Reconstruction of natural images from responses of primate retinal ganglion cells

Nora Brackbill[1], Colleen Rhoades[2], Alexandra Kling[3], Nishal P. Shah[4], Alexander Sher[5], Alan M. Litke[5], E.J. Chichilnisky[3]

1. Department of Physics, Stanford University, Stanford, United States
2. Department of Bioengineering, Stanford University, Stanford, United States
3. Department of Neurosurgery, Stanford School of Medicine, Stanford, United States
   Department of Ophthalmology, Stanford University, Stanford, United States
   Hansen Experimental Physics Laboratory, Stanford University, Stanford, United States
4. Department of Electrical Engineering, Stanford University, Stanford, United States
5. Santa Cruz Institute for Particle Physics, University of California, Santa Cruz, Santa Cruz, United States

Abstract

The visual message conveyed by a retinal ganglion cell (RGC) is often summarized by its spatial receptive field, but in principle should also depend on other cells' responses and natural image statistics. To test this idea, linear reconstruction (decoding) of natural images was performed using combinations of responses of four high-density macaque RGC types, revealing consistent visual representations across retinas. Each cell's visual message, defined by the optimal reconstruction filter, reflected natural image statistics, and resembled the receptive field only when nearby, same-type cells were included. Reconstruction from each cell type revealed different and largely independent visual representations, consistent with their distinct properties. Stimulus-independent correlations primarily affected reconstructions from noisy responses. Nonlinear response transformation slightly improved reconstructions with either ON or OFF parasol cells, but not both. Inclusion of ON-OFF interactions enhanced reconstruction by emphasizing oriented edges, consistent with linear-nonlinear encoding models. Spatiotemporal reconstructions revealed similar spatial visual messages.

Introduction

The brain uses visual information transmitted by retinal neurons to make inferences about the external world. Traditionally, the visual signal transmitted by an individual retinal ganglion cell (RGC) to the brain has been summarized by its spatial profile of light sensitivity, or receptive field (RF), measured with stimuli such as spots or bars (Chichilnisky, 2001; Kuffler, 1953; Lettvin et al., 1959). Although it is intuitively appealing, there are several reasons why this description may not reveal how the spikes from a RGC contribute to the visual representation in the brain under natural viewing conditions, in which objects and scenes create distinctive structure. In particular, the strong spatial correlations in natural images (Ruderman & Bialek, 1994) mean that the response of a single RGC contains information about visual space well beyond its RF. Thus, across the RGC population, the responses of many individual cells could contain information about the same region of visual space, and it is not obvious how the brain could exploit this potentially redundant information (Puchalla et al., 2005). Complicating this issue is the fact that there are roughly twenty RGC types, each covering all of visual space with their RFs, and each with different spatial, temporal, and chromatic selectivity properties (Dacey et al., 2003). Furthermore, RGCs show both stimulus-induced and stimulus-independent correlated activity, within and across cell types (Greschner et al., 2011; Mastronarde, 1983), which could substantially influence the encoding of the stimulus (Meytlis et al., 2012; Pillow et al., 2008; Zylberberg et al., 2016). For these reasons, the visual message transmitted by a RGC to the brain is not fully understood.

One way to understand how each RGC contributes to naturalistic vision is to determine how a natural image can be reconstructed from the light-evoked responses of the entire RGC population. This analysis approach mimics the challenge faced by the brain: using sensory inputs for making inferences about the environment (Bialek et al., 1991; Rieke et al., 1997). In the simplest case of linear reconstruction, the visual message of an individual RGC can be summarized by its optimal reconstruction filter, i.e. its contribution to the reconstructed image. Linear reconstruction has been used to estimate the temporal structure of a spatially uniform stimulus from the responses of salamander RGCs, revealing that the reconstruction filters varied widely and depended heavily on the other RGCs included in the reconstruction (Warland et al., 1997). However, no spatial information was explored, and only a small number of RGCs of unknown types were examined. A later study linearly reconstructed spatiotemporal natural

movies from the activity of a category of linearly-responding neurons in the cat LGN (Stanley et al., 1999). However, neurons from many recordings were pooled, without cell type identification or spatial coverage indicating a complete population. More recently, several studies have used nonlinear and machine learning methods for reconstruction, but have been limited to simulation (Parthasarathy et al., 2017), small data sets from salamander retina with unidentified cell types and incomplete spatial coverage (Zhang et al., 2020), or reconstruction of moving dot patterns from rat retina without clear cell types (Botella-Soler et al., 2018). Thus, it remains unclear what spatial visual message primate RGCs convey to the brain, in the context of natural scenes and the full neural population.

We performed linear reconstruction of flashed natural images using the responses of hundreds of RGCs in macaque retina, collected simultaneously using large-scale, multi-electrode recordings. These recordings provided access to the visual signals of nearly complete populations of ON and OFF parasol cells, as well as locally complete populations of ON and OFF midget cells, the four numerically dominant RGC types that provide high resolution visual information to the brain (Dacey et al., 2003). Fifteen recordings were analyzed, enabling comparison of the visual message across a range of response properties. Examination of the reconstruction filters revealed that the dependence of a given RGC's visual message on the responses of other RGCs extended to the spatial domain in natural viewing conditions due to the statistics of natural scenes. Reconstruction from different cell types revealed that they conveyed different (but expected) features of the visual scene, and the spatial information carried by one type was largely unaffected by the contributions of other types. Two simple tests of nonlinear reconstruction indicated that more elaborate processing of retinal signals in the brain may improve the fidelity of visual perception. Finally, full spatiotemporal reconstruction with dynamic scenes revealed similar spatial visual messages, suggesting that the above findings generalize to natural vision.

Results

Large-scale multi-electrode recordings were used to characterize light responses in complete populations of peripheral macaque retinal ganglion cells (RGCs; Chichilnisky & Kalmar, 2002; Field et al., 2010; Frechette et al., 2005; Litke et al., 2004). The classical receptive field (RF) of each cell was measured by reverse correlation between its spike train and a spatiotemporal white noise stimulus, resulting in a spike-triggered average (STA) stimulus that summarized the spatial, temporal and chromatic properties of the light response (Chichilnisky, 2001). Clustering of these properties revealed multiple identifiable and complete cell type populations (see Methods; Chichilnisky & Kalmar, 2002; Dacey, 1993; DeVries & Baylor, 1997; Field et al., 2007; Frechette et al., 2005), including the four numerically dominant RGC types in macaque: ON parasol, OFF parasol, ON midget, and OFF midget. The RFs of each identified type formed an orderly lattice, consistent with the mosaic organization of each RGC type known from anatomical studies (Wässle et al., 1983).

Responses to natural images were then characterized by displaying static, grayscale images from the ImageNet database, which contains a wide variety of subjects including landscapes, objects, people, and animals (Fei-Fei et al., 2010). Each image was displayed for 100ms, separated by 400ms of spatially uniform illumination with intensity equal to the mean across all images (Figure 1A). This timing resulted in a strong initial response from both parasol and midget cells, and a return to maintained firing rates prior to the onset of the subsequent image. For each image, the population response was quantified as a vector of RGC spike counts in the 150ms window after image onset (Figure 1B; window chosen to optimize reconstruction performance; see Methods). The stimulus (*S,* dimensions: number of images by number of pixels) was reconstructed from the recorded ON and OFF parasol and midget cell responses *(R,* dimensions: number of images by number of cells) using a linear model, $S = RW$. The optimal weights for the linear model *(W,* dimensions: number of cells by number of pixels) were calculated using least squares regression,

$$W_{ls} = (R^T R)^{-1} R^T S. \quad (1)$$

The weights were then used to reconstruct a held-out set of test images. Reconstruction performance was measured by comparing only the areas of the original and reconstructed

images covered by the RF mosaic for each RGC type included (see Methods). Pearson's linear correlation coefficient (ρ) was used and reported as the primary metric, but mean squared error (MSE) and the structural similarity (SSIM; Wang et al., 2004) were also calculated, and showed the same trends. In what follows, reconstruction "from RGCs" is used as a shorthand to indicate reconstruction from their recorded responses, as described above.

The basic characteristics of spatial linear reconstruction were evaluated by reconstructing images from the responses of populations of ON and OFF parasol cells in 15 recordings from 9 monkeys. In each case, both cell types formed complete or nearly complete mosaics with uniform coverage, indicating that nearly every cell of each type over the electrode array was recorded (see Figures 1C and 2). Thus, the reconstructions revealed the full representation in these RGC populations. In each recording, reconstruction performance varied considerably across the set of test images (Figure 1D, ρ = 0.76 +/- 0.12 across n = 2250 images from 15 recordings) but was similar for repeated presentations of the same image (standard deviation across repeats = 0.014). Reconstruction performance was also similar for presentations of the same image in different recordings (standard deviation across recordings = 0.039), despite differences in the population responses and the properties of the RF mosaics (Figure 2). The reconstructed images themselves were also very similar across recordings (ρ = 0.90 +/- 0.06, across 150 images and 66 pairs of recordings; Figure 2). The minor differences in performance between recordings were strongly correlated with the average RF size in each recording (ρ = 0.7). Qualitatively, large scale image structure seemed to be well captured, but fine details were not. These results indicate that the image structure and the spatial resolution of the RGC population, rather than response variability, were primarily responsible for variation in reconstruction performance across images and recordings.

To further probe the role of the spatial resolution of the RGC population, the reconstructed images were compared to smoothed images, created by convolving the original images with a Gaussian matching the average parasol cell RF size for each recording (see Figure 1E, bottom row). Broadly, the smoothed images provided a good approximation to the images obtained by reconstruction. On average, the reconstructed image (averaged across trials) was more similar to the smoothed image (ρ=0.91) than to the original image (ρ=0.78). The residuals from reconstruction and smoothing, obtained by subtracting the original image, were also similar

(ρ=0.83), suggesting that reconstruction and smoothing captured and discarded similar features of the original images. While smoothed images do not represent a strict upper limit on reconstruction performance, this analysis further indicates that the RGC density is a dominant factor in image reconstruction.

An alternate measure of population response was also tested, spike latency, which has been shown to convey more stimulus information than spike counts in salamander RGCs in certain conditions (Gollisch & Meister, 2008). Here, the RGC response was defined as the time from the image onset to the time of the first spike. This latency response measure did not improve performance overall (reconstruction from ON and OFF parasol cell responses: $\Delta\rho$ = -0.10 +/- 0.12 across 4500 images from 15 recordings; reconstruction from ON and OFF midget cell responses: -0.16 +/- 0.19 across 3300 images from 11 recordings), although it did improve performance for reconstruction from ON parasol cells alone in two recordings ($\Delta\rho$ = 0.04 +/- 0.12 across 600 images from two recordings) and from ON midget cells alone in one recording ($\Delta\rho$ = 0.02 +/- 0.1 across 300 images). Overall, the limited performance of the latency measure may be due to high maintained firing rates in the mammalian retina (see Figure 1B), which make it difficult to identify the first stimulus-driven spike.

