An Image Reconstruction Framework for Characterizing Early Vision

Ling-Qi Zhang[1]*, Nicolas P. Cottaris[1], and David H. Brainard[1]
[1]Department of Psychology, University of Pennsylvania
*lingqiz@sas.upenn.edu

Abstract

We developed an image-computable observer model of the early visual system that operates on fully naturalistic input, based on a framework of Bayesian image reconstruction from retinal cone mosaic excitations. Our model extends previous work on ideal observer analysis and the evaluation of performance beyond psychophysical discrimination tasks, takes into account the statistical regularities of our visual environment, and provides a unifying framework for answering a wide range of questions regarding early vision. Using the error in the reconstruction as a metric, we analyzed the variations of the number of different photoreceptor types on human retina as an optimal design problem. In addition, the reconstructions allow both visualization and quantification of information loss due to physiological optics and cone mosaic sampling, and how these vary with eccentricity. Furthermore, in simulations of color deficiencies and interferometric experiments, we found that the reconstructed images provide a reasonable proxy for directly modeling subjects' percepts. Lastly, we used the reconstruction-based observer for the analysis of psychophysical threshold, and found notable interactions between spatial frequency and chromatic direction in the resulting spatial contrast sensitivity function. Our method should be widely applicable to many experiments and practical applications in which early vision plays an important role.

Introduction

Visual perception begins at the retina, which takes sensory measurements of the light incident at the eyes. This initial representation is then transformed by computations that support perceptual inferences about the external world. Even these earliest sensory measurements, however, do not preserve all of the information available in the light signal. Factors such as optical aberrations, spatial and spectral sampling by the cone mosaic, and noise in the cone excitations all limit the information available downstream.

One approach to understanding the implications of such information loss is ideal observer analysis, which evaluates the optimal performance on psychophysical discrimination tasks. This allows for quantification of the limits imposed by features of early vision, as well as predictions of the effect of variation in these features (Geisler 1989, 2011). Ideal observer analysis separates effects due to the early visual representation from inefficiencies in the processes that mediate the discrimination decisions themselves. Such analyses have often been applied to analyze performance for simple artificial stimuli, assuming that the stimuli to be discriminated are known exactly (Banks, Geisler, and Bennett 1987; Davila and Geisler 1991) or known statistically with some uncertainty (Pelli 1985; Geisler 2018). The ideal observer approach has been extended to consider decision processes that learn aspects of the stimuli being discriminated, rather than being provided with these a priori, and extended to handle discrimination and estimation tasks with naturalistic stimuli (Burge and Geisler 2011, 2014; Singh et al. 2018; Chin and Burge 2020; Kim and Burge 2020). For a recent review see Burge (2020); also see Tjan and Legge (1998) and Cottaris et al. (2019, 2020).

It is generally accepted that the visual system has internalized the statistical regularities of natural scenes, so as to take advantage of these regularities in the inference computations (Attneave 1954; Field 1987; Shepard 1987; Knill, Kersten, and Yuille 1996). This motivates interest in extending ideal observer analysis to apply to fully naturalistic input, while incorporating the statistical regularities of natural scenes (Burge 2020). Here we pursue an approach to this goal that, in addition, extends the evaluation of performance to a diverse set of objectives.

We developed a method that, under certain assumptions, optimally reconstructs images from noisy cone excitations, with the excitations generated from an accurate image-computable model of the early visual system (Cottaris et al. 2019, 2020). The image reconstruction approach provides us with a unified framework for characterizing the information loss due to various factors in early vision. In the next sections, we show analyses that: 1) use image reconstruction error as an information metric to understand the retinal mosaic “design” problem, such as the implications of the different allocations of retinal photoreceptor types; 2) allow both visualization and quantification of information loss due to physiological optics and cone mosaic sampling, their variations with changing eccentricity, as well as due to color deficiency; 3) combine the image reconstruction approach with analysis of psychophysical discrimination, providing an

way to incorporate into such analyses the assumption that our visual system takes into account the statistical regularities of natural images.

