Neuron

Article

Individual variability of neural computations in the primate retina

Highlights

- Hundreds of recordings of macaque retina reveal individual variation in neural coding
- A low-dimensional manifold captures this variability and preserves known invariances
- Manifold reveals sex-related differences in firing rates and temporal integrations
- Manifold efficiently characterizes the neural code in new macaque and human retina

Authors

Nishal P. Shah, Nora Brackbill,
Ryan Samarakoon, ..., Yoram Singer,
Jonathon Shlens, E.J. Chichilnisky

Correspondence

bhaishahster@gmail.com

In brief

Many studies of neural coding identify features that are consistent across animals and treat individual variation as noise. Using an interpretable representation of hundreds of population recordings in primate retina, Shah et al. reveal structures in individual variability, co-variability of different cell types, sex-related differences, and similarities between macaque and human.

Shah et al., 2022, Neuron *110*, 698–708
February 16, 2022 © 2021 Elsevier Inc.
https://doi.org/10.1016/j.neuron.2021.11.026

CellPress

CellPress

Neuron

Article

Individual variability of neural computations in the primate retina

Nishal P. Shah,[1,2,4,9,*] Nora Brackbill,[2,3] Ryan Samarakoon,[2,4] Colleen Rhoades,[2,5] Alexandra Kling,[2,4] Alexander Sher,[6] Alan Litke,[6] Yoram Singer,[7] Jonathon Shlens,[8] and E.J. Chichilnisky[2,4]

[1]Department of Electrical Engineering, Stanford University, Stanford, CA 94305, USA
[2]Hansen Experimental Physics Lab, Stanford University, Stanford, CA 94305, USA
[3]Department of Physics, Stanford University, Stanford, CA 94305, USA
[4]Department of Neurosurgery, Stanford University, Stanford, CA 94305, USA
[5]Department of Bioengineering, Stanford University, Stanford, CA 94305, USA
[6]University of California, Santa Cruz, Santa Cruz, CA 95064, USA
[7]WorldQuant, LLC, 1700 E Putnam Ave., Old Greenwich, CT 06870, USA
[8]Google Brain, 1600 Amphitheatre Pkwy., Mountain View, CA 94043, USA
[9]Lead contact
*Correspondence: bhaishahster@gmail.com

https://doi.org/10.1016/j.neuron.2021.11.026

SUMMARY

Variation in the neural code contributes to making each individual unique. We probed neural code variation using ~100 population recordings from major ganglion cell types in the macaque retina, combined with an interpretable computational representation of individual variability. This representation captured variation and covariation in properties such as nonlinearity, temporal dynamics, and spatial receptive field size and preserved invariances such as asymmetries between On and Off cells. The covariation of response properties in different cell types was associated with the proximity of lamination of their synaptic input. Surprisingly, male retinas exhibited higher firing rates and faster temporal integration than female retinas. Exploiting data from previously recorded retinas enabled efficient characterization of a new macaque retina, and of a human retina. Simulations indicated that combining a large dataset of retinal recordings with behavioral feedback could reveal the neural code in a living human and thus improve vision restoration with retinal implants.

INTRODUCTION

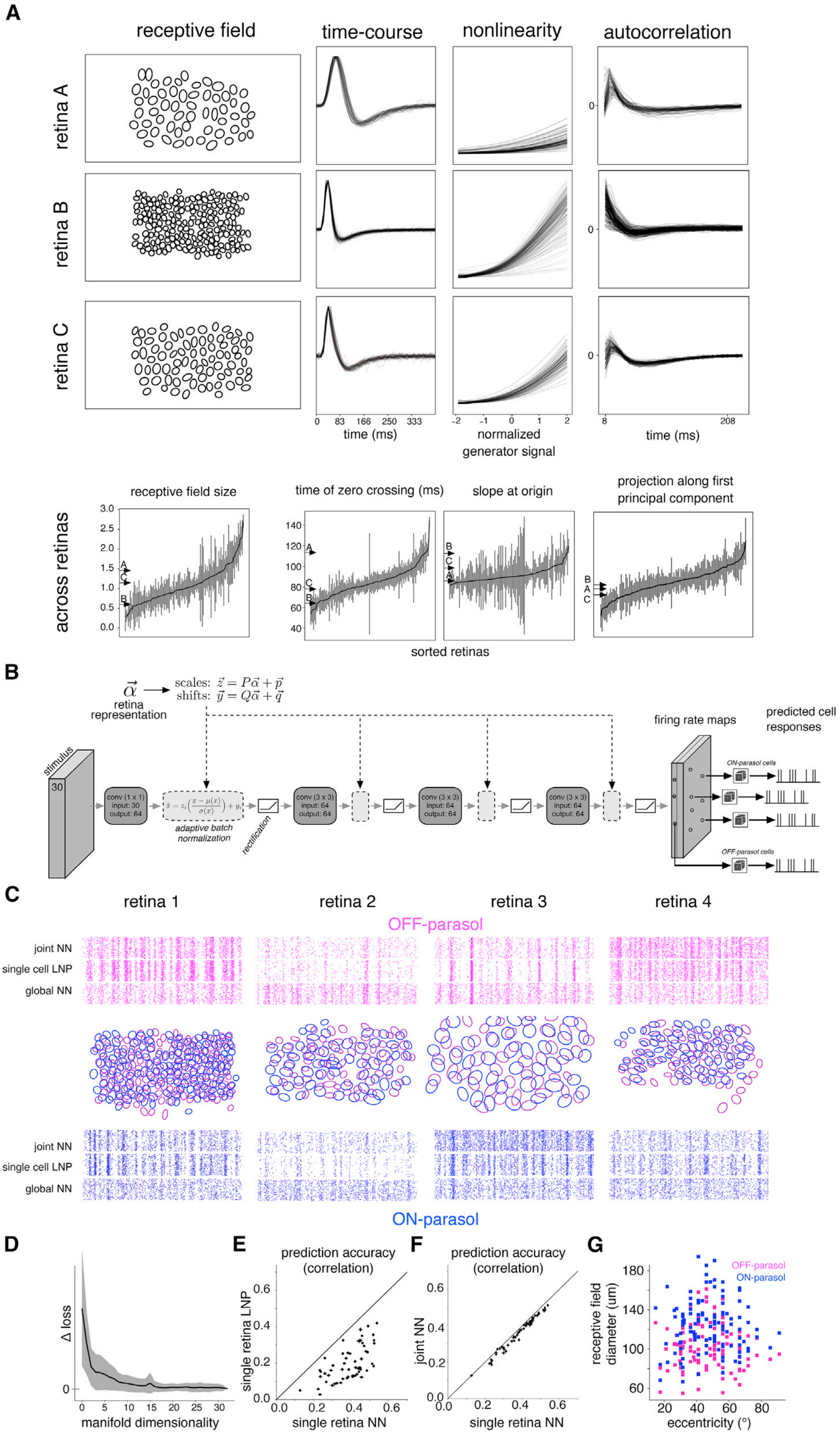
An emerging frontier in biomedicine is the understanding of variability between individuals, with implications ranging from the mathematical modeling of living systems to ethics and personalized medicine. In neuroscience, differences in mental function among individuals are substantial, yet little is known about the underlying variation in the information processing performed by neural circuits, particularly in species similar to humans and at spatial and temporal scales of neural computations. Furthermore, biological variability is frequently obscured by inevitable experimental variability, which can severely limit the ability to test biological hypotheses decisively. These factors have led to a large gap in our understanding of variability in the neural code and its implications for translational medicine and neuroengineering.

Two technical challenges have limited our understanding in animal models relevant to humans: high-resolution, large-scale physiological recordings from many individuals; and methods for deciphering variability in complex circuit-level computations. In this article, we present a novel method for analyzing variability in the neural code—both biological and experimental—and apply that method to a unique, large-scale dataset gathered over a decade of recordings in the macaque retina. The results reveal striking differences in male and female retinas, structure in variability related to lamination of dendrites, and novel implications for neuroengineering to replace retinas damaged by disease.

RESULTS

Modeling the shared features and individual variability of neural coding

Large-scale multi-electrode recordings from retinal ganglion cells (RGCs) were performed from isolated macaque monkey retinas, in which the functional properties of diverse cell types have been extensively studied (Litke et al., 2004; Frechette et al., 2005; Field et al., 2007; Greschner et al., 2014; Rhoades et al., 2019). In total, 21,626 RGCs were analyzed in 112 recordings from 75 retinas of 66 animals. These data, gathered over a decade of experimentation, exhibited significant neural coding variability across recordings, presumably reflecting a mixture of biological and experimental variation (see below). As a baseline to summarize that variability, response properties in each recording were captured by the parameters of a linear-nonlinear-Poisson (LNP) encoding model. This widely used



(legend on next page)

model (Chichilnisky, 2001) captures light-evoked responses in RGCs using a spatiotemporal linear filter applied to the stimulus, followed by an output nonlinearity and stochastic spike generation. Receptive field sizes of two known RGC types—On and Off parasol—exhibited substantial variation across recordings. This variability was evident across retinal eccentricities but was also present at a given eccentricity (Figures 1A, first column, and 1G). Diversity was also seen in the kinetics of light responses (Figure 1A, second column), the shape of the output nonlinearity (Figure 1A, third column), and the autocorrelation function (Figure 1A, fourth column).

The structure and covariation of these response properties was explored using a flexible model that combined shared and recording-specific parameters. The shared component was a multilayered convolutional neural network (CNN), an extension of the LNP model, consisting of multiple alternating stages of spatiotemporal filtering, normalization, and rectification. The rectification captured nonlinear spatial integration (Shah et al., 2020), and the convolutional structure captured the known translational invariance of visual signals in each cell type (cells of the same type at different locations have very similar response properties (Chichilnisky and Kalmar, 2002). The model output consisted of one firing-rate map for each cell type. To predict a given cell's responses, the model-predicted firing rate was read from the map at the cell's location (Figure 1B). Because of the translational-invariance constraint, the proposed model cannot capture differences between cells belonging to the same cell type, resulting in low prediction accuracy compared with models that allow for cell-specific parameters, such as single-cell LNP (Chichilnisky, 2001) or other state-of-the-art models (McIntosh et al., 2016; Batty et al., 2017). However, the translational-invariance constraint enabled the proposed neural network architecture to predict responses across recordings with different numbers and spatial arrangements of cells. When trained using On and Off parasol cell responses in each retina separately, the CNN model exhibited performance superior to the single-retina LNP model with common parameters for all cells of a given type, as expected given its more flexible structure (Figure 1E). However, when the CNN model was trained on multiple retinas together, it failed to capture the responses of retinas with low firing rates, highly modulated responses, or other features that varied among recordings (Figure 1C, rasters).