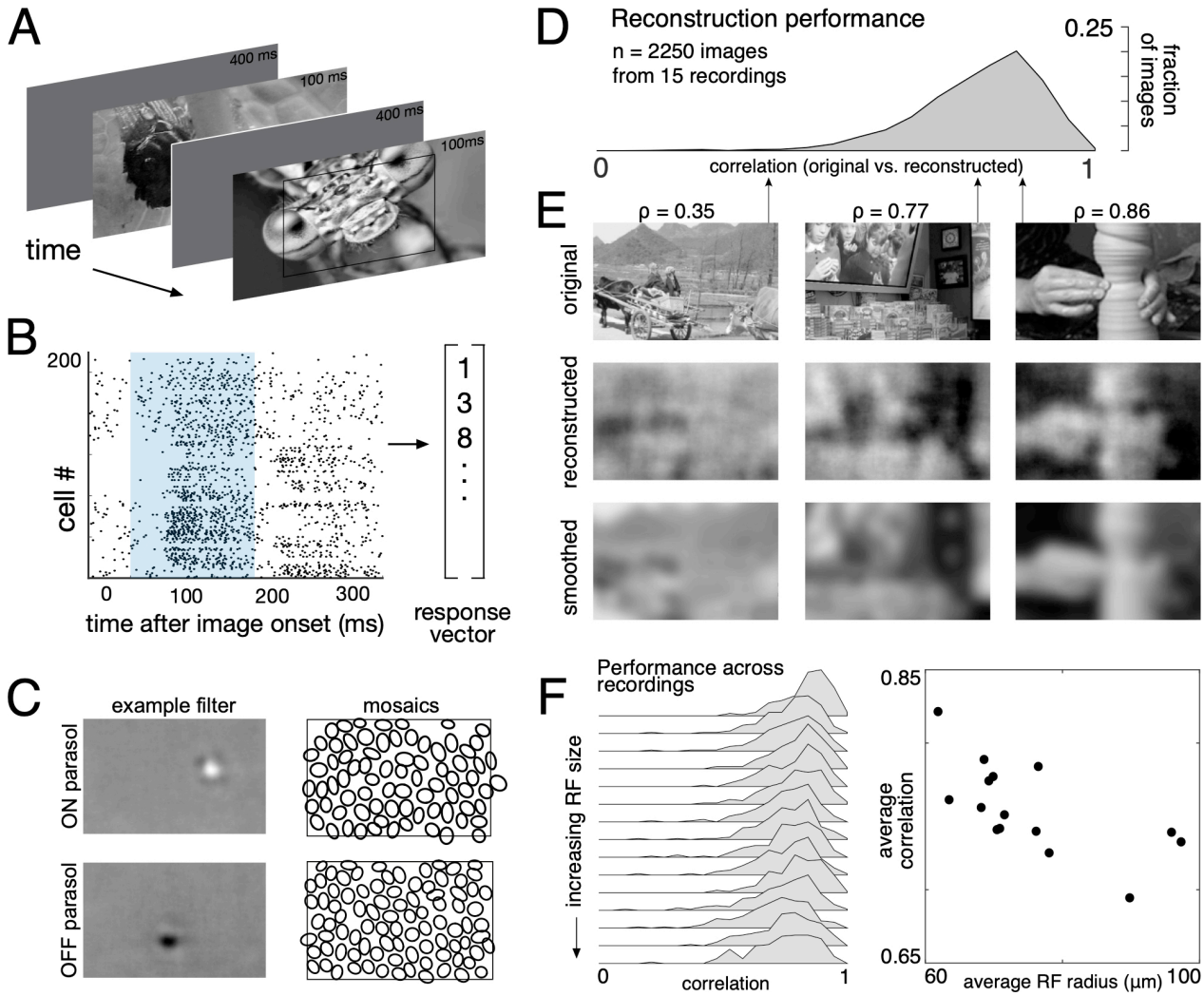

Figure 1: Linear reconstruction from ON and OFF parasol cell responses. A) Visual stimulus: static images from the ImageNet database were flashed for 100 ms, with 400ms of gray between. The thin black rectangle indicates the central image region shown in C and E. B) Example population response: each entry corresponds to the number of spikes from one RGC in a 150ms window (shown in blue) after the image onset. C) Left: Examples of reconstruction filters for an ON (top) and OFF (bottom) parasol cell. Right: RF locations for the entire population of ON (top) and OFF (bottom) parasol cells used in one recording. D) Reconstruction performance (correlation) across all recordings. E) Example reconstructions for three representative scores (middle row), compared to original images (top row) and smoothed images (bottom row), from the same recording shown in C. F) Reconstruction performance across 15 recordings. Left: Distributions of scores across images for each recording, ordered by average receptive field (RF) size. Right: Average reconstruction performance vs. average RF radius ($\rho = 0.7$).

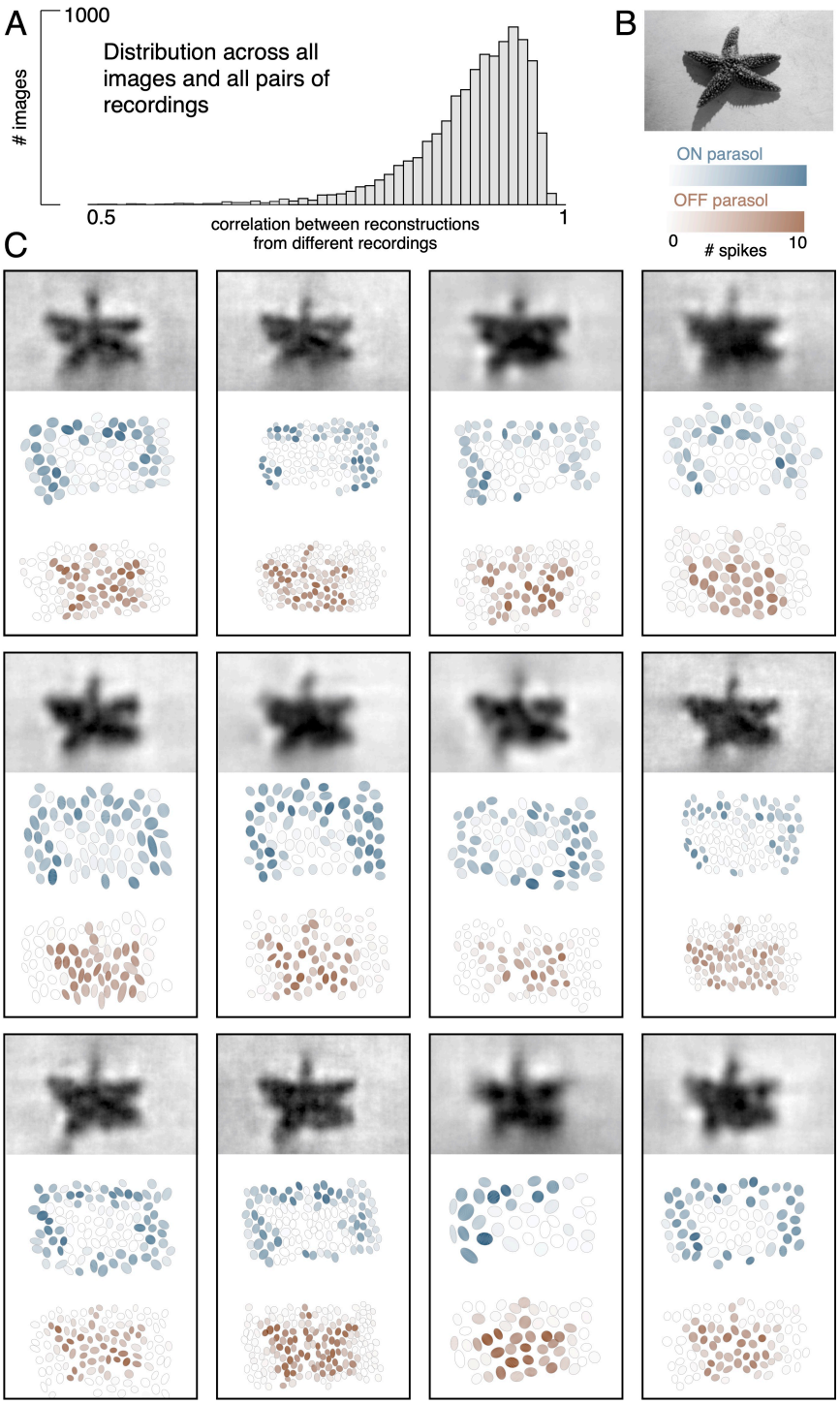

Figure 2: Visual representation across retinas. A) Distribution of correlation between reconstructed images from different recordings, across 150 images and 66 pairs of recordings. B) Example image. C) Across 12 recordings, reconstructed images (top, averaged across trials), ON (middle, blue) and OFF (bottom, orange) parasol responses, shown as the mosaic of Gaussian RF fits shaded by the spike count in response to this image.

The visual message conveyed by retinal ganglion cells

To understand how the visual message conveyed by a single RGC depends on the signals transmitted by others, reconstruction was performed from the responses of a given cell alone or with other cells of the same type. Cells of the same type have very similar response properties (Chichilnisky & Kalmar, 2002), but non-overlapping RFs that form a mosaic tiling visual space. When a single cell was used for reconstruction, its reconstruction filter (Figure 3A, top) exhibited much larger spatial spread than that of its spatial RF (Figure 3A, bottom, measured with white noise; see Methods), or of the spatially localized filter obtained in the full population reconstruction described above (Figure 1C). The full width at half maximum of the average single cell filter was roughly four times the average RF width (3.6 +/- 1.4 across 15 recordings). As additional RGCs of the same type were included in reconstruction, the spatial spread of the primary cell's reconstruction filter was progressively reduced, leveling off to slightly wider than the average RF with the inclusion of the 6 nearest neighbors (1.3 +/- 0.2 across 15 recordings; average filters shown in Figure 3C, widths shown in Figure 3D).

Both the spatial spread of the single cell reconstruction filter and its reduction in the context of the neural population can be understood by examining how the optimal filters (Equation 1) depend on stimulus (S) and response (R) statistics. $R^T R$ is a matrix that represents correlations in the activity of different RGCs. $R^T S$ is a matrix of unnormalized, spike-triggered average (STA) images, one for each RGC. These natural image STAs were broad (Figure 3A, top), reflecting the strong spatial correlations present in natural scenes (Figure 3B). For reconstruction from a single cell's responses, $R^T R$ is a scalar, and therefore the single cell reconstruction filter is directly proportional to the natural image STA. However, in the case of population reconstruction, $R^T R$ is a matrix that shapes the reconstruction filter based on the activity of other cells. Specifically, each cell's population filter is a linear mixture of its own natural image STA and those of the other cells in the reconstruction, weighted negatively based on the magnitude of their correlated activity. This mixing resulted in the reduction in the spatial spread of the RGC's filter when nearby cells of the same type were included (Figure 3C).

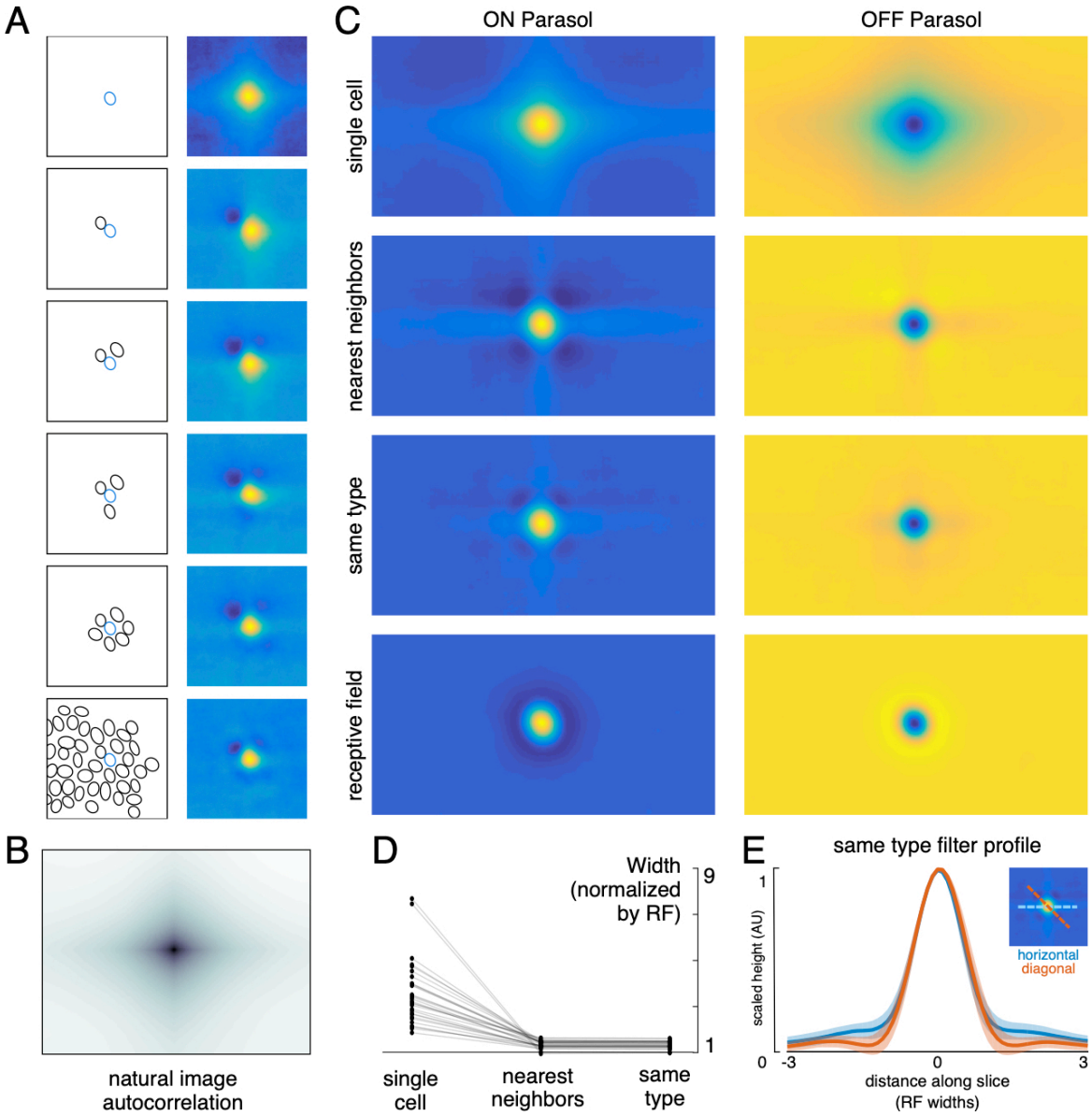

Figure 3: Effect of the population on the visual message. A) The reconstruction filter of a single cell as more neighboring cells are included in the reconstruction. Left: receptive fields (RFs) of cells in reconstruction, with the primary cell indicated in blue. Right: Filter of the primary cell. B) Autocorrelation structure of the natural images used here. C) Average ON (left) and OFF (right) parasol cell filters for a single recording. From top to bottom: reconstruction from a single cell, reconstruction from that cell plus all nearest neighbors, reconstruction from that cell plus all cells of the same type, and that cell's RF. D) Filter width, normalized by the RF width. E) Profiles of the same type filters in the horizontal (orange) and diagonal (blue) directions. Average (bold) +/- standard deviation (shaded region) across recordings.

When the complete population of RGCs of the same type was included in the reconstruction, the resulting spatially localized filters were similar to the RFs obtained with white noise stimuli (ρ = 0.78 +/- 0.10, n = 997 ON and 1228 OFF parasol cells from 15 recordings). However, some natural image spatial structure remained and was consistent across recordings, cells, and cell types. Most strikingly, the reconstruction filters exhibited broad vertical and horizontal structure (Figure 3C,E). This is a known feature of natural scenes (Girshick et al., 2011), and is present in the data used here (Figure 3B).