Results

We developed a Bayesian method to reconstruct images from sensory measurements, which we describe briefly here (see *Methods* for details). We begin with a forward model that expresses the relation between an image and its visual representation at a well-defined stage in the visual pathway. Here that stage is the excitations of the photoreceptors of the cone mosaic, so that our model accounts for blur in retinal image formation and spatial and spectral sampling by the retinal cone mosaic. The approach is general, however, and may be applied to other sites in the visual pathways (see e.g., Naselaris et al. 2009; Parthasarathy et al. 2017). Our forward model is implemented within the open-source software package ISETBio (isetbio.org; **Fig. 1A-C**) which encapsulates the probabilistic relationship between the stimulus (i.e., pixel values of a displayed RGB image) and the cone excitations (i.e., trial-by-trial photopigment isomerizations). ISETBio simulates the process of displaying an image on the monitor (**Fig. 1A**), the wavelength-dependent optical blur of the human eye and spectral transmission through the lens and the macular pigment (**Fig. 1B**), as well as the interleaved spatial and chromatic sampling of the retinal image by the L, M and S cones (**Fig. 1C**). Noise in the cone signals is characterized by a Poisson process. The forward model allows us to compute the *likelihood* function. The likelihood function represents the probability that an observed pattern of cone excitations was produced by any given image.

To obtain a *prior* over natural images, we applied independent components analysis (ICA, see *Methods*) to a large dataset of natural images (Russakovsky et al. 2015), and fit an exponential probability density function to the individual component weights (**Fig. 1D**). The prior serves as our description of the statistical structure of natural images.

Given the likelihood function, prior distribution, and an observed pattern of cone excitations, we can then obtain a reconstruction of the original image stimulus by applying Bayes rule to find the posterior probability of any image given that pattern. We take the reconstructed image as the one that maximizes the *a posteriori* probability (MAP estimate, see *Methods*) (**Fig. 1D**).

