ARTICLE **Communicated by Haim Sompolinsky**

Functional Diversity in the Retina Improves the Population Code

Michael J. Berry II
berry@princeton.edu
Princeton Neuroscience Institute and Department of Molecular Biology, Princeton University, Princeton, NJ 08544, U.S.A.

Felix Lebois
felix.lebois@yahoo.fr
Department of Physics, Ecole Normale Supérieure, 75005 Paris, France

Avi Ziskind
aziskind@Princeton.edu
Department of Physics, Princeton University, Princeton, NJ 08544, U.S.A.

Rava Azeredo da Silveira
rava@ens.fr
Princeton Neuroscience Institute, Princeton University, Princeton, NJ 08544, U.S.A.; Department of Physics, Ecole Normale Supérieure, 75005 Paris; Laboratoire de Physique Statistique, Ecole Normale Supérieure, PSL Research University, 75231 Paris; Université Paris Diderot Sorbonne Paris Cité, 75031 Paris; Sorbonne Universités UPMC Université Paris 6, 75005 Paris, France; CNRS

Within a given brain region, individual neurons exhibit a wide variety of different feature selectivities. Here, we investigated the impact of this extensive functional diversity on the population neural code. Our approach was to build optimal decoders to discriminate among stimuli using the spiking output of a real, measured neural population and compare its performance against a matched, homogeneous neural population with the same number of cells and spikes. Analyzing large populations of retinal ganglion cells, we found that the real, heterogeneous population can yield a discrimination error lower than the homogeneous population by several orders of magnitude and consequently can encode much more visual information. This effect increases with population size and with graded degrees of heterogeneity. We complemented these results with an analysis of coding based on the Chernoff distance, as well as derivations

A.Z. is currently at SRI International, Princeton, NJ U.S.A.; F.L. is currently at Saint-Gobain, Paris, France.

Neural Computation **31, 270–311** (2019)
doi:10.1162/neco_a_01158
© 2018 Massachusetts Institute of Technology

of inequalities on coding in certain limits, from which we can conclude that the beneficial effect of heterogeneity occurs over a broad set of conditions. Together, our results indicate that the presence of functional diversity in neural populations can enhance their coding fidelity appreciably. A noteworthy outcome of our study is that this effect can be extremely strong and should be taken into account when investigating design principles for neural circuits.

1 Introduction

Neurons are complex objects, bewildering in their anatomical and functional diversity. Neuroscience has managed to bring order to this chaos by recognizing that functional properties are often organized into spatial maps, where local neighborhoods have similar tuning, and that within a local neighborhood, the shapes of neurons often come in stereotyped patterns that can be divided into cell types. Still, local neural circuits are generally made up of neurons from many cell types, and, hence, a high degree of functional heterogeneity is present in them. To cite just two examples from visual areas, the retina is tiled by as many as 40 types of ganglion cells, so that any spot in visual space is monitored by a large number of ganglion cell types (Azeredo da Silveira & Roska, 2011; Baden et al., 2016; Robles, Laurell, & Baier, 2014; Seung & Sumbul, 2014). In turn, the diversity of their light responses results from the diversity in presynaptic circuits, made up of a dozen types of bipolar cells (Ghosh, Bujan, Haverkamp, Feigenspan, & Wassle, 2004) and over 25 types of amacrine cells (MacNeil & Masland, 1998; Marc et al., 2013). Primary visual cortex has great local diversity in receptive field shapes and spatial frequency tuning (Bonin, Histed, Yurgenson, & Reid, 2011; Ringach, Shapley, & Hawken, 2002), as was apparent even in the earliest studies (Hubel & Wiesel, 1962). In addition to the functional diversity that comes from dividing neurons into cell types, there is also variability within a given cell type, which comes from fluctuations in the morphology and chemical makeup of neurons, as well as fluctuations in the connectivity of local circuits (Asari & Meister, 2012; Brenowitz & Regehr, 2007; Dobrunz & Stevens, 1997, 1999; Prinz, Bucher, & Marder, 2004). Thus, the processing of information within local circuits is subject to both the heterogeneity that exists among cell types and the "random heterogeneity" coming from fluctuations in connectivity and biochemical processing within the same cell type.

One is naturally led to ask in which ways the heterogeneity among neurons participates in information processing. On the one hand, one can argue that a population of identical cells would favor coding by allowing the transmission of a well-averaged signal related to the tuning properties of the neurons. In this picture, random heterogeneity is a bug that results from

a developmental inability to generate perfectly ordered circuits. But the fact that there exist perfectly ordered neural circuits in nature, such as the invertebrate ommatidia, and that in vertebrates, there appears to be a higher degree of heterogeneity in higher brain areas together indicate that this argument is too simplistic. On the other hand, one expects that heterogeneity can benefit coding because it endows a neural population with a broader range of "information sensors."

In recent years, a number of studies have substantiated this latter line of thought. Analyses of studies of visual (Chelaru & Dragoi, 2008; Kastner, Baccus, & Sharpee, 2015; Osborne, Palmer, Lisberger, & Bialek, 2008), auditory (Holmstrom, Eeuwes, Roberts, & Portfors, 2010), and olfactory (Tripathy, Padmanabhan, Gerkin, & Urban, 2013) neurons have demonstrated that heterogeneity can enhance the processing of information by allowing for combinatorial codes. Theoretical work has also shown that splitting retinal ganglion cells into ON and OFF subpopulations is more efficient than splitting them into same-polarity subpopulations with different thresholds (Gjorgjieva, Sompolinsky, & Meister, 2014). Heterogeneity has also been shown to affect the dynamics of neural assemblies and thereby improve their coding properties (Hunsberger, Scott, & Eliasmith, 2014; Lengler, Jug, & Steger, 2013; Mejias & Longtin, 2012). In the case of coding with a population of broadly tuned neurons, cell-to-cell variability can counteract the harmful effects of correlated noise on the coding precision (Ecker, Berens, Tolias, & Bethge, 2011; Shamir & Sompolinsky, 2006; Wilke et al., 2001). Finally, increased heterogeneity of neural activity has been shown to correlate with better behavioral performance in a visual image recognition task (Montijn, Goltstein, & Pennartz, 2015).

Here, we provide complementary demonstrations that heterogeneity can benefit the coding of information appreciably. We analyze the activity of large populations of retinal ganglion cells in response to both artificial and natural stimuli, and we show that their heterogeneity is responsible for improved coding of visual information. A new aspect of this result, which we emphasize here, is that this effect—namely, the enhancement of coding accuracy due to heterogeneity—can be very large quantitatively and, hence, is a factor to take into consideration for understanding both the performance and design of neural codes. We also show that the Chernoff distance, another useful measure of coding fidelity (Kang & Sompolinsky, 2001), is appreciably larger in heterogeneous than homogeneous populations. Finally, we examine simple models that help us explore the mechanisms by which heterogeneity favors coding. We formulate mathematical arguments that demonstrate that heterogeneity is favorable quite generally. These arguments do not rely on particular forms of the tuning properties of neurons or other specific assumptions. Together, our results suggest that functional heterogeneity is not a bug but a feature of neural population codes.

2 Results

2.1 Choice of Stimuli and Population Responses. Any analysis of the role of heterogeneity in a population code must make some choice about the class of stimuli or experimental conditions to be studied. We start our analysis of the retinal population code with the case of spatially uniform stimulation. The virtue of this choice is that under these visual conditions, every retinal ganglion cell experiences exactly the same input. Thus, any differences in the response among cells can be attributed entirely to their heterogeneity. We use random flicker stimulation, where we sample a broad distribution of all possible temporal patterns of light intensity without imposing potentially limiting conditions from the start. Specifically, we randomly draw a value of light intensity from a gaussian distribution on a fast timescale (30 ms).

Under this stimulus ensemble, the response of retinal ganglion cells depends on the history of the stimulus going back over several hundred milliseconds into the past (Chichilnisky, 2001; Fairhall et al., 2006; Warland, Reinagel, & Meister, 1997). Given a frame time of 8.33 ms, this implies that the relevant stimulus is a vector of about 30 or more light intensity values. The entropy of these stimulus patterns is large enough that the same stimulus essentially never repeats under realistic experimental conditions. Thus, we can assume that the stimulus preceding every time bin in the neural response is different. Our task will be to use the activity of the retinal population to distinguish among stimuli. For the case of two discrete stimuli, this task amounts to distinguishing the set of firing probabilities in the population elicited at one point in time, t_1, which we call the "target" stimulus, from that elicited at another point in time, t_2, called the "distracter" (see Figure 1). We measure the firing probability in small time bins (20 ms) for each neuron and we denote the set of firing probabilities in response to the target stimulus as $\{p_i\}$ and those in response to the distracter stimulus as $\{q_i\}$, where p_i and q_i are the firing rates of the ith cell in the two conditions, respectively.

Figure 2 shows the firing rate of populations of retinal ganglion cells under stimulation by either a 30 sec segment of spatially uniform flicker or a 120 sec natural movie clip. As has been reported before, the firing rate for an individual ganglion cell was vanishing at most points in time and then rose and fell rapidly in a sparse set of firing events (Berry, Warland, & Meister, 1997). The firing rate averaged over the entire population was highly heterogeneous across time (see Figure 2B). In addition, there were appreciable differences among cells in their overall firing rate, averaged across the entire stimulus ensemble. This led to a broad distribution of overall firing rates across cells, ranging from a lower bound of 0.01 spikes/sec (below which we do not trust our spike sorting) up to almost 10 spikes/sec; most cells had an overall firing rate of less than 1 spike/sec (median rate for natural movies = 0.27 spikes/sec; see Figure 2C). As a result of these two properties, the

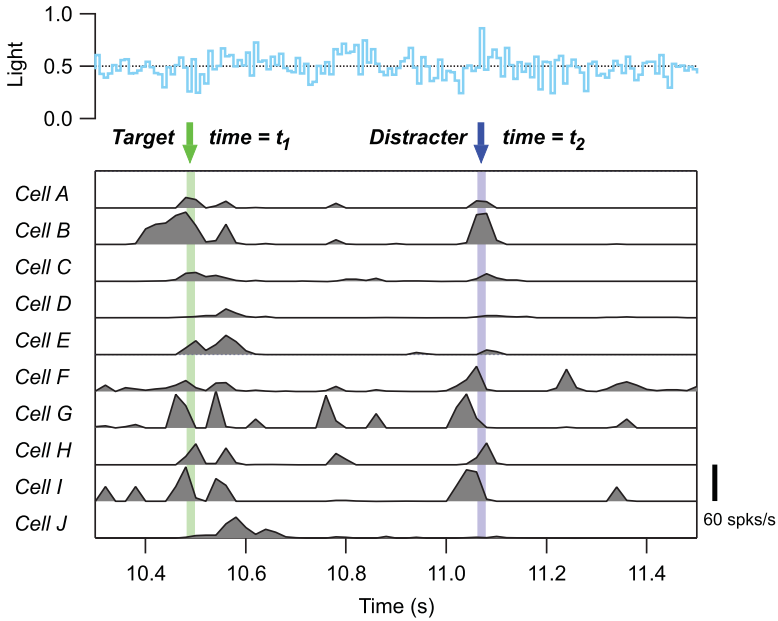

Figure 1: Stimulus discrimination task. (Top) Light intensity versus time for spatially uniform flicker. (Bottom) Firing rates for 10 example ganglion cells, obtained as averages over 300 repeated trials of the same stimulus segment. Colored arrows illustrate a choice of two time bins with population activity patterns in response to target and distracter stimuli, respectively.

distribution of firing rates across cells and time bins was quite broad, with a prominent peak at 0, as the cells were mostly quite sparse, and a tail extending up to over 100 spikes/sec that approximately followed an exponential function (see Figure 2D). (This nearly exponential dependence has also been observed in the visual cortex. As this distribution has maximum entropy at a fixed mean, it has been suggested that this distribution represents a form of efficient coding using the firing rates of individual cells (Baddeley et al., 1997).) Similar forms of sparseness and heterogeneity have been observed in other neural systems as well (Chechik et al., 2006; Shoham, O'Connor, & Segev, 2006; Weliky, Fiser, Hunt, & Wagner, 2003). Together, these observations suggest that the retinal data that we analyze here have a structure to their population code similar to that in other brain regions.

2.2 How Relevant Is Heterogeneity for Population Coding? In order to quantify the computational relevance of heterogeneity, we compared the performance of our measured neural population against an equivalent homogeneous population. Because the coding performance must increase as larger populations or higher overall firing rates are considered, we

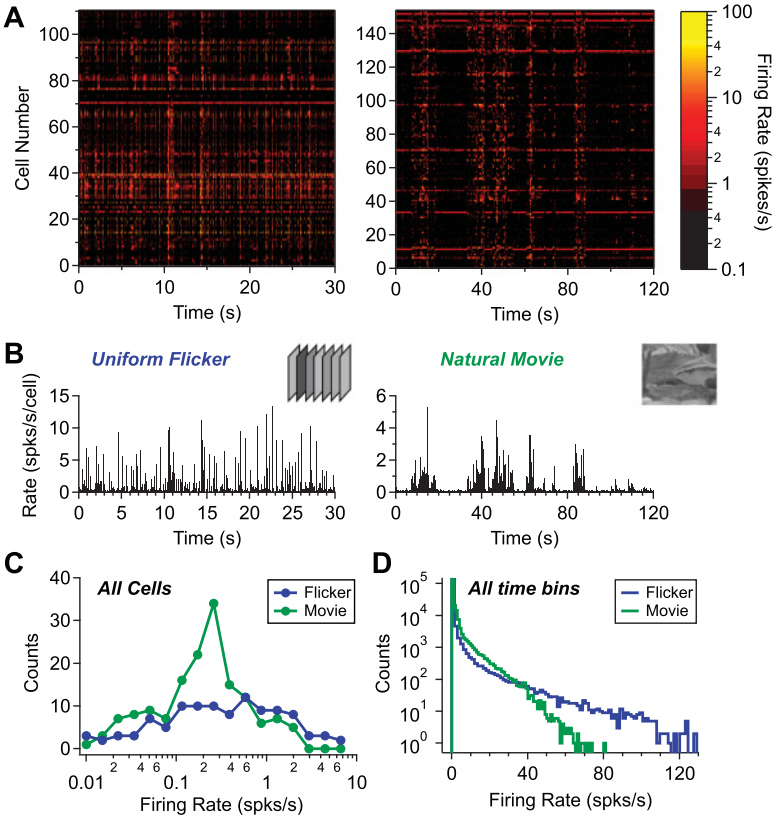

Figure 2: Heterogeneity of neural activity patterns. (A) Matrix of firing rates (color scale) across time (x-axis) and cell identity (y-axis) for spatially uniform flicker (left) and a natural movie (right). (B) Firing rate averaged across cells in each time bin (i.e., the population PSTH) for spatially uniform flicker (left, stimulus illustrated in inset) and a natural movie (right, stimulus illustrated in inset). (C) Histogram of average firing rates (log scale) compiled across cells. (D) Histogram of firing rates (log counts) compiled across time bins and cells.

constructed our equivalent homogeneous populations so as to always have the same number of cells and average number of spikes as our real populations. Hence, every neuron in the homogeneous population had a firing rate equal to the average firing rate of the neurons in our measured population. Specifically, the firing probability in the homogeneous population given the target stimulus was $\bar{p} = \frac{1}{N}\sum_i p_i$, and similarly for distracter stimuli.