To capture the variation of light response properties across recordings in a compact and tractable way, data from all recordings were used to learn the ~100,000 parameters of the shared CNN, whereas data from each recording were used to estimate a few recording-specific parameters (see below). A linear transformation of the recording-specific parameters determined the mean and standard deviation of signals in the channels in each layer of the CNN. The collection of these recording-specific parameters was interpreted as a *manifold of neural-coding variability.* When learned using 71 recordings, the low-dimensional manifold captured variations in background firing rate, sustained versus transient dynamics, and response nonlinearities (Figure 1C, rasters). The ability to simultaneously predict responses across multiple retinas saturated at ~15 dimensions of the learned manifold (Figure 1D), much less than the total number of CNN parameters, and the performance of the joint model based on the manifold was only slightly less than that of a CNN model trained for each retina separately (Figure 1F). Thus, a simple, low-dimensional representation can efficiently and accurately capture the biological and experimental variability of retinal computation.

Neural coding manifold smoothly captures systematic variation across recordings

The learned manifold smoothly captured variations in the neural code. For a given recording, a greater perturbation in manifold location led to a greater decrease in response-prediction accuracy (Figure 2G). The manifold geometrically represented variation in several light-response properties, including firing rate, receptive field size, time course, output nonlinearity, and spike train autocorrelation. This was observed by projecting the average response property for each retina onto its principal

Figure 1. Modeling variability in the neural code

(A) Variability of response properties across recordings. Spatial receptive fields (first column), time of zero-crossing of temporal filter (second column), nonlinearity (third column), and variation in autocorrelation function (fourth column) for Off parasol cells in three representative recordings (rows, same y axis for each response property across recordings). The last row represents the range of response parameters across 122 recordings, sorted according to their population means and error bars corresponding to the robust standard deviation. Arrows indicate the values of the chosen retinas in the first three rows.

(B) Architecture of the neural network for capturing response variation. The visual stimulus is passed through multiple layers of convolution with spatial filters, with adaptive batch normalization and rectification at each layer, producing two firing-rate maps (one each for On and Off parasol cell types). The Poisson firing rate for each cell is read off from the value at the cell's location in the firing-rate map of its cell type. Retina-specific tuning of responses is performed by adjusting the means and standard deviation of the activation values at each layer, determined by a linear transformation of the retina's location in the low-dimensional manifold.

(C) Response prediction across four representative training retinas (columns). The receptive field mosaics are shown for each retina (middle row) along with response predictions for a randomly selected Off parasol (top row) and On parasol cell (bottom row). Rasters (60 trials) for predicted responses to a 3-s-long white noise stimulus using the LNP model; neural network model, trained jointly on multiple retinas, with retina-specific parameters (15 dimensional manifold, joint NN) or without them (no manifold, global NN).

(D) Model error (log likelihood) on test stimuli with varying manifold dimensionality; dimensions = 0 indicates no retina-specific adaptation (global NN). Error bars indicate the standard deviation across retinas.

(E) Prediction accuracy, averaged across cells, for different retinas (points), using a neural network trained on data from each individual retina (x axis) and an LNP model with shared parameters across cells of a given type in each retina (y axis). Prediction accuracy is measured as correlation between predicted firing rate and recorded responses smoothed with a Gaussian filter (σ: 11 ms).

(F) Similar to (E), comparing predictions from the neural network with a 15-dimensional manifold, trained jointly (y axis) versus trained on each retina separately (x axis).

(G) Range of eccentricities (x axis) and the average receptive-field sizes (y axis) for On parasol (blue) and Off parasol (magenta) cells across 122 recordings used in this study.

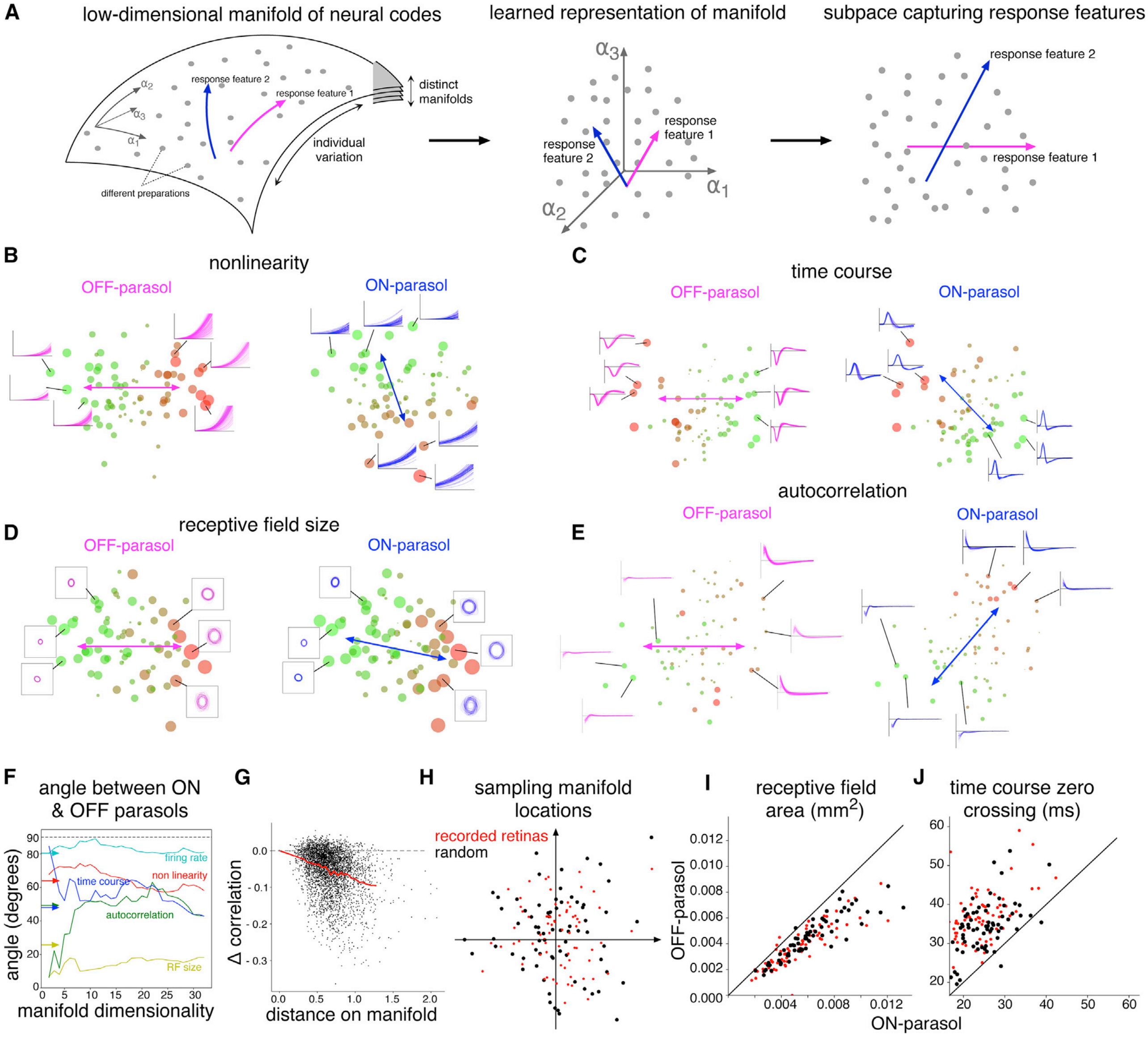


Figure 2. Geometry of manifold

(A) Summary of the steps in subsequent analysis. Left: schematic of the manifold representation of variability. Each recording is summarized by its neural encoding function, indicated by a point (gray) in space of all possible neural encoding functions. The observed neural encoding functions lie in a low-dimensional manifold (depicted as a curved surface but could be more than two dimensional) within this space. Different manifolds (other surfaces) would potentially correspond to different properties that are invariant across recordings. The training procedure learns a coordinate system (α) within the manifold. Middle: directions corresponding to response features can be identified in the learned coordinate system for representing the manifold. Right: geometry of the subspace corresponding to the identified directions lead to interpretation of the variations.

(B) Manifold locations for 95 recordings (points), projected onto the two-dimensional (2D) subspace given by the first-principal components of variation in the output nonlinearity for Off and On parasol cells. Size and color of dots indicate deviation from the mean. Colored lines indicate the direction of maximum nonlinearity variation for On parasol (blue) and Off parasol (magenta) cells. Insets: lines show output nonlinearities for all cells in representative retinas.

(C–E) Similar to (B), for time course, receptive-field size, and autocorrelation, respectively.

(F) Angle between On and Off parasol directions for particular response properties, as a function of manifold dimension. Arrows indicate the cosine inverse of Spearman rank correlation computed directly between the response properties.

(G) Change in response-prediction accuracy (y axis) as the manifold location is perturbed from the learned location. Each black dot represents a different perturbation; the red line is the average.

(H) Random manifold locations (red) were sampled by adding noise to the learned retina-specific locations (black). Responses to a white noise stimulus of 100 cells of each type (On and Off parasol) were sampled from random locations in the visual field using the firing-rate maps for the two cell types associated with these sampled manifold locations. These simulated responses were then used to fit a LNP model.

(I and J) Relationship between On and Off parasol cells for receptive-field area and zero crossing of response time course, respectively, for recorded (red) and randomly sampled (black) retinas.

component across all recordings and then identifying the manifold direction with maximum correlation to that projection. The Spearman rank correlation between the principal component projections and the projections along the identified manifold direction was significantly higher than the value observed in random permutations of the data ($p < 0.05$ for all response properties) (Figures 2B–2F).