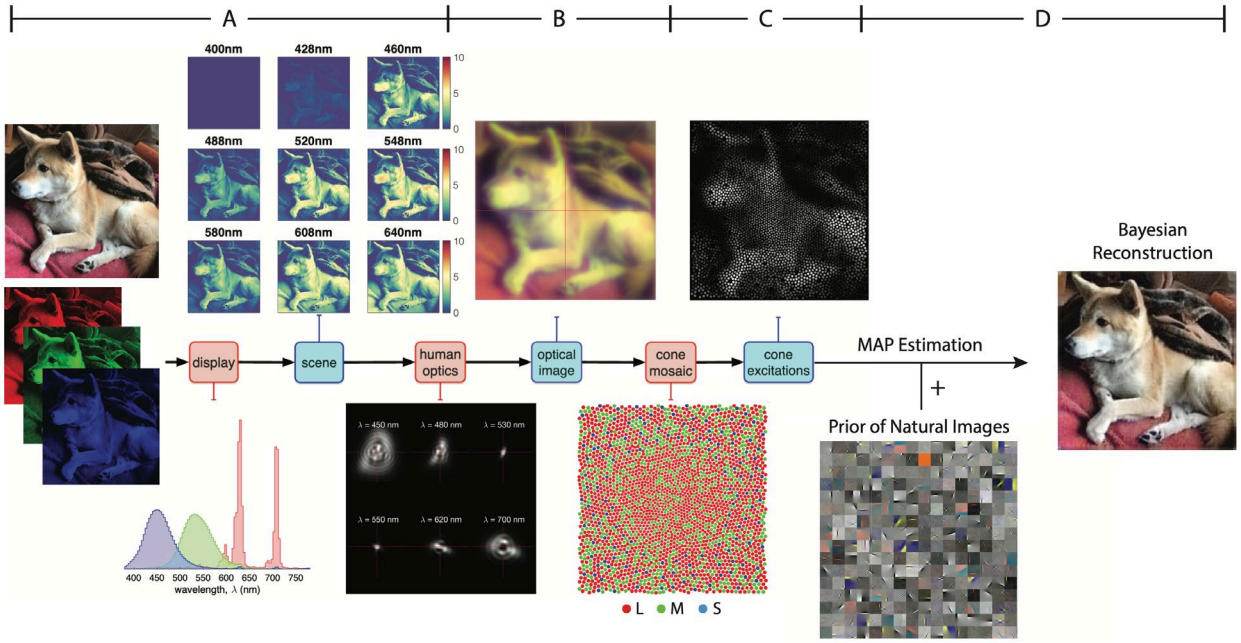

Figure 1: Model of Early Vision and Bayesian Reconstruction from Cone Mosaic Excitation. (A) The visual stimulus, in our case a natural image in RGB format, is displayed on a simulated monitor, which generates a hyperspectral scene representation of that image. **(B)** The hyperspectral image is blurred with a set of wavelength-dependent point-spread functions typical of human optics. We also account for spectral transmission through the lens and the macular pigment. This process produces the retinal image at the photoreceptor plane. **(C)** The retinal image is then sampled by a realistic cone mosaic, which generates cone excitations (isomerizations) for each cone. The trial-by-trial variability in the cone excitations is modeled as a Poisson process. **(D)** Our Bayesian reconstruction method takes the pattern of cone excitations as input and estimates the original stimulus (RGB image) based on the likelihood function and a statistical model (prior distribution) of natural images (See *Methods*).

Basic Properties of the Reconstructions:

To understand the consequences of early visual encoding, we need to study the interaction between the likelihood function (i.e., our model of early vision) and the statistics (i.e., prior distribution) of natural images. There are strong constraints on the statistical structure of natural images, such that natural images occupy only a small manifold within the space of all possible images. The properties of early vision produce ambiguities with respect to what image is displayed when only the likelihood function is considered, but if these can be resolved by taking advantage of the statistical regularities of the visual environment, they should in principle, not prohibit effective visual perception. To illustrate this point, consider the simple example of discrete signal sampling: Based on the sampled signal, one cannot distinguish between the original signal from all its possible aliases (Bracewell 1986). However, with the prior knowledge that the original signal contains only frequencies below the Nyquist frequency of the sampling array, this ambiguity is resolved.

To highlight the importance of prior information while holding the likelihood function fixed, we can vary a parameter γ that adjusts the weight of the log-prior term in the reconstruction objective function (see *Methods*). Explicitly manipulating γ reveals the effect of the prior on the reconstruction (**Fig. 2**). When γ is small, the reconstruction is corrupted by the noise and the ambiguity of early vison (**Fig. 2A, 2B**). When γ is large, the effect of prior mean attraction produces desaturation and over-smoothing (**Fig. 2E**) in the reconstruction. For the rest of our simulations, the value of γ is determined on the training set by a cross-validation procedure that minimizes the reconstruction error unless specified otherwise (**Fig. 2C**).