To gain more intuition into this approach to formalizing the question, we can imagine a population of retinal ganglion cells with either the same or different contrast tuning functions. Then the question is whether one can

better discriminate two contrasts using the heterogeneous or homogeneous population. Similarly for the visual cortex, we can imagine a population of neurons with either the same or different orientation tuning being used to discriminate between two different orientations. The answer to these questions will depend on the choice of contrasts and orientations, so we must carry out the calculation for many different choices of stimulus pairs. Of course, in our design, the two different stimuli do not differ by a single, known parameter, like contrast or orientation. However, this choice has the benefit of improving the generality of our possible conclusions because they apply over a broad range of realistic visual conditions rather than to a single, tightly controlled experimental task.

For a given pair of stimuli at times (t_1, t_2), we constructed a maximum likelihood decoder to distinguish the population activity elicited by the target stimulus, $\{p_i\}$ from the activity elicited by the distracter, $\{q_i\}$ (see section 4). A similar calculation was performed for the matched homogeneous population. Unsurprisingly, the resulting error rates depended very strongly on the particular choice of (t_1, t_2), as the neural activity elicited by some stimuli was quite high, while for many other stimuli, the population was not very active. To organize our results, we plotted the error rate as a function of the difference in firing probability for target and distractor stimuli, $\Delta \equiv |\bar{p} - \bar{q}|$, as we expect that the error rate should depend strongly on this quantity (see Figure 3). Indeed, under spatially uniform flicker, the error for the homogeneous population, ε_1, ranged from close to 0.5 (chance level) down to less than 10^{-3} for pairs of time points with very different firing probabilities (see Figure 3A, open red circles). The error rate for the real, heterogeneous population, ε_N, was strikingly lower, often by several orders of magnitude (see Figure 3A, solid blue circles). In order to focus more specifically on the difference in error rate between the homogeneous and heterogeneous populations, we calculated the ratio of the error rate for each pair of stimuli sampled, $\varepsilon_1/\varepsilon_N$. This ratio varied from close to one all the way up to almost 10^6 (see Figure 3B). (Note that the maximum measurable error ratio was limited by the numerical methods that we used to sample errors.) We found this extreme difference to be surprising and noteworthy. A similar pattern of the error rate, as well as similar values of the error, was found under stimulation by natural movie clips (see Figures 3C and 3D), indicating that this result is not specific to spatially uniform flicker.

We note that the finite sampling from which we estimate the experimental firing probability can, in itself, give rise to an apparent benefit of heterogeneity. One reason is that even if the true firing probability of a cell is identical for two stimuli, noise in the neural response will cause the estimated firing probabilities to differ. Another reason is that finite sampling will slightly enhance the degree of heterogeneity. One way to estimate the significance of this effect is to create a matched homogeneous neural response for two stimuli and resample the firing probabilities of all the cells via bootstrapping. The resulting resampled population will have different

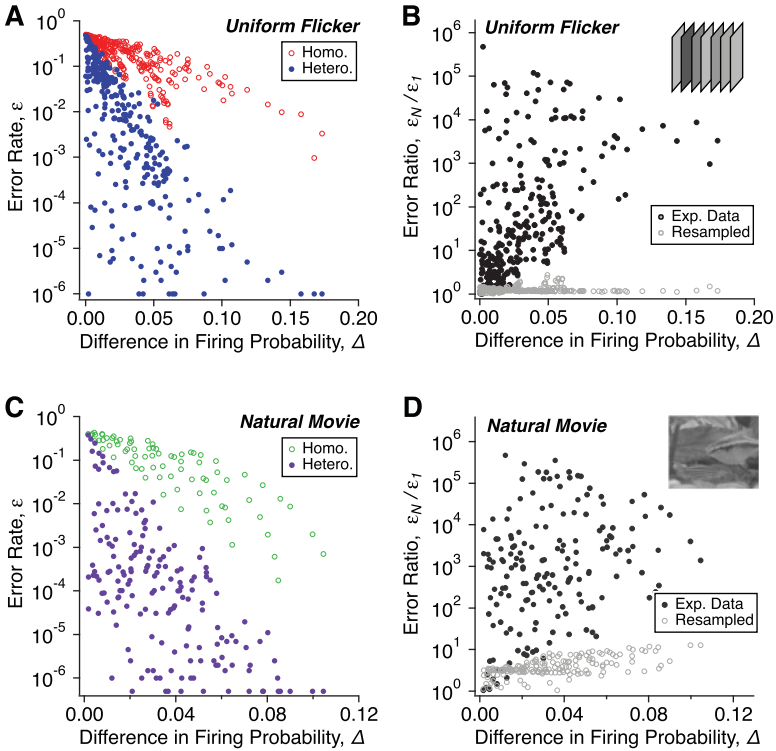

Figure 3: Discrimination error for heterogeneous versus homogeneous neural populations. (A) Error rate (on a log scale) plotted against the difference in average firing probability, Δ, for the real, heterogeneous population, ε_N, (blue circles) and the matched, homogeneous population, ε_1, (red open circles) under spatially uniform flicker. Each circle corresponds to a choice of stimulus pair. (B) Error ratio for the heterogeneous population ($\varepsilon_1/\varepsilon_N$; black dots) and error ratio due to finite sampling ($\varepsilon_1/\varepsilon_{resampled}$; gray dots) plotted on a log scale against the difference in firing probability, Δ, for the spatially uniform stimulus ensemble. (C, D) Same as panels A and B but for the natural movie stimulus ensemble.

estimated firing probabilities for all of the cells, and hence should have a lower discrimination error due to finite sampling alone. We carried out this procedure using 10 bootstrap resamples for each stimulus pair (see section 4) and calculated the ratio of the error for the homogeneous population and the error for the resampled homogeneous population. For spatially uniform flicker, where we had 300 stimulus repeats, this effect was very small (see Figure 3B, gray circles). For the natural movie, where we had only 70 repeats, this ratio was slightly larger but still far below the effect of real

heterogeneity in the overwhelming majority of instances (see Figure 3D gray circles).

While the error rates in the homogeneous population were strongly determined by the difference in firing probabilities, Δ, the error rates in the heterogeneous population varied widely for a given value of Δ, especially those near zero. Inspection revealed that pairs of stimuli with very low error and Δ close to zero had neural activity with large but similar means for target and distracter stimuli. Such patterns of neural activity were nearly indistinguishable for the matched homogeneous population, but because different neurons were active in each activity pattern, the error was very low in the real, heterogeneous population.

Another way of quantifying the effect of heterogeneity is to calculate the mutual information per cell between pairs of target and distracter stimuli (see section 4). Carrying out this analysis, we found that the information for the matched homogeneous population had a substantial range, even at a given firing difference, Δ (see Figure 4A). But, again, the information for the real, heterogeneous population was larger and had an even greater range at a given Δ. The results for the natural movie clip were similar to those for spatially uniform flicker (see Figure 4B). Plotting the heterogeneous information, I_N, versus the homogeneous information, I_1, showed that heterogeneity always improved the mutual information per cell, sometimes by large factors (see Figures 4C and 4D). As for the discrimination error, we estimated the effect that finite sampling has on increasing the mutual information by resampling the firing probabilities of the matched homogeneous population, which increased the information only marginally above the homogeneous value (see Figures 4C and 4D, gray dots). For many pairs of stimuli, the heterogeneous information was several orders of magnitude larger than the homogeneous information. This was especially true for activity patterns with nearly the same average firing probability (see Figures 4E and 4F), similar to our results for the discrimination error (see Figure 3).

Even at large Δ, where there was substantial information in the homogeneous population, heterogeneity often enhanced the information per cell by factors of two or more. We emphasize that since this information is calculated per cell, even "small" multiplicative factors such as these imply an appreciable enhancement in the information contained in the entire population.

Next, we asked whether the effect of heterogeneity was greater when the population activity was higher or lower. To this end, we displayed the ratio of information in heterogeneous versus homogeneous populations in a color scale given by the average firing probability, $^1/_2(\bar{p} + \bar{q})$ (see Figures 4E and 4F). We found that the ratio was systematically higher when the population activity was greater. This result can be interpreted by noting that neural activity in the ganglion cell population is sparse. As a result, the most common case is one in which both cells are silent in both time bins. In this case, there is no discriminability. When neural activity is higher, fewer

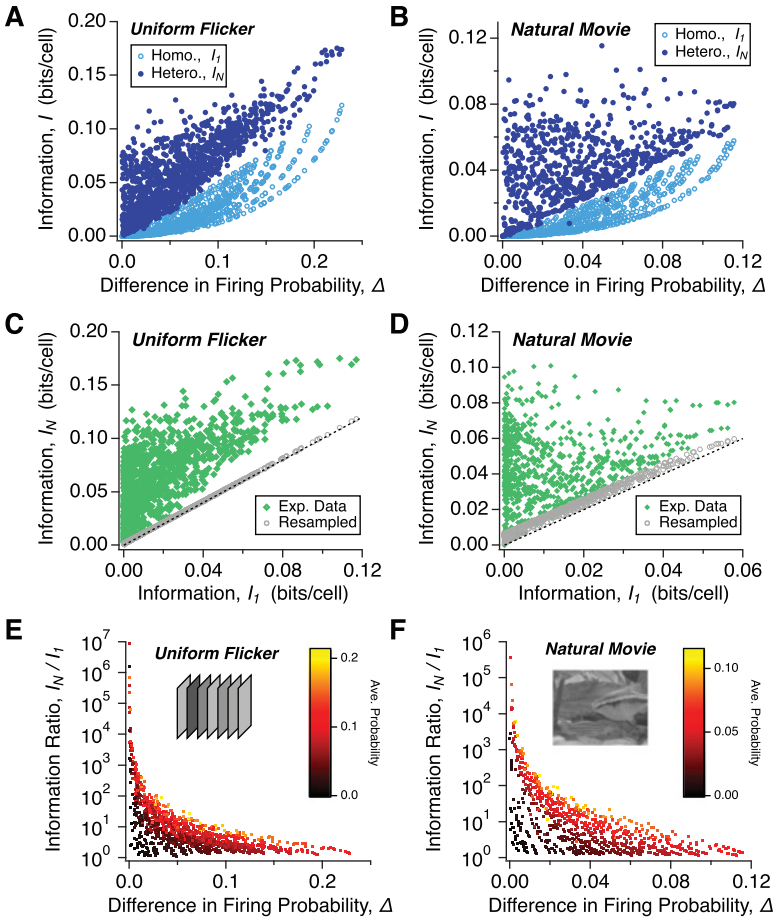

Figure 4: Mutual information for heterogeneous versus homogeneous neural populations. (A) Mutual information plotted against the difference in average firing probability, Δ, for the real, heterogeneous population, I_N, (solid circles), and the matched, homogeneous population, I_1, (open circles) for spatially uniform flicker; each circle corresponds to a choice of stimulus pair. (B) Same as panel A but for the natural movie stimulus ensemble. (C) Mutual information for the real, heterogeneous population, I_N (solid diamonds), and for resampled data (open diamonds) plotted against that for the homogeneous population, I_1, for spatially uniform flicker. (D) Same as panel C but for the natural movie stimulus ensemble. (E) Information ratio, $(I_N - I_{bias})/I_1$, plotted against the difference in average firing probability, Δ, for spatially uniform flicker. Color scale indicates average firing probability, $^1/_2(\bar{p} + \bar{q})$. (F) Same as panel D but for the natural movie stimulus ensemble.

cells have zero firing for both stimuli, and hence the population information is higher. This comparison also addresses the question of the importance of heterogeneity when the stimulus is well tuned to drive neural responses. In this case, average neural activity would be higher, leading to a stronger effect of heterogeneity.

Another important question is whether the effects that we report depend strongly on individual cells having perfectly reliable responses. First, cells that apparently come with perfect reliability, $p_i = 1$ and $q_i = 0$, or vice versa, are an artifact of finite data and how we estimate probabilities from those data. No cell is truly perfectly reliable. More broadly, one way to probe the role of highly reliable cells quantitatively is to calculate the maximum information encoded by a cell and compare it against the sum of information across cells. If a single cell typically dominates the population information, the maximum will be nearly equal to the sum. What we found instead was that the maximum was roughly 0.1 of the sum for stimulus pairs with the largest information. So, overall, information about different pairs of stimuli was broadly distributed throughout the population. It is still possible that a small number of cells do dominate the discriminability for particular pairs of stimuli. But single-cell coding cannot account for our conclusions.

