrinsic biophysical diversity decorrelates neuronal firing while increasing information content. *Nat Neurosci* 13(10): 1276–1282.
2. Galán RF, Ermentrout GB, Urban NN (2008) Optimal time scale for spike-time reliability: theory, simulations, and experiments. *J Neurophysiol* 99(1):277–283.
3. Pillow JW, et al. (2008) Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature* 454(7207):995–999.
4. Pillow JW, Ahmadian Y, Paninski L (2011) Model-based decoding, information estimation, and change-point detection techniques for multineuron spike trains. *Neural Comput* 23(1):1–45.
5. Warland DK, Reinagel P, Meister M (1997) Decoding visual information from a population of retinal ganglion cells. *J Neurophysiol* 78(5):2336–2350.
6. Schneidman E, Bialek W, Berry MJ, 2nd (2003) Synergy, redundancy, and independence in population codes. *J Neurosci* 23(37):11539–11553.

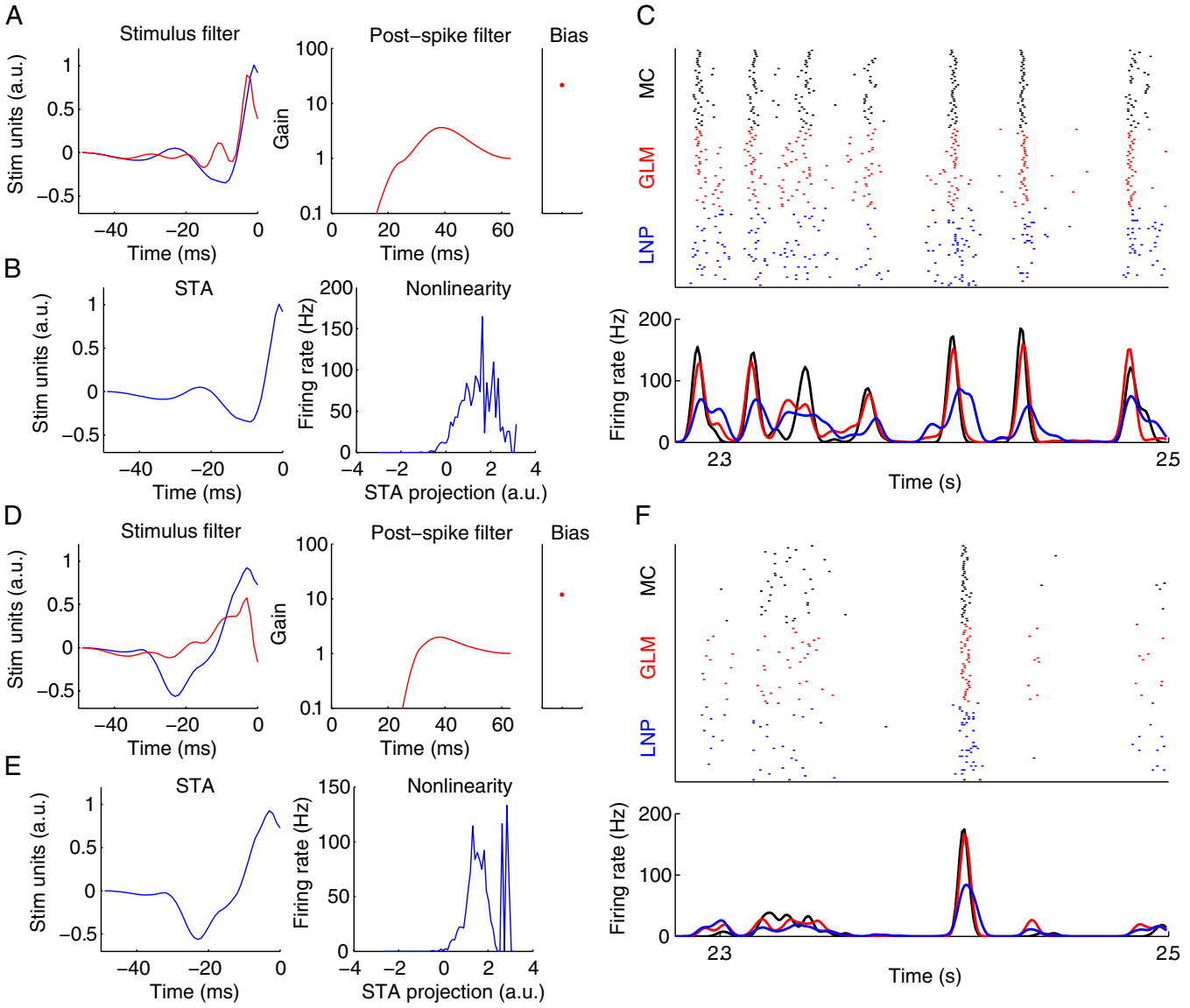

Fig. S1. Comparison of GLM and LNP model parameters and prediction accuracy. (*A*) GLM parameters (red) and spike triggered average current (STA, *Left*, blue) for the neuron in Fig. 1*A*. (*B*) LNP parameters for same neuron. (*C*) Experimental MC, GLM, and LNP rasters (*Top*) and PSTH (*Bottom*). Note that GLM spikes replicate the MC more precisely than the LNP model. (*D*–*F*) As in *A*–*C* but for neuron in Fig. 1*B*.