In addition, the visual scene was more uniformly covered by the reconstruction filters than by the RFs (Figure 4A,C). Coverage was defined as the proportion of pixels that were covered by exactly one cell's filter. The extent of each filter was defined by a relative magnitude threshold, whose level was set separately for the filters and for the RFs to maximize the resulting coverage value. Across both the ON and OFF parasol cells in 12 recordings, the average coverage was 0.62 +/- 0.06 for the RFs and 0.78 +/- 0.03 for the reconstruction filters (Figure 4C). By comparison, expanded RFs, scaled around each RGC's center location to match the average filter width, led to a small reduction in coverage (0.57 +/- 0.06) due to increased overlap. This indicates that the filters are not simply broader versions of the RF, but rather that they are distorted relative to the RFs to fill gaps in the mosaic.

To understand how the features of the filters that differed from the RF affected the reconstructed images, reconstruction was performed using the spatial RFs in place of the linear filters (scaled to minimize MSE, see Methods; Figure 4B). This manipulation reduced the reconstruction performance by 24% ($\Delta\rho$ = -0.12 +/- 0.09 across 4800 images from 8 recordings; Figure 4D), primarily in the lower spatial frequencies, which also contain most of the power in the original images (Figure 4E). The resulting images were noticeably less smooth in appearance than the optimally reconstructed images, and exhibited structure resembling the RGC mosaic (Figure 4B). Thus, although the reconstruction filters generally resembled the RFs, the additional spatial structure related to natural images and the spatial arrangement of RGCs led to smoother reconstructed images. These features may help explain the high consistency in reconstruction performance across many retinas (see above; Figure 2).

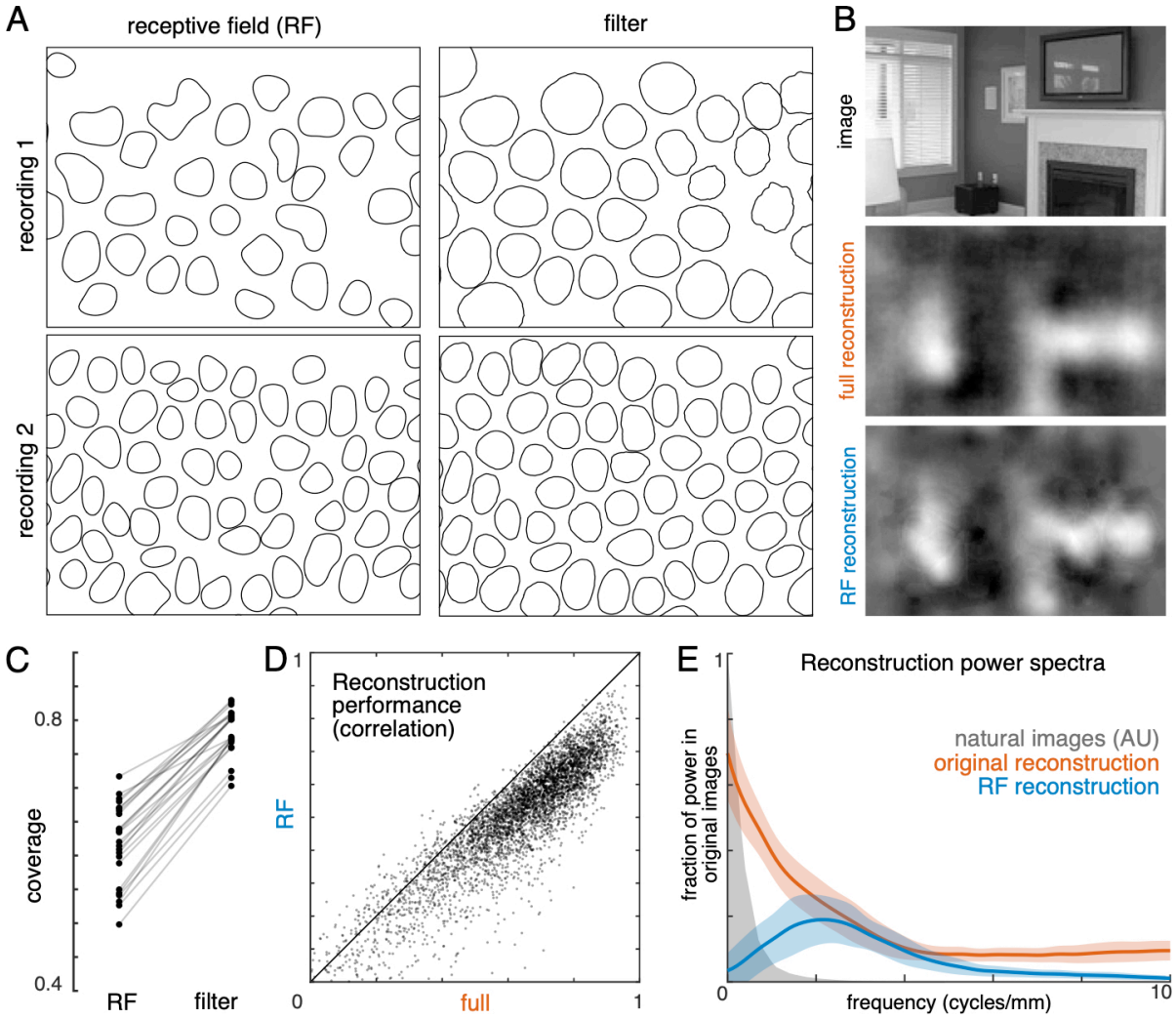

Figure 4: Effect of visual message on reconstruction. A) Receptive field (RF, left) and reconstruction filter (right) contours for two sample recordings. B) Reconstruction of an image (top) using the full, fitted filters (middle) and using scaled RFs (bottom). C) Comparison of RF and filter coverage across recordings. D) Comparison of reconstruction performance using scaled RFs or using full, fitted filters, across n = 4800 images from 8 recordings. E) Power in the reconstructed images (as a fraction of power in the original image) using fitted filters (orange) or scaled RFs (blue). Average (bold) +/- standard deviation (shaded region) across 8 recordings. The original power structure of the natural images is shown in gray and has arbitrary units.

Distinct contributions of major cell types

The visual message transmitted by RGCs of a particular type could additionally be affected by the other cell types encoding the same region of visual space (Warland et al., 1997). To test this possibility, reconstructions were performed using the responses of a given RGC alone, or in combination with each of the four major cell type populations. For each recording, the features of the average reconstruction filters for a given cell type were compared (Figure 5A). Inclusion of any one cell type reduced the magnitude of the cell's reconstruction filter, relative to the single-cell reconstruction condition (Figure 5B, left). This can be understood by noting that the entries in $(R^T R)^{-1}$, which mix the natural image STAs to produce the reconstruction filters, have the opposite sign of the response correlations. As expected, the correlations were positive for same-polarity cells and negative for opposite-polarity cells (not shown; Greschner et al., 2011; Mastronarde, 1983). Therefore, the cell's reconstruction filter was reduced in magnitude by positively weighted cells of the opposite polarity (ON or OFF), and by negatively weighted cells of the same polarity.

As discussed previously, inclusion of the remaining cells of the same type substantially reduced the spatial extent of each parasol cell's filter (Figure 3). However, this did not occur when cells of other types were included in reconstruction instead (Figure 5B, right, top two rows). Specifically, the inclusion of the midget cells with the same polarity only slightly reduced the spatial extent of the cell's filter, and inclusion of opposite polarity cells of either type had little effect. This is likely because the other cell types provide roughly uniform coverage, whereas the remaining cells of the same type have a gap in the location of the primary cell, resulting in significant shaping by the immediately neighboring cells. In summary, the spatial structure of the visual message of a single parasol cell is primarily influenced by neighboring cells of the same type, and is largely unaffected by cells of other types.

The filters for the midget cells were also shaped by the inclusion of the remaining cells of the same type (Figure 5A, second column), and were largely unaffected by the inclusion of opposite polarity cells of either type. However, unlike parasol cells, midget cell filters were significantly affected by the inclusion of the same-polarity parasol cells (Figure 5A, third column). This asymmetry may be due to the fact that parasol cells tended to have much

stronger responses to the natural images than midget cells. Thus, the interpretation of the visual signal from a midget cell does depend somewhat on the signals sent by the same-polarity parasol cell population.

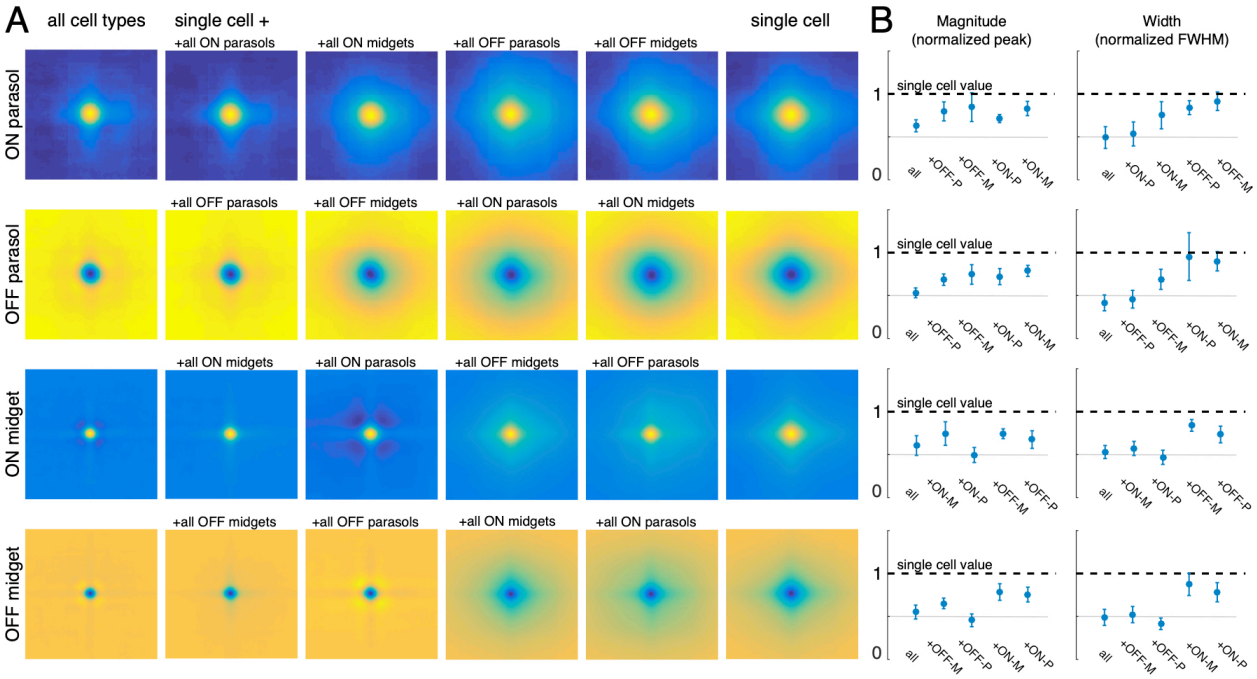

Figure 5: Effect of other cell types on the visual message. A) Average reconstruction filters for ON parasol (top row), OFF parasol (second row), ON midget (third row), and OFF midget (bottom row) cells for one recording. Left to right: including all cell types, all cells of the same type, all cells of the same polarity but opposite class, all cells of the opposite polarity but the same class, all cells of opposite polarity and class, and no other cell types. B) Comparison of magnitude (left) and width (right) of average reconstruction filters across conditions, normalized by the features of the single cell filter. Average +/- standard deviation across recordings is plotted (parasol: n = 11 recordings, midget: n = 5 recordings). Rows correspond to cell types as in A.

The image features represented by each cell type were revealed by analysis of the reconstructed images. In particular, the separate contributions of ON and OFF type cells, and of the parasol and midget cell classes, were investigated.

To estimate the contribution of ON and OFF types, reconstruction was first performed with either ON or OFF parasol cells alone (Figure 6A,B). The reconstruction performance was similar in the two cell types, and was less accurate than reconstruction using both together (Figure 6C, both: $\rho = 0.76 +/- 0.12$, ON: $\rho = 0.64 +/- 0.16$, OFF: $\rho = 0.67 +/- 0.14$, across $n = 2250$ images from 15 recordings). Reconstruction using just ON cells failed to accurately capture intensity variations in dark areas of the image, while reconstruction with just OFF cells failed to capture variations in light areas of the image (for pixels values above the mean value: $\rho = 0.57$ for ON and 0.26 for OFF, for pixel values below the mean value: $\rho = 0.31$ for ON and 0.68 for OFF). Only a narrow middle range of pixel intensities were effectively encoded by both types (Figure 6D). This is consistent with their known output nonlinearities, which suppress responses to stimuli of the non-preferred contrast, and therefore limit linear reconstruction in that range. Thus, both ON and OFF cells were necessary to reconstruct the full range of image contrasts. Surprisingly, reconstruction using the responses of both cell types seemed to encode darker contrast more accurately than lighter contrast (Figure 6D, bottom panel, black curve). This could reflect the fact that ON cells exhibit a more linear contrast-response relationship (Chichilnisky & Kalmar, 2002) and less reconstruction performance difference between preferred and non-preferred contrast, and/or the fact that the natural image distribution is skewed towards darker pixel values (Figure 6D, bottom panel, gray distribution), potentially placing greater weight on the accurate reconstruction of these values.