We can extend the generality of our results by considering the Chernoff distance, as a measure quantifying coding fidelity in neural populations (Kang, Shapley, & Sompolinsky, 2004; Kang & Sompolinsky, 2001). Specifically, it describes the asymptotic limit of the mutual information that activity in a large neural population represents about an ensemble of stimuli. In this limit, the information is dominated by the distance between the "closest" pair of stimuli. Thus, the Chernoff distance can also be interpreted as a measure of coding that is defined between two stimuli. This measure ranges from zero, when no discriminability is possible, to infinity, when discrimination is perfect. Furthermore, the Chernoff distance is readily calculated in our case, the asymptotic limit (see section 4).

We found that the Chernoff distance was systematically much larger for the fully heterogeneous population than the matched homogeneous population (see Figures 5A and 5B), consistent with our results on both the discrimination error and the mutual information (see Figures 3 and 4). Part of the reason for this consistency is the fact that the Chernoff distance tracks both the error and the information (see Figures 5C and 5D). Because the Chernoff distance is closely correlated with both quantities, its evaluation helps confirm that these different measures yield consistent results on the benefit of heterogeneity in neural populations.

2.3 Coding Fidelity for Graded Levels of Heterogeneity. So far, we have compared the real heterogeneous neural population to a matched population where every neuron had identical stimulus tuning. While there certainly are some contexts in which it has been fruitful to analyze neural populations in terms of their average firing rate—for example, integration

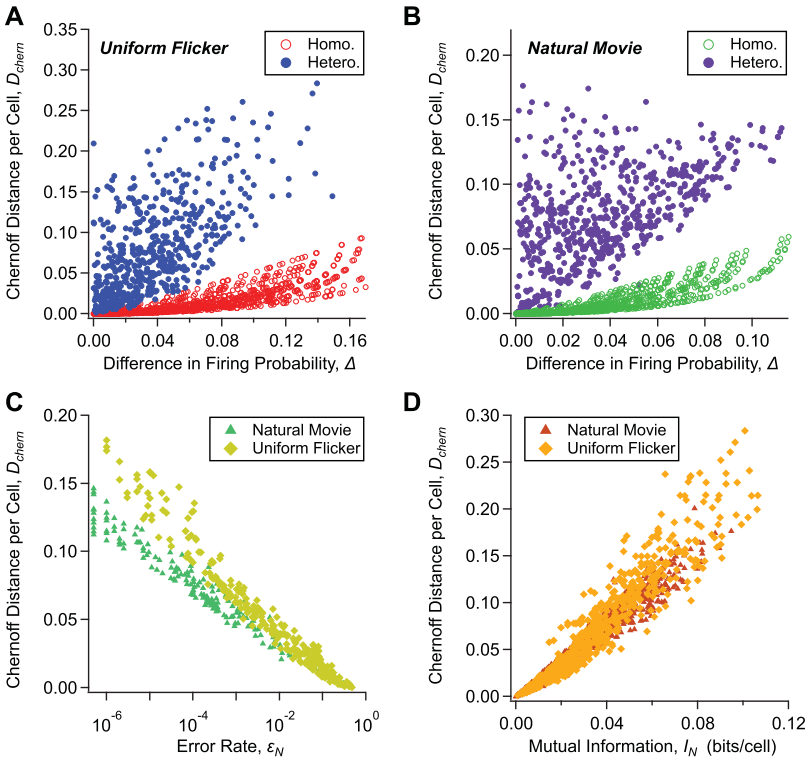

Figure 5: Chernoff distance. (A, B) Chernoff distance per cell, D_{chern}, plotted against the difference in average firing probability, Δ, for the real heterogeneous population (blue circles) and the matched homogeneous population (red open circles) under spatially uniform flicker (A) and natural movie stimulation (B). Each circle corresponds to a choice of stimulus pair. (C) Chernoff distance per cell plotted against the error rate for the heterogeneous population, ε_N. (D) Chernoff distance per cell plotted against the mutual information for the heterogeneous population, I_N.

of sensory evidence in cortical area LIP (Roitman & Shadlen, 2002)—this is a somewhat caricatured limit. For instance, many classic studies of neural coding, such as the discrimination of the direction of random dot motion in cortical area MT (Newsome, Britten, & Movshon, 1989), have divided the neural population into two pools: neurons with the stimulus tuned to their peak or preferred direction versus neurons with tuning in the antipreferred direction. It is thus of interest to compare a fully heterogeneous population to populations with a coarser form of heterogeneity, as in the case of a population divided into preferred and antipreferred pools.

We can accomplish this by defining one pool ("preferred") as all of the neurons that have a higher firing probability for the target stimulus, $p_i \geq q_i$, and the other pool ("antipreferred") as the remainder of the population (see Figure 6A). We then form a matched neural population with two pools by computing the average firing probability for target and distracter for the N_1 neurons in pool 1, $\bar{p}^{(1)}$ and $\bar{q}^{(1)}$, respectively, and similarly for the N_2 neurons in pool 2, $\bar{p}^{(2)}$ and $\bar{q}^{(2)}$ (depicted in Figure 6A by crosses). In this case, the two-pool population can be characterized by the spike count in pool 1, k_1, and that in pool 2, k_2, allowing us to calculate the discrimination error exactly (see section 4).

We can further subdivide the neural population into four pools. Here, we take the neurons in the preferred pool and divide them into equal groups having the top half versus the bottom half of the firing probabilities to the target stimulus (see Figure 6B). Similarly, the antipreferred pool can be divided into equal groups of neurons having the top half and the bottom half of firing probabilities for the distracter stimulus, respectively. The state of the four-pool neural population is uniquely described by four spike count variables, $\{k_1, k_2, k_3, k_4\}$, again allowing an exact calculation of the discrimination error. More generally, we can divide the neural population into any even number of pools, L, using an analogous method. For instance, we can form an eight-pool population by dividing the preferred pool into four groups with rank-ordered quartiles of target firing probabilities and the same for the antipreferred pool. For populations with more than four pools, it becomes unwieldy to perform the exact computation, so we instead relied on the same Monte Carlo sampling methods used for the real, fully heterogeneous population.

We found that the discrimination error decreased in a graded fashion as we increased the degree of heterogeneity. Figure 5C shows example plots of error versus the number of pools subdividing the population, L, for several different choices of stimulus pairs. Although the error for the homogeneous case takes a range of values (as seen in Figure 3), the error decreased continuously as we divided the population into more pools.

This led us to compute another statistic, the improvement factor, λ, defined as the multiplicative factor by which the average error decreases when the number of pools is increased by a factor of 2, $\lambda(L) \equiv \varepsilon_{L/2}/\varepsilon_L$. The improvement factor was relatively large when dividing the homogeneous population into 2 pools ($\lambda \sim 5.5$), then settled down to a value $\lambda \sim 2$–3 up to $L = 32$ pools, and finally rose again as full heterogeneity was obtained (see Figure 6D). This behavior implies that the error decreased roughly as a power law function of the pool number L with an exponent in the range of two to three.

We also calculated the mutual information per cell as a function of the degree of heterogeneity. We found that it rose gradually and monotonically from less than 0.02 bit/cell for the homogeneous population to more than 0.06 bit/cell for the real, fully heterogeneous population (see Figure 6E).

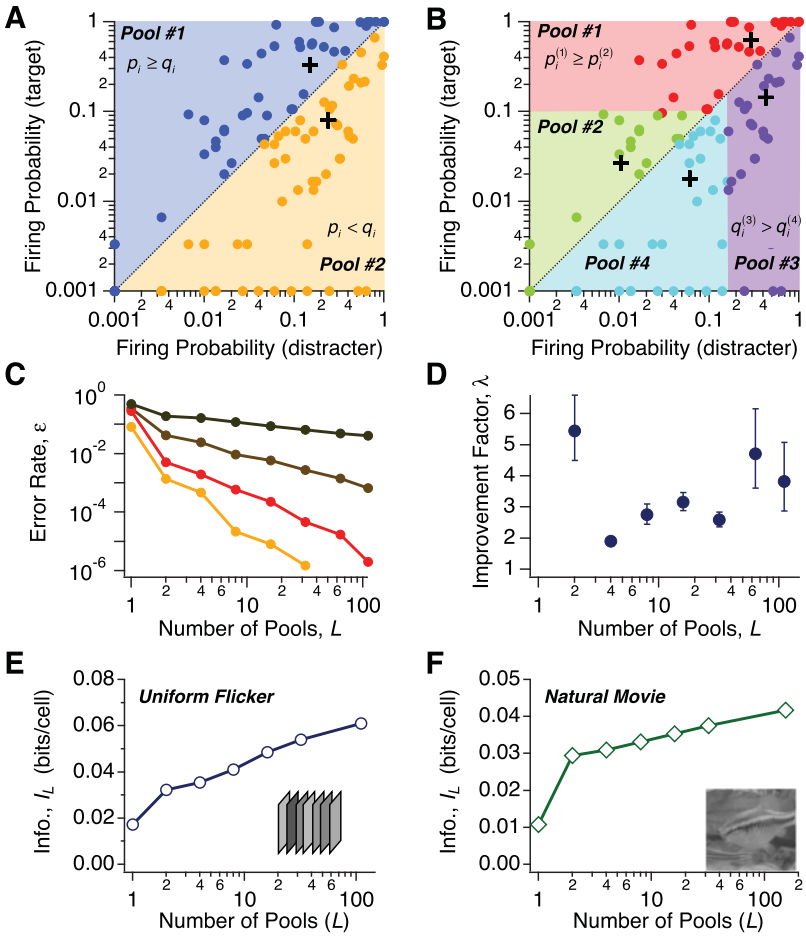

Figure 6: Coding performance for graded levels of heterogeneity. (A) Schematic showing how matched, two-pool populations were formed from the measured neural activity patterns. Each point is the firing probability $\{p_i, q_i\}$ for target and distracter stimuli, respectively; colored regions correspond to the set of firing probabilities falling into a given pool; crosses depict the average firing probabilities within each pool. (B) Same as panel A but for a subdivision into four pools. (C) Error rate (on a log scale) plotted against the number of pools, L, for four different choices of stimulus pairs (shown in different colors). (D) Improvement factor, $\lambda(L) \equiv \varepsilon_{L/2}/\varepsilon_L$, plotted against the number of pools, L. (E) Average mutual information per cell plotted against the number of pools, L, for spatially uniform stimulation. (F) Same as panel E but for the natural movie stimulus ensemble.

A similar trend was obtained under natural movie stimulation, with somewhat lower overall values, presumably due to the lower average firing rate of ganglion cells for this stimulus ensemble (see Figure 6F; 0.91 ± 0.43 spikes/sec for spatially uniform stimulation versus 0.42 ± 0.18 spikes/sec for the natural movie clip). In fact, under natural movie stimulation, the information increased linearly with $\log(L)$. Together, these results indicate that increasing the degree of heterogeneity in the neural population increases the fidelity of the population code, in a graded fashion.

2.4 Heterogeneity Arising from Different Cell Types. The natural interpretation of the pools that we have formed is that they correspond to neurons having similar tuning properties. For instance, in the case of motion direction discrimination in area MT, neurons in the preferred and antipreferred pools have opposite direction selectivity and hence belong to different direction columns. In the retina, an obvious method of choosing functional pools is to assign ganglion cells of the same functional type to a single pool. Following previous classification methods for the salamander retina, we divided the ganglion cells into eight functional types based on their reverse correlation under spatially uniform flicker (Marre et al., 2012; Segev, Puchalla, & Berry, 2006; Warland et al., 1997) (see Figures 7A and 7B).

We can explore graded levels of heterogeneity by splitting the neural population into successively more refined pools, dividing the population according to increasingly fine criteria of selectivity. For instance, it is natural to split the entire population into two pools formed from ON and OFF cells. Next, we can form four pools by dividing the OFF cells into fast, medium, and slow OFF, which have been distinguished in previous studies (Chen et al., 2013; Keat, Reinagel, Reid, & Meister, 2001; Warland et al., 1997). We can also form more than 8 pools by further splitting the 8 main cell types into 16 or 32 types (see section 4). We note that salamander ganglion cells have never previously been divided into 16 or 32 cell types, and we are not claiming that we are providing evidence that the salamander truly possesses this many functional types of ganglion cells. But we have two motivations for performing this analysis. First, the most current estimates of the total number of ganglion cell types in several mammalian species are in the range of 20 to even 40 types (Baden et al., 2016; Seung & Sumbul, 2014), considerably more than the 7 or 8 types typically described in the salamander. Second, this allows us to study greater levels of heterogeneity corresponding to finer gradations of the functional differences among neurons.

Similar to the results of our analysis of the discrimination error as a function of the number of neural pools, L, here the error rate decreased monotonically as we split the retinal population into more cell types. For two to eight cell types, the error ratio λ was modest, but significantly greater than one (see Figure 7C). Interestingly, when we split the population into 16 or 32 cell types, the error ratio was substantially larger. The mutual information per cell, as before, increased monotonically with the number of cell

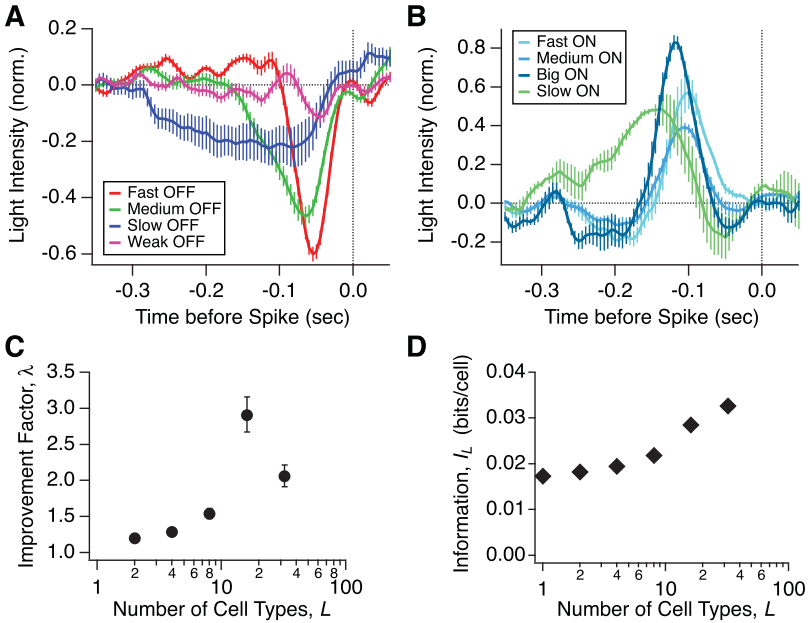

Figure 7: Heterogeneity defined by cell types and its impact on coding. (A) Reverse correlation during spatially uniform flicker averaged across all cells of the same function type (shown in different colors) for OFF ganglion cells; error bars represent standard error. (B) Same as panel A but for ON ganglion cells. (C) Improvement factor for the discrimination error, λ, plotted against the number of cell types, for spatially uniform flicker. (D) Average mutual information per cell plotted against the number of cell types, for spatially uniform flicker.

types, again showing the largest changes at 16 and 32 cell types (see Figure 7D). This analysis suggests that the presence of a large number of cell types among the retinal ganglion cells serves a beneficial purpose for encoding visual information.