The geometry of the manifold also captured co-variation in response properties of different cell types across recordings. For both firing rate and response nonlinearity, the large angles between the manifold directions for On and Off parasol cells (84° and 66°, respectively) were consistent with low Spearman-rank correlation in these response properties (0.17 and 0.44, respectively). Conversely, for response autocorrelation, time course, and receptive field size, the small angles between the directions for On and Off cells (49°, 54°, and 13°, respectively) were consistent with the larger Spearman-rank correlation (0.65, 0.67, and 0.90 respectively). Although high correlation is expected for receptive field size based on variation in the eccentricity of different recordings, the varying degrees of covariance for time-course, autocorrelation, output-nonlinearity, and firing rate have not been previously reported. Hence, the manifold presents an intuitive, geometric representation of variation and covariation of response properties.

The manifold also captured known *invariances* in retinal encoding, i.e., properties that were consistent across retinas. To assess whether these invariances were present at many intermediate manifold locations not directly sampled in the experiments, the manifold locations of recorded retinas were perturbed using Gaussian noise with a standard deviation equal to the median nearest-neighbor separation (Figure 2H). Light responses were then generated using these randomly sampled manifold locations, and the encoding properties were summarized by fitting a LNP model. For both the recorded and simulated responses, Off parasol cells had consistently smaller receptive fields (RFs) (Figure 2I) and slower time courses (Figure 2J) than On parasol cells, consistent with previously reported asymmetries (Chichilnisky and Kalmar, 2002).

To the degree that these systematic variations represent differences among animals, a simple prediction is that recordings from the same animal should be closer on the manifold than recordings from different animals. The results confirmed that prediction (paired t test, $p < 0.01$ across manifold dimensions). Note, however, that animal-specific variations in experimental procedures could also produce that result (see Discussion), motivating a deeper look at variations in the manifold that likely reflect true biological differences.

Manifold reveals covariation associated with retinal connectivity and male-female differences

The RGC types that receive synaptic input from bipolar cells at similar depths in the inner plexiform layer (IPL) showed greater covariation in their response properties across recordings than other RGC types did. To examine covariation between cell types, the On and Off midget cell types were included with the On and Off parasol cell types considered thus far. In 85 recordings (53 macaques), the similarity of three response properties—firing rate, nonlinearity, and time course—across different pairs of cell types was measured either directly or in the manifold (Figures 3A–3C). Using both methods, the greatest correlation in these physiological properties was observed between cell-type pairs with the same polarity (On or Off), consistent with the lamination of On and Off cells in the inner and outer IPL, respectively (Figure 3D). Moreover, for cell-type pairs with opposite polarities, a greater correlation in physiological properties was observed for the On-parasol/Off-parasol pair than for the On-midget/Off-midget pair, consistent with the lamination of parasol cells closer to the middle of the IPL (Figures 3A–3D). These observations support the approach of studying the response properties of newer cell types (such as On and Off smooth-monostratified cells) (Rhoades et al., 2019) after normalizing to the properties of more-commonly studied cells with similar synaptic inputs, thereby minimizing the effects of inter-retina variability.

Surprisingly, recordings from male and female retinas were separated in the manifold ($d' = 1.8$ for 15 dimensional manifold, Figure 3E) in a way that was not explained by variations in experimental factors (Figure 3I). For both On and Off parasol cells, the differences between male and female retinas were partially attributable to differences in firing rate and speed of temporal filtering. This was determined by computing the direction separating the means of male and female retinas in the manifold and determining the angle between that direction and the directions most aligned with variation in firing rate and response time course (cosine[angle] $\sim$ 0.5 for both; Figure 3H). In principle, the observed male-female differences could be confounded by variation in experimental methods, such as dissection procedure (isolation from the retinal pigment epithelium [RPE]) or temperature (31°–36°C across recordings). Because higher recording temperatures were associated with isolation from the RPE for technical reasons, the directions in the manifold representing dissection method and temperature variation were aligned (cosine[angle] $\sim$ 0.57; Figure 3H). Compared with firing rate and response time course, these experimental factors were less aligned to the direction of sex separation (cosine[angle] $\sim$ 0.26; Figure 3H), suggesting that sex differences were probably not due to differences in these experimental methods. To eliminate experimental factors more rigorously, the separation of males and females in the manifold was measured after conditioning the data in several ways. For each condition, a bootstrap rank test was performed to test whether the mean locations of male and female recordings differed (see Method details). Significant separation ($p < 0.05$) was observed for the more-numerous recordings with RPE-isolated dissections (37 male, 15 female) and high temperature ($\geq$33.5°C) (32 male, 12 female), whereas the separation was not significant ($p > 0.1$) for less-numerous dissections with the RPE attached (12 male, 4 female) and lower temperatures (<33.5°C) (19 male, 5 female) (Figure 3I). For the RPE-isolated retinas, the male recordings exhibited higher firing rates (Figure 3F) and faster temporal integration (Figure 3G) ($p < 0.01$, see Method details).

Although the conditioning on specific experimental conditions above revealed statistically significant differences between males and females, the level of significance was less when compared with all the recordings (Figure 3I), potentially because of a reduction in the number of samples when analysis was restricted to a particular set of experimental conditions. The

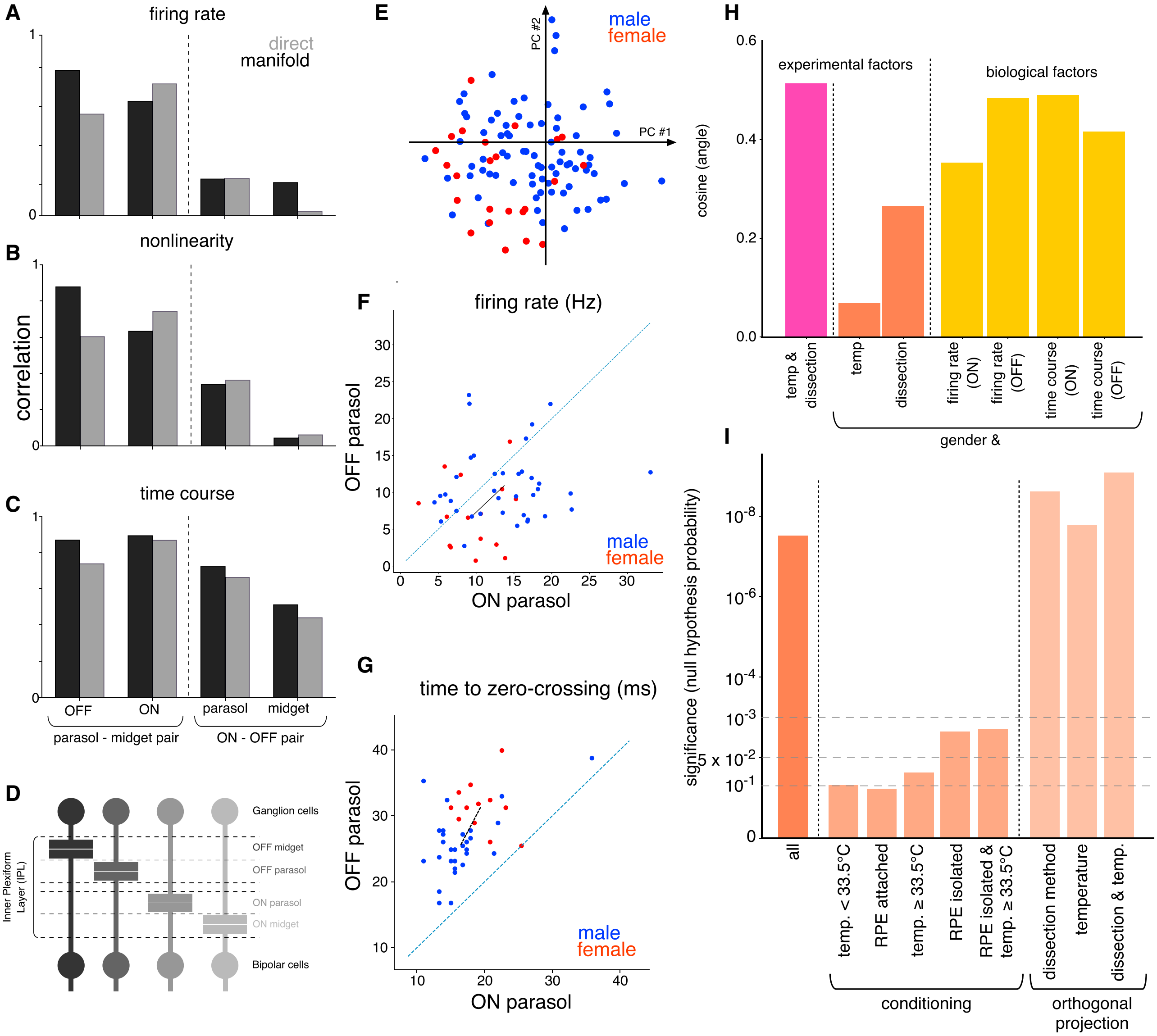


Figure 3. Biological factors underlying variability

(A–C) Relation between the first principal component of (A) firing rate, (B) nonlinearity, and (C) response time course variation for different pairs of cell types. The relationship is either measured directly using Spearman rank correlation or the cosine of the angle between corresponding directions in the 15-dimensional manifold.

(D) Distinct lamination depths for the bipolar-ganglion cell synapse for different ganglion cell types (Wassle and Boycott, 1991).

(E) Two-dimensional principle component analysis (PCA) projection of manifold locations for recordings from male (blue) and female (red) retinas.

(F) The average firing rate for On parasol (y axis) and Off parasol (x axis) cells for recordings with isolated RPE. The mean manifold location of male (blue) and female (red) recordings were different ($p < 0.01$ for bootstrap and $p < 0.05$ for hierarchical bootstrap). Black line joins the mean male and female locations.