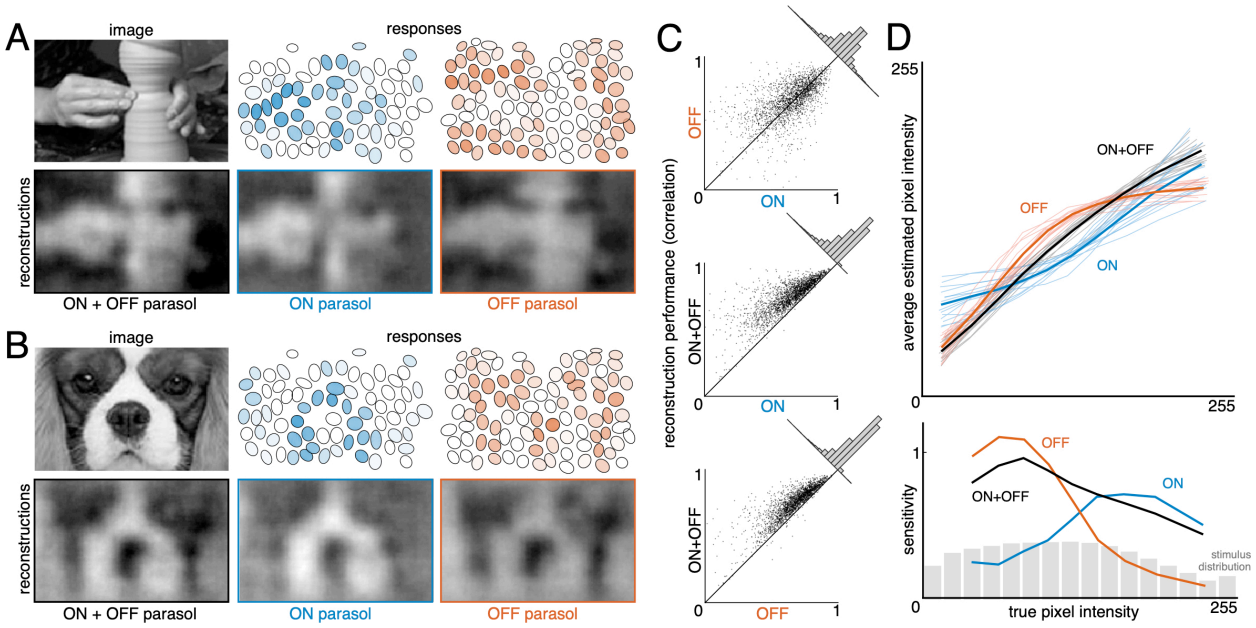

Figure 6: Contributions of ON and OFF parasol cells. A,B) Example images, responses, and reconstructions from ON and OFF parasol cells. Top left: original image. Top right: Parasol cell mosaics shaded by their response value (ON - blue, middle, OFF - orange, right). Bottom left: reconstruction from both cell types. Bottom right: reconstruction from just ON (blue, middle) or just OFF (orange, right) parasol cells. C) Reconstruction performance for ON vs. OFF (top), both vs. ON (middle), and both vs. OFF (bottom), with n = 2250 images from 15 recordings. D) Average reconstructed pixel intensity (top) and sensitivity (bottom, defined as Δaverage reconstructed pixel intensity/Δtrue pixel intensity) vs. true pixel intensity for ON (blue), OFF (orange), or both (black). Individual recordings are shown in the top plot, with the average in bold.

To estimate the contributions of parasol and midget cell classes, reconstruction was performed using either the responses of one class alone (ON and OFF parasol, or ON and OFF midget), or using both classes (Figure 7A,B). As expected, reconstruction using both parasol and midget cells was more accurate than using either alone, and accuracy was similar across the two classes (Figure 7C, both: $\rho = 0.81 +/- 0.10$, parasol: $\rho = 0.77 +/- 0.12$, midget: $\rho = 0.73 +/- 0.13$, across $n = 1050$ images from 7 recordings). In addition, images reconstructed from midget cell responses contained more high frequency spatial information, consistent with their higher density (Figure 7D). This suggests that midget cells convey finer spatial structure in the image, while both parasol and midget cells captured integrated intensity information.

The above analysis obscures the significantly different temporal responses properties of these two cell classes. In particular, parasol cells have more transient responses (De Monasterio, 1978; De Monasterio & Gouras, 1975; Gouras, 1968) which may allow them to convey information more rapidly than midget cells. To test this possibility, image reconstruction was performed using spikes collected over increasing windows of time after the image onset. The reconstruction performance of parasol cells improved quickly and reached 95% of peak reconstruction performance at 80 +/- 20 ms, while the performance of midget cells improved more slowly, and reached 95% performance at 114 +/- 17 ms (across 7 recordings; Figure 7E). This difference indicates that a spatiotemporal reconstruction will be necessary to fully reveal the distinct contributions of these two classes (see below).

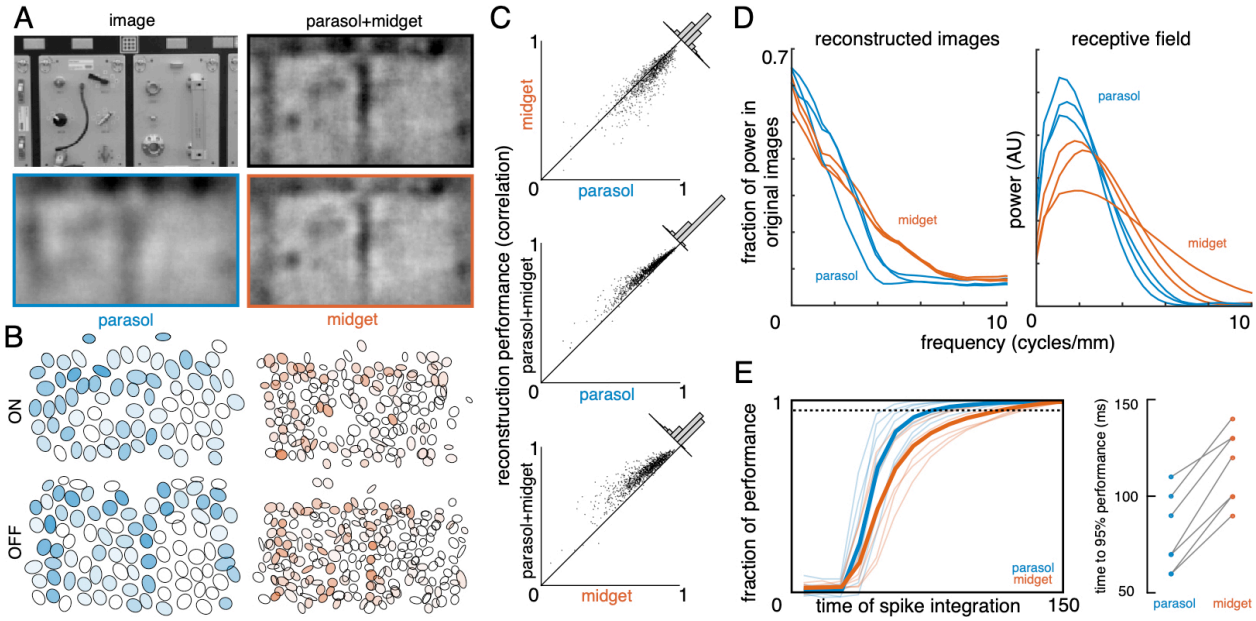

Figure 7: Contributions of the parasol and midget cell classes. A) Example image and reconstructions for parasol and midget cells. Top left: original image. Top right: reconstruction with parasol and midget cells (black). Bottom left: reconstruction with only parasol cells (blue). Bottom right: reconstruction with only midget cells (orange). B) Cell type mosaics shaded by their response values, for ON (top) and OFF (bottom) parasol cells (left, blue) and midget cells (right, orange). C) Reconstruction performance for midget vs. parasol (top), both vs. parasol (middle), and both vs. midget (bottom). D) Power in the reconstructed images as a fraction of power in the original image (left) and receptive fields (right) for parasol (blue) and midget (orange) cells. E) Left: Fraction of peak reconstruction performance with increasing spike integration times for parasol (blue) and midget (orange) cells, with averages across recordings shown in bold. Dotted line indicates 95% performance. Right: Time to 95% performance for parasol vs. midget reconstructions across 7 recordings.

The effect of correlated firing

The above results indicate that the visual message of each RGC, and the contributions of each cell type, are shaped by correlated activity. However, these analyses do not distinguish between stimulus-induced (signal) correlations, and stimulus-independent (noise) correlations that arise from neural circuitry within and across cell types in the primate retina (Greschner et al., 2011; Mastronarde, 1983).

To test the effect of noise correlations, reconstruction performance was evaluated on repeated presentations of test images. This performance was compared to a control condition in which the responses of each cell were independently shuffled across trials to remove noise correlations while preserving single-cell statistics and signal correlations. The reconstruction filters (computed from unshuffled training data) were then used to reconstruct the test images, using either the shuffled or unshuffled responses. In principle, shuffling could increase or decrease reconstruction accuracy. Because the reconstruction filters incorporate correlated activity in training data (Equation 1), any deviation from this correlation structure in the test data could reduce performance. On the other hand, if the noise correlations produced spatial structure in the reconstructions that obscured the structure of the natural images, their removal could lead to enhanced reconstruction performance. The relative influence of these competing effects could also depend on the overall fidelity of the reconstruction.

Accordingly, the shuffling manipulation was tested in three scenarios. In the first, RGC responses were calculated by counting spikes in the 150ms window after image onset, as above. In the second, the effect of correlations was tested at their intrinsic timescale (~10ms; DeVries, 1999; Mastronarde, 1983; Meister et al., 1995; Shlens et al., 2006), by counting spikes in fifteen 10ms bins, and reconstructing with this multivariate response vector instead of the scalar spike count. In the third, the overall fidelity of reconstruction was reduced by counting spikes only in the 10ms bin that had the highest average firing rate, which corresponded to the window from 50ms to 60ms after image onset.

Reconstruction using the first two approaches had similar unshuffled performance (ρ = 0.76 +/- 0.12 and 0.75 +/- 0.12 respectively), and low variation across trials (standard deviation across repeats = 0.015). In these scenarios, shuffling had a minimal, detrimental effect on reconstruction (across 3 recordings with 27 repeats of 150 test images: (1) $\Delta\rho$ = -0.0004 +/- 0.0015; $|\Delta\rho|$ = 0.0012 +/- 0.0010; (2) $\Delta\rho$ = -0.0007 +/- 0.0019; $|\Delta\rho|$ = 0.0014 +/- 0.0014). In each case, the magnitude of the change in correlation represented about 10% of the variation in correlation across trials (Figure 8). Thus, noise correlations had overall limited impact in these conditions, and the time scale of analysis was relatively unimportant.

Reconstruction using the third approach, however, had lower unshuffled performance (ρ=0.64 +/- 0.14), and higher variation across trials (standard deviation across repeats = 0.039). In this case, shuffling led to a larger and consistent increase in reconstruction performance (Δρ = 0.0070 +/- 0.0073; lΔρl = 0.0074 +/- 0.0069). Although the increase was small in absolute terms, it represented a larger fraction of the variation in correlation across trials (20%), which was also higher in this scenario (Figure 8). Thus, noise correlations play a more significant and detrimental role when the retinal signal is noisy, even if reconstruction is designed to take noise correlations into account. These considerations suggest that in low-fidelity situations, noise correlations in the RGC population can more strongly obscure the structure of natural images.

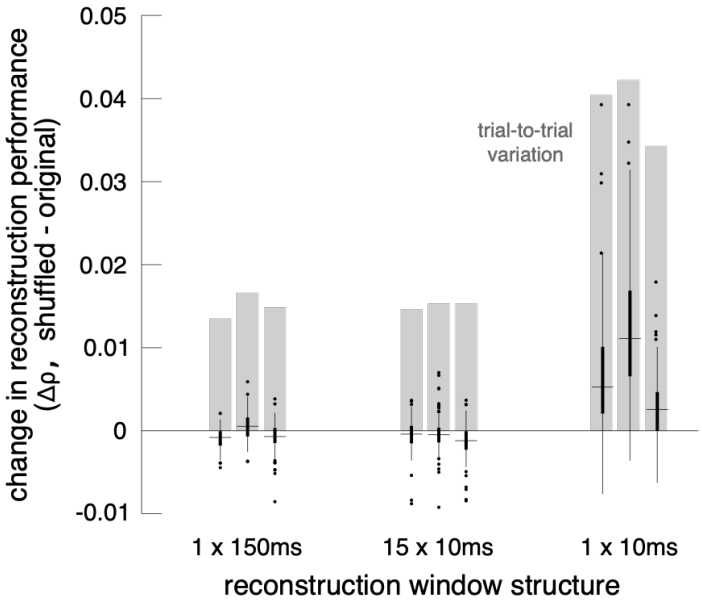

Figure 8: Effect of noise correlations. The change in reconstruction performance (Δρ) when using shuffled data for three scenarios: one 150ms window, fifteen 10ms windows, and one 10ms window. Black bars show median +/- interquartile range for three recordings (each shown separately). Grey bars show the standard deviation in the reconstruction performance across trials.

Nonlinear reconstruction

Linear reconstruction provides an easily interpretable estimate of the visual message, but it may limit the quality of reconstruction by not extracting all of the information available in the neural responses, and may differ greatly from how the brain processes the retinal input. Therefore, two simple extensions of linear reconstruction were tested: transformation of the responses using a static nonlinearity, and inclusion of interaction terms between nearby cells.