2.5 Scaling of the Discrimination Error with Population Size. It is interesting to ask how the effect of heterogeneity varies with the size of the neural population, both because this allows us to relate our study to many others that have involved fewer neurons and because it gives us some expectation for what might be observed with even larger populations. We studied this trend by randomly selecting subsets of our recorded ganglion cells and calculating the discrimination error. We first chose two stimulus pairs, one with moderately low discrimination error (see Figure 8A) and the other with very low error (see Figure 8B). In each case, we carried out this calculation for different degrees of heterogeneity (different colors). For

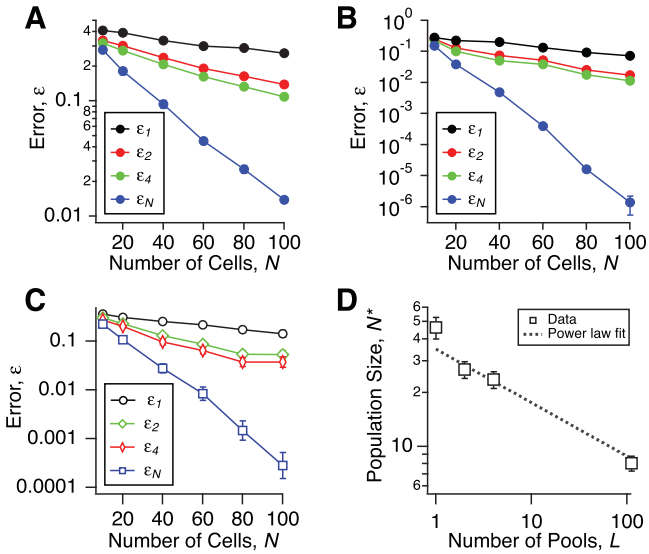

Figure 8: Scaling of the heterogeneity effect with population size. (A, B) Error rate (on a log scale) plotted against the number of cells, N, in subsets of the population, for two different choices of stimulus pair. Error bars represent standard error across 30 random subset selections. (C) Discrimination error (on a log scale) plotted against the number of cells, N, in subsets of the population, for different levels of heterogeneity (1-,2-,4-pool and full heterogeneity; shown as colors); error bars represent standard error across different choices of stimulus pair. (D) Characteristic population scale, N^*, plotted against the number of pools, L; error bars represent standard error across different choices of stimulus pair.

both examples, the error decreased approximately exponentially with increasing population size, N. But notably, the rate of this decrease depended appreciably on the degree of heterogeneity, with steeper slopes for greater heterogeneity. Averaging similar calculations over many choices of stimulus pair, we found that this effect was robust (see Figure 8C). (Because the discrimination error was more naturally distributed on a logarithmic than linear scale, all averages over error rates here and elsewhere in the letter were geometric means, not arithmetic means.)

A similar qualitative trend emerged from all the analyses: the rate at which the discrimination error decreased as a function of population size, N, was steeper for greater degrees of heterogeneity, parameterized by the number of pools, L, into which the population was divided. Given the large number of neurons available in any brain area, trends as a function of population size are important properties of the population neural code. In all

cases, the functional dependence appeared to follow a simple exponential form. As we will see in the following section, this behavior is expected in populations of independent neurons. Thus, we define a characteristic population size, N^*, by fitting the discrimination error, $\varepsilon(N)$, to an exponential form, $\exp(-N/N^*)$. This scale, N^*, measures the number of neurons that must be added to the population for the error to be reduced by a factor of e.

As the behavior of the error depended on the choice of stimulus pair as well as the number of pools, L, we calculated the characteristic size individually for each condition. In order to see how the error scaled with the degree of heterogeneity, we averaged values of N^* across all pairs of stimuli for the same value of L. The characteristic size varied widely from $N^* = 46.4 \pm 6.5$ (mean $\pm$ SE, $n = 55$) for the matched homogeneous populations down to $N^* = 8.1 \pm 0.77$ for full heterogeneity (see Figure 8D). After a steeper drop from one pool to two pools, the value of N^* was fit well by a power law form with an exponent, gamma $= -0.29 \pm 0.04$. Another way of thinking about this effect is as follows. If the number of neurons that process the population code is constrained, then it is advantageous to break up the population into pools with distinct functional properties. Equivalently, heterogeneity has an amplifying effect on the code, since by reducing N^*, it enhances the effective size of the population. When there are more distinct functional pools in a population, the extra performance gained by adding each neuron is boosted.

2.6 Why Does Heterogeneity Improve the Population Code? In all the examples studied in our analysis of neural data, heterogeneity reduced the error in discriminating between two stimuli and increased the mutual information per neuron. The consistency of this result naturally led us to think that the beneficial effect of heterogeneity might not be a fortuitous consequence of the statistics of neural activity in retinal ganglion cells under particular visual conditions, but might instead be a rather general property of population neural codes. To explore the effect of heterogeneity in greater generality, we examined neural population coding theoretically, from different perspectives and using different models that we describe hereafter.

2.6.1 Simple Illustration: Homogeneous versus Two-Pool Populations. We begin with a homogeneous neural population of N neurons, where the firing probability is p for the target stimulus and q for the distracter. In this simple situation, the state of the population is defined by the spike count, k, the number of neurons that fire among the N cells in the population, and we can easily write down its probability distribution (see equations 9a and 9b). Using this result, we can calculate the discrimination error as a function of N, which decayed exponentially for large N (see Figure 9A). As the firing probabilities, (p, q), are varied, the trend remains exponential, but the rate of decay changes. Similar to our analysis of real data described above, we were led to define a characteristic population size, $N^*(p, q)$, as the inverse

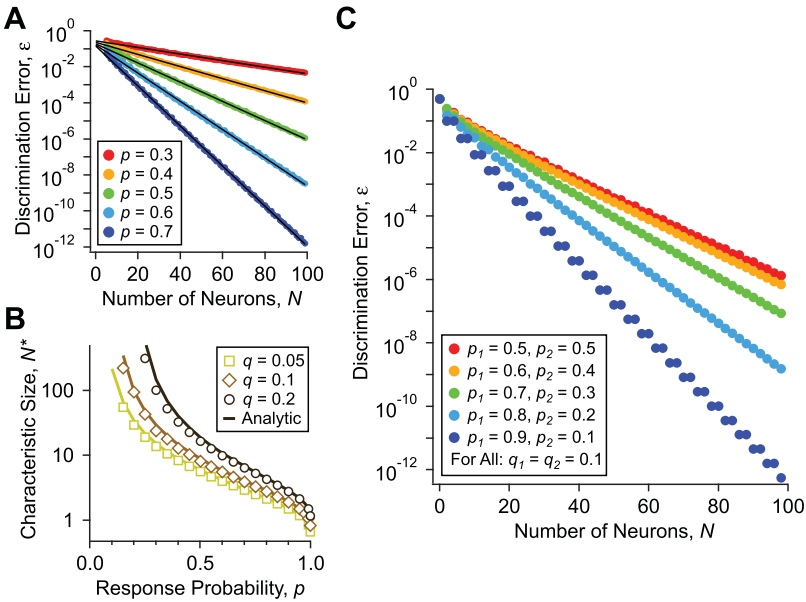

Figure 9: Analysis of one- and two-pool models. (A) Discrimination error of a homogeneous (one-pool) neural population, ε_1, plotted as a function of the number of neurons, N, for different choices of firing probabilities, p and q (shown in different colors); exponential curve fits shown as black lines. (B) Characteristic population scale, N^*, plotted as a function of the target firing probability, p, with curves representing different values of the distracter firing probability, q (gray-scale symbols); analytic approximation of N^* from equation 4.11 (gray-scale lines). (C) Discrimination error of a heterogeneous (two-pool) neural population, ε_2, plotted as a function of the number of neurons, N, for different choices of firing probabilities, p_1, p_2, and q (shown in different colors).

exponential rate. This function has a nontrivial and strong dependence on the firing probabilities (p, q), with values ranging from less than 1 for $p \sim 1$ and $q \sim 0$ and diverging as $p = q$ (see Figure 9B). We can derive an analytic formula for the quantity N^*, which corresponds to the characteristic system size beyond which homogeneous coding becomes faithful, as a function of the firing probabilities p and q (see section 4). The analytic form agrees well with direct numerical calculations (see Figure 9B, solid lines).

Next, we introduced heterogeneity by splitting the homogeneous population into two pools. For clarity, we change only the firing probability for the target stimulus, $p \to (p_1, p_2)$, while keeping the mean number of spikes the same, $p = {}^1\!/_2(p_1 + p_2)$. Examining the trend of error versus the number of neurons, N, we found that the rate of decay was the shallowest when $p_1 = p_2$ (the homogeneous case) and became increasingly steep as the

difference between p_1 and p_2 increased (see Figure 9C). This effect was particularly striking for $p_1 = 0.9$, $p_2 = 0.1$ (blue points); in this case, the neurons in the second pool offered no help at all in the stimulus discrimination, as the firing probability for the distracter was $q = 0.1$. One way of understanding this result is by reference to the behavior of $N^*(p, q)$: this function is strongly nonlinear, such that the increased separation between p_1 and q for one neural pool more than compensates for the decreased separation between p_2 and q in the other pool. In other words, the enhancement of the coding performance of one pool with neurons having better separated firing rates (from the firing rate in response to the distracter stimulus) generically exceeds the suppression of the coding performance in the other pool having more similar firing rates.

2.6.2 Suppression of the Discrimination Error by Neuron-to-Neuron Variability. Intuition suggests that the above argument carries over to cases with more general forms of heterogeneity. Here, we investigate one such case, in which heterogeneity takes the form of neuron-to-neuron variability. While we find that, again, heterogeneity favors coding, the approach provides us with a complementary picture of why this is true.

We consider two population models. The first is homogeneous, with each neuron firing with probability p in response to the target stimulus and probability q in response to the distracter stimulus. In the second model population, we perturb the homogeneous firing probabilities as

$$\begin{aligned} p_i &= p + \delta p_i, \\ q_i &= q + \delta q_i, \end{aligned} \tag{2.1}$$

where we assume $\delta p_i \ll p$ and $\delta q_i \ll q$, and where the label $i = 1, \ldots, N$ runs over all the neurons in the population. For a fair comparison, we further assume that the perturbations leave the total population firing probability unchanged, that is, we assume

$$\sum_{i=1}^{N} \delta p_i = \sum_{i=1}^{N} \delta q_i = 0. \tag{2.2}$$

The spike count in the population, in the homogeneous case, is given by equations 2.9a and 2.9b for the target and distracter, respectively. In the perturbed system, because of the variability among neurons, the decision boundary has to be considered in the N-dimensional space of the population activity. However, we can obtain an upper bound to the error if we reduce the problem to one of spike count coding, that is, if we treat the spike count as the coding variable. In the perturbed model population, spike counts are distributed as

$$\frac{P_{hetero}(k)}{P_{homo}(k)} = 1 - \left[\frac{k(k-1)}{N(N-1)} - \frac{2pk}{N} + p^2\right] \sum_{i=1}^{N} (\delta p_i)^2 + O\left(\delta p^3\right) \tag{2.3}$$

(see section 4). In this one-dimensional approximation to our problem, the coding precision can be quantified by the variances of the two distributions corresponding to the homogeneous and heterogeneous cases: the narrower the distribution, the smaller the discrimination error. When we calculate the variances of these two distributions, using expressions for the moments of the binomial distribution, we obtain a ratio:

$$\frac{Var_{hetero}}{Var_{homo}} = 1 - \left[1 + \frac{2p(1-p)}{N}\right] \sum_{i=1}^{N} (\delta p_i)^2 + O\left(\delta p^3\right). \tag{2.4}$$

Notice that this ratio is less than one. Thus, a perturbative amount of heterogeneity in the firing probabilities of individual neurons always suppresses the width of the distribution of spike counts and in turn suppresses the discrimination error.

2.6.3 Enhancement of the Mutual Information by Neuron-to-Neuron Variability. The benefit of heterogeneity can similarly be seen by considering the behavior of the mutual information. Again, we compare a homogeneous population to a heterogeneous population. But here, we consider general neuron-to-neuron variability in firing rate; in particular, the magnitude of this variability need not be small. A limit in which one can derive a powerful and general rule is the case where the target stimulus is rare—namely, it occurs with probability $P(T) = \rho$, with $\rho \ll 1$. This is the case, for example, if one is trying to recognize one target stimulus versus all other stimuli (Schwartz, Macke, Amodei, Tang, & Berry, 2012) or if one is trying to recognize a target stimulus class that is a small subset of all possible stimuli (such as one person's face versus any other person in a large group of individuals).