(G) Similar to (F) for the time course of STA, with separation of male and female retinas ($p < 0.01$ for bootstrap and $p < 0.05$ for hierarchical bootstrap) (Saravanan et al., 2020).

(H) Cosine of the angle (y axis) between the manifold directions corresponding to different pairs of factors, which are either biological (sex, firing rate, or time of zero crossing of time course) or experimental (temperature or dissection—whether retinal pigment epithelium [RPE] was attached or isolated).

(I) Degree of separation of male and female recordings measured using a resampling test, for all the recordings, conditioned on the subset of recordings with specific dissection procedure or temperature or all recordings with locations projected orthogonal to directions for dissection procedure and temperature variation.

manifold made it possible to separate experimental variations more efficiently, without reducing the number of data points. To accomplish that, the data were projected onto axes in the manifold orthogonal to the two identified directions of experimental variability, namely dissection method and temperature variation (a separate analysis showed that two directions sufficed to capture these two types of variability; not shown). This projection increased the statistical significance of separation between the male and female retinas ($p < 10^{-6}$) (Figure 3I). Thus, the geometry of the manifold makes it possible to examine statistical trends in the data efficiently in spite of potential experimental confounds.

Manifold generalizes to a novel recording

The manifold permitted efficient response modeling of a new, previously unseen retina by leveraging trends in the large dataset of retinas used for learning the model. Response modeling can be performed efficiently by identifying the manifold location of a new retina in several ways, using limited data. In the absence of any new data, the "expected" manifold location for a new retina can be obtained by merely *averaging* the locations of all training retinas. In the case of a degenerated retina (in which no light-evoked response can be recorded), partial information, such as the firing rate of recorded neurons, can be identified from spontaneous activity. In this case, the manifold location can be *approximated* by averaging the locations of training retinas that have similar firing rates. Finally, in the presence of measured light-evoked responses, gradient descent can be used to *optimize* the manifold location based on the likelihood of the data, leveraging the training retinas as prior information (Figure 4A).

These three approaches were examined using a model trained with 71 recordings and tested with 24 recordings. The proximity of manifold locations identified by approximation and optimization suggested that these approaches accurately capture properties of the new retina (Figure 4B). To examine the efficiency benefits of using the manifold, the optimization approach with limited data was examined in detail. The manifold location converged quickly as the recording duration increased, with ~3 min of light-response data producing a location similar to that produced by ~30 min of data (Figure 4C). Although optimization of the manifold location predicted light responses with similar accuracy as that in training the full model (along with the CNN) when tested with a large amount of data (15–30 min), the manifold approach more-accurately predicted light responses when the data were limited (~8 s) (Figure 4D).

The low dimensionality of the manifold enabled efficient generalization to a new retina, but, in principle, it could also reduce accuracy. On varying the manifold dimensionality, the ability to predict responses to previously unseen retinas saturated at ~15 dimensions (Figure 4E, black line), a value that did not change with the number of retinas used for training (Figure 4E, colored lines). Hence, in addition to the previous observation on generalization to new stimuli within the collection of retinas used for learning (Figure 1D), the low-dimensional manifold is able to generalize in another way—to new, previously unseen retinas.

These findings suggest that the manifold may aid in translating our understanding of the macaque retina to the human retina, an important goal for biomedical research. Recent work (Kling et al., 2020; Soto et al., 2020) has shown that the receptive-field properties of the four numerically dominant RGC types (On and Off parasol and midget) are similar to those of their macaque counterparts. To test whether light responses in the human retina fall within the range observed across many macaque retinas, the manifold location of a single human retina was identified using the three operations described above (averaging, approximation, and optimization). For each method, the estimated manifold location of the human retina was well within the span of manifold locations of many macaque retinas (Figures 4B and 4C), and the responses of a new retina were predicted with similar accuracy and efficiency in the two species (Figure 4D). Hence, the manifold reveals that the function of certain RGC types in the human retina can be understood as falling within the range of variation of macaque retinas.

Given that the macaque retinal code translates accurately to humans, it may provide a valuable tool for the development of an advanced artificial retina for treating vision loss. However, a challenging first step in restoring vision with such a device is to identify how the neurons in a blind retina should encode visual stimuli with the implant, based on the particular neural code of the individual receiving the implant. In this setting, the retina is no longer light sensitive, so the healthy neural code cannot be identified directly. However, the human subject could potentially report the similarity of artificially induced images to a verbally described object. The neural encoding that produces the most-accurate perception could then be identified by estimating the location of the retina on the low-dimensional manifold, using a psychophysical discrimination task. Such a task was simulated using the following iterative procedure: (1) sample a few of the plausible manifold locations, (2) use each of these manifold locations to predict retinal responses for a particular visual stimulus, (3) stimulate with the implanted device to produce each of these responses, (4) ask the subject which stimulus produced a visual sensation that most closely matches a verbal description, and (5) update the set of plausible manifold locations.

To illustrate the feasibility of this procedure, the above steps were simulated, assuming (for simplicity) an artificial retina that has perfect cellular selectivity of stimulation. For updating the plausible locations, the perceptual accuracy was assumed to be governed by the Kullbeck-Leibler divergence between the distribution of neural responses associated with the tested manifold location and the distribution of responses associated with the true location. The set of plausible locations identified by the procedure converged to the true location (given by optimization; see above) in a small (<20) number of steps (see Method details; Figure 4F). Hence, the low-dimensional manifold could provide valuable efficiency in translational applications.

DISCUSSION

This work presented a new method to model individual variability in the neural code of the macaque retina using data from a large number of recordings. The approach employed a deep neural network that captured the shared response properties among retinas, along with a low-dimensional manifold that captured differences between them. The manifold preserved the known invariances in neural encoding and captured the inter-dependence of response properties across cell types. It also revealed previously unknown sex-related differences in neural encoding and provided more efficient ways of testing hypotheses on complex datasets in the face of both biological and experimental variability. Finally, the low-dimensional manifold enabled novel

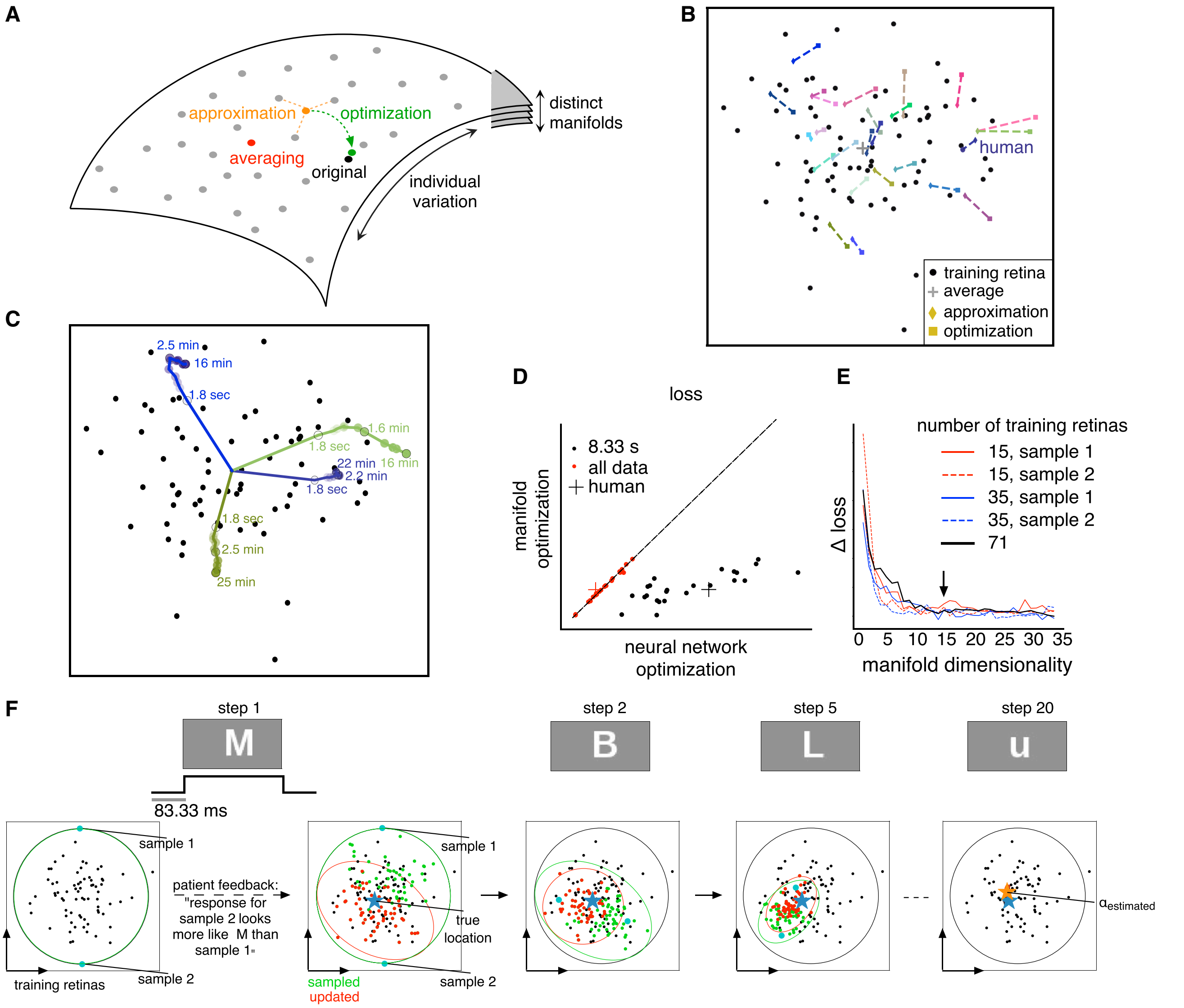


Figure 4. Generalization to new, previously unseen retinas

(A) Efficient identification of manifold location. Training retinas (gray points) are used to define the manifold. Then, the manifold location of a new retina is determined in one of several ways: averaging (red), compute the mean location of all training retinas; approximation (orange), compute the mean location of a subset of training retinas with specific features similar to the new retina; and optimization (green), gradient descent on the manifold location to maximize the prediction accuracy for measured light responses of the new retina.