In the first case, the response was first transformed using a static nonlinearity, and linear regression was performed between the transformed response and the natural images. The stimulus estimate S_{NL} is given by $S_{NL} = f(R) \cdot W$, where W is a matrix of reconstruction weights, and $f(R)$ is the point-wise nonlinear transform of the population response vector R. This is equivalent to inverting a linear-nonlinear (LN) encoding model of the form $R = g(K \cdot S)$, where g is the inverse of f, and K is a different set of weights (although it is not necessarily true that a nonlinear encoder requires an equivalent nonlinear decoder for optimum performance, see Rieke et al., 1997 for a full discussion). A common form of the LN model uses an exponential nonlinearity, $g = exp()$; therefore, the inverse function $f = log()$ was used for reconstruction. A square root transformation was also tested, and yielded similar results as below (not shown).

The relationship to pixel values was more linear for the transformed RGC responses than for the original responses (Δlinear fit RMSE = -1.9 +/- 1.5 across n = 2225 cells from 15 recordings; Figure 9A,B), indicating that this inverse function captured at least some of the nonlinearity in retinal signals. The nonlinear transformation slightly increased reconstruction accuracy when using the responses of ON or OFF parasol cells alone (across 15 recordings with 300 images each: ON parasol: Δρ = 0.017 +/- 0.053; OFF parasol: Δρ = 0.015 +/- 0.035; Figure 9C), but had little effect when using the responses of both ON and OFF parasol cells together (Δρ = -0.0007 +/- 0.033; Figure 9C). A likely reason for this is that the relationship between the true pixel values and the pixel values reconstructed using the original, untransformed responses was already approximately linear when using both cell types, but not when using just one cell type (Figure 6).

Nonlinear interactions between the signals from different cells could also potentially increase reconstruction performance. To test this idea, the products of spike counts in pairs of neighboring cells were added as predictors in the reconstruction. Neighbors were defined as cells with RF centers that were within 1.5 times the median nearest neighbor distance between RF centers of the cells of the same type. For parasol cells, this definition resulted in roughly 6 ON and 6 OFF neighbors per cell, as expected (see Figure 2). Including these interactions resulted in a modest increase in reconstruction accuracy ($\Delta\rho$ = 0.0093 +/- 0.023, across 3 recordings with 300 test images each). Including only interactions of a specific cell type pairing revealed that the primary contribution was from ON-OFF pairs (ON-OFF: $\Delta\rho$ = 0.0089 +/- 0.019; ON-ON: $\Delta\rho$ = 0.0021 +/- 0.010; OFF-OFF: $\Delta\rho$ = 0.0024 +/- 0.013; Figure 9D). The reconstruction filters associated with these interaction terms typically had biphasic structure oriented along the axis between the cells (Figure 9E,F), suggesting that the improvement in reconstruction may come primarily from using the joint activation of partially overlapping ON and OFF cells to capture edges in the visual scene.

In order to understand the origin of these biphasic filters, the responses of pairs of ON and OFF parasol cells were simulated and used for reconstruction. Mosaics of ON and OFF parasol cells were generated by placing circular, 2D Gaussian receptive fields along a jittered grid (the results were robust to the details of the mosaic generation). The stimulus was filtered using these RFs, and then a rectifying nonlinearity was applied to produce the response (rectified linear, exponential, and sigmoidal functions all produced qualitatively similar results). Then, as above, the simulated responses of both cell types and the products of the responses of neighboring cells were used to reconstruct natural images. The spatial reconstruction filter corresponding to the interaction term between nearby ON and OFF cells was biphasic and qualitatively similar to the reconstruction filters observed in reconstructions from real data (Figure 9E). This was also the case for responses simulated using a nonlinear-linear-nonlinear (NLN) cascade, where the stimulus was rectified before the first summation step (not shown). However, this was not the case for responses simulated using a linear model, without any rectification. In this case, the filter corresponding to the interaction term had no clear structure (not shown). Therefore, some degree of nonlinear encoding is necessary to explain the spatially biphasic nature of the interaction term filters, and a simple linear-nonlinear or NLN encoding model is sufficient.

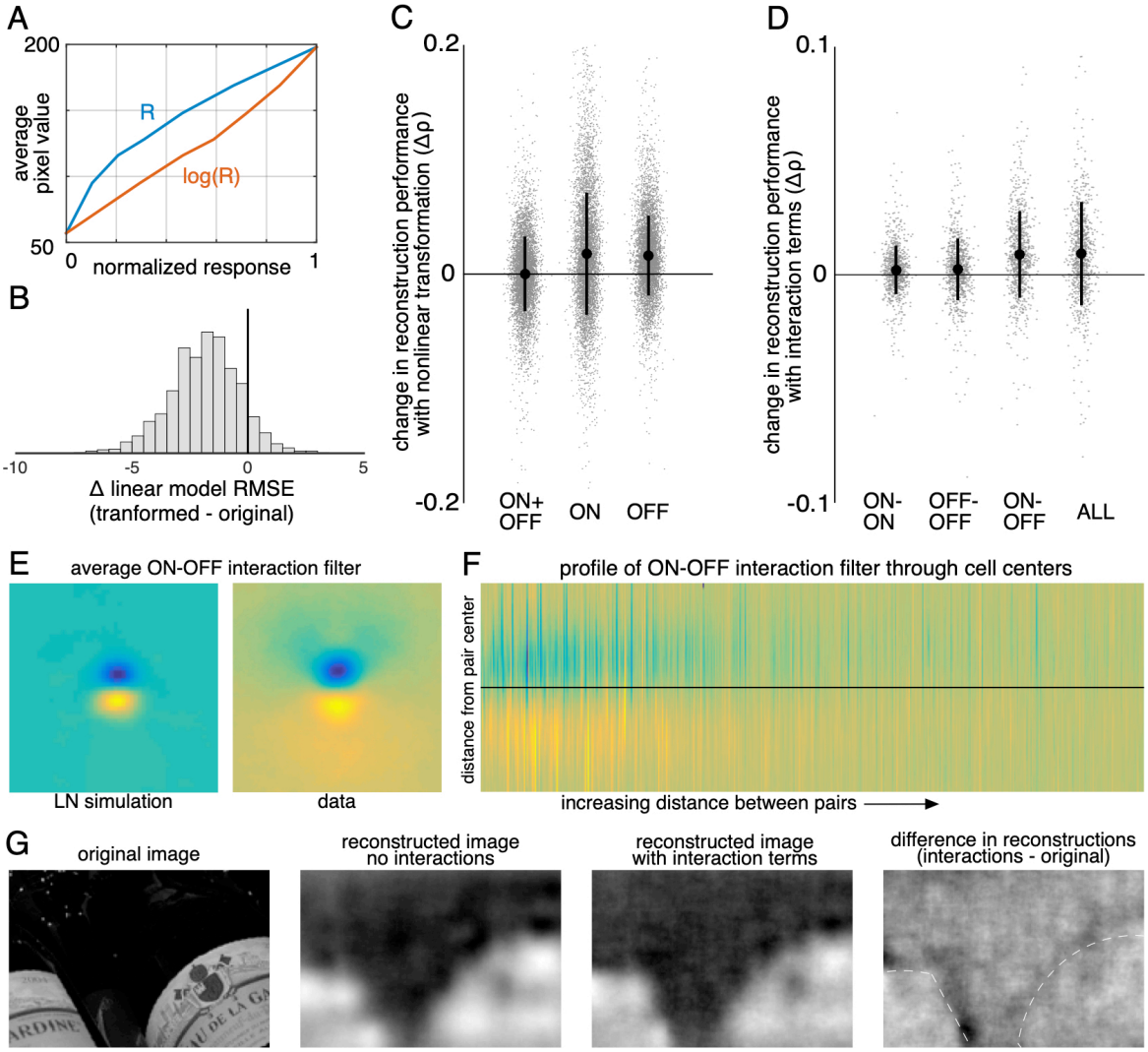

Figure 9: Nonlinear reconstruction. A) Average pixel value in receptive field center vs. original response (blue) and transformed response (orange). B) Distribution (across n = 2225 cells from 15 recordings) of the change in RMSE of a linear model (mapping from response to pixel value) when using the transformed response. C) Change in reconstruction performance (correlation) when using transformed responses (log(R)) for reconstruction with either ON and OFF parasol cells, only ON parasol cells, or only OFF parasol cells. Individual images (n = 300 from each of the 15 recordings) are plotted in gray with jitter in the x-direction. The black bars represent mean +/- standard deviation, and the standard error is smaller than the central dot. D) Change in reconstruction performance (correlation) when including interaction terms. Individual images (n = 300 from each of the 3 recordings) are plotted in gray with jitter in the x-direction. The black bars represent mean +/- standard deviation, and the standard error is smaller than the central dot. E) Average reconstruction filters corresponding to ON-OFF type interactions, centered and aligned along the cell-to-cell axis, for simulation (left) and data (right). F) 1D Profiles of all ON-OFF interaction filters through the cell-to-cell axis, sorted by distance between the pair. G) Example image (left), reconstructions with and without interaction terms (middle), and difference between the reconstructions, with dotted lines indicating edges (right).

Spatial information in a naturalistic movie

In natural vision, a continuous stream of retinal responses is used to make inferences about the dynamic external world. Therefore, the reconstruction approach above – using the accumulated spikes over a fraction of a second to reconstruct a flashed image – could fail to capture important aspects of normal vision. To test whether the above results extend to spatiotemporal reconstructions, a naturalistic movie, consisting of a continuous stream of natural images with simulated eye movements superimposed, was reconstructed from the spike trains of RGCs. The spike trains were binned at the frame rate of the movie (120Hz), and linear regression was performed between the frames of the movie and the RGC responses in 15 bins following each frame, resulting in a spatiotemporal reconstruction filter for each RGC.

A spatial summary of the filter for each cell was obtained by extracting a time course from the spatial pixels that were significantly different from the background (examples shown in Figure 10A; see Methods). This spatial filter was highly correlated with the spatial reconstruction filters obtained in the preceding analysis with flashed images (ρ = 0.87 +/- 0.07, n = 351 parasol cells from 3 recordings; Figure 10B). The dynamic filters displayed high space-time separability (explained variance from first component = 0.78 +/- 0.18). This suggests that the essential spatial features of the visual message observed above extend to the spatiotemporal regime. In addition, the reconstructed movie frames were qualitatively similar to the reconstructions of the static images (ρ between sum over fixation and static reconstruction = 0.72 +/- 0.19 across 120 images from 3 recordings, Figure 10C).

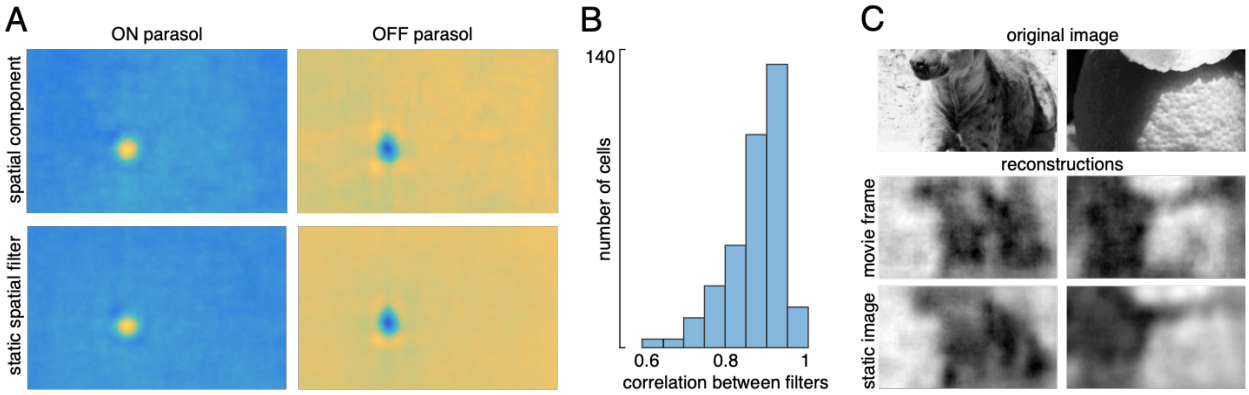

Figure 10: Spatiotemporal reconstruction. A) Examples of the spatial components extracted from the spatiotemporal reconstruction filter (top) and the static spatial reconstruction filters (bottom) for an ON (left) and OFF (right) parasol cell. B) Correlation between spatial component and static filter across n=351 cells from 3 recordings. C) Example reconstructions of movie frames and of static images.

Discussion

Reconstruction of natural images was used to investigate the spatial information transmitted to the brain by complete populations of primate retinal ganglion cells (RGCs). The quality of the reconstructions was consistent across retinas and primarily limited by the density of RGCs. The optimal interpretation of the spikes produced by a RGC – i.e. its visual message – depended not only on its encoding properties, but also on the statistics of natural scenes and the spatial arrangement of neighboring RGCs. These factors enabled smoother natural image reconstructions from the RGC population than would be expected from the RFs alone. In addition, each cell type conveyed different and largely independent features of the visual scene, consistent with their distinct encoding properties. The impact of noise correlations among RGCs was small, but in some conditions constituted a substantial proportion of overall noise. Using a nonlinear transformation to linearize ON and OFF parasol cell responses before reconstruction had little effect, suggesting that the two cell types together provide a fairly linear visual representation of intensity, although individually they do not. Interaction terms between overlapping ON and OFF type cells slightly improved reconstructions by emphasizing sharp transitions between bright and dark areas. Finally, a limited test of spatiotemporal reconstruction indicated that these results may generalize to natural vision.