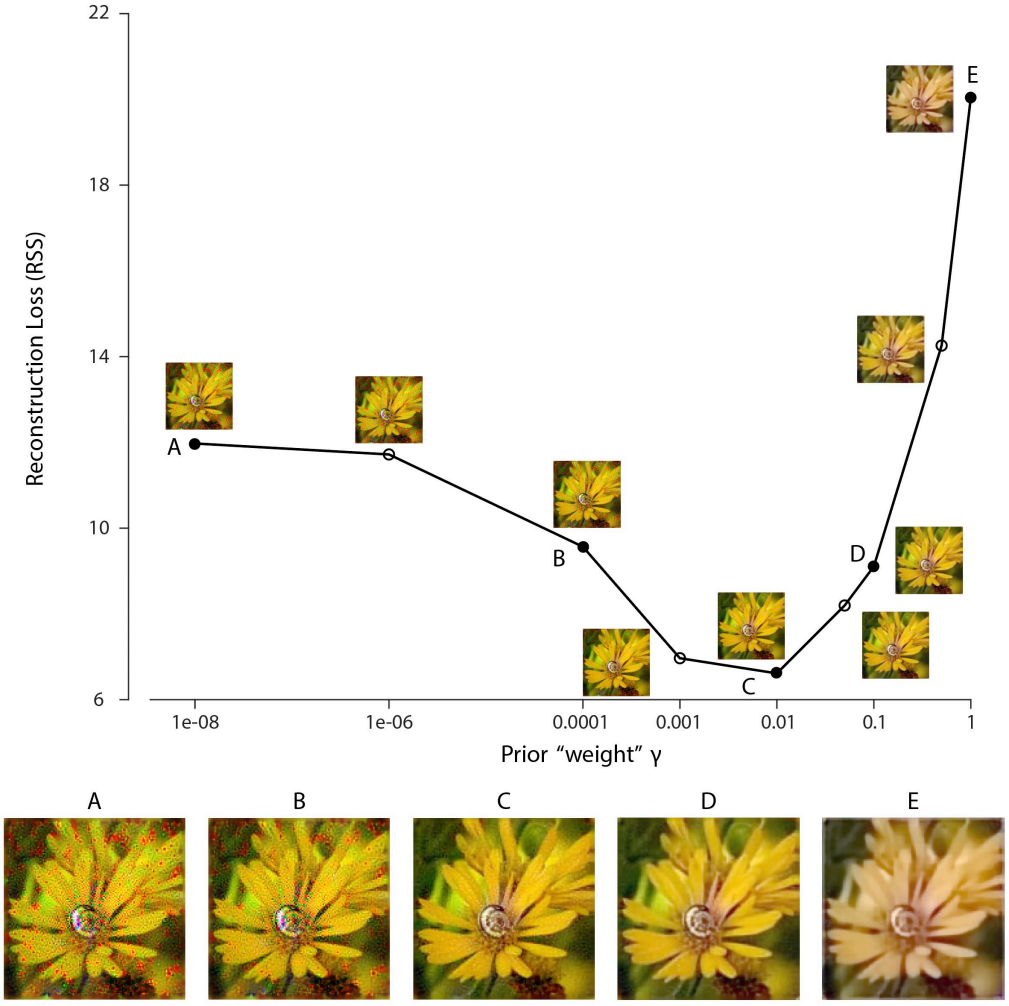

Figure 2: Effect of Prior Weight on Reconstructed Image. Reconstruction error for an example natural image using a 1-deg foveal mosaic and root sum of squared distance (RSS, y-axis) in the pixel space as the error metric, as a function of weight γ on the log-prior term (x-axis, see *Methods*) in the reconstruction objective function. The reconstructed image obtained with each particular γ value is shown alongside each corresponding point. Image **C** corresponds to the value of γ obtained through the cross-validation procedure (see *Methods*). The images at the bottom are magnified versions of a subset of the images for representative γ values, as indicated by the solid dots in the plot.

To further elucidate properties of the Bayesian reconstruction, especially the interaction between the likelihood and prior, we plotted a few representative images in a log-prior, log-likelihood coordinate system, given a particular instance of cone excitations (**Fig. 3**). The optimal reconstruction generated by MAP has both a high prior probability and likelihood value as expected (**Fig. 3A**). In fact, for our reconstruction algorithm, there should not exist any image above the $\gamma x + y = c$ line that goes through **A** (solid line, **Fig. 3**), otherwise the optimization routine has failed to find the global optimum. The original image stimulus (ground truth) has a slightly lower likelihood value, mainly due to noise present in the cone excitations, and also a slightly lower prior probability, presumably due to the fact that our prior is only an approximation to the true natural image distribution (**Fig. 3B**). The detrimental effect of noise becomes the most prominent in the maximum likelihood estimate (MLE, **Fig. 3C**): Noise in the cone response is interpreted as true variation in the original image stimulus, thus slightly increasing the likelihood value but also creating artifacts. Such artifacts are penalized by the prior in other reconstructions. Furthermore, even without the presence of noise, other factors in early visual processing (e.g., **Fig. 1B, C**) will still cause loss of information and ambiguity for the reconstruction. This is reflected as a set of images that lie on the equal likelihood line (**Fig. 3D**): There exist an infinite set of variations in the image (stimulus) that have no effect on the value of the likelihood function (i.e., variations within the null space of the linear likelihood render matrix, see *Methods*). Thus, the cone excitations provide no information to distinguish between images that differ by such variations. However, as with the case of noise, variations inconsistent with natural images are discouraged by the prior. Other corruptions of the image, such as addition of white noise in the RGB pixel