(B and C) Identified manifold locations using averaging (+; same for all retinas), approximation (orange), and optimization (□) for testing retinas (colors, each pair joined with a line). Black points correspond to the locations of training retinas. (C) Optimized manifold location for three retinas (colored lines), with varying duration of recorded responses. Training retina locations (black points) and locations identified by averaging (+) are shown. For (B) and (C), the 15-dimensional manifold is projected into two dimensions that capture firing rate variation (same as Figure 2E)

(D) Response prediction loss with optimization of manifold location (y axis) versus loss with learning all neural network parameters (x axis), using either 8.33 ms (red) or 15–30 min (black) of data.

(E) Convergence of prediction loss with optimized manifold location, averaged across 24 testing retinas (y axis) as a function of the number of manifold dimensions (x axis). Loss measured as negative log likelihood, averaged across cells. Colored lines indicate loss obtained using fewer training retinas.

(F) Simulation of the discrimination task used to identify manifold location in the retina of a blind person. At each step, the visual stimulus is a letter from English alphabet (top row). Two-dimensional projection of the manifold with locations of training retinas (black dots) is shown (bottom row). Posterior over the set of feasible manifold locations approximated using a Gaussian (red circle). At each step, random manifold locations are sampled from the posterior, corresponding responses are reproduced using the artificial retina, and feedback from the subject is used to update the posterior using Monte Carlo methods (all samples in green, accepted samples in red). In ~20 steps, the estimated manifold location (orange) converges to the true underlying location (blue).

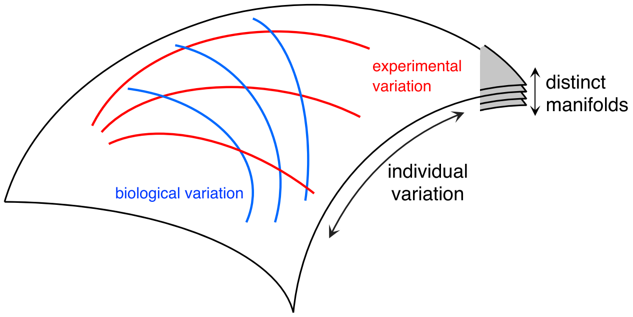


Figure 5. Sub-manifolds associated with variations in experimental and biological factors

Each submanifold is defined by the manifold locations spanned by the variation in one factor, keeping the other factors fixed. In the illustration, red (blue) curves correspond to variation in a particular experimental (biological) factor, keeping other factors fixed. Different settings of the fixed factors correspond to a transformation of the submanifold (different curves) for a given varying factor.

methods for identifying the neural code in a new recording, paving the way for exploiting large-scale datasets to optimize artificial vision with a retinal implant.

Note that, in the present data, biological and experimental variability for the most part could not be definitively separated. At least three sources of *in vivo* biological variability in the neural code could be present in the data: variation between animals, differences between the two eyes, and variation across retinal locations in a given eye. However, these sources of variability are inherently confounded with variation in three experimental procedures: euthanasia, eye removal, and *ex vivo* recording, respectively. In addition, some of the possible factors of inter-individual variation, such as the history of experiments and medical procedures throughout the lifespan of the animal, are neither entirely biological nor entirely experimental. Thus, although the relative importance of these three sources of variability could be studied further, for the most part, it is not possible to reliably isolate biological and experimental variation in the present data. An exception is the observed difference in retinal encoding between males and females: there were no known associations between the *ex vivo* experimental procedures used and the sex of the animal. Although there is no way to definitively exclude the possibility of experimental differences, these analyses strongly suggest that the male-female differences observed reflect true biological differences. Additionally, another biological factor, the proximity of dendritic lamination of specific cell types, is associated with *covariation* in measured response properties across preparations.

Importantly, the tools developed here make it possible to capture and analyze *both* experimental and biological variability in a single framework. This is an asset for understanding the neural code more completely using experimental data, which are always imperfect. For example, the low-dimensional manifold made it possible to project the data into a subspace orthogonal to known confounds and, thus, to control for them without conditioning on specific experimental variables, thus retaining all the available data and statistical power.

The manifold framework can be used to understand biological and experimental variability in a principled manner. This is an extension of the idea that individual variability creates a manifold in the neural response space, with different manifolds corresponding to different possible invariances across recordings (Figure 2A). Isolated changes in a particular experimental or biological factor (keeping the other factors fixed) could be conceptualized as producing a sub-manifold (Figure 5). Although the data used for this article are not sufficient to study these sub-manifolds, we see evidence of additional structure within the manifold, such as the varying levels of covariation of response properties across cell-types and their relation to dendritic lamination.

For certain scientific questions it is not necessary to identify the underlying cause of variability. For example, attributing the varying degrees of covariation of different response properties (temporal kinetics, nonlinearity, and autocorrelation) and cell types (On versus Off and parasol versus midget) to either biological or experimental factors is not crucial for understanding the dependence on a biological variable like lamination depth. For this observation, the presence of either biological or experimental variability can be understood as a ‘‘perturbation experiment’’ that reveals the interdependence of features of neural responses, which presumably has a biological origin.

The neural coding differences observed between male and female retinas add to a large literature on sex-based differences in brain structure and function (Cahill, 2006; Choleris et al., 2018; Poplin et al., 2018). In the retina, genetic differences between male and female primates (including humans) produce different variation in cone photopigment spectral sensitivities and, thus, different color vision, across the population (National Eye Institute, 2019). However, to our knowledge, differences in neural coding between males and females have not been reported, perhaps because of the lack of appropriate physiological recordings and/or analysis tools.

Several recent studies have shown that neural responses in the retina, which have often been described using pseudo-linear models (Chichilnisky, 2001; Pillow et al., 2008), can be modeled more accurately using artificial neural networks (McIntosh et al., 2016; Batty et al., 2017). However, the function of these complex models can be difficult to interpret. In the present work, the machine-learning model was used to more accurately capture shared (and complex) aspects of neural response, but information about variability between recordings was summarized in a simple, interpretable, low-dimensional manifold (Schneidman et al., 2001). In principle, such an approach could be used for other applications in which a neural computation of interest is captured using a simple, low-dimensional manifold, whereas a machine-learning model improves accuracy by capturing the possibly complex aspects of the computation that are of less immediate interest.

The manifold of neural coding variability may also be useful in other neuroengineering applications, such as motor prostheses. The goal of a motor prosthesis is usually to read out the neural activity in a paralyzed person to control a computer cursor or a robotic limb (Bensmaia and Miller, 2014). Similar to the problem of identifying the neural encoding in a blind person, identifying the neural mapping in a paralyzed person is limited by the absence of simultaneous neural recordings and limb trajectory measurements and, thus, may benefit from leveraging existing data that capture the diversity of neural coding across individuals. Specifically, a manifold of inter-individual variation may be useful for identifying neural decoding in a person with an implant, perhaps using a task involving imagined movements to identify the manifold location of a particular person (Shenoy and Carmena, 2014). The manifold may also be useful for dealing with the challenge of variability over time in chronic recordings (Chestek et al., 2011).

The neural coding manifold may also be useful for harnessing brain plasticity, which could improve vision with an artificial retina. Indeed, present-day retinal implants make little attempt to reproduce the neural code of the retina and, thus, implicitly rely heavily on plasticity to compensate for device limitations (Beyeler et al., 2017). In motor prostheses, it has been shown that the brain can more easily adjust its activity to accommodate perturbations in the artificial neural decoder if those perturbations lie in a low-dimensional manifold (Golub et al., 2018). In the case of an artificial retina, we hypothesize that the brain may more readily learn to interpret the neural activity produced by the implanted device if the visual encoding that it uses is designed to lie within the manifold of retinal coding variability.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - ○ Lead contact
  - ○ Materials availability
  - ○ Data and code availability
- METHOD DETAILS
  - ○ Recordings
  - ○ Visual stimuli and cell type identification
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - ○ Linear-nonlinear-Poisson model
  - ○ Neural network model
  - ○ Notation
  - ○ Model architecture
  - ○ Model fitting
  - ○ Variation of neural coding on the manifold
  - ○ Relating the manifold location to biological and experimental factors of variation
  - ○ Invariance of neural coding on the manifold
  - ○ Estimation of the manifold location of a previously unseen retina
  - ○ Averaging
  - ○ Approximation
  - ○ Optimization
  - ○ Discrimination task

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.neuron.2021.11.026.

ACKNOWLEDGMENTS

We thank J. Carmena, K. Bankiewicz, T. Moore, W. Newsome, M. Taffe, T. Albright, E. Callaway, H. Fox, R. Krauzlis, S. Morairty, and the California National Primate Research Center for access to macaque retinas. We thank Kristy Berry, K. Williams, B. Morsey, J. Frohlich, and M. Kitano for accumulating and providing macaque retina metadata. The human eye was provided by Donor Network West (San Ramon, CA). We are thankful for the cooperation of Donor Network West and all of the organ and tissue donors and their families, for giving the gift of life and the gift of knowledge, by their generous donations. We thank K. Shenoy, G. Field, F. Rieke, W. Newsome, E. Simoncelli, S. Mitra, V. Gupta, K. Talwar, V. Feldman, A. Gogliettino, E. Wu, S. Madugula, M. Zaidi, and the entire Stanford Artificial Retina team for helpful discussions. We thank Google internship and Student Researcher programs (NPS), Research to Prevent Blindness Stein Innovation award, Wu Tsai Neurosciences Institute Big Ideas, NIH NEI R01-EY021271, NIH NEI R01-EY029247, and NIH NEI P30-EY019005 (E.J.C.); NSF IGERT grant 0801700 (N.B.); NIH NEI F31EY027166 (C.R.); and NSF GRFP DGE-114747 (N.B. and C.R.) for funding this work. We thank Lauren Jepson, Peter Li, Martin Greshner, Greg Field, Jeff Gautier, Alex Heitman, Lauren Grosberg, Geroges Goetz, and Sasidhar Madugula for participating in data collection in older experiments.