The results show that the dependence of a given RGC's visual message on the responses of other RGCs, which was demonstrated previously in the temporal domain using a spatially uniform stimulus (Warland et al., 1997), extends to the spatial domain in natural viewing conditions. For decades, the spatial visual message of a RGC has been estimated using its receptive field, measured with artificial stimuli. However, due to spatial correlations in natural scenes, a single RGC's response contains information about the stimulus far beyond its RF. In this light, it is at first surprising that the visual message is spatially localized and similar to the classical RF (Figure 3A,C). However, nearby regions of visual space are already covered by the neighboring RGCs of the same type, and the substantial redundant information in adjacent cells is seemingly not utilized to represent the image structure. Still, the visual messages retain some natural scene structure that differentiates them from the RFs. This structure results in smoother reconstructions and more uniform coverage of visual space than the coverage provided by the RF mosaic (Figure 4). The significance of natural scene statistics for interpreting the neural code has also been suggested in the visual cortex (Naselaris et al.,

2009), and can be used as a prior to improve image estimates in multi-step reconstruction methods (Parthasarathy et al., 2017).

The visual message of each RGC was largely unaffected by RGCs of the other major types. This indicates that the major primate RGC types, despite covering the same region of visual space, conveyed different stimulus features, consistent with their distinct spatial and temporal encoding properties (Figures 6 and 7). However, this separation was clearer for the ON and OFF types than for the parasol and midget cell classes, as the midget cell filters were influenced by the inclusion of same-polarity parasol cells. Alternate reconstruction schemes, particularly incorporating temporal information, may further reveal the interactions between these cell classes. Although it is not clear why the retina separates visual information into separate cell type channels, the roughly linear intensity representation by ON and OFF cell types together is consistent with suggestions that encoding by multiple cell types with nonlinear response properties could enable relatively simple linear reconstruction by downstream areas (Gjorgjieva et al., 2019). Notably, however, the inclusion of interaction terms between nearby cells of different types did slightly improve reconstructions, and the resulting oriented spatial filters were explained by the simple output nonlinearities in RGCs (Figure 9). As new cell types are identified and characterized (Puller et al., 2015; Rhoades et al., 2019), their contributions to vision may be revealed by these linear and simple nonlinear reconstruction approaches.

The visual representation may also be impacted by stimulus-independent noise correlations, which occur both within and across cell types (DeVries, 1999; Greschner et al., 2011; Mastronarde, 1983; Meister et al., 1995). Some studies have shown that taking noise correlations into account can improve the fidelity of reconstruction (Cafaro & Rieke, 2010; Ganmor et al., 2015; Pillow et al., 2008; Puchalla et al., 2005; Zylberberg et al., 2016), but others have indicated that it has little effect (Botella-Soler et al., 2018; Meytlis et al., 2012; Nirenberg et al., 2001). In the present work, the impact of noise correlations on reconstruction was limited by the low trial-to-trial variability (total noise) in responses (and thus in reconstructions). In a low-fidelity reconstruction, performed by reducing the spike integration time window to 10ms to match the timescale of noise correlations, the effect of noise correlations remained small in absolute terms. However, it amounted to ~20% of the trial-to-trial variation in reconstruction performance. While this is not a strict bound, it represents roughly how much improvement

could be expected. It is also possible that the results would be affected by removing noise correlations from both the training and testing data, but evaluating this possibility would require longer repeated presentations of training stimuli than were performed here.

Although linear least squares spatial reconstruction from recorded RGC spike counts enabled investigation of the RGC visual messages, alternative approaches are worth exploring. Different measures of response, such as latency (Gollisch & Meister, 2008; Gütig et al., 2013) and relative activity (Portelli et al., 2016), have been shown to convey more stimulus information than spike counts for non-primates under some conditions, although this was not the case in the present data. Also, although the present results generalized to a limited analysis of natural movies, it is likely that investigation of spatial, temporal and chromatic information together will further reveal the roles of spike timing and multiple cell types in visual signaling and will be necessary for a complete view of the neural code. In addition, recent studies have indicated that nonlinear and deep learning models could improve reconstruction performance for static images, moving patterns, and naturalistic movies (Botella-Soler et al., 2018; Parthasarathy et al., 2017; Zhang et al., 2020). While these approaches make the visual message more difficult to define, they could be used to extract richer information potentially present in RGC responses. Models that are both interpretable and nonlinear could also be used to further investigate the visual message (Pillow et al., 2008). Finally, because pixel-wise mean squared error does not accurately reflect the perceived quality of the visual representation, other metrics for optimization and evaluation of reconstruction should be explored (Wang et al., 2002, 2004).

By projecting neural responses into a common stimulus space, reconstruction enabled direct comparison and evaluation of the visual signals transmitted downstream. The large collection of recordings used here revealed a consistent visual representation across retinas, in spite of differences in mosaics, firing rates, and cell spacing that make comparing the neural response itself almost impossible. The information contained in the retinal signal limits the information available to later visual areas, so the results presented here could inform studies of visual processing in the LGN, V1, and other brain structures. For example, the biphasic nature of the interaction term filters supports the hypothesis that orientation selectivity in the cortex results from pairs of nearby ON and OFF RGCs (Paik & Ringach, 2011; Ringach, 2007). In addition, comparing reconstructions from different visual areas using a standard measurement – the

reconstructed image – could help inform how information about the external world is represented at various stages of the visual system.

Using reconstruction to understand the signals transmitted by neurons may be increasingly important in future efforts to read and write neural codes using brain-machine interfaces (BMIs). In the retina, certain types of blindness can be treated with implants that use electrical stimulation to activate the remaining retinal neurons (Goetz & Palanker, 2016), and the visual messages described in the present work can help guide the selection of optimal stimulation patterns (Goetz & Palanker, 2016; Golden et al., 2019; Shah et al., 2019). Reconstruction can also be used to compare the evoked visual representation with the representation produced by natural neural activity. Outside the visual system, many BMIs rely on reconstruction to read out and interpret neural activity, e.g. controlling prosthetic limbs using activity recorded in the motor cortex (Lawhern et al., 2010; Vargas-Irwin et al., 2010). While these studies typically focus on the ability to perform specific tasks, the present results suggest that a deeper examination of the reconstruction filters could further reveal contributions of different cells and cell types in diverse modalities.

Materials and Methods

Experimental methods

Multielectrode array recordings

A *ex vivo* multielectrode array preparation was used to obtain recordings from the major types of primate RGCs (Chichilnisky & Kalmar, 2002; Field et al., 2010; Frechette et al., 2005; Litke et al., 2004). Briefly, eyes were enucleated from terminally anesthetized macaques used by other researchers in accordance with institutional guidelines for the care and use of animals. Immediately after enucleation, the anterior portion of the eye and vitreous were removed in room light, and the eye cup was placed in a bicarbonate-buffered Ames' solution (Sigma, St. Louis, MO). In dim light, pieces of retina roughly 3 mm in diameter and ranging in eccentricity from 7 to 17 mm (6-12 mm temporal equivalent eccentricity; Chichilnisky & Kalmar, 2002) or 29-56 degrees (Dacey & Petersen, 1992; Perry & Cowey, 1985), were placed RGC side down on a planar array consisting of 512 extracellular microelectrodes covering a 1.8 mm x 0.9 mm region (roughly 4x8° visual field angle). In all but one preparation, the retinal pigment epithelium (RPE) was left attached to allow for photopigment regeneration and to improve tissue stability, but the choroid (up to Bruch's membrane) was removed to allow oxygenation and maintain even thickness. For the duration of the recording, the preparation was perfused with Ames' solution (30-34° C, pH 7.4) bubbled with 95% O_2, 5% CO_2. The raw voltage traces recorded on each electrode were bandpass filtered, amplified, and digitized at 20kHz (Litke et al., 2004). Spikes from individual neurons were identified by standard spike sorting techniques, and only spike trains from cells exhibiting a 1ms refractory period were analyzed further (Field et al., 2007; Litke et al., 2004).

Visual stimulation

The visual stimulus was produced by a 120Hz, gamma-corrected, CRT monitor (Sony Trinitron Multiscan E100; Sony, Tokyo, Japan), which was optically reduced and projected through the mostly-transparent array onto the retina at low photopic light levels (2000, 1800, and 800 isomerizations per second for the L, M and S cones respectively at 50% illumination; see Field et al., 2009, 2010). The total visual stimulus area was 3.5 by 1.75 mm, which extended well beyond the recording area.

A 30-minute spatiotemporal white noise stimulus was used to characterize RGC responses and to periodically assess recording quality (Chichilnisky, 2001). The stimulus was updated at either 30 or 60 Hz, and consisted of a grid of pixels (spacing ranged from 44 to 88μm across recordings). For each update, the intensities for each of the three monitor primaries at each pixel location were chosen randomly from a binary distribution.

Natural images from the ImageNet database (Fei-Fei et al., 2010) were converted to grayscale values. On a scale of 0 to 1, the mean image intensity was 0.45. The natural images were displayed at either 320x160 pixels, with each pixel measuring 11x11 μm on the retina, or at 160x80 pixels, with each pixel measuring 22x22 μm on the retina. The images were displayed for 100ms each (12 frames at 120Hz), separated by spatially uniform gray at intensity 0.45 for 400 ms, chosen to ensure a return to the average firing rates. The images were displayed in blocks of 1000, interleaved with a repeated set of 150 test images. Stimulation durations ranged from 5 to 40 blocks.

Dynamic movies consisted of the same set of images, each displayed for 500ms with eye movements simulated as Brownian motion with a diffusion constant of $10\mu m^2$/frame, selected to roughly match recorded eye movements from humans (Kuang et al., 2012; Van Der Linde et al., 2009) and primate (Z.M. Hafed and R.J. Krauzlis, personal communication, June 2008). After 500ms, a new image appeared, with no gray screen between image presentations, and again was jittered. Each recording consisted of 5000 images, for a total of 300,000 frames of stimulation.

Cell type classification

The spike triggered average (STA) stimulus for each neuron was computed from the response to the white noise stimulus (Chichilnisky, 2001), to reveal the spatial, temporal, and chromatic properties of the light response. Cell type identification was performed by identifying distinct clusters in the response properties, including features of the time course and the spike train autocorrelation function extracted via principal components analysis, and the spatial extent of the receptive field (RF; Chichilnisky & Kalmar, 2002; Dacey, 1993; DeVries & Baylor, 1997; Field et al., 2007; Frechette et al., 2005). This analysis revealed multiple identifiable and complete cell type populations. In particular, the four major types, ON and OFF parasol and midget cells,

were readily identifiable by their temporal properties, RF size, density, and mosaic organization (see Rhoades et al., 2019 for a more detailed discussion). Recorded populations of parasol cells formed nearly complete mosaics over the region of retina recorded; recorded midget cell populations were less complete.

Linear reconstruction

Linear regression

Reconstruction filters were fitted using linear regression, as described in Results. The responses of every RGC were included in the regression for every pixel; restricting the filters to a local area did not improve reconstructions. Prior to regression, the distribution of each cell's responses and the pixel values at each location were centered around 0 (i.e. the mean over samples was subtracted in each case). The length of time over which spikes were counted after the image onset was chosen to optimize reconstruction performance (tested in 10ms intervals from 10ms to 200ms; see Figure 7E). For the spike latency comparison, a maximum time of 150ms was assigned to cells that had not yet spiked.

Convergence of estimates

For all recordings, reconstruction performance obtained with half of the data was typically 95-98% of the reconstruction performance obtained with the full data (Figure M1). Both an L2-penalty on filter coefficients and applying a singular value cutoff when calculating the pseudoinverse of the response matrix (Golden et al., 2019; Strang, 1980) were tested as methods for optimizing performance with limited data. However, neither improved reconstruction performance.

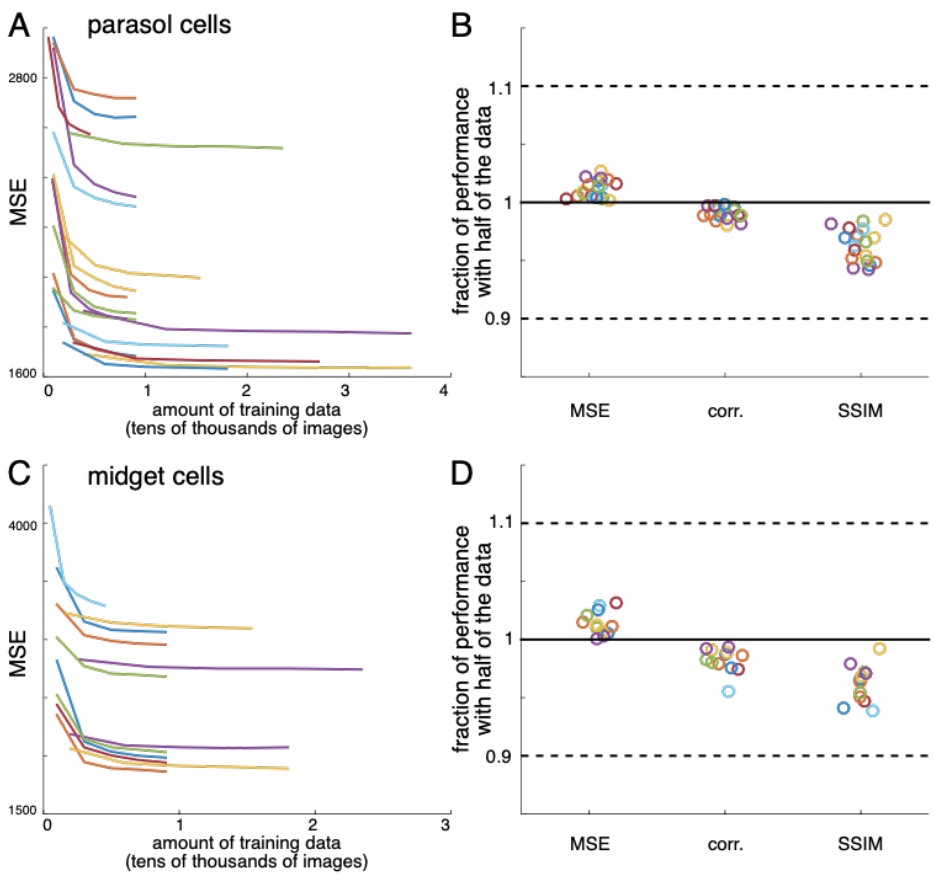

Figure M1: Convergence of estimates. A) Performance of reconstructions from parasol cell responses as a function of the amount of training data, for 19 recordings (colors). B) Fraction of performance of reconstructions from parasol cell responses (MSE, correlation, and SSIM) achieved with half of the training data for each recording. C,D) Same as A,B for reconstructions from midget cell responses for 12 recordings (colors).