With these assumptions, we can write down the mutual information as

$$I_{hetero} = \rho \sum_{i=1}^{N} \left[p_i \ln\left(\frac{p_i}{q_i}\right) + (1-p_i)\ln\left(\frac{1-p_i}{1-q_i}\right)\right] + O\left(\delta p^2\right), \tag{2.5}$$

where p_i is the firing probability of neuron i in response to the target stimulus and q_i is the firing rate of neuron i in response to the distracter stimulus (see section 4). The corresponding mutual information for the homogeneous population is

$$I_{homo} = \rho \sum_{i=1}^{N} \left[\bar{p} \ln \left(\frac{\bar{p}}{\bar{q}} \right) + (1 - \bar{p}) \ln \left(\frac{1 - \bar{p}}{1 - \bar{q}} \right) \right] + O\left(\delta p^2\right). \tag{2.6}$$

One can then show that the terms in these equations obey the inequality (see section 4):

$$\sum_{i=1}^{N} p_i \ln \left(\frac{p_i}{q_i} \right) \geq N \, \bar{p} \ln \left(\frac{\bar{p}}{\bar{q}} \right). \tag{2.7}$$

This implies that in the case of a rare target stimulus at least, the mutual information between stimulus and response is always larger in a heterogeneous population than in a homogeneous population, no matter the form of the neuron-to-neuron heterogeneity. The generality of the argument supports the intuition on the generic benefit of heterogeneity for population neural coding.

3 Discussion

We have studied the role of functional heterogeneity in a population neural code using two mutually reinforcing approaches. First, we have analyzed experimental data from multielectrode recordings of populations of over 100 retinal ganglion cells. Here, we found that the error for discriminating between pairs of visual stimuli was often many orders of magnitude lower for the real heterogeneous population compared to a matched homogeneous population having the same number of cells and spikes (see Figure 3). Similarly, the mutual information about stimulus identity was also enhanced, often by more than an order of magnitude for the heterogeneous population versus the homogeneous one (see Figure 4). This heterogeneity effect depended strongly on the population size, with greater improvement for larger population sizes (see Figure 8).

Second, we have analyzed theoretically the fidelity of the neural population code in a simple model and in two broad limits. In one limit, we considered any arbitrary but small perturbation of each cell's firing rate away from a perfectly homogeneous population. We showed that this perturbation always decreases the discrimination error, regardless of the initial firing rates. In the other limit, we considered any possible set of firing rates within a neural population, but in the case in which the target stimulus was rare. Here, heterogeneity always increased the mutual information, that is, the information contained in the population about whether the target was present. These analytic proofs substantially increase the generality of the finding that heterogeneity benefits the neural population code.

In our analyses, we used a flexible framework in which the characteristics of the neural population were summarized by the set of firing

probabilities for all cells, in response to a target stimulus, $\{p_i\}$ and to a distracter stimulus, $\{q_i\}$. These simple properties are readily measured in experiment and can be defined for any pair of stimuli or conditions. But at the same time, this approach does not address the potential trade-off implicit in the design of a neural circuit that attempts to achieve heterogeneous responses across an entire stimulus ensemble. Our results therefore complement previous studies that have considered this latter problem by assuming a specific form of tuning curve or receptive field model (Ecker et al., 2011; Kastner et al., 2015; Shamir & Sompolinsky, 2006; Wilke et al., 2001).

One strength of our approach is that we do not have to make any explicit assumptions about the response functions of neurons. For the benefit of tractability and concreteness, previous studies have often used models of the neural response that are incomplete or inaccurate. For instance, an orientation tuning curve for a V1 neuron does not contain any prediction about how the neuron will respond to a stimulus that is not an oriented grating, and the linear-nonlinear (LN) model of a ganglion cell's receptive field breaks down under many visual conditions (Barlow & Levick, 1965; Chen, Chou, Park, Schwartz, & Berry II, 2014; Clark, Benichou, Meister, & Azeredo da Silveira, 2013; DeVries, 2000; Olveczky, Baccus, & Meister, 2003; van Hateren, Ruttiger, Sun, & Lee, 2002). Many of the ways in which the real light responses of retinal ganglion cells deviate from simplified models, like the LN model, introduce additional heterogeneity among neurons. For instance, spatial hot spots within each receptive field reduce the redundancy of spatial information distributed among ganglion cells with similar, overlapping receptive fields (Soo, Schwartz, Sadeghi, & Berry, 2011), and realistic variations in the receptive field shape break the symmetry among ganglion cells in the same mosaic (Liu, Stevens, & Sharpee, 2009). Our results therefore imply that many of the complexities of neural circuits that are not captured by even state-of-the-art functional models can potentially play a positive role in improving the fidelity of the population neural code.

Another notable difference between our results and previous ones is the sheer effect size that we have observed: over 10-fold increases in the mutual information per neuron and over 10^5-fold decrease in the discrimination error. One major source of this discrepancy is that we have analyzed larger populations than most previous studies have. This matters, because we have shown that the effect of heterogeneity depends strongly on the number of neurons in the population. There is every indication that this trend continues for even larger populations, making heterogeneity an even more relevant property for the realistically large neural populations that operate in many local neural circuits.

Other factors are more technical: we report the effect for discrimination between specific pairs of stimuli rather than for the average information over an entire stimulus ensemble. Since the effect of heterogeneity is, of course, negligible when neurons do not fire, averages that include frequent periods of silence will make the effect appear smaller than it is during

periods of substantial neural activity. In any case, the large effect sizes that we observe point to an even greater potential than previously appreciated for neural circuits to use functional diversity in encoding information. In addition, we assessed the true coding fidelity of neural populations using maximum likelihood decoders based on measured firing rates. Conversely, studies that use cross-validated decoders cannot infer error rates that are smaller than the inverse number of trials (see section 4). However, we did estimate how much of the heterogeneity effect was due to finite sampling in our measurement of each cell's firing rate by resampling the responses of homogeneous neural populations to estimate realistic levels of heterogeneity due simply to finite sampling of our measured neural responses and then recalculating the coding fidelity. We found that our results for the effect of heterogeneity in measured neural activity were significantly larger than for the degree of heterogeneity due to finite sampling.

3.1 Limitations of the Current Study. While our study has made strides in demonstrating a wide range of circumstances in which heterogeneity is beneficial to the population neural code, we have left unexplored two important directions. First, we have disregarded noise correlation and its possible role in coding. This is a broad topic that has been the subject of many previous studies. Using a framework of parameterized tuning curves and Fisher information to evaluate the fidelity of coding in neural populations, several important studies have found that heterogeneity in the tuning curves can help reduce the deleterious impact of positive noise correlation or synchrony (Ecker et al., 2011; Padmanabhan & Urban, 2010; Shamir & Sompolinsky, 2006; Wilke et al., 2001). Another study added realistic levels of noise correlation to experimentally measured tuning curves and found that the benefit of heterogeneity survived (Osborne et al., 2008). Yet another study explicitly added heterogeneity to the distribution of pairwise correlations and found that this enhanced coding fidelity (Azeredo da Silveira & Berry, 2014). Taken together, these results suggest that the beneficial effects of heterogeneity will extend to the case in which a correlation structure among neurons is included, and in fact the benefits may even be enhanced.

Second, we have treated the response of each neuron as a firing probability in a small time bin (here 20 ms). The choice of this time bin is appropriate for the retinal code, as ganglion cells have roughly this level of temporal precision (Berry et al., 1997; Uzzell & Chichilnisky, 2004; van Rossum, O'Brien, & Smith, 2003). In such a small time bin, most ganglion cells fire zero or one spike. While it is possible for a cell to fire two or more spikes, most of the coding power of the neural population is contained in the binary response of each neuron (see Schwartz et al., 2012, where discrimination errors were compared for a binary code versus a spike count code). However, one limitation of this approximation is that each neuron has a fixed Poisson level of noise. It would be interesting to see how the effect of heterogeneity may vary with non-Poisson noise statistics.

3.2 What Is the Purpose of So Many Ganglion Cell Types? One of the puzzles about the organization of the vertebrate retina is why there are so many different types of ganglion cells with overlapping receptive fields. Of course, some ganglion cell types project to unique brain centers and carry out qualitatively distinct visual computations, like ON direction-selective cells that project only to the accessory optic system (Vaney, Peichl, Wassle, & Illing, 1981) in the brain stem and convey an error signal corresponding to retinal image slip to the cerebellum, relevant to adjusting the gain of the vestibulo-ocular reflex (Raymond, Lisberger, & Mauk, 1996). Other cell types appear to have a clear function, such as the M1 melanopsin-containing cell, which measures light level over a long integration time (Berson, 2003) and projects not only to the suprachiasmatic nucleus, where it helps entrain the circadian rhythm, but also to other brain regions, like the superior colliculus (Hattar et al., 2006). However, many ganglion cell types project to the two major visual brain centers: the lateral geniculate nucleus and the superior colliculus (Berson, 2008; Dacey, Peterson, Robinson, & Gamlin, 2003). Visual information encoded by these different cell types are then combined by downstream neural circuits, for example, in the primary visual cortex, yielding a modified code representing the same region of visual space (Berson, 2008; Rodieck, 1979). So the question remains: Why are there so many ganglion cell types?

Our work offers one possible interpretation: a multiplicity of cell types helps to form a heterogeneous population code that can represent visual information more faithfully or over a broader range of stimulus patterns, as compared to a neural code using the same number of less diverse neurons. Specifically, we found that when we divided our recorded ganglion cell population into more and more cell types, the fidelity of the population code increased (see Figure 7). A functionally broad array of "sensors" allows a diverse population to capture more aspects of the input. By contrast, in a less diverse population, noise in the output of the subsets of identical sensors is averaged out more thoroughly. Our results indicate that this trade-off is biased significantly in favor of diversity. Because the effect of heterogeneity is so strong, it overcompensates for the deleterious effect of increased noise.

One mechanism to achieve functional diversity in neural populations is that of developmental noise: neurons can have the same genetic program that determines their synaptic contacts, but various sources of biophysical noise can still cause some degree of variability in the cell's synaptic circuit. However, this mechanism might not be sufficient to harness the full benefits of functional heterogeneity; components of, for instance, thermal noise acting on a scale much more modest than that of the neuron will sum to a small collective effect due to the law of large numbers. Instead, a better mechanism might be to have a set of different developmental programs that force neurons within the population to specialize their function and thereby achieve greater heterogeneity (see Figure 10). Given the broad range of

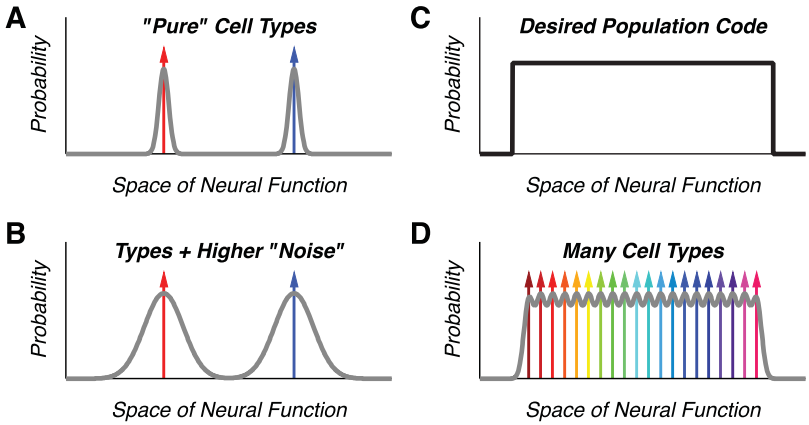

Figure 10: Illustration of the benefits of functional heterogeneity. (A) Illustration of a local neural circuit that contains two well-separated cell types. The probability of finding a cell with a given set of functional properties (gray curve) is given by two distributions centered on the mean of each of the two cell types (red and blue arrows). (B) With higher developmental noise, there is greater scatter of functional properties around the mean of the two cell types. (C) Given the benefits of heterogeneity, the desired choice of functional properties within the neural population is a broad distribution (black curve). (D) A good approximation (gray curve) to the desired distribution of functional properties (black curve, panel C) combines developmental noise with multiple cell types, each having a different mean functional characteristic (different colored arrows).

conditions in which heterogeneity benefits the population neural code, developmental noise can then be expected to provide additional benefit even in a population divided into many cell types. And, in fact, we found that the experimental, fully heterogeneous population substantially outperformed a code with 32 cell types (in Figure 6C, the visual information encoded was 0.061 bit/cells for $L = 111$ cells, while in Figure 7D, the information was 0.032 bits/cell for $L = 32$ cell types). Developmental noise may thus be capitalized on by downstream circuits to represent information at a finer resolution.

4 Methods

4.1 Multielectrode Recording. We used a multielectrode array to record spike trains from large populations of retinal ganglion cells in the larval tiger salamander, a method that has been described elsewhere (Marre et al., 2012; Puchalla, Schneidman, Harris, & Berry, 2005). In brief, we euthanized animals according to institutional standards (IACUC protocol 1828: rapid decapitation following ice water anesthesia), dissected the retina out

of the eye, cut a piece of a size roughly one-third of the entire retina, and placed the tissue ganglion-side down against the array. Retinas were held in place with a dialysis membrane that was mounted on a gantry that allowed precise vertically displacements by turning a screw. Oxygenated Ringer's solution was perfused over the tissue to keep it alive for many hours. Spike sorting was carried out with a custom-written algorithm (Marre et al., 2012).

4.2 Visual Stimulation. Visual stimuli were generated on a computer monitor whose light was focused on the photoreceptor layer of the retina (Puchalla et al., 2005). The mean light level was 11 $\mathrm{mW/m^2}$, which corresponds to phototopic vision in the salamander. Spatially uniform flicker consisted of light intensity values that were randomly drawn from a gaussian distribution every 8.33 ms. The width of gaussian distribution defined a temporal contrast of 33% of the mean. A 30 sec segment was repeated 300 times. The natural movie consisted of fish swimming in a tank against a background of aquatic plants—a visual environment that the larval tiger salamander encounters in its natural life cycle. Example frames of a similar movie have appeared elsewhere (Tkacik et al., 2014). A 120 sec segment of this movie was repeated 70 times.

4.3 Maximum Likelihood Decoding. The activity of a neural population of N cells in a given time bin is denoted by $R = \{r_i\}$, where r_i is the activity of neuron i. The firing probability in each time bin was estimated from the peristimulus time histogram (PSTH) over many repeated presentations of the same stimulus (see above). In our treatment, we considered only binary neural activity in a single time bin, $r_i = (0, 1)$. This is expected to be a good approximation for small time bins, Δt, such as the 20 ms bins used in this study. In fact, a previous study found that even with 100 ms time bins, this binary approximation captured most of each ganglion cell's visual information about a spatial coding task (Schwartz et al., 2012).