AUTHOR CONTRIBUTIONS

N.B., A.K., C.R., N.P.S., J.S., and E.J.C. collected the experimental data; A.S. and A.L. developed and supported the MEA hardware and software; N.P.S., E.J.C., J.S., and Y.S. conceived the analysis; N.P.S. performed the analysis; R.S. helped with data curation; N.P.S. and E.J.C. wrote the manuscript; all authors edited the manuscript; and E.J.C. supervised the project.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: January 5, 2021
Revised: August 10, 2021
Accepted: November 20, 2021
Published: December 20, 2021

REFERENCES

Angueyra, J.M., and Rieke, F. (2013). Origin and effect of phototransduction noise in primate cone photoreceptors. Nat. Neurosci. *16*, 1692–1700.

Batty, E., Brackbill, N., Heitman, A., Sher, A., Litke, A., Chichilnisky, E.J., and Paninski, L. (2017). Multilayer recurrent network models of primate retinal ganglion cell responses. In Proceedings of the 5th International Conference on Learning Representations https://openreview.net/pdf?id=HkEI22jeg.

Bensmaia, S.J., and Miller, L.E. (2014). Restoring sensorimotor function through intracortical interfaces: progress and looming challenges. Nat. Rev. Neurosci. *15*, 313–325.

Beyeler, M., Rokem, A., Boynton, G.M., and Fine, I. (2017). Learning to see again: biological constraints on cortical plasticity and the implications for sight restoration technologies. J. Neural Eng. *14*, 051003.

Cahill, L. (2006). Why sex matters for neuroscience. Nat. Rev. Neurosci. *7*, 477–484.

Chestek, C.A., Gilja, V., Nuyujukian, P., Foster, J.D., Fan, J.M., Kaufman, M.T., Churchland, M.M., Rivera-Alvidrez, Z., Cunningham, J.P., Ryu, S.I., and Shenoy, K.V. (2011). Long-term stability of neural prosthetic control signals from silicon cortical arrays in rhesus macaque motor cortex. J. Neural Eng. *8*, 045005.

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

Chichilnisky, E.J., and Kalmar, R.S. (2002). Functional asymmetries in On and Off ganglion cells of primate retina. J. Neurosci. *22*, 2737–2747.

Choleris, E., Galea, L.A.M., Sohrabji, F., and Frick, K.M. (2018). Sex differences in the brain: Implications for behavioral and biomedical research. Neurosci. Biobehav. Rev. *85*, 126–145.

Dumoulin, V., Shlens, J., and Kudlur, M. (2017). A learned representation for artistic style. arXiv, arXiv:1610.07629v5 https://arxiv.org/abs/1610.07629.

Field, G.D., and Chichilnisky, E.J. (2007). Information processing in the primate retina: Circuitry and coding. Annu. Rev. Neurosci. *30*, 1–30.

Field, G.D., Sher, A., Gauthier, J.L., Greschner, M., Shlens, J., Litke, A.M., and Chichilnisky, E.J. (2007). Spatial properties and functional organization of small bistratified ganglion cells in primate retina. J. Neurosci. *27*, 13261–13272.

Frechette, E.S., Sher, A., Grivich, M.I., Petrusca, D., Litke, A.M., and Chichilnisky, E.J. (2005). Fidelity of the ensemble code for visual motion in primate retina. J. Neurophysiol. *94*, 119–135.

Golub, M.D., Sadtler, P.T., Oby, E.R., Quick, K.M., Ryu, S.I., Tyler-Kabara, E.C., Batista, A.P., Chase, S.M., and Yu, B.M. (2018). Learning by neural reassociation. Nat. Neurosci. *21*, 607–616.

Greschner, M., Field, G.D., Li, P.H., Schiff, M.L., Gauthier, J.L., Ahn, D., Sher, A., Litke, A.M., and Chichilnisky, E.J. (2014). A polyaxonal amacrine cell population in the primate retina. J. Neurosci. *34*, 3597–3606.

Kingma, D.P., and Ba, J. (2014). Adam: A method for stochastic optimization. arXiv, arXiv:1412.6980v9 https://arxiv.org/abs/1412.6980.

Kling, A., Gogliettino, A.R., Shah, N.P., Wu, E.G., Brackbill, N., Sher, A., Litke, A.M., Silva, R.A., and Chichilnisky, E.J. (2020). Functional organization of midget and parasol ganglion cells in the human retina. bioRxiv. https://doi.org/10.1101/2020.08.07.240762.

Litke, A.M., Bezayiff, N., Chichilnisky, E.J., Cunningham, W., Dabrowski, W., Grillo, A.A., Grivich, M., Grybos, P., Hottowy, P., Kachiguine, A.M., et al. (2004). What does the eye tell the brain?: Development of a system for the large-scale recording of retinal output activity. IEEE Trans. Nucl. Sci. *51*, 1434–1440.

McIntosh, L.T., Maheswaranathan, N., Nayebi, A., Ganguli, S., and Baccus, S.A. (2016). Deep learning models of the retinal response to natural scenes. Adv. Neural Inf. Process. Syst. *29*, 1369–1377.

National Eye Institute (2019). Color blindness (National Institutes of Health). https://www.nei.nih.gov/learn-about-eye-health/eye-conditions-and-diseases/color-blindness.

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

Poplin, R., Varadarajan, A.V., Blumer, K., Liu, Y., McConnell, M.V., Corrado, G.S., Peng, L., and Webster, D.R. (2018). Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. Nat. Biomed. Eng. *2*, 158–164.

Rhoades, C.E., Shah, N.P., Manookin, M.B., Brackbill, N., Kling, A., Goetz, G., Sher, A., Litke, A.M., and Chichilnisky, E.J. (2019). Unusual physiological properties of smooth monostratified ganglion cell types in primate retina. Neuron *103*, 658–672.e6.

Rousseeuw, P.J., and Croux, C. (1993). Alternatives to the median absolute deviation. J. Am. Stat. Assoc. *88*, 1273–1283.

Saravanan, V., Berman, G.J., and Sober, S.J. (2020). Application of the hierarchical bootstrap to multi-level data in neuroscience. Neuron. Behav. Data Anal. Theory *3*, 13927.

Schnapf, J.L., Nunn, B.J., Meister, M., and Baylor, D.A. (1990). Visual transduction in cones of the monkey Macaca fascicularis. J. Physiol. *427*, 681–713.

Schneidman, E., Brenner, N., Tishby, N., de Ruyter van Steveninck, R.R., and Bialek, W. (2001). Universality and individuality in a neural code. Volume 13. In Advances in Neural Information Processing Systems, T.K. Leen, T.G. Dietterich, and V. Tresp, eds. (MIT Press), pp. 159–165.

Shah, N.P., Brackbill, N., Rhoades, C., Kling, A., Goetz, G., Litke, A.M., Sher, A., Simoncelli, E.P., and Chichilnisky, E.J. (2020). Inference of nonlinear receptive field subunits with spike-triggered clustering,. Elife *9*, e45743.

Shenoy, K.V., and Carmena, J.M. (2014). Combining decoder design and neural adaptation in brain-machine interfaces. Neuron *84*, 665–680.

Soto, F., Hsiang, J.C., Rajagopal, R., Piggott, K., Harocopos, G.J., Couch, S.M., Custer, P., Morgan, J.L., and Kerschensteiner, D. (2020). Efficient coding by midget and parasol ganglion cells in the human retina. Neuron *107*, 656–666.e5.

Wässle, H., and Boycott, B.B. (1991). Functional architecture of the mammalian retina. Physiol. Rev. *71*, 447–480.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Macaque retina	UC Davis Primate Research Center	N/A
Macaque retina	SRI	N/A
Macaque retina	Stanford University	N/A
Macaque retina	Salk Institute	N/A
Macaque retina	The Scripps Research Institute	N/A
Macaque retina	University of California, Berkeley	N/A
Macaque retina	University of California, San Francisco	N/A
Macaque retina	Washington National Primate Research Center	N/A
Chemicals, peptides, and recombinant proteins		
Ames' medium	Sigma-Aldrich	Cat# 1420
Software and algorithms		
MGL	Gardner Lab	http://stanford.idm.oclc.org/doku.php/mgl/overview
MATLAB	Mathworks	N/A
Python	https://www.python.org/	N/A
Intaglio	Purgatory Design	N/A
Custom spike sorting software	Chichilnisky Lab	N/A
Tensorflow	https://www.tensorflow.org/	N/A

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Nishal P. Shah (bhaishahster@gmail.com).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- A subset of the data used in this study is available as Data S1. Due to the immense size of the complete set of data, additional data used in this study will be shared by the lead contact upon reasonable request.
- Code is available in Data S1.
- Any additional information required to reanalyze the data reported in this work paper is available from the lead contact upon request.

METHOD DETAILS

Recordings

Preparation and recording methods are described elsewhere (Chichilnisky and Kalmar, 2002; Litke et al., 2004; Frechette et al., 2005). Briefly, eyes were enucleated from terminally anesthetized macaque monkeys (*M. Mulatta* or *M. Fascicularis*) used by other experimenters in accordance with institutional guidelines for the care and use of animals. Immediately after enucleation, the anterior portion of the eye and the vitreous were removed in room light. A single human eye was obtained from a 29 year-old Hispanic male donor, rendered brain-dead from head and chest trauma in a car accident, through Donor Network West, San Ramon, CA. There were no known or suspected problems with the health of the retina. The eye was removed immediately after cross-clamp. In both species, the eye was stored in darkness in oxygenated Ames' solution (Sigma, St. Louis, MO) at 33°C pH 7.4. Segments of isolated or RPE-attached peripheral retina (approximately 3mm x 3mm, taken from 6-15mm temporal equivalent eccentricity (Chichilnisky and Kalmar, 2002) were placed flat, RGC side down, on a planar array of 512 extracellular microelectrodes arranged in an isosceles triangular lattice. The electrode spacing was 60 μm in each row, and the array covered a rectangular region measuring 1800 μm x 900 μm. While recording, the retina was perfused with Ames' solution (31-36°C; typically 32°C for RPE attached and 34°C for RPE isolated dissections), bubbled with 95% O_2 and 5% CO_2, pH 7.4. The recordings were performed over the course of 16 years (2005-2021), with the experimental procedure, rig setup (including electrode arrays, optics, displays, and software) remaining largely constant. However, there were changes in experimental details such as the personnel, dissection details, sources of animals, and selection of recordings for experimentation. The source of eyes also changed substantially from labs in the San Diego area to the labs in the San Francisco Bay Area in 2013, and the frequency of RPE-attached recordings increased steadily over the years.