Image region selection

Reconstruction performance was calculated over the image regions covered by the RFs of the recorded RGCs. To define this area, the spatial profile of each RF was fitted with a two-dimensional elliptical Gaussian (Chichilnisky & Kalmar, 2002), and any pixel within two standard deviations was considered covered (Figure M2). For each analysis, pixels were only included if they were covered by at least one of each cell type used in that analysis. Two analyses used a manually selected, rectangular central image region instead of the mosaic coverage logic above: the comparison across recordings (Figure 2), and the spatial frequency analyses (Figures 4 and 6).

Error metrics

The primary measures of reconstruction performance, mean squared error (MSE) and the correlation coefficient, were calculated between the original and reconstructed image, across all included pixels (as defined above). Note that linear least squares regression, which was used to obtain the filters, by definition minimizes MSE on the training data, but does not necessarily maximize the correlation coefficient. In addition, an alternative measure more closely related to perceptual difference between images, the structural similarity (SSIM; Wang et al., 2004), was calculated across the whole image (parameters: radius = 22 μm, exponents = [1 1 1]), and then averaged across the included pixels (see above) for each image. In all cases, similar trends were observed with each metric.

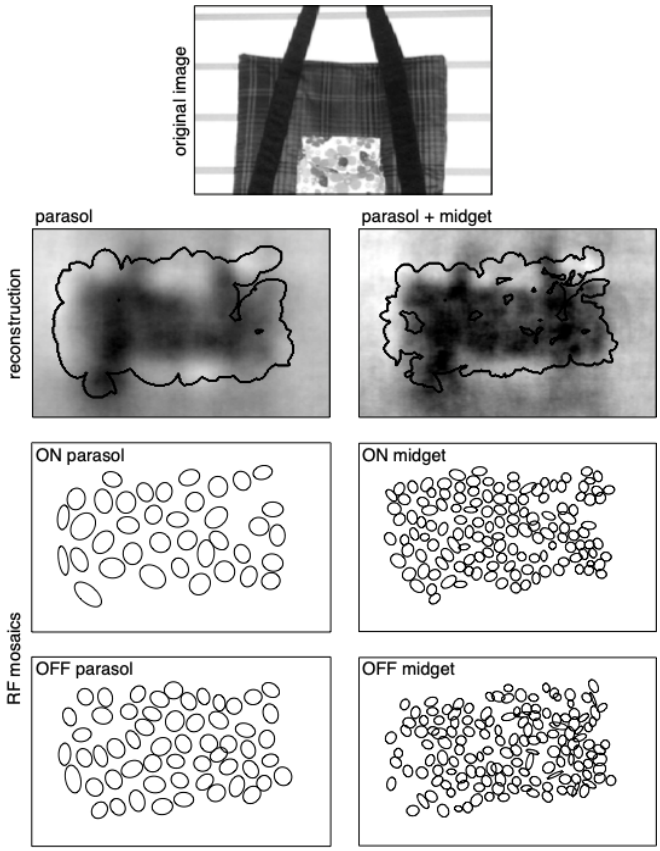

Figure M2: Mosaic coverage and region of analysis. Reconstruction performance on a sample image (top) is measured by comparing the regions inside the contours shown on the reconstructions in the second row. These contours were obtained using the receptive field mosaics (bottom two rows) of parasol cells, or of both parasol and midget cells, as described. The bounding boxes mark the extent of the visual stimulus.

Filter analysis

Average filter calculations

The spatial receptive field (RF; used in Figures 3 and 4) was extracted from the full spatial, temporal, and chromatic spike-triggered average (STA; used for cell type classification as described above) as follows. First, the values at each pixel location and time in the STA were summed across the color channels. Significant pixels were identified as those with an absolute maximum value (across time) of more than 5 times the robust standard deviation of all the pixels in the STA (Freeman et al., 2015). Averaging across these significant pixels resulted in a single time course. The inner product of this time course with the time course of each pixel in the STA was then computed, resulting in a spatial RF.

Average RFs (Figure 3) were calculated by first upsampling the spatial RFs (with linear interpolation) to match the resolution of the reconstruction filters (across recordings, scaling ranged from 2-8x), then aligning the RF centers (obtained by fitting a 2D Gaussian to the RF as described above) and averaging. Average reconstruction filters (Figure 3) were not upsampled, but otherwise were calculated the same way. The average RFs and filters shown in Figure 3C were calculated separately for each recording, cell type and condition. A one-dimensional profile through the center of each average reconstruction filter was used to calculate full width at half maximum (Figure 3D,E). This calculation was robust to the angle of the profile. The average filters in Figure 5 only included cells in regions with locally dense populations of all four major cell types (defined by the number of nearby cells of each type).

Receptive field reconstruction

Reconstruction from receptive fields (RFs; Figure 4) was performed as follows. Each image was estimated as a sum of RFs, weighted by the RGC response and a fitted scale factor. These scale factors were calculated by minimizing the MSE between the true and estimated images as follows:

$$a^* = argmin_a \sum_{i=1}^{n_{images}} (\widehat{S_i} - S_i)^2 ; \quad \widehat{S_i} = \sum_{c=1}^{n_{cells}} F_c \cdot R_{i,c} \cdot a_c \qquad (2)$$

where S is the stimulus, R is the response, F is the RF, and a is the scale factor, calculated using linear least squares regression (as described above). In this case, each pixel in each image was considered a separate sample, and was modeled as a linear combination of the image responses of all RGCs multiplied by the respective values of their RFs at that pixel. Therefore, the outputs were a vector with length equal to N, the number of images times the number of pixels in each image. The input (regressor) matrix had dimensions (N x number of cells), and the weight vector a had dimensions (number of cells x 1).

Analysis of cell type contributions

ON and OFF parasol cells

Images were reconstructed from the responses of either ON or OFF parasol cells and performance was calculated, as described above. The relationship between true and reconstructed pixel value (Figure 6D) was calculated for each recording by first binning the true pixel values by percentile, resulting in bins with equal numbers of samples. Then, for each bin, the average true pixel value and the average of the corresponding reconstructed pixel values were calculated. The sensitivity was defined as the change in average reconstructed pixel value divided by the change in true pixel value across bins. The observed trends were not dependent on the number of bins.

Parasol and midget cell classes

Images were reconstructed from the responses of either parasol or midget cell classes (including both ON and OFF types) and performance was calculated, as described above. The power spectra for the reconstructed images, original images, and average RFs (Figure 7D) were calculated by discrete Fourier transform. The temporal properties of the parasol and midget classes (Figure 7E) were compared by gradually increasing the length of the window over which spikes were counted after image onset, from 10ms to 150 ms (in 10ms increments). For each window size, the reconstruction filters were refitted, and the performance was calculated as described above.

For these analyses, only the recordings with the highest midget cell coverage were used, defined by the fraction of pixels included in a parasol cell analysis that would also be included

in a midget cell analysis (see *Image region selection* above). 7 recordings were included for measuring reconstruction performance (Figure 7C) and comparing temporal properties (Figure 7E). Only 3 of those were also included in the spatial frequency analysis (Figure 7D), which required complete or nearly complete mosaics.

Analysis of noise correlations

Noise correlation analysis (Figure 8) was limited to the 3 recordings with the most repeated presentations of the same set of test images (27 repeats each). For each of the three scenarios described in Results, reconstruction filters were fitted on a single repeat of training data, and then tested using either shuffled or unshuffled testing data. The testing data was shuffled by randomly permuting each RGC's responses independently across repeated presentations of the same image. Reconstruction performance on the test data was measured as described earlier.

Interaction terms

Only the three recordings with the most training data were included (at least 25,000 training images each; the same subset was used for the noise correlation analysis). For the simulation in Figure 9E, ON and OFF parasol mosaics were generated by placing two-dimensional circular Gaussian RFs (Chichilnisky & Kalmar, 2002) on a jittered grid. The standard deviation of the Gaussian was one fifth of the grid spacing, although the results were robust to the details of the RF mosaic generation. The mean pixel value over all images was subtracted from each image before filtering. The nonlinear-linear-nonlinear cascade was simulated by thresholding the image according to the preferred polarity for each cell type prior to filtering.

Spatiotemporal reconstruction

Each frame of the spatiotemporal movie was reconstructed using the RGC spikes recorded during that frame and the following frames. Therefore, each RGC included in the reconstruction was fitted with a full-rank, spatiotemporal reconstruction filter. The spikes were binned at the frame rate of the movie, and a filter length of 15 frames (125ms) was selected to optimize performance. A spatial summary of the spatiotemporal filter (Figure 10A,B) was calculated as described above for spatial RFs. The spacetime separability of the filters was

calculated using the explained variance from the first component of a singular value decomposition (limited to a spatially local region to reduce the effects of the many low-magnitude, noisy pixels outside the primary filter peak). Three recordings that contained responses to both static, flashed natural images and dynamic, spatiotemporal natural movies were included. 2400 consecutive movie frames were withheld from fitting for comparison of movie frame and static image reconstructions (Figure 10C).

Acknowledgements

This work was supported by NSF IGERT 0801700 (N.B.), NSF GRFP DGE-114747 (N.B., C.R.), NEI F31EY027166 (C.R.), Pew Charitable Trusts Fellowship in Biomedical Sciences (A.S.), donation from John Chen (A.M.L.), NIH R01EY017992, NIH NEI R01-EY029247, NSF/NIH CRCNS Grant IIS-1430348 (E.J.C.), and the Wu Tsai Neurosciences Institute. We thank Fred Rieke for helpful suggestions on the manuscript; Jill Desnoyer and Ryan Samarakoon for technical assistance; Sasi Madugula, Eric Wu and Alex Gogliettino for discussions and feedback; and Corinna Darian-Smith and Tirin Moore (Stanford), Jose Carmena and Jack Gallant (UC Berkeley), Jonathan Horton (UCSF), and the UC Davis Primate Center for access to primate retinas.

Competing interests

The authors declare no competing interests.

References

Bialek, W., Rieke, F., de Ruyter van Steveninck, R. R., & Warland, D. (1991). Reading a neural code. *Science*, *252* (5014), 1854–1857. https://doi.org/10.1126/science.2063199

Botella-Soler, V., Deny, S., Martius, G., Marre, O., Tkačik, G. (2018). Nonlinear decoding of a complex movie from the mammalian retina. PLoS Computational Biology, 14(5), e1006057. https://doi.org/10.1371/journal.pcbi.1006057

Cafaro, J., & Rieke, F. (2010). Noise correlations improve response fidelity and stimulus encoding. *Nature*, *468* (7326), 964–967. https://doi.org/10.1038/nature09570

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

Chichilnisky, E. J., & Kalmar, R. S. (2002). Functional asymmetries in ON and OFF ganglion cells of primate retina. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *22*(7), 2737–2747. https://doi.org/10.1523/JNEUROSCI.22-07-02737.2002

Dacey, D. M. (1993). The mosaic of midget ganglion cells in the human retina. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *13*(12), 5334–5355. https://doi.org/10.1523/JNEUROSCI.13-12-05334.1993

Dacey, D. M., & Petersen, M. R. (1992). Dendritic field size and morphology of midget and parasol ganglion cells of the human retina. *Proceedings of the National Academy of Sciences of the United States of America*, *89*(20), 9666–9670. https://doi.org/10.1073/pnas.89.20.9666

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. https://doi.org/10.1016/S0896-6273(02)01143-1

De Monasterio, F. M. (1978). Properties of concentrically organized X and Y ganglion cells of macaque retina. In *Journal of Neurophysiology* (Vol. 41, Issue 6, pp. 1394–1417). https://doi.org/10.1152/jn.1978.41.6.1394

De Monasterio, F. M., & Gouras, P. (1975). Functional properties of ganglion cells of the rhesus monkey retina. *The Journal of Physiology*, *251*(1), 167–195. https://doi.org/10.1113/jphysiol.1975.sp011086

DeVries, S. H. (1999). Correlated firing in rabbit retinal ganglion cells. *Journal of Neurophysiology*, *81*(2), 908–920. https://doi.org/10.1152/jn.1999.81.2.908