We used a maximum likelihood decoding rule to distinguish the target stimulus, denoted T, from the ensemble of distracter stimuli, denoted D: $P(T|R) > P(D|R) =$ "target." Bayes' rule can be used to invert these conditional probabilities and express them as quantities that can be measured experimentally:

$$P\,(R|T) = \prod_{i}^{cells} (p_i)^{r_i}\ \ (1 - p_i)^{(1-r_i)} \quad \text{and} \tag{4.1a}$$

$$P\,(R|D) = \prod_{i}^{cells} (q_i)^{r_i}\ \ (1 - q_i)^{(1-r_i)}. \tag{4.1b}$$

For the case of the matched homogeneous population, these expressions reduce to simpler forms that depend only on k, the number of spikes in the

neural population:

$$P_{homo}(k|T) = \frac{N!}{k!\,(N-k)!}\,\bar{p}^k\,(1-\bar{p})^{N-k} \quad \text{and} \tag{4.2a}$$

$$P_{homo}(k|D) = \frac{N!}{k!\,(N-k)!}\,\bar{q}^k\,(1-\bar{q})^{N-k}\,, \tag{4.2b}$$

where the combinatorial factor counts how many activity patterns have a given value of k spikes. Because our goal was to quantify the effects of heterogeneity in the firing rates of neurons, we ignored correlations among neurons in formulating these probability distributions. As shown in previous studies, heterogeneity can be beneficial for populations of correlated neurons (Ecker et al., 2011; Shamir & Sompolinsky, 2006; Wilke et al., 2001) and diversity in pairwise correlations can itself be beneficial (Azeredo da Silveira & Berry, 2014).

The decoding rule for the homogeneous decoder simplifies greatly. If we assume $\bar{p} > \bar{q}$, we find (without loss of generality) a decoding rule $k \geq k^* \rightarrow$ "target," $k < k^* \rightarrow$ "distracter," where k^* is the spike count that solves $P(k^*|T) = P(k^*|D)$. This then defines the error rates for misses and false alarms:

$$P_{miss} = \sum_{k=k^*}^{N} P(k|D) \quad \text{and} \quad P_{false\,alarm} = \sum_{k=0}^{k^*-1} P(k|T), \tag{4.3}$$

which then allows us to define the total probability of error:

$$P_{error} = \tfrac{1}{2}\left(P_{miss} + P_{false\,alarm}\right).$$

The error for the homogeneous population could be calculated exactly, because the neural activity pattern could be reduced to a single variable, k. But for the heterogeneous case, the decision boundary between target and distracter stimuli was complicated. Exact numerical solution was not always possible, as it required iterating over all 2^N activity states in the population. Instead, we performed numerical sampling, where we used the distribution of neural activity given the target, $P(R|T)$, to generate M samples of neural activity. For each of these, we applied the decoding rule. A subset of these activity patterns was erroneously categorized as coming from the distracter distribution; this fraction defined the miss rate, P_{miss}. A similar procedure was carried out by starting with the distracter distribution, $P(R|D)$, generating M additional samples, and defining the false alarm rate, $P_{false\,alarm}$. Sampling continued until 1000 errors of each type occurred. However, for stimulus discriminations with low error, this process was stopped at 10^6 samples, implying that the lowest error rate that we could sample was $0.5 \cdot 10^{-6}$.

Notice that our method does not rely on cross-validated decoders. The reason is that error rates smaller than $\sim 1/\#$ trials cannot be estimated by cross-validation methods. This means that the estimate of the error rate will be artificially constrained by the practical details of our neurophysiology experiment. These details do not apply to how the animal uses its own neural populations, as the animal potentially has access to much longer sampling periods. Because we are interested in what is the true coding fidelity of neural populations, we constructed maximum likelihood decoders based on the measured firing rates of each cell. Of course, with any finite sample of measured neural responses, there will be uncertainty in the firing rate of each cell. We address this issue with bootstrap resampling, as described in section below 4.5.

We also quantified coding performance using mutual information. For neuron i having a firing probability p_i for the target and q_i for the distracter, the mutual information between its response and whether the stimulus was T or D is given by

$$\begin{aligned} I(r_i; S) &= H\left[\tfrac{1}{2}(p_i + q_i)\right] - \tfrac{1}{2}H\left[p_i\right] - \tfrac{1}{2}H\left[q_i\right] \quad \text{where} \\ H\left[p_i\right] &= -p_i \log_2 p_i - (1 - p_i)\log_2(1 - p_i) \quad \text{and} \end{aligned} \tag{4.4}$$

we are assuming $P(T) = P(D) = 0.5$. For the average information per cell in the entire neural population, we simply averaged over the mutual information conveyed by each cell:

$$I(R; S) = \frac{1}{N}\sum_{i}^{cells} I(r_i; S). \tag{4.5}$$

4.4 Selection of Stimulus Pairs. For many purposes, we wanted to average over a selection of many pairs of time points representing the population neural response to different pairs of visual stimuli. As it was not practical for us to sum over all possible pairs, we selected a representative subset of times covering the full range of average population firing rates. The resulting averages allow for significant comparisons among conditions, such as number of pools L or neurons N, but should not be interpreted as true ensemble averages for each stimulus conditions.

4.5 Correction for Sampling Bias. Even if all neurons had exactly the same underlying firing probability for a given stimulus, we would observe some heterogeneity in their estimated firing probabilities due to finite sampling. This heterogeneity, while spurious, would appear to benefit coding. In order to evaluate the significance of this effect, we started with a matched homogeneous population and used bootstrap resampling to generate an apparently heterogeneous population. Specifically, if each cell has an average firing probability of p for a given stimulus and if our experiment

had M repeated trials of this stimulus, then the total spike count would be $n_{count} = Mp$. We expect that this resampling procedure will generate fluctuations in the apparent spike count $\sim$sqrt(n_{count}), which will change the firing probability by $\sim$sqrt(p/M) in each cell, while preserving the same average firing probability. We then calculated the discrimination error and mutual information between stimuli for these apparently heterogeneous neural populations.

4.6 Discrimination Error in a Homogeneous Population of Neurons. We consider two stimuli that elicit firing probabilities p and q in each neuron, respectively. In such a homogeneous population, information is encoded in the total spike count, distributed according to equations 4.2a and 4.2b. If the two stimuli considered occur with the same prior probability, the spike count for which their posterior probabilities are equal, k^*, is obtained by equating the likelihoods in equations 4.2a and 4.2b, from which we obtain

$$p^{k^*}(1-p)^{N-k^*} = q^{k^*}(1-q)^{N-k^*} \tag{4.6}$$

and

$$k^* = \frac{\ln(1-p) - \ln(1-q)}{\ln(q(1-p)) - \ln(p(1-q))}. \tag{4.7}$$

In order to evaluate the miss and false alarm rates, we have to calculate the probability weight of the tails of the two distributions beyond this threshold. We approximate the sum over the spike count, k, by an integral and define an auxiliary variable, $\kappa = k/N$, which we use to express the form of the spike count distribution at large N. Applying Stirling's approximation to the factorials in equations 4.2a, we find that the tail of the distribution is dominated by an exponential decay,

$$\exp\left(-N\left[\kappa \ln\left(\frac{\kappa}{p}\right) + (1-\kappa)\ln\left(\frac{1-\kappa}{1-p}\right)\right]\right). \tag{4.8}$$

From this, it follows that the dominant term in the discrimination error behaves exponentially with N, as exp($-N/N^*$), with

$$N^* = \left[\kappa^* \ln\left(\frac{\kappa^*}{p}\right) + (1-\kappa^*)\ln\left(\frac{1-\kappa^*}{1-p}\right)\right]^{-1}, \tag{4.9}$$

where $\kappa^* = k^*/N$.

4.7 Model of Neuron-to-Neuron Variability. We consider a population of neurons in which the firing probabilities in response to the target (resp. the distracter) are perturbations about a homogeneous case, as

$$\begin{aligned} p_i &= p + \delta p_i, \\ q_i &= q + \delta q_i, \end{aligned} \tag{4.10}$$

where we assume $\delta p_i \ll p$ and $\delta q_i \ll q$, where the label $i = 1, \ldots, N$ runs over all the neurons in the population and the perturbation is constrained by the identities

$$\sum_{i=1}^{N} \delta p_i = \sum_{i=1}^{N} \delta q_i = 0. \tag{4.11}$$

In the homogeneous case, the distribution of the spike count, k, is calculated from equations 4.2a and 4.2b. In the perturbed system, the analogous quantity can be calculated in a straightforward manner, in two steps. (Since the calculation is the same for the target and the distracter cases, we derive for the former case only.) First, we calculate the probability—call it $\pi(i_1, \ldots, i_k)$—that neurons $i_1, \ldots, i_k$ be active and the others silent:

$$\pi\left(i_1, \ldots, i_k\right) = \prod_{\alpha=1}^{k} \left(p + \delta p_{i_\alpha}\right) \prod_{\beta=k+1}^{N} \left(1 - p - \delta p_{i_\beta}\right). \tag{4.12}$$

For a perturbative result, we expand this quantity in orders of δp; because of our constraint on the total firing probability, ultimately the terms that are linear in δp will sum to zero, so we consider this quantity up to second order in δp. After expanding and rearranging the terms in the expansion, we obtain the expression

$$\begin{aligned} \pi\left(i_1, \ldots, i_k\right) &= p^k (1-p)^{N-k} + \left[O\left(\delta p\right) \text{terms}\right] \\ &+ \frac{1}{2} p^{k-2} (1-p)^{N-k-2} \left[\sum_{\alpha,\alpha'=1}^{k} \delta p_{i_\alpha} \delta p_{i_{\alpha'}} - (1-p)^2 \sum_{\alpha=1}^{k} \left(\delta p_{i_\alpha}\right)^2 \right. \\ &\left. - p^2 \sum_{\beta=k+1}^{N} \left(\delta p_{i\beta}\right)^2 \right] + O\left(\delta p^3\right). \end{aligned} \tag{4.13}$$

Second, we sum this quantity over all possible choices of k neurons to obtain the probability—call it $P_{\text{hetero}}(k)$—that any k neurons in the population be active. For this, we have to count the number of occurrences of each of the second-order terms. Consider first the term

$$\sum_{\alpha,\alpha'=1}^{k} \delta p_{i_\alpha} \delta p_{i_{\alpha'}} = \sum_{\substack{\alpha,\alpha'=1 \\ \alpha\neq\alpha'}}^{k} \delta p_{i_\alpha} \delta p_{i_{\alpha'}} + \sum_{\alpha=1}^{k} (\delta p_{i_\alpha})^2. \tag{4.14}$$

By symmetry, we have

$$\sum_{\substack{\text{all choices} \\ i_1,\ldots,i_k}} \left(\sum_{\substack{\alpha,\alpha'=1 \\ \alpha\neq\alpha'}}^{k} \delta p_{i_\alpha} \delta p_{i_{\alpha'}} \right) = A \times \sum_{\substack{i,j=1 \\ i\neq j}}^{N} \delta p_i \, \delta p_j \tag{4.15}$$

and

$$\sum_{\substack{\text{all choices} \\ i_1,\ldots,i_k}} \left(\sum_{\alpha=1}^{k} (\delta p_{i_\alpha})^2 \right) = B \times \sum_{i=1}^{N} (\delta p_i)^2, \tag{4.16}$$

where A and B are numerical prefactors to be determined. This is done easily by noticing that the first sum has a total of $\binom{N}{k} k(k-1)$ terms and the second sum has a total of $\binom{N}{k} k$ terms, so that

$$A = \binom{N}{k} \frac{k\,(k-1)}{N\,(N-1)} \tag{4.17}$$

and

$$B = \binom{N}{k} \frac{k}{N}. \tag{4.18}$$

We can further simplify the second sum by noting that

$$\begin{aligned} \sum_{\substack{i,j=1 \\ i\neq j}}^{N} \delta p_i \, \delta p_j &= \sum_{i,j=1}^{N} \delta p_i \, \delta p_j - \sum_{i=1}^{N} (\delta p_i)^2 \\ &= \left(\sum_{i=1}^{N} \delta p_i \right)^2 - \sum_{i=1}^{N} (\delta p_i)^2 \\ &= -\sum_{i=1}^{N} (\delta p_i)^2 \end{aligned} \tag{4.19}$$

because of our constraint on the total firing probability. These evaluations take care of the first two sums that make up the second-order term in

$P_{\text{hetero}}(k)$. From similar combinatorial bookkeeping, we can write the third sum as

$$\sum_{\substack{\text{all choices}\\ i_{k+1},\ldots,i_N}} \left(\sum_{\beta=k+1}^{N} \left(\delta p_{i_\beta}\right)^2 \right) = \binom{N}{k} \frac{(N-k)}{N} \sum_{i=1}^{N} (\delta p_i)^2 . \tag{4.20}$$

Note that the first-order terms sum to zero. Finally, we rearrange the second-order terms and use equations 4.2 for $P_{\text{homo}}(k)$ to obtain equations 2.3.