Voltage signals on each electrode were bandpass filtered (80Hz - 2kHz), amplified, and digitized at 20 kHz (Litke et al., 2004). A custom spike-sorting algorithm was used to identify and segregate spikes from distinct cells (Litke et al., 2004). Briefly, candidate spike events were detected using a threshold on each electrode, and voltage waveforms on the electrode and six neighboring electrodes in the 4ms period surrounding the time of the spike were extracted. Candidate neurons were identified by clustering the waveforms using a Gaussian mixture model. Candidate neurons were retained only if the assigned spikes exhibited a 1 ms refractory period and had a stable firing rate for the entire duration of recording. Duplicate spike trains were identified by temporal cross-correlation and removed. For each cell, the autocorrelation function of the recorded spike train was computed as the correlation coefficient for different time lags.

Visual stimuli and cell type identification

Visual stimuli were delivered using the optically reduced image of a CRT monitor refreshing at 120 Hz and focused on the photoreceptor outer segments. The optical path passed through a glass coverslip, transparent dialysis membrane, and Ames' solution, or through the mostly transparent electrode array and the retina before reaching the photoreceptors. The relative emission spectrum of each display primary was measured with a spectroradiometer (PR-701, PhotoResearch) after passing through the optical elements between the display and the photoreceptors. The total power of each display primary was measured with a calibrated photodiode (UDT Instruments). The mean photoisomerization rates for the cone photoreceptors were estimated by computing the inner product of the primary power spectra with the spectral sensitivity of each cone type, and multiplying by the effective collecting area of primate cones ($\sim$0.6 μm^2) (Schnapf et al., 1990; Angueyra and Rieke, 2013), resulting in photoisomerization rates of approximately 800–2200, 800–2200, 400–900 for the long-, middle- and short-wavelength sensitive cones, respectively. The stimulus pixel size on the retina was either 41.6 microns (8 monitor pixels), 52 microns (10 monitor pixels) or 83.2 microns (16 monitor pixels). A new white noise frame was drawn at refresh rates of 60 Hz or 30 Hz. The pixel contrast (difference between the maximum and minimum intensities divided by the sum) was 96% for each display primary, with mean intensity equal to 50% of the maximum for each primary. The white noise stimulus either modulated the three display primaries independently, or coherently, at each spatial location.

In each recording, RGCs were classified into distinct types using properties of the spatial receptive field and response time course obtained from the spike-triggered average (STA) stimulus (Chichilnisky, 2001; Chichilnisky and Kalmar, 2002; Field and Chichilnisky, 2007). A two-dimensional Gaussian fit to the spatial receptive field was used for determining the center location (Chichilnisky and Kalmar, 2002). Analysis was restricted to recordings with stable firing rates and nearly complete tiling of ON and OFF parasol cell receptive field mosaics. For Figure 3, only recordings that also had nearly complete ON and OFF midget cell mosaics were used. For model fitting, both the visual stimulus and spike times were binned at 8.33ms (120Hz), and the visual stimulus was upsampled to 8 monitor pixels, resulting in a common 80x40 pixel grid across recordings.

QUANTIFICATION AND STATISTICAL ANALYSIS

Linear-nonlinear-Poisson model

The Linear-Nonlinear-Poisson (LNP) model consists of a linear spatio-temporal filter followed by a point nonlinearity (Chichilnisky, 2001). A filter that is separable in space and time was used (Chichilnisky and Kalmar, 2002), which is equivalent to a cascade of a spatial filter and a temporal filter. These filters were estimated for each cell in each recording as follows. First, the STA was computed by averaging the stimulus preceding spikes, over all pixel locations and 250ms (30 frames at 120Hz) prior to the spike. Next, the spatial filter was computed by choosing the STA frame with the single largest pixel magnitude. The spatial filter was restricted to a rectangular window around the receptive field. The receptive field was defined as the set of pixels with absolute magnitude greater than 2.5σ, contiguous with the strongest pixel, where σ is the robust standard deviation (Rousseeuw and Croux, 1993) of pixels in the STA, an estimate of the measurement noise. Next, the temporal filter was identified by averaging the time course of all pixels in the receptive field. Finally, the output nonlinearity was estimated by fitting a 5th order polynomial to the relationship between the observed responses and the generator signal, which was computed by filtering the stimulus with the estimated spatial and temporal filters (Chichilnisky, 2001).

Neural network model

A convolutional neural network was used to predict RGC responses across multiple recordings simultaneously. Below, the model architecture and the fitting procedure are described in detail.

Notation

The model $f(S,\ \alpha_i,\ C_i)$, takes as its input the visual stimulus S, recording-specific information about the collection of recorded cells C_i, and the recording-specific manifold location α_i, and yields as its output the predictions for recording-specific response R_i. For simplicity, the following description is for a single sample stimulus-response pair, but it can be extended to multiple stimuli and corresponding responses in a straightforward manner.

The recent history of the visual stimulus is given by $S\in \boldsymbol{R}^{d_x \times d_y \times d_z}$, where **R** is the set of real numbers, $d_x \times d_y$ are the spatial dimensions (80 × 40), and $d_t$is the number of time bins (30). Stimuli presented at different spatial or temporal resolutions were upsampled or downsampled to these dimensions.

The recording specific manifold location is given by $\alpha_i \in \boldsymbol{R}^n$, where n is the manifold dimensionality.

The recording-specific information about recorded cells is given by $C_i\ =\ \{x(c),\ y(c),\ t(c)\}_{c=1}^{c=|C_i|}$, where each cell c is described by its receptive location $[x(c),\ y(c)]$ in the $d_x\ \times d_y$ visual space and its cell type $t(c)$. For models with only two cell types, $t(c)\in\ \{0,\ 1\}$ (representing ON and OFF parasol cell types). For models with four cell types, $t(c)\in\{0,\ 1,\ 2,\ 3\}$ (representing ON parasol, OFF parasol, ON midget, and OFF midget cell types).

The responses are given by $R_i \in \boldsymbol{Z}_+{}^{|C_i|}$, where $\boldsymbol{Z}_+$ denotes non-negative integers and $|C_i|$ is the number of cells in recording i. Responses were binned at the same resolution as the stimulus (8.33ms).

Model architecture

The model $f(S,\ \alpha_i,\ C_i)$ passes the visual stimulus S through a multilayered convolutional neural network, with each layer consisting of a convolution (stride 1), retina-specific normalization and a point-wise (softplus) nonlinearity (see Figure 1B). The model output is a Poisson firing rate in each time bin. This firing rate is used to predict the responses R_i. The number of channels and the filter sizes were chosen by cross-validation, as described below. Recording-specific normalization and challenges associated with predicting responses for varying numbers of cells across recordings are also described below.

Recording-specific normalization is inspired by previous work (Dumoulin et al., 2017), in which a translationally-invariant affine transformation of the layer activations adapts the model to each recording. The scale and shift coefficients for this affine transform are determined linearly from the manifold location α_i. Let $\widehat{a}(x,\ y,\ l,\ t)$be the activation after convolution at location $x,\ y$ in channel l of layer t. First, the mean μ and standard deviation σ across samples in a batch are computed, and used to calculate normalized activations: $\tilde{a}(x,\ y,\ l,\ t)\ =\ \frac{\widehat{a(x,\ y,\ l,\ t)-\mu}}{\sigma}$. Next, using the manifold location α_i, a learned affine transform determines the desired mean $(\tilde{\mu}\ =P\alpha_i\ \ +p)$ and standard deviation $(\tilde{\sigma}\ =Q\alpha_i\ \ +q)$ for each layer. Finally, the normalized activations are transformed to give recording-specific activations $a(x,\ y,\ l,\ t)\leftarrow\tilde{a}(x,\ y,\ l,\ t)\sigma_{\widetilde{l,\ t}}+\mu_{\widetilde{l,\ t}}$. Note that the retina-specific scales and shifts are the same for each location in visual space, preserving the translational invariance of convolutional networks and reflecting the homogeneous response properties of the RGCs of a single type.

A potential challenge is that the number of recorded neurons, and hence the number of outputs of $f(.)$, is variable across recordings. To address this issue, the model predicts multiple response maps, one for each cell-type, with the same spatial dimensions as the visual stimulus. The response for each cell is read off from its cell location in the response map of the corresponding cell type. Specifically, $f(.)$ outputs $m_i(x,y)$, which corresponds to firing rate map of cell-type i. Thus for a cell with type $t(c)$and centered at $x(c),\ y(c)$ the firing rate is given by $m_{t(c)}[x(c),\ y(c)]$.

Model fitting

Estimation of recording-specific parameters (α_i) and the shared parameters are performed by maximizing the log-likelihood of observed responses, summed across all the cells, recordings and stimuli. This is performed by stochastic gradient descent, where at each step, a randomly sampled batch of stimuli and corresponding responses from a particular recording are used to update the shared and the corresponding recording-specific parameters. The batch size was 250 and the updates were performed using the Adam update algorithm (Kingma and Ba, 2014) with learning rate of 0.1. For each recording used for fitting the model, the first 4 min of white noise data were used for testing and the remainder was used for training. The duration of the stimulus varied from 15-90 min (median 30 min) across recordings. A model with 4 layers, 3 × 3 or 1 × 1 filter size, 64 channels per layer and a 15 dimensional manifold was chosen based on cross validation and used for subsequent analysis (see Figure 1B for architecture).