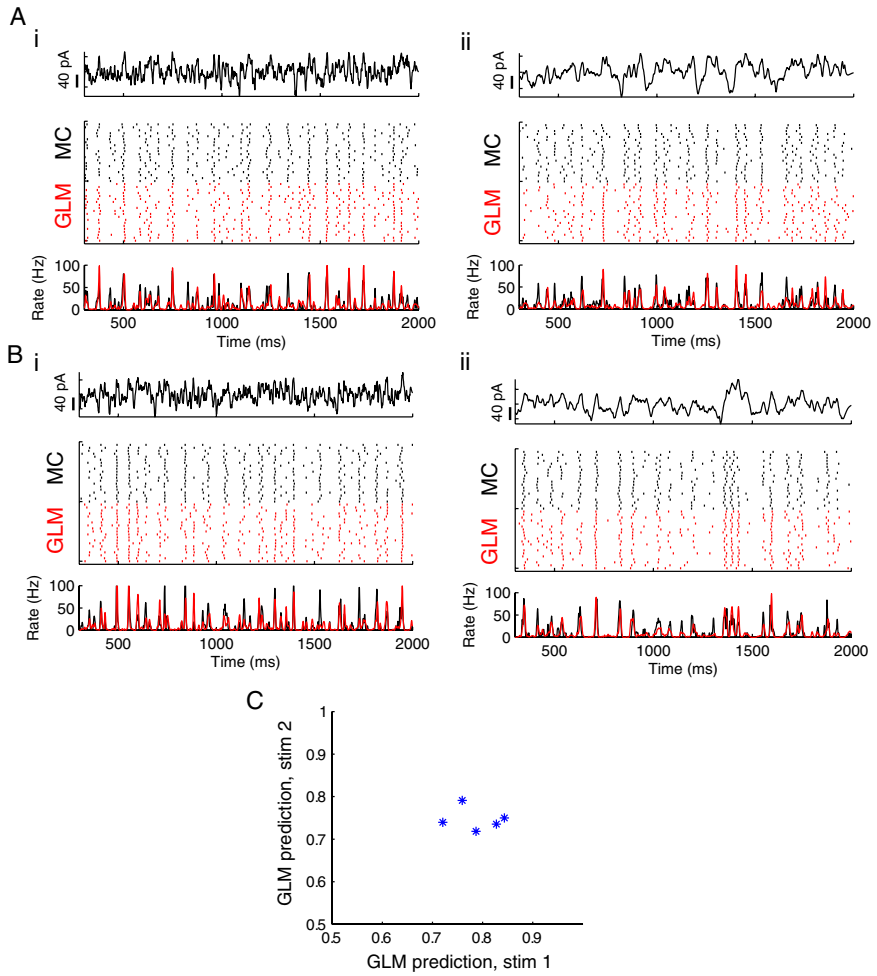

Fig. S2. GLM modeling procedure is effective at fitting neuron responses to different stimulus types and temporal correlation structures. (*A*) Stimulus and spike rasters for a recorded MC (black) and simulated model responses (GLM, red) for a high-frequency stimulus (*i*, white noise convolved with alpha function with τ = 3 ms, stim 1) or a low-frequency stimulus (*ii*, white noise convolved with alpha function with τ = 10 ms, stim 2). Entire stimulus and spike response shown to illustrate temporal correlation difference between stimuli in *i* and *ii*. (*B*) As in *A*, but for a different mitral cell. (*C*) GLM models accurately predict neuron responses to each of these two stimulus types (n = 5 MCs). Prediction accuracy computed as the correlation coefficient between MC and model PSTH.