4.8 Mutual Information for a Heterogeneous Population of Neurons. We consider the case of a rare target stimulus, which occurs with probability ρ, so that

$$\begin{aligned} P(T) &= \rho \\ P(D) &= 1-\rho, \end{aligned} \tag{4.21}$$

where T and D refer to the target and distracter stimuli, respectively, and $\rho \ll 1$. To first order in ρ, the mutual information between neural response, R, and stimulus, S, becomes proportional to the Kullback-Leibler divergence, as follows:

$$\begin{aligned} I(R;S) &= \sum_{R,S} P(R,S) \ln \frac{P(R,S)}{P(R)\,P(S)} \\ &= \sum_{R,S} P(R|S)\,P(S) \ln \frac{P(R|S)}{P(R)} \\ &= \sum_{R,S} P(R|S)\,P(S) \ln \left(P(R|S)\right) - \sum_{R} P(R) \ln \left(P(R)\right) \\ &= \sum_{R} \left[\rho\, P(R|T) \ln \left(P(R|T)\right) + (1-\rho)\; P(R|D) \ln \left(P(R|D)\right)\right] \\ &\quad - \sum_{R} \left[\rho\, P(R|T) + (1-\rho)\, P(R|D)\right] \ln \left(\left[\rho\, P(R|T)\, right.\right. \\ &\quad \left. + (1-\rho)\, P(R|D)\right]\right). \end{aligned} \tag{4.22}$$

By expanding in orders of ρ and using the normalization of the probability, we reduce this expression to the Kullback-Leibler form:

$$I(R;S) = \rho \sum_{R} P(R|T) \ln \left(\frac{P(R|T)}{P(R|D)} \right) + O(\rho^2). \tag{4.23}$$

For a population of N independent neurons, each of which fires zero or one spike, we can use equations 4.1a and 4.1b for $P(R|T)$ and $P(R|D)$. The mutual information then takes the simple form

$$I(R;S) = \rho \sum_{i=1}^{N} \left[p_i \ln\left(\frac{p_i}{q_i}\right) + (1-p_i)\ln\left(\frac{1-p_i}{1-q_i}\right)\right] + O(\rho^2). \tag{4.24}$$

Now we show that this quantity is minimized by the uniform choice in which all p_i's are equal and all q_i's are equal. We keep the total number of spikes fixed so that we assume, throughout the constraints

$$\sum_{i=1}^{N} p_i = N\,p,$$
$$\sum_{i=1}^{N} q_i = N\,q, \tag{4.25}$$

where p and q are the mean response to the target (resp. distracter) stimulus in the homogeneous population. Since the two terms entering the expression of the mutual information are symmetric, it will be sufficient to consider only one of them; if the latter is minimized for the case of uniform p_i's, then the same immediately applies to the full expression.

In order to complete the demonstration, it is instructive to consider the quantity $\sum_{i=1}^{N}(p_i/p)\ln(q_i/q)$. By taking first- and second-order derivatives of this quantity (while respecting the above constraints), we obtain in a straightforward manner that it is maximized for the choice $p_i/p = q_i/q$, that is, the inequality

$$\sum_{i=1}^{N} \frac{p_i}{p}\ln\left(\frac{q_i}{q}\right) \le \sum_{i=1}^{N} \frac{p_i}{p}\ln\left(\frac{p_i}{q}\right) \tag{4.26}$$

holds for all choices of the parameters that satisfy our constraints. Finally, by rearranging the heterogeneous terms on the left-hand side and the homogeneous terms on the right-hand side, we obtain the bound

$$\sum_{i=1}^{N} p_i \ln\left(\frac{p_i}{q_i}\right) \ge N\,p\ln\left(\frac{p}{q}\right), \tag{4.27}$$

sometimes referred to as the *log sum inequality*. This completes the demonstration that the mutual information between neural response and stimulus

is smallest in the case of a uniform population, to first order in the probability of occurrence of a rare target stimulus.

4.9 Chernoff Distance. The Chernoff distance is a measure that describes how the mutual information of a large population of neurons encoding discrete stimuli approaches the entropy of the stimulus as population size increases (Kang & Sompolinsky, 2001). It depends on the least discriminable of all possible pairs of stimulus values. Because of this dependence on a single pair of stimulus values, the Chernoff distance is also related to the discrimination error. Thus, it is a quantity that conceptually unifies the quantification of coding fidelity using error and information metrics. If we denote two stimuli by A and B, then the Chernoff distance between them, in terms of population activity, is given by

$$D_\alpha\left(A\|B\right) \equiv -\sum_{i=1}^{N} \ln \left\langle \exp\, \alpha S_i\left(r_i, A, B\right)\right\rangle_{r_i|B}, \quad \text{where} \tag{4.28}$$

$$S_i\left(r_i, A, B\right) \equiv \ln \frac{p(r_i|A)}{p(r_i|B)}, \quad \text{and} \tag{4.29}$$

$$\frac{\partial D_\alpha}{\partial \alpha} = 0. \tag{4.30}$$

For the case of two stimuli,

$$S_i(r_i, A, B) = \delta_{r_i,1} \ln\, p_i + \delta_{r_i,0} \ln(1-p_i) - \delta_{r_i,1} \ln q_i + \delta_{r_i,0} \ln(1-q_i). \tag{4.31}$$

Substituting this expression into equation 4.28, we get

$$\left\langle \exp\, \alpha S_i\left(r_i, A, B\right)\right\rangle_{r_i|B} = q_i \exp\left(\alpha \ln \frac{p_i}{q_i}\right) + (1-q_i) \exp\left(\alpha \ln \frac{1-p_i}{1-q_i}\right) \tag{4.32}$$

and

$$D_\alpha(A\|B) \equiv -\sum_{i=1}^{N} \ln\left[p_i^\alpha q_i^{(1-\alpha)} + (1-p_i)^\alpha (1-q_i)^{(1-\alpha)}\right]. \tag{4.33}$$

We solved equation 4.30 numerically to find the extremal value, α^*, and then substituted into equation 4.33 to get the average Chernoff distance per cell:

$$D_{chem}(A\|B) = D_{\alpha^*}(A\|B)/N. \tag{4.34}$$

In our case, the Chernoff distance has a simple relationship to the mutual information:

$$l = 1 - \exp(-D_{\alpha^*}). \tag{4.35}$$

In the limit of $p \to 1$ and $q \to 0$, or vice versa, the Chernoff distance diverges. Of course, this is not a conceptual problem because with finite sampling, one cannot have confidence that $p \to 1$ or $q \to 0$. However, the divergence in the Chernoff distance was not numerically well behaved, in the sense that choosing $p = 1 - \delta$ would have values that depended strongly on δ. For this reason, we left these stimulus pairs out of Figure 5. Such divergences were quite rare. In the spatially uniform stimulus, there were 526 out of 166,500 time points with $p = 1$, which is 0.3% of all times; in the natural movie, there were only 66 out of 924,000 times points with $p = 1$, which is 0.007% of all times.

Appendix A: Cell Types

Following previous studies, we used the reverse correlation to random flicker stimulation to group ganglion cells into functional types (Marre et al., 2012; Segev et al., 2006). This study differs somewhat from previous studies in that we used spatially uniform flicker, which engages both the receptive field center and surround, while previous studies mostly used checkerboard flicker and found the temporal profile of the center alone. This difference could well have influenced our results.

Previous studies have found six to eight functional types. In a similar vein, we identified eight types here (see Figures 7A and 7B). These included fast, medium, and slow OFF, as well as fast, medium, and slow ON, as have been found in most previous studies. We also identified an OFF type with a reverse correlation barely different from average, which we called a "weak OFF" type. This may correspond to weak receptive field cells seen previously (Marre et al., 2012). Unlike previous studies, we also found an ON type with an uncommonly large reverse correlation, which we named "big ON." The separation of the ganglion cell population into these eight functional types can be visualized by plotting the average reverse correlation of all cells of the same type along with error bars showing the standard error at each time point (see Figures 7A and 7B). The fact that these standard errors are clearly well separated at many time points serves as an a postieriori justification for our classification.

In order to study greater degrees of heterogeneity, we further split these 8 functional types into as many as 32 types (see Figure 11). This was accomplished by grouping together sets of ganglion cells with exceptionally similar reverse correlations or with qualitatively unusual features, such as a double-peaked, monophasic structure (e.g., type II double). The purpose

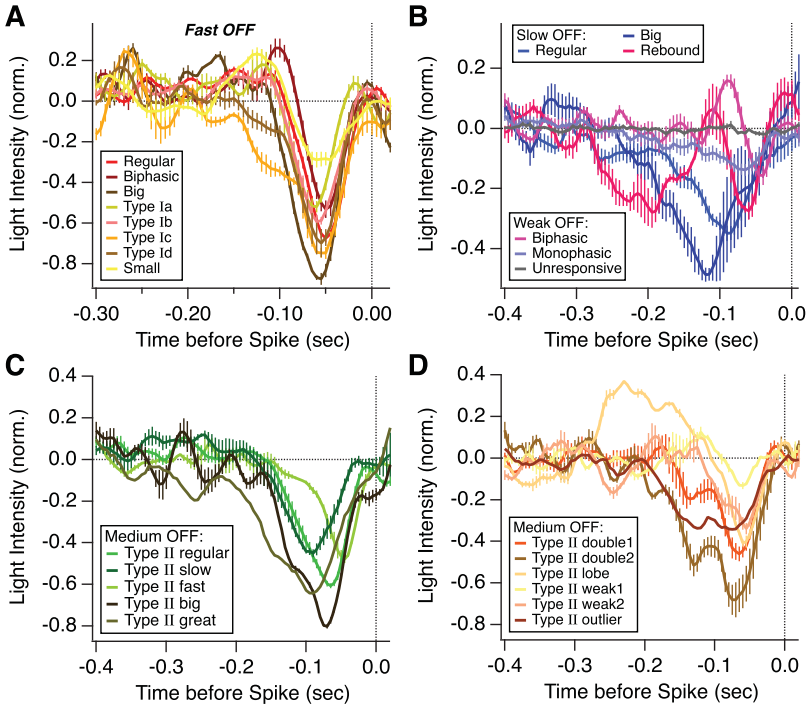

Figure 11: Definition of fine cell types. Each panel shows the reverse correlation for spatially uniform flicker, averaged across all cells of the same functional type (shown in different colors); errors bars represent standard error. (A) Eight subtypes of fast OFF cells. (B) Three subtypes of slow OFF cells, along with three subtypes of weak OFF cells. (C, D) (Five, six), subtypes of medium OFF cells.

was to further subdivide the neural population. When we plotted the average reverse correlation of cells within the same fine cell type, we found that their standard error was well separated from that of other fine cell types at many time points, as shown for our primary classification of 8 cell types. Here, we do not present any further evidence that these are true cell types that generalize across multiple retinas; indeed, some "types" comprise just a single cell.

Fast OFF cells were subdivided into eight subtypes (see Figure 11A): regular (or monophasic) OFF ($n = 9$); biphasic OFF cells ($n = 7$), which have been described before; big OFF cells ($n = 3$), with a large-amplitude reverse correlation; type Ia ($n = 4$), with a late peak in the reverse correlation; type Ib ($n = 6$), with a slightly larger reverse correlation; type Ic ($n = 3$), with a late shoulder to its reverse correlation; type Id ($n = 2$), with a smaller shoulder; and small OFF ($n = 2$), with a smaller-amplitude reverse correlation.

Medium OFF cells were subdivided into 11 subtypes (see Figures 11C and 11D): type II regular ($n = 8$); type II slow ($n = 3$), with a longer latency peak in the reverse correlation; type II fast ($n = 2$), with a shorter latency peak; type II big ($n = 2$), with a larger-amplitude reverse correlation; type II great ($n = 1$) with a larger and broader reverse correlation; type II double1 ($n = 3$), with a second narrow peak in the reverse correlation; type II double 2 ($n = 2$), similar to double1 but with a larger amplitude; type II lobe ($n = 2$), with a pronounced, late biphasic peak in the reverse correlation; type II weak1 ($n = 2$), with a very small amplitude and somewhat biphasic reverse correlation; type II weak2 ($n = 2$), with a small amplitude, monophasic reverse correlation; type II outlier ($n = 1$), with a double-peaked reverse correlation.

Slow OFF cells were subdivided into three subtypes (see Figure 11B): regular ($n = 5$), big ($n = 4$), and rebound ($n = 3$), with a pronounced, second peak in the reverse correlation. Weak OFF cells were divided into three subtypes (see Figure 11B): biphasic ($n = 7$), monophasic ($n = 5$), and unresponsive ($n = 2$). ON cells had fast ($n = 3$), medium ($n = 8$), and big ON ($n = 5$) subtypes, as before, but the remaining cells were subdivided into three outliers (data not shown): slow ON1, slow ON2, and biphasic ON.

In order to form 16 cell types, we merged several fine cell types together into four fast OFF types, four medium OFF types, three slow OFF types, one weak OFF type, and four ON types. For fast OFF cells, types Ia and Ib were merged ($n = 10$); types Ic, Id, and big were merged ($n = 10$); and regular OFF ($n = 9$) and biphasic OFF ($n = 7$) remained the same. For medium OFF cells, types II regular, II slow, and II fast were merged into type II core ($n = 13$); types II double1, II double2, II big, and II outlier were merged into type II double ($n = 8$); types II weak1, II weak II, and II great were merged into type II other ($n = 5$); type II lobe ($n = 2$) remained the same. For slow OFF cells, regular and big types were merged into slow OFF type ($n = 9$); slow OFF rebound type ($n = 2$) remained the same; three outlier cells were split away into slow OFF outlier type. Weak OFF cells were all merged together ($n = 14$), and ON cells remained the same.

Acknowledgments

M.B. acknowledges support from NEI grant EY014196 and NSF grant 1504977, and R.AdS. acknowledges support from Princeton University through the Global Scholars Program and from the CNRS through UMR 8550.

References

Asari, H., & Meister, M. (2012). Divergence of visual channels in the inner retina. *Nat. Neurosci.*, *15*(11), 1581–1589. doi:10.1038/nn.3241

Azeredo da Silveira, R., & Berry, M. J. II. (2014). High-fidelity coding with correlated neurons. *PLoS Comput. Biol.*, *10*(11), e1003970. doi:10.1371/journal.pcbi.1003970

Azeredo da Silveira, R., & Roska, B. (2011). Cell types, circuits, computation. *Curr. Opin. Neurobiol.*, *21*(5), 664–671. doi:10.1016/j.conb.2011.05.007

Baddeley, R., Abbott, L. F., Booth, M. C., Sengpiel, F., Freeman, T., Wakeman, E. A., & Rolls, E. T. (1997). Responses of neurons in primary and inferior temporal visual cortices to natural scenes. *Proc. Biol. Sci.*, *264*(1389), 1775–1783. doi:10.1098/rspb.1997.0246

Baden, T., Berens, P., Franke, K., Roman Roson, M., Bethge, M., & Euler, T. (2016). The functional diversity of retinal ganglion cells in the mouse. *Nature*, *529*(7586), 345–350. doi:10.1038/nature16468

Barlow, H. B., & Levick, W. R. (1965). The mechanism of directionally selective units in rabbit's retina. *J. Physiol.*, *178*(3), 477–504.