Variation of neural coding on the manifold

The following steps were used to test whether the manifold captured variations in neural response properties across recordings. First, the manifold direction that was maximally correlated to the variations of a particular response feature was identified by linear regression. Second, recordings were projected along this direction, and the Spearman rank correlation with the response property was measured. Statistical significance was measured with a permutation test, where the null distribution was generated by permuting the recordings with the manifold locations fixed. In Figure 2, the Spearman rank correlation and its statistical significance was measured for the first principal component projection of various response properties such as firing rate (ON: 0.97, $p < 0.0001$ and OFF: 0.95, $p < 0.0001$) receptive field size (ON: 0.78, $p < 0.0001$; OFF: 0.76, $p < 0.0001$), time course (ON: 0.83, OFF: 0.85; $p < 0.0001$), output non-linearity of the LNP model (Spearman rank correlation for ON: 0.92; OFF: 0.92; $p < 0.0001$) and auto-correlation (ON: 0.84, $p < 0.001$ OFF: 0.62, $p < 0.05$). The interdependence between response properties was either measured directly in raw data using Spearman rank correlation, or with the angle between the corresponding manifold directions. Pairwise correlations between all response properties considered (firing rate, output nonlinearity, time course, autocorrelation & RF area across ON and OFF parasol cells) was captured in a ~6 dimensional space (90% of variance captured). A potential confound for these analyses, the dependence of manifold location on response instability during the course of recording, was eliminated by observing that the manifold locations were stable when estimated with subsets of data at the beginning and end of the recording.

Relating the manifold location to biological and experimental factors of variation

The estimated model was examined to reveal the dependence of manifold location on various biological and experimental factors. For analysis of sex differences, only the subset of recordings (102) from one species (M. Mulatta) were used. Based on the documented medical history, the animals with a gender specific condition (e.g., uterine fibroids) were not considered for this analysis. First, the separation between male and female recordings was measured by computing the d' value of the projection of the two distributions onto the difference in the means. The d' value observed (~1.8) indicated that the sex-based differences were not large on an individual basis. Second, a bootstrap test was performed to test whether the mean locations of the male and female recordings were statistically distinguishable. The distance between the mean manifold locations of male and female retinas was measured and compared to a null distribution of distances generated by resampling (with replacement) of the manifold locations. The null distribution was fitted with a normal distribution and the significance level was measured as the probability mass greater than the observed distance in data. Because multiple recordings were frequently recorded from the same animal, a hierarchical variant of this bootstrap test was also performed, in which resampling was performed according to the hierarchical structure of the data (Saravanan et al., 2020), by first sampling an animal and then sampling the manifold location of one of the recordings from that animal, both with replacement. Hierarchical bootstrap is more conservative and biased toward accepting the null hypothesis (Saravanan et al., 2020). Mean manifold locations for male and female retinas were significantly different ($p < 10^{-6}$ for bootstrap and $p < 0.05$ for hierarchical bootstrap). Identical tests were applied for assessing male-female differences in firing rate and the speed of temporal filtering ($p < 0.01$ for bootstrap and $p < 0.05$ for hierarchical bootstrap for both quantities).

Systematic differences in visual encoding properties were not observed with respect to animal age, laterality of the eye, temporal/nasal location, circadian time of dissection, month/season of the experiment and the lab from where the retina was obtained. Note that it is important to be careful when testing multiple statistical hypotheses using the same dataset. For the present data, the male-female differences were highly significant (p < 10^-6) using the marginalization method, even when compared to a lower Bonferroni-corrected threshold corresponding to multiple (~10) hypothesis tests.

Invariance of neural coding on the manifold

The ability of the manifold to preserve previously reported invariances of the neural code was tested as follows. First, random manifold locations were sampled by perturbing the learned locations of training retinas with a Gaussian noise of standard deviation equal to their median nearest-neighbor distance. Second, ON and OFF parasol firing rate maps in response to a white noise stimulus were computed using neural networks corresponding to these locations. Third, the Poisson firing rates for 200 cells (100 of each type) with random receptive field locations were read off from the firing rate maps. Finally, these responses were sampled and used to estimate a Linear-Nonlinear-Poisson model, which served as an interpretable summary of neural encoding captured by the manifold location. Comparison of average receptive field size and the zero crossing time of the temporal filter revealed known invariances between ON and OFF parasol cells (Figures 2J and 2K).

Estimation of the manifold location of a previously unseen retina

By fixing the shared parameters after learning, and estimating the recording-specific representation on the manifold, the trained model was adapted to predict responses in a new, previously unseen recording. Based on the amount of data available, several methods can be employed to identify the manifold location (Figure 4). These methods are described below in detail.

Averaging

When no data about the new retina are available, the simplest approach is to average the manifold location of all the retinas used for training.

Approximation

This is similar to averaging, but only using the subset of training retinas with similar response properties as the new retina. For Figure 4, locations of five training retinas with the most similar firing rate function were used for approximation.

Optimization

When light response data are available, the manifold location α_i of retina *i* was determined by Bayesian inference. Bayesian inference combines a Gaussian prior $[P_{prior}(\alpha) \sim N(\mu_{\text{prior}},\ \sigma_{\text{prior}})]$ over manifold locations determined from the training retinas and the likelihood

$[P(R_i^n \mid S^n, \alpha_i)]$ of collection of stimulus-response data $\{S^n, R_i^n\}_{n=1}^{n=N}$for the new retina i. The posterior $P(\alpha_i | \{S^n, R_i^n\})$ was maximized using gradient ascent (learning rate 0.1):

$$\alpha_i^* = \operatorname{argmax}_\alpha P(\alpha_i | \{S^n, R_i^n\}) = \operatorname{argmax}_\alpha \sum_n \log P(R_i^n | S^n, \alpha_i) + \log P_{\text{prior}}(\alpha_i)$$

Discrimination task

When light responses are unavailable, as in the case of a blind person implanted with a retinal prosthesis, the manifold location can in principle be estimted using a discrimination task. The effectiveness of this proposed approach was validated in simulation. For a given visual stimulus, the discrimination task involves using the implanted retinal prosthesis to reproduce responses corresponding to two manifold locations and asking the subject to select the response that yields perception most closely matching a verbally-described stimulus. Multiple rounds of this task are used to update the posterior on manifold locations.

The discrimination task was simulated under the assumption that the perceptual difference of the responses generated by hypothetical retinas at two manifold locations $\alpha_1$and α_2 for a stimulus S is equal to the KL-divergence between the corresponding response distributions $P(R|S, \alpha_1)$ and $P(R|S, \alpha_2)$. Given α_{true} as the true underlying manifold location, the subject's feedback $Y(\alpha_{true}, \alpha_1, \alpha_2) = 0$ if

$$D_{KL}[P(R|S, \alpha_1) \| P(R|S, \alpha_{\text{true}})] \leq D_{KL}[P(R|S, \alpha_2) \| P(R|S, \alpha_{\text{true}})]$$

and $Y(\alpha_{\text{true}}, \alpha_1, \alpha_2) = 1$ otherwise. For simplicity, sampled responses were used to compute an approximation of the KL-divergence:

$$D_{KL}\left[P(R|S, \alpha_1) \| P(R|S, \alpha_{\text{true}})\right] \approx \sum_{R_i \sim P(R|S, \alpha_1)} log\left[\frac{P(R_i|S, \alpha_1)}{P(R_i|S, \alpha_{\text{true}})}\right].$$

Hence, the posterior over manifold location after k steps of the task is given by:

$$P_{\text{posterior}}\left[\alpha | \{S_k, \alpha_{1,k}, \alpha_{2,k}, Y_k\}_{k=1}^{k=K}\right] \propto \prod_{k=1}^{k=K} P(Y_k | S_k, \alpha_{1,k}, \alpha_{2,k}) P_{\text{prior}}(\alpha)$$

where the prior is estimated from training retinas as a Gaussian distribution: $P_{\text{prior}}(\alpha) \sim N(\mu_{\text{prior}}, \sigma_{\text{prior}})$.

In the simulations, the visual stimulus S consisted of letters of the English alphabet, flashed for 100 ms and preceded and succeeded by 50 ms of gray screen. At each step, a Gaussian approximation of the posterior $P_{\text{posterior}}(\alpha) \sim N(\mu_{\text{posterior}}, \sigma_{\text{posterior}})$ was maintained, and updated using Monte-Carlo sampling. In summary, the steps for the k^{th}iteration of the algorithm are as follows:

1. Sample symmetric $\alpha_{1,k}$, $\alpha_{2,k}$ around posterior mean: $\alpha_{1,k} \sim P_{posterior}(\alpha)$; $\alpha_{2,k} = 2\mu_{posterior} - \alpha_{1,k}$.
2. Sample an English letter and a target stimulus S_k.
3. Sample responses $R_{1,k} \sim P(R|S_k, \alpha_{1,k})$; $R_{2,k} \sim P(R|S_k, \alpha_{2,k})$.
4. Get patient feedback $Y(\alpha_{true}, \alpha_{1,k}, \alpha_{2,k})$, based on an estimate of the KL divergence using sampled responses $R_{1,k}$, $R_{2,k}$.
5. Update the posterior of plausible manifold locations:

a. Sample N retina locations $\alpha_i \sim P_{posterior}(\alpha)$ for $i \in [1, \cdots, N]$.
b. For the set of sampled manifold locations, find the subset that matches user feedback, i.e., with $Y(\alpha_i, \alpha_{1,l}, \alpha_{2,l}) = Y(\alpha_{true}, \alpha_{1,l}, \alpha_{2,l})$ for all l $(=1, ..., K)$ previous steps. Let this subset of be $\{\tilde{\alpha}_j\}$.
c. Update the Gaussian approximation to posterior distribution with $\mu_{posterior} = <\tilde{\alpha}_j>$ and $\sigma^2_{posterior} = <\tilde{\alpha}_j \tilde{\alpha}_j{}^T> - \mu_{posterior}\mu^T_{posterior}$.

For the results shown in Figure 4F, α_{true} was set as the result of optimizing the manifold location using light-evoked responses. In the simulations, the posterior distribution converged in ~20 steps, suggesting that the low-dimensional manifold can be used for efficiently identifying the expected neural code in a blind person. However, the amount of noise in a human subject would probably be substantially higher, which would imply the need for a larger number of steps to identify the true manifold location, changes to the estimator of KL divergence and the method to update the posterior of α.