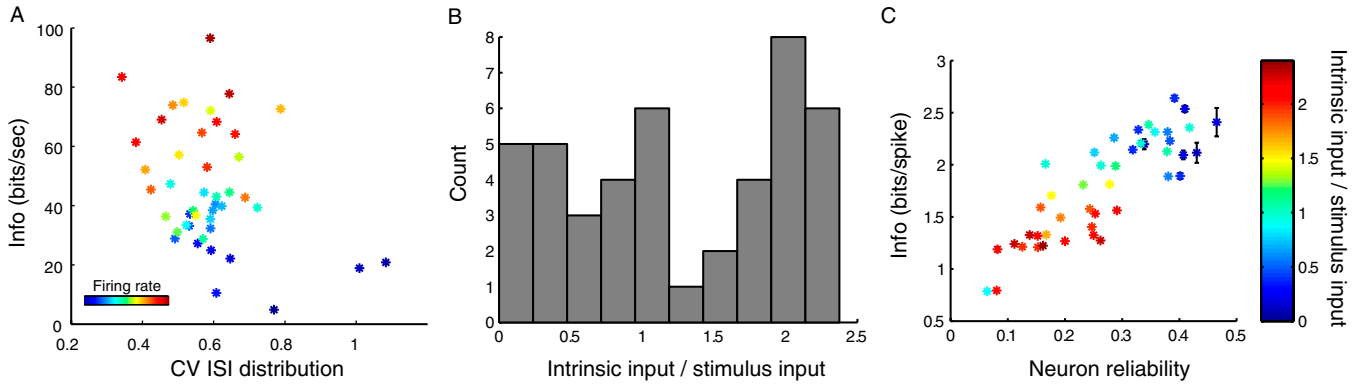

Fig. S3. Relationship between neuronal information rate and other spiking characteristics. (*A*) Neuron information rate as a function of coefficient of variation (CV) of the interspike interval distribution (ISI) (CV ISI distribution). Neurons with higher information rates tend to have lower CVs, indicating that they fire more regularly (i.e., less Poisson-like). (*B*) Histogram of the relative ratio of neuron spiking due to stimulus driven or intrinsic components (bias plus history term) as indicated from the GLM. Neurons vary from being driven entirely by stimulus to being driven primarily by intrinsic components. (*C*) Same as Fig. 2*D* but neurons are colored by ratio of intrinsic to stimulus input (as in *B*). Neurons that are primarily stimulus driven are both more reliable and encode more information per spike.

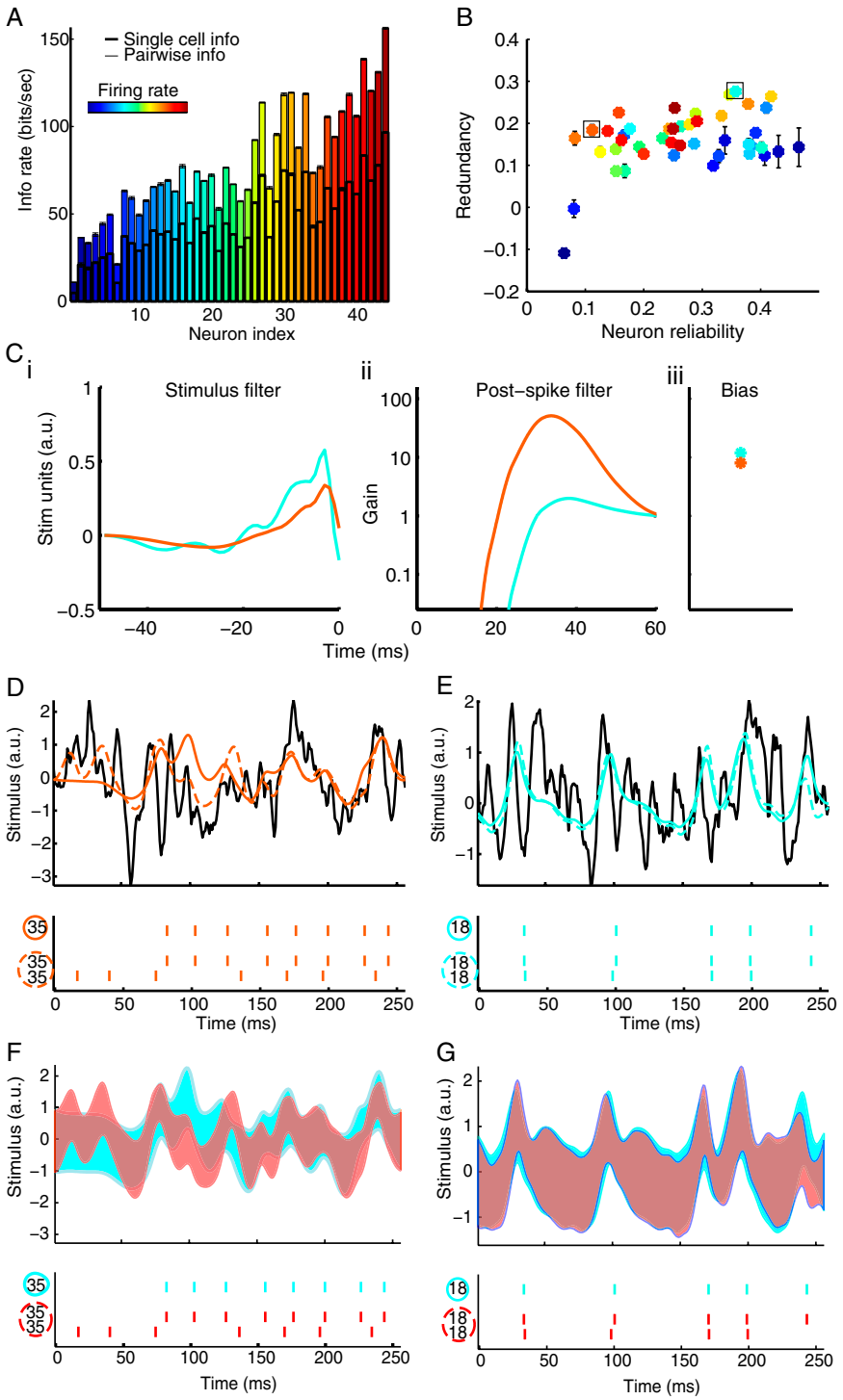

Fig. S4. Two copies of the same neuron are more informative than a single neuron alone. (*A*) Stimulus-response mutual information for single neurons (thick line) and populations of two copies of the same cell (thin line). In all cases, two neurons convey more information than a single neuron alone. (*B*) Informational redundancy of homogeneous pairs versus single neuron populations plotted as a function of single neuron reliability. Boxes show neurons 18 and 35, which are also highlighted in Fig. 2*B*. (*C*) GLM parameters for boxed neurons in *B*. These two neurons were chosen as an example because they have very different GLM parameters which cause them to spike differently in response to the same stimulus. Neuron 18 (cyan) fires primarily as a result of the stimulus whereas neuron 35 (orange) has a large postspike rebound current causing it to fire relatively periodically. (*D* and *E*) Example stimulus and reconstructions for neurons 35 and 18. Note that for neuron 35, an unreliable neuron, there is an improved stimulus representation when another copy of 35 is added; this is effect is less pronounced for neuron 18, which fires more reliably. (*F* and *G*) As in *D* and *E* but showing uncertainty in stimulus reconstruction. Blue shaded trace indicates neuron 18 and red trace indicates neuron 35. Stimulus not shown for clarity. Error bars represent 1 SD of the Gaussian estimate of the decoded posterior distribution. For both neurons, the stimulus representation is less uncertain with neuron pairs versus single neurons, highlighting the benefit of redundancy or pooling representations over neurons.

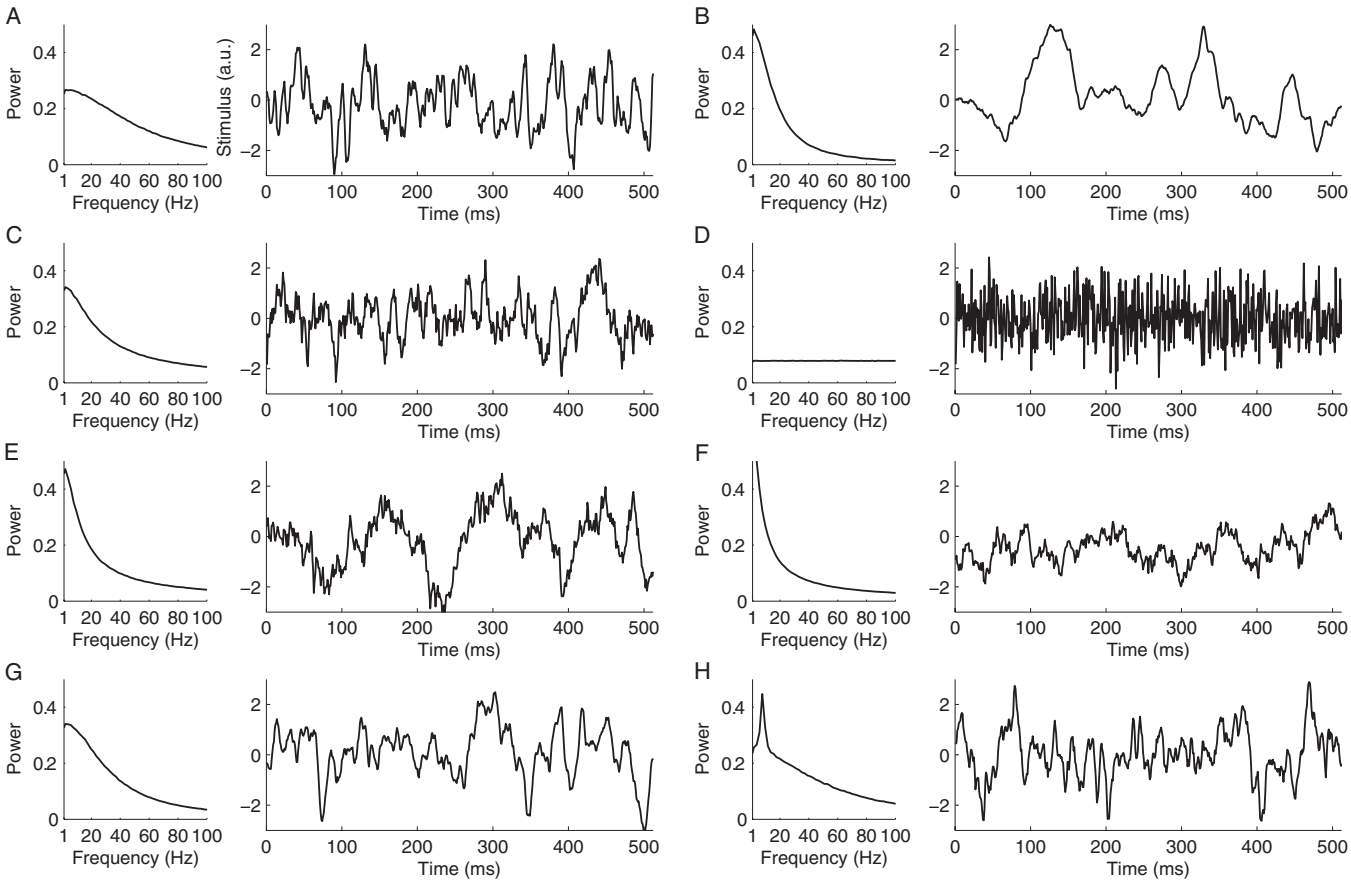

Fig. S5. Examples of the eight stimulus statistics used in this study. (*A*) White noise convolved with alpha function with $\tau = 3$ ms. (*B*) White noise convolved with alpha function with $\tau = 10$ ms. (*C*) Ornstein–Ulhenbeck process with $\tau = 10$ ms. (*D*) Pure white-noise stimulus (max frequency = 500 Hz). (*E*) Ornstein–Ulhenbeck process with $\tau = 20$ ms. (*F*) Ornstein–Ulhenbeck process with $\tau = 40$ ms. (*G*) White noise convolved with alpha function with $\tau = 5$ ms. (*H*) Naturalistic sniffing stimulus generated by combining 8-Hz oscillation with Ornstein–Ulhenbeck process with $\tau = 10$ ms.

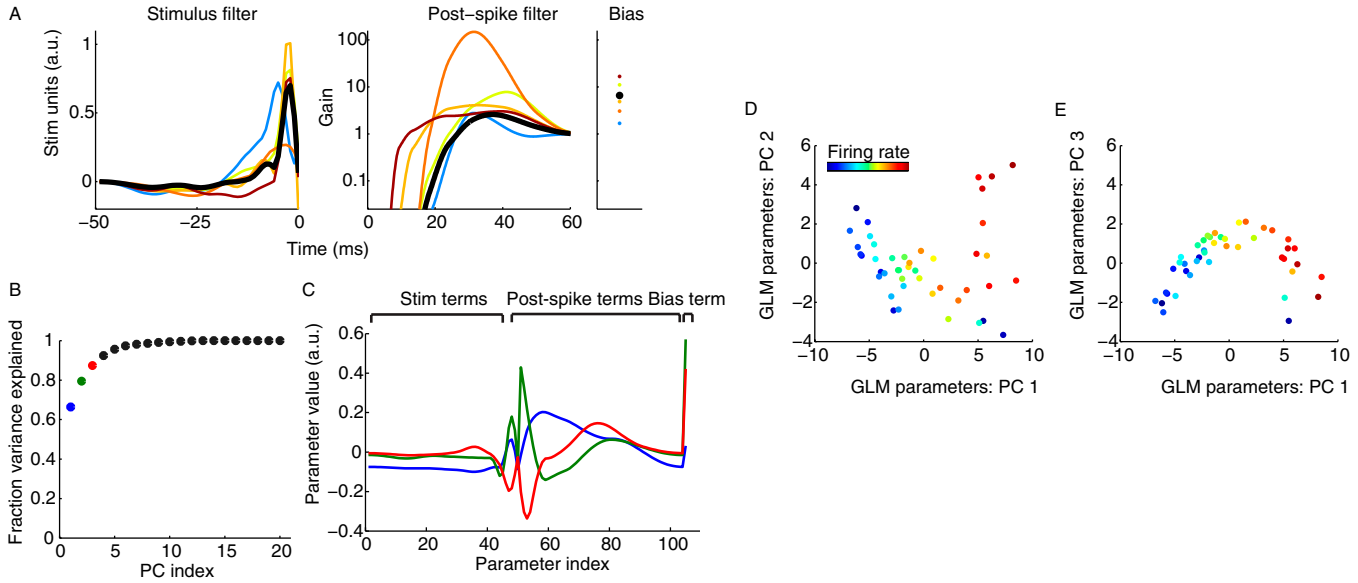

Fig. S6. Decomposition of neuronal GLM parameter space into a small number of principal components. (*A*) GLM parameters for an example five-neuron population (colors) and mean GLM parameters across all neurons (black). (*B*) Percent variance explained for each subsequent PC. The first three principal components explain 85% of variance among GLM parameters. (*C*) Visualization of the first three PCs computed from GLM parameters. (*D* and *E*) Projection of neurons (dots) into space defined by PCs 1 and 2 (*D*) or PCs 1 and 3 (*E*). This analysis plots neurons such that those with similar GLM parameters are plotted close to one another. The computed PCs largely reflect differences among postspike and bias terms, and to a lesser extent stimulus filters. Neurons with negative PC1 tend to have low firing rates, low-amplitude stimulus filters (relative to the mean across neurons) and longer refractory periods with less of a tendency to burst 20–40 ms following a spike and vice versa for positive PC1. Neurons with high PC2 tend to have high baseline excitability, very short refractory periods, and increased amplitude stimulus filters relative to the mean.

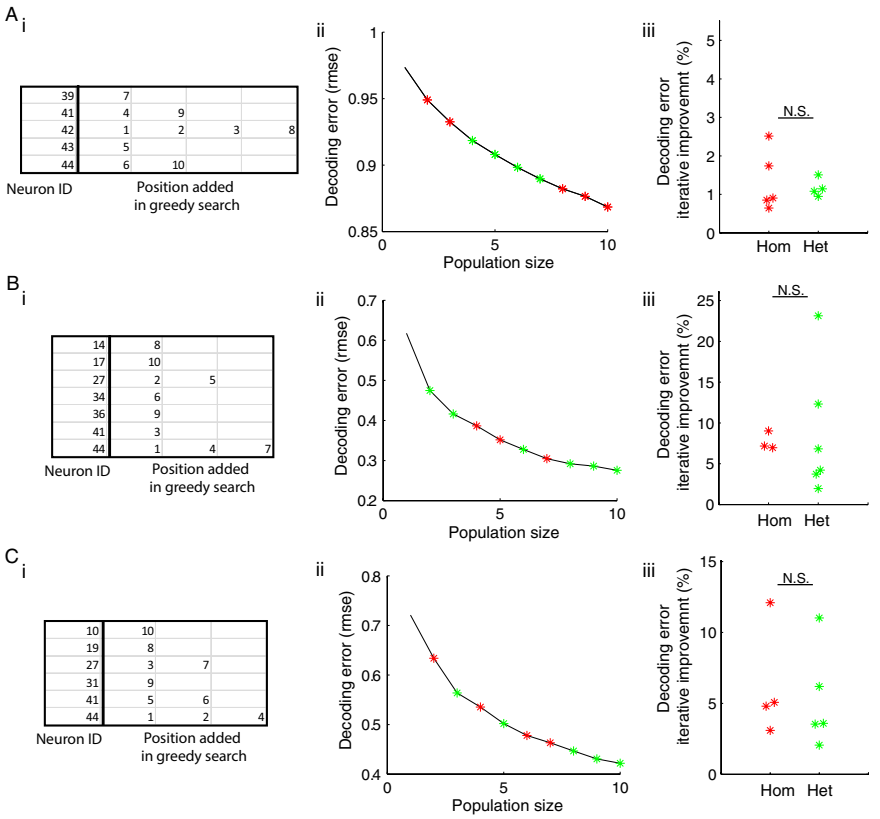

Fig. S7. Selection of homogeneous neurons during greedy search does not preferentially occur after coding improvements have saturated. (*A, i*) Detailed analysis of neurons added and their selected position during greedy search algorithm. Neuron population selected to best encode a white-noise stimulus. (*ii*) Decoding error as a function of population size during greedy search optimization procedure. Colored asterisks indicate whether neuron added at *n*th iteration is a homogeneous neuron (red, i.e., a copy of selected neuron is already in the population) or heterogeneous (green). Same population as shown in *i*. (*iii*) Same data as *ii*, plotted as iterative improvement in decoding error when adding an additional neuron, broken down by if neuron added is a homogeneous neuron (red) or heterogeneous neuron (green). There is no significant difference ($P > 0.05$) in decoding error improvement between homogeneous and heterogeneous groups. (*B* and *C*) As in *A*, but for a low-frequency stimulus (*B*, white noise convolved with alpha function with $\tau = 10$ ms) or a high-frequency stimulus (*C*, white noise convolved with alpha function with $\tau = 3$ ms).

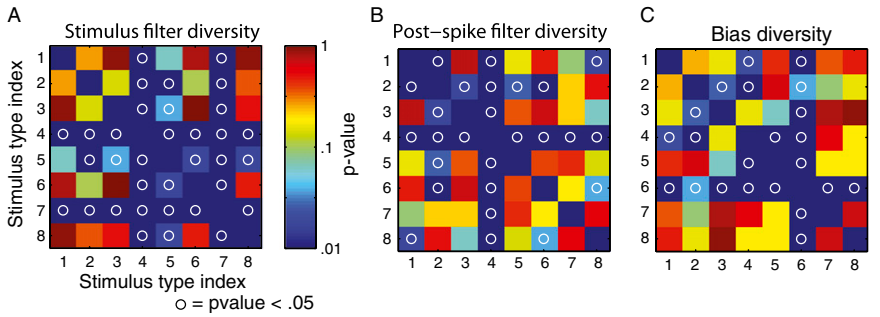

Fig. S8. Companion figure to Fig. 4*E*, showing that populations selected through the greedy search procedure to best represent different stimulus types tend to have different amounts of GLM parameter diversity. (*A–C*) Pairwise comparisons between the amounts of population diversity resulting when populations are optimized to best represent different stimulus types. Population diversity values are first computed by running the greedy search algorithm multiple times for each stimulus type ($n = 10$). Colored squares indicate *P* value for statistical test comparing population diversity between each pair of stimulus types. Circles between stimulus pairs indicates that populations are significantly different in terms of their diversity (Wilcoxon, $P < 0.05$). Stimuli are numbered as in Fig. S5 (1): White noise convolved with alpha function with $\tau = 3$ ms (2). White noise convolved with alpha function with $\tau = 10$ ms (3). Ornstein–Ulhenbeck process with $\tau = 10$ ms (4). Pure white-noise stimulus (max frequency = 500 Hz) (5). Ornstein–Ulhenbeck process with $\tau = 20$ ms (6). Ornstein–Ulhenbeck process with $\tau = 40$ ms. (7) White noise convolved with alpha function with $\tau = 5$ ms. (8) Naturalistic sniffing stimulus generated by combining 8-Hz oscillation with Ornstein–Ulhenbeck process with $\tau = 10$ ms.

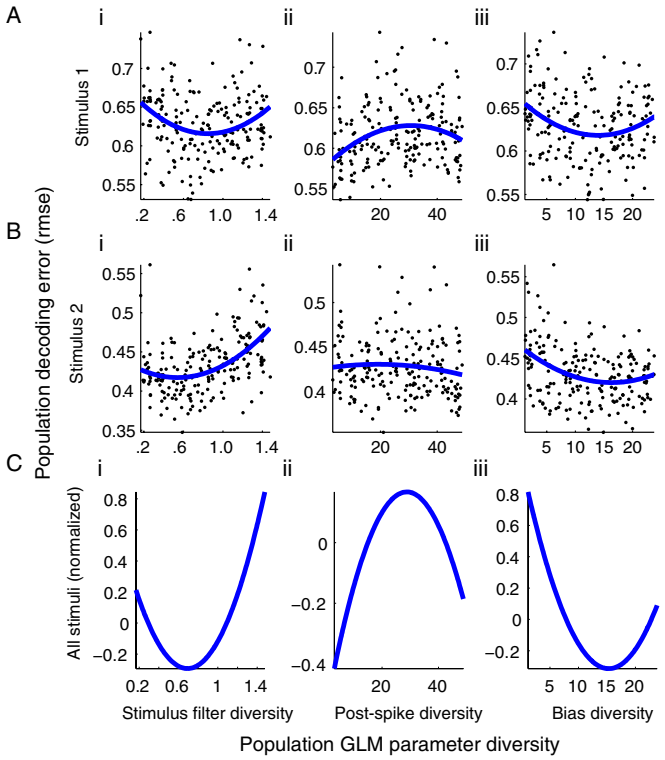

Fig. S9. Evidence for a U-shaped relationship between population diversity and decoding error. (*A* and *B*) Decoding error for five-neuron populations (black dots) as a function of population diversity along stimulus filters (*i*); postspike filters (*ii*); and bias parameters (*iii*) for stimulus 1 (*A*, as in Fig. 3*A*) or stimulus 2 (*B*, as in Fig. 3*B*). (*i–iii*) Two-hundred heterogeneous populations were drawn such that populations of varying diversity (from superdiverse through subdiverse) were sampled with equal probability (*Materials and Methods*; $n = 5$ neurons per population). Blue line shows fit of a quadratic polynomial, used to test for expected U-shaped relationship. In all cases, the regression coefficient associated with the quadratic term of the polynomial fit was positive and significant ($P < 0.01$), except for *A, ii* and *B, ii*, indicating that reconstruction error is minimized at an intermediate values of stimulus filter and bias diversity. The reason why there does not appear to be a concave-up U-shaped relationship for postspike filters is due to sampling confounds: low postspike diversity populations tend to have higher firing rates than high postspike diversity populations. (*C*) As in *A*, but showing U curves averaged across each of the eight stimuli. In this case, the decoding error was first normalized to *z* scores before performing the regression, allowing comparison across stimuli.

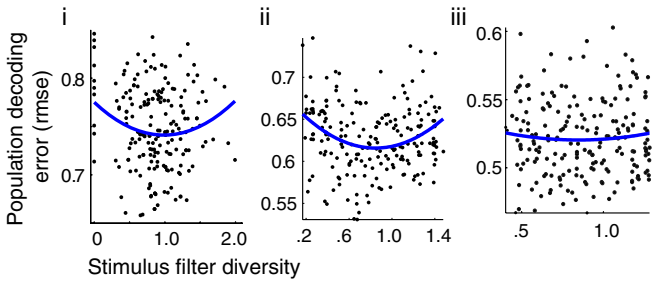

Fig. S10. Lack of a substantial U-shaped relationship between population diversity and decoding error for larger sized populations. Decoding error for neuron populations (black dots) of size $n = 2$ neurons (*i*, data reanalyzed from Fig. 2); 5 neurons (*ii*, same as Fig. 4*F*); or 10 neurons (*iii*) per population as a function of population diversity along the GLM stimulus filter dimension for stimulus 1 (as in Fig. 3*A*). (*i–iii*) Two-hundred heterogeneous populations were drawn such that populations of varying diversity (from superdiverse through subdiverse) were sampled with equal probability (*Materials and Methods*). Blue line shows fit of a quadratic polynomial, used to test for expected U-shaped relationship. The simplest explanation for why the observed U-shaped curve effect gets weaker with more neurons per population is that each of these larger populations have saturated in their ability to represent the stimulus. In this case, it matters less whether the populations are diverse (or not) because there are enough neurons in each population to effectively represent the stimulus.

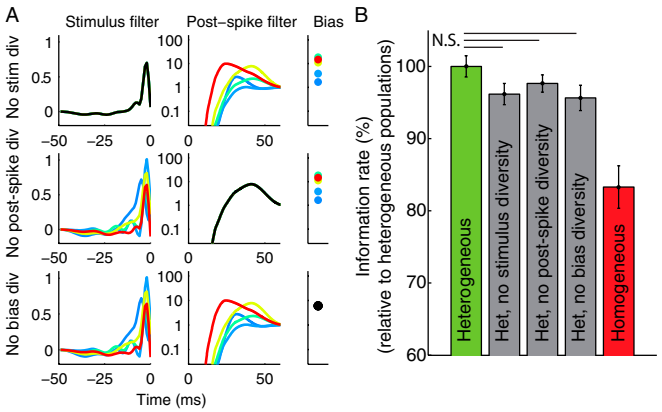

Fig. S11. Benefit of neuron variability does not depend on a single GLM model dimension. (*A*) Example five-neuron populations where population variability in a single GLM dimension (stimulus, postspike, and bias) has been eliminated (*Top*, *Middle*, and *Bottom*, respectively). (*B*) Mean information rates for heterogeneous, homogeneous, and stimulus, postspike, and bias-reduced diversity populations. Information rates computed relative to heterogeneous populations ($n = 5$ neurons per population, 200 populations per condition, stimulus is high-frequency stimulus, white noise convolved with alpha function with $\tau =$ 3 ms). None of the reduced diversity populations were significantly different from random heterogeneous populations ($P > 0.05$, Wilcoxon, N.S.) indicating that the coding benefits of diversity do not rely upon a single GLM dimension. This figure suggests that the representational advantage of neuron variability is not specifically tied to any one of the three GLM dimensions.