Berry, M. J., Warland, D. K., & Meister, M. (1997). The structure and precision of retinal spike trains. *Proc. Natl. Acad. Sci. U.S.A.*, *94*(10), 5411–5416.

Berson, D. M. (2003). Strange vision: Ganglion cells as circadian photoreceptors. *Trends Neurosci.*, *26*(6), 314–320. doi:10.1016/S0166-2236(03)00130-9

Berson, D. M. (2008). Retinal ganglion cell types and their central projections. In R. H. Masland and T. D. Albright (Eds.), *The senses: A comprehensive reference* (pp. 491–519). Amsterdam: Elsevier.

Bonin, V., Histed, M. H., Yurgenson, S., & Reid, R. C. (2011). Local diversity and fine-scale organization of receptive fields in mouse visual cortex. *J. Neurosci.*, *31*(50), 18506–18521. doi:10.1523/JNEUROSCI.2974-11.2011

Brenowitz, S. D., & Regehr, W. G. (2007). Reliability and heterogeneity of calcium signaling at single presynaptic boutons of cerebellar granule cells. *J. Neurosci.*, *27*(30), 7888–7898. doi:10.1523/JNEUROSCI.1064-07.2007

Chechik, G., Anderson, M. J., Bar-Yosef, O., Young, E. D., Tishby, N., & Nelken, I. (2006). Reduction of information redundancy in the ascending auditory pathway. *Neuron*, *51*(3), 359–368. doi:10.1016/j.neuron.2006.06.030

Chelaru, M. I., & Dragoi, V. (2008). Efficient coding in heterogeneous neuronal populations. *Proc. Natl. Acad. Sci. U.S.A.*, *105*(42), 16344–16349. doi:10.1073/pnas.0807744105

Chen, E. Y., Chou, J., Park, J., Schwartz, G., & Berry II, M. J. (2014). The neural circuit mechanisms underlying the retinal response to motion reversal. *Journal of Neuroscience*, *34*, 15557–15575.

Chen, E. Y., Marre, O., Fisher, C., Schwartz, G., Levy, J., da Silviera, R. A., & Berry, M. J., II. (2013). Alert response to motion onset in the retina. *J. Neurosci.*, *33*(1), 120–132. doi:10.1523/JNEUROSCI.3749-12.2013

Chichilnisky, E. J. (2001). A simple white noise analysis of neuronal light responses. *Network*, *12*(2), 199–213.

Clark, D. A., Benichou, R., Meister, M., & Azeredo da Silveira, R. (2013). Dynamical adaptation in photoreceptors. *PLoS Comput. Biol.*, *9*(11), e1003289. doi:10.1371/journal.pcbi.1003289

Dacey, D. M., Peterson, B. B., Robinson, F. R., & Gamlin, P. D. (2003). Fireworks in the primate retina: In vitro photodynamics reveals diverse LGN-projecting ganglion cell types. *Neuron*, *37*(1), 15–27.

DeVries, S. H. (2000). Bipolar cells use kainate and AMPA receptors to filter visual information into separate channels. *Neuron*, *28*(3), 847–856.

Dobrunz, L. E., & Stevens, C. F. (1997). Heterogeneity of release probability, facilitation, and depletion at central synapses. *Neuron*, *18*(6), 995–1008.

Dobrunz, L. E., & Stevens, C. F. (1999). Response of hippocampal synapses to natural stimulation patterns. *Neuron*, *22*(1), 157–166.

Ecker, A. S., Berens, P., Tolias, A. S., & Bethge, M. (2011). The effect of noise correlations in populations of diversely tuned neurons. *J. Neurosci.*, *31*(40), 14272–14283. doi:10.1523/JNEUROSCI.2539-11.2011

Fairhall, A. L., Burlingame, C. A., Narasimhan, R., Harris, R. A., Puchalla, J. L., & Berry, M. J. II. (2006). Selectivity for multiple stimulus features in retinal ganglion cells. *J. Neurophysiol.*, *96*(5), 2724–2738. doi:10.1152/jn.00995.2005

Ghosh, K. K., Bujan, S., Haverkamp, S., Feigenspan, A., & Wassle, H. (2004). Types of bipolar cells in the mouse retina. *J. Comp. Neurol.*, *469*(1), 70–82. doi:10.1002/cne.10985

Gjorgjieva, J., Sompolinsky, H., & Meister, M. (2014). Benefits of pathway splitting in sensory coding. *J. Neurosci.*, *34*(36), 12127–12144. doi:10.1523/JNEUROSCI.1032-14.2014

Hattar, S., Kumar, M., Park, A., Tong, P., Tung, J., Yau, K. W., & Berson, D. M. (2006). Central projections of melanopsin-expressing retinal ganglion cells in the mouse. *J. Comp. Neurol.*, *497*(3), 326–349. doi:10.1002/cne.20970

Holmstrom, L. A., Eeuwes, L. B., Roberts, P. D., & Portfors, C. V. (2010). Efficient encoding of vocalizations in the auditory midbrain. *J. Neurosci.*, *30*(3), 802–819. doi:10.1523/JNEUROSCI.1964-09.2010

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiol.*, *160*, 106–154.

Hunsberger, E., Scott, M., & Eliasmith, C. (2014). The competing benefits of noise and heterogeneity in neural coding. *Neural Comput.*, *26*(8), 1600–1623. doi:10.1162/NECO_a_00621

Kang, K., Shapley, R. M., & Sompolinsky, H. (2004). Information tuning of populations of neurons in primary visual cortex. *J. Neurosci.*, *24*(15), 3726–3735. doi:10.1523/JNEUROSCI.4272-03.2004

Kang, K., & Sompolinsky, H. (2001). Mutual information of population codes and distance measures in probability space. *Phys. Rev. Lett.*, *86*(21), 4958–4961. doi:10.1103/PhysRevLett.86.4958

Kastner, D. B., Baccus, S. A., & Sharpee, T. O. (2015). Critical and maximally informative encoding between neural populations in the retina. *Proc. Natl. Acad. Sci. U.S.A.*, *112*(8), 2533–2538. doi:10.1073/pnas.1418092112

Keat, J., Reinagel, P., Reid, R. C., & Meister, M. (2001). Predicting every spike: A model for the responses of visual neurons. *Neuron*, *30*(3), 803–817.

Lengler, J., Jug, F., & Steger, A. (2013). Reliable neuronal systems: The importance of heterogeneity. *PLoS One*, *8*(12), e80694. doi:10.1371/journal.pone.0080694

Liu, Y. S., Stevens, C. F., & Sharpee, T. O. (2009). Predictable irregularities in retinal receptive fields. *Proc. Natl. Acad. Sci. U.S.A.*, *106*(38), 16499–16504. doi:10.1073/pnas.0908926106

MacNeil, M. A., & Masland, R. H. (1998). Extreme diversity among amacrine cells: Implications for function. *Neuron*, *20*(5), 971–982.

Marc, R. E., Jones, B. W., Watt, C. B., Anderson, J. R., Sigulinsky, C., & Lauritzen, S. (2013). Retinal connectomics: Towards complete, accurate networks. *Prog. Retin. Eye Res.*, *37*, 141–162. doi:10.1016/j.preteyeres.2013.08.002

Marre, O., Amodei, D., Deshmukh, N., Sadeghi, K., Soo, F., Holy, T. E., & Berry, M. J. II. (2012). Mapping a complete neural population in the retina. *J. Neurosci.*, *32*(43), 14859–14873. doi:10.1523/JNEUROSCI.0723-12.2012

Mejias, J. F., & Longtin, A. (2012). Optimal heterogeneity for coding in spiking neural networks. *Phys. Rev. Lett.*, *108*(22), 228102.

Montijn, J. S., Goltstein, P. M., & Pennartz, C. M. (2015). Mouse V1 population correlates of visual detection rely on heterogeneity within neuronal response patterns. *Elife*, *4*, e10163. doi:10.7554/eLife.10163

Newsome, W. T., Britten, K. H., & Movshon, J. A. (1989). Neuronal correlates of a perceptual decision. *Nature*, *341*(6237), 52–54. doi:10.1038/341052a0

Olveczky, B. P., Baccus, S. A., & Meister, M. (2003). Segregation of object and background motion in the retina. *Nature*, *423*(6938), 401–408. doi:10.1038/nature01652

Osborne, L. C., Palmer, S. E., Lisberger, S. G., & Bialek, W. (2008). The neural basis for combinatorial coding in a cortical population response. *J. Neurosci.*, *28*(50), 13522–13531. doi:10.1523/JNEUROSCI.4390-08.2008

Padmanabhan, K., & Urban, N. N. (2010). Intrinsic biophysical diversity decorrelates neuronal firing while increasing information content. *Nat. Neurosci.*, *13*(10), 1276–1282. doi:10.1038/nn.2630

Prinz, A. A., Bucher, D., & Marder, E. (2004). Similar network activity from disparate circuit parameters. *Nat. Neurosci.*, *7*(12), 1345–1352. doi:10.1038/nn1352

Puchalla, J. L., Schneidman, E., Harris, R. A., & Berry, M. J. (2005). Redundancy in the population code of the retina. *Neuron*, *46*(3), 493–504. doi:10.1016/J.Neuron.2005.03.026

Raymond, J. L., Lisberger, S. G., & Mauk, M. D. (1996). The cerebellum: A neuronal learning machine? *Science*, *272*(5265), 1126–1131.

Ringach, D. L., Shapley, R. M., & Hawken, M. J. (2002). Orientation selectivity in macaque V1: Diversity and laminar dependence. *J. Neurosci.*, *22*(13), 5639–5651. doi:20026567

Robles, E., Laurell, E., & Baier, H. (2014). The retinal projectome reveals brain-area-specific visual representations generated by ganglion cell diversity. *Curr. Biol.*, *24*(18), 2085–2096. doi:10.1016/j.cub.2014.07.080

Rodieck, R. W. (1979). Visual pathways. *Annu. Rev. Neurosci.*, *2*, 193–225. doi:10.1146/annurev.ne.02.030179.001205

Roitman, J. D., & Shadlen, M. N. (2002). Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. *J. Neurosci.*, *22*(21), 9475–9489.

Schwartz, G., Macke, J., Amodei, D., Tang, H., & Berry, M. J., II. (2012). Low error discrimination using a correlated population code. *J. Neurophysiol.*, *108*(4), 1069–1088. doi:10.1152/jn.00564.2011

Segev, R., Puchalla, J., & Berry, M. J., II. (2006). Functional organization of ganglion cells in the salamander retina. *J. Neurophysiol.*, *95*(4), 2277–2292. doi:10.1152/jn.00928.2005

Seung, H. S., & Sumbul, U. (2014). Neuronal cell types and connectivity: Lessons from the retina. *Neuron*, *83*(6), 1262–1272. doi:10.1016/j.neuron.2014.08.054

Shamir, M., & Sompolinsky, H. (2006). Implications of neuronal diversity on population coding. *Neural Comput.*, *18*(8), 1951–1986. doi:10.1162/neco.2006.18.8.1951

Shoham, S., O'Connor, D. H., & Segev, R. (2006). How silent is the brain: Is there a "dark matter" problem in neuroscience? *J. Comp. Physiol. A: Neuroethol. Sens. Neural Behav. Physiol.*, *192*(8), 777–784. doi:10.1007/s00359-006-0117-6

Soo, F. S., Schwartz, G. W., Sadeghi, K., & Berry, M. J. (2011). Fine spatial information represented in a population of retinal ganglion cells. *Journal of Neuroscience*, *31*(6), 2145–2155. doi:10.1523/Jneurosci.5129-10.2011

Tkacik, G., Marre, O., Amodei, D., Schneidman, E., Bialek, W., & Berry, M. J., II. (2014). Searching for collective behavior in a large network of sensory neurons. *PLoS Comput. Biol.*, *10*(1), e1003408. doi:10.1371/journal.pcbi.1003408

Tripathy, S. J., Padmanabhan, K., Gerkin, R. C., & Urban, N. N. (2013). Intermediate intrinsic diversity enhances neural population coding. *Proc. Natl. Acad. Sci. U.S.A.*, *110*(20), 8248–8253. doi:10.1073/pnas.1221214110

Uzzell, V. J., & Chichilnisky, E. J. (2004). Precision of spike trains in primate retinal ganglion cells. *J. Neurophysiol.*, *92*(2), 780–789. doi:10.1152/jn.01171.2003

van Hateren, J. H., Ruttiger, L., Sun, H., & Lee, B. B. (2002). Processing of natural temporal stimuli by macaque retinal ganglion cells. *Journal of Neuroscience*, *22*(22), 9945–9960.

van Rossum, M. C., O'Brien, B. J., & Smith, R. G. (2003). Effects of noise on the spike timing precision of retinal ganglion cells. *J. Neurophysiol.*, *89*(5), 2406–2419. doi:10.1152/jn.01106.2002

Vaney, D. I., Peichl, L., Wassle, H., & Illing, R. B. (1981). Almost all ganglion cells in the rabbit retina project to the superior colliculus. *Brain Res.*, *212*(2), 447–453.

Warland, D. K., Reinagel, P., & Meister, M. (1997). Decoding visual information from a population of retinal ganglion cells. *J. Neurophysiol.*, *78*(5), 2336–2350.

Weliky, M., Fiser, J., Hunt, R. H., & Wagner, D. N. (2003). Coding of natural scenes in primary visual cortex. *Neuron*, *37*(4), 703–718.

Wilke, S. D., Thiel, A., Eurich, C. W., Greschner, M., Bongard, M., Ammermuller, J., & Schwegler, H. (2001). Population coding of motion patterns in the early visual system. *J. Comp. Physiol. A*, *187*(7), 549–558.

Received July 27, 2017; accepted September 28, 2018.