DeVries, S. H., & Baylor, D. A. (1997). Mosaic arrangement of ganglion cell receptive fields in rabbit retina. *Journal of Neurophysiology*, *78*(4), 2048–2060. https://doi.org/10.1152/jn.1997.78.4.2048

Fei-Fei, L., Deng, J., & Li, K. (2010). ImageNet: Constructing a large-scale image database. In *Journal of Vision* (Vol. 9, Issue 8, pp. 1037–1037). https://doi.org/10.1167/9.8.1037

Field, G. D., Gauthier, J. L., Sher, A., Greschner, M., Machado, T. A., Jepson, L. H., Shlens, J., Gunning, D. E., Mathieson, K., Dabrowski, W., Paninski, L., Litke, A. M., & Chichilnisky, E. J. (2010). Functional connectivity in the retina at the resolution of photoreceptors. *Nature*, *467*(7316), 673–677. https://doi.org/10.1038/nature09424

Field, G. D., Greschner, M., Gauthier, J. L., Rangel, C., Shlens, J., Sher, A., Marshak, D. W., Litke, A. M., & Chichilnisky, E. J. (2009). High-sensitivity rod photoreceptor input to the blue-yellow color opponent pathway in macaque retina. *Nature Neuroscience*, *12*(9), 1159–1164. https://doi.org/10.1038/nn.2353

Field, G. D., Sher, A., Gauthier, J. L., Greschner, M., Shlens, J., Litke, A. M., & Chichilnisky, E. J. (2007). Spatial properties and functional organization of small bistratified ganglion cells in primate retina. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *27*(48), 13261–13272. https://doi.org/10.1523/JNEUROSCI.3437-07.2007

Frechette, E. S., Sher, A., Grivich, M. I., Petrusca, D., Litke, A. M., & Chichilnisky, E. J. (2005). Fidelity of the ensemble code for visual motion in primate retina. *Journal of Neurophysiology*, *94*(1), 119–135. https://doi.org/10.1152/jn.01175.2004

Freeman, J., Field, G. D., Li, P. H., Greschner, M., Gunning, D. E., Mathieson, K., Sher, A., Litke, A. M., Paninski, L., Simoncelli, E. P., & Chichilnisky, E. J. (2015). Mapping nonlinear receptive field structure in primate retina at single cone resolution. *eLife*, *4*. https://doi.org/10.7554/eLife.05241

Ganmor, E., Segev, R., & Schneidman, E. (2015). A thesaurus for a neural population code. *eLife*, *4*. https://doi.org/10.7554/eLife.06134

Girshick, A. R., Landy, M. S., & Simoncelli, E. P. (2011). Cardinal rules: visual orientation perception reflects knowledge of environmental statistics. In *Nature Neuroscience* (Vol. 14, Issue 7, pp. 926–932). https://doi.org/10.1038/nn.2831

Gjorgjieva, J., Meister, M., & Sompolinsky, H. (2019). Functional diversity among sensory neurons from efficient coding principles. *PLoS Computational Biology*, *15*(11), e1007476. https://doi.org/10.1371/journal.pcbi.1007476

Goetz, G. A., & Palanker, D. V. (2016). Electronic approaches to restoration of sight. *Reports on Progress in Physics*, *79*(9), 096701. https://doi.org/10.1088/0034-4885/79/9/096701

Golden, J. R., Erickson-Davis, C., Cottaris, N. P., Parthasarathy, N., Rieke, F., Brainard, D. H., Wandell, B. A., & Chichilnisky, E. J. (2019). Simulation of visual perception and learning with a retinal prosthesis. *Journal of Neural Engineering*, *16*(2), 025003. https://doi.org/10.1088/1741-2552/aaf270

Gollisch, T., & Meister, M. (2008). Rapid Neural Coding in the Retina with Relative Spike Latencies. In *Science* (Vol. 319, Issue 5866, pp. 1108–1111). https://doi.org/10.1126/science.1149639

Gouras, P. (1968). Identification of cone mechanisms in monkey ganglion cells. In *The Journal of Physiology* (Vol. 199, Issue 3, pp. 533–547). https://doi.org/10.1113/jphysiol.1968.sp008667

Greschner, M., Shlens, J., Bakolitsa, C., Field, G. D., Gauthier, J. L., Jepson, L. H., Sher, A., Litke, A. M., & Chichilnisky, E. J. (2011). Correlated firing among major ganglion cell types in primate retina. *The Journal of Physiology*, *589*(Pt 1), 75–86. https://doi.org/10.1113/jphysiol.2010.193888

Gütig, R., Gollisch, T., Sompolinsky, H., & Meister, M. (2013). Computing complex visual features with retinal spike times. *PloS One*, *8*(1), e53063. https://doi.org/10.1371/journal.pone.0053063

Kuang, X., Poletti, M., Victor, J. D., & Rucci, M. (2012). Temporal encoding of spatial information during active visual fixation. *Current Biology: CB*, *22*(6), 510–514. https://doi.org/10.1016/j.cub.2012.01.050

Kuffler, S. W. (1953). Discharge patterns and functional organization of mammalian retina. *Journal of Neurophysiology*, *16*(1), 37–68. https://doi.org/10.1152/jn.1953.16.1.37

Lawhern, V., Wu, W., Hatsopoulos, N., & Paninski, L. (2010). Population decoding of motor cortical activity using a generalized linear model with hidden states. *Journal of Neuroscience Methods*, *189*(2), 267–280. https://doi.org/10.1016/j.jneumeth.2010.03.024

Lettvin, J., Maturana, H., McCulloch, W., & Pitts, W. (1959). What the Frog's Eye Tells the Frog's Brain. In *Proceedings of the IRE* (Vol. 47, Issue 11, pp. 1940–1951). https://doi.org/10.1109/jrproc.1959.287207

Litke, A. M., Bezayiff, N., Chichilnisky, E. J., Cunningham, W., Dabrowski, W., Grillo, A. A., Grivich, M., Grybos, P., Hottowy, P., Kachiguine, S., Kalmar, R. S., Mathieson, K., Petrusca, D., Rahman, M., & Sher, A. (2004). What does the eye tell the brain?: Development of a system for the large-scale recording of retinal output activity. In *IEEE Transactions on Nuclear Science* (Vol. 51, Issue 4, pp. 1434–1440). https://doi.org/10.1109/tns.2004.832706

Mastronarde, D. N. (1983). Interactions between ganglion cells in cat retina. *Journal of Neurophysiology*, *49*(2), 350–365. https://doi.org/10.1152/jn.1983.49.2.350

Meister, M., Lagnado, L., & Baylor, D. A. (1995). Concerted signaling by retinal ganglion cells. *Science*, *270*(5239), 1207–1210. https://doi.org/10.1126/science.270.5239.1207

Meytlis, M., Nichols, Z., & Nirenberg, S. (2012). Determining the role of correlated firing in large populations of neurons using white noise and natural scene stimuli. *Vision Research*, *70*, 44–53. https://doi.org/10.1016/j.visres.2012.07.007

Naselaris, T., Prenger, R. J., Kay, K. N., Oliver, M., & Gallant, J. L. (2009). Bayesian reconstruction of natural images from human brain activity. *Neuron*, *63*(6), 902–915. https://doi.org/10.1016/j.neuron.2009.09.006

Nirenberg, S., Carcieri, S. M., Jacobs, A. L., & Latham, P. E. (2001). Retinal ganglion cells act largely as independent encoders. *Nature*, *411*(6838), 698–701. https://doi.org/10.1038/35079612

Paik, S.-B., & Ringach, D. L. (2011). Retinal origin of orientation maps in visual cortex. *Nature Neuroscience*, *14*(7), 919–925. https://doi.org/10.1038/nn.2824

Parthasarathy, N., Batty, E., Falcon, W., Rutten, T., Rajpal, M., Chichilnisky, E. J., & Paninski, L. (2017). *Neural Networks for Efficient Bayesian Decoding of Natural Images from Retinal Neurons*. https://doi.org/10.1101/153759

Perry, V. H., & Cowey, A. (1985). The ganglion cell and cone distributions in the monkey's retina: implications for central magnification factors. *Vision Research*, *25*(12), 1795–1810. https://doi.org/10.1016/0042-6989(85)90004-5

Pillow, J. W., Shlens, J., Paninski, L., Sher, A., Litke, A. M., Chichilnisky, E. J., & Simoncelli, E. P. (2008). Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature*, *454*(7207), 995–999. https://doi.org/10.1038/nature07140

Portelli, G., Barrett, J. M., Hilgen, G., Masquelier, T., Maccione, A., Di Marco, S., Berdondini, L., Kornprobst, P., & Sernagor, E. (2016). Rank Order Coding: a Retinal Information Decoding Strategy Revealed by Large-Scale Multielectrode Array Retinal Recordings. *eNeuro*, *3*(3). https://doi.org/10.1523/ENEURO.0134-15.2016

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. https://doi.org/10.1016/j.neuron.2005.03.026

Puller, C., Manookin, M. B., Neitz, J., Rieke, F., & Neitz, M. (2015). Broad Thorny Ganglion Cells: A Candidate for Visual Pursuit Error Signaling in the Primate Retina. In *Journal of Neuroscience* (Vol. 35, Issue 13, pp. 5397–5408). https://doi.org/10.1523/jneurosci.4369-14.2015

Rhoades, C. E., Shah, N. P., Manookin, M. B., Brackbill, N., Kling, A., Goetz, G., Sher, A., Litke, A. M., & Chichilnisky, E. J. (2019). Unusual Physiological Properties of Smooth Monostratified Ganglion Cell Types in Primate Retina. *Neuron*, *103*(4), 658–672.e6. https://doi.org/10.1016/j.neuron.2019.05.036

Rieke, F., Warland, D., de Ruyter van Steveninck, R. R., Bialek, W. (1997). *Spikes: Exploring the Neural Code*. Bradford Books.

Ringach, D. L. (2007). On the Origin of the Functional Architecture of the Cortex. In *PLoS ONE* (Vol. 2, Issue 2, p. e251). https://doi.org/10.1371/journal.pone.0000251

Ruderman, D. L., & Bialek, W. (1994). Statistics of natural images: Scaling in the woods. In *Physical Review Letters* (Vol. 73, Issue 6, pp. 814–817). https://doi.org/10.1103/physrevlett.73.814

Shah, N. P., Madugula, S., Grosberg, L., Mena, G., Tandon, P., Hottowy, P., Sher, A., Litke, A., Mitra, S., & Chichilnisky, E. J. (2019). Optimization of Electrical Stimulation for a High-Fidelity Artificial Retina. In *2019 9th International IEEE/EMBS Conference on Neural Engineering (NER)*. https://doi.org/10.1109/ner.2019.8716987

Shlens, J., Field, G. D., Gauthier, J. L., Grivich, M. I., Petrusca, D., Sher, A., Litke, A. M., & Chichilnisky, E. J. (2006). The structure of multi-neuron firing patterns in primate retina. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *26*(32), 8254–8266. https://doi.org/10.1523/JNEUROSCI.1282-06.2006

Stanley, G. B., Li, F. F., & Dan, Y. (1999). Reconstruction of natural scenes from ensemble responses in the lateral geniculate nucleus. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *19*(18), 8036–8042. https://doi.org/10.1523/JNEUROSCI.19-18-08036.1999

Strang, G. (1980). *Linear Algebra and Its Applications*.

Van Der Linde, I., Rajashekar, U., Bovik, A. C., & Cormack, L. K. (2009). DOVES: a database of visual eye movements. *Spatial Vision*, *22*(2), 161–177. https://doi.org/10.1163/156856809787465636

Vargas-Irwin, C. E., Shakhnarovich, G., Yadollahpour, P., Mislow, J. M. K., Black, M. J., & Donoghue, J. P. (2010). Decoding complete reach and grasp actions from local primary motor cortex populations. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *30*(29), 9659–9669. https://doi.org/10.1523/JNEUROSCI.5443-09.2010

Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004). Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing: A Publication of the IEEE Signal Processing Society*, *13*(4), 600–612. https://doi.org/10.1109/TIP.2003.819861

Wang, Z., Bovik, & Lu, L. (2002). Why is image quality assessment so difficult? In *IEEE International Conference on Acoustics Speech and Signal Processing*. https://doi.org/10.1109/icassp.2002.1004620

Warland, D. K., Reinagel, P., & Meister, M. (1997). Decoding visual information from a population of retinal ganglion cells. *Journal of Neurophysiology*, *78*(5), 2336–2350. https://doi.org/10.1152/jn.1997.78.5.2336

Wässle, H., Peichl, L., & Boycott, B. B. (1983). Mosaics and territories of cat retinal ganglion cells. *Progress in Brain Research*, *58*, 183–190. https://doi.org/10.1016/S0079-6123(08)60019-9

Zhang, Y., Jia, S., Zheng, Y., Yu, Z., Tian, Y., Ma, S., Huang, T., & Liu, J. K. (2020). Reconstruction of natural visual scenes from neural spikes with deep neural networks. *Neural Networks: The Official Journal of the International Neural Network Society*, *125*, 19–30. https://doi.org/10.1016/j.neunet.2020.01.033

Zylberberg, J., Cafaro, J., Turner, M. H., Shea-Brown, E., & Rieke, F. (2016). Direction-Selective Circuits Shape Noise to Ensure a Precise Population Code. *Neuron*, *89*(2), 369–383. https://doi.org/10.1016/j.neuron.2015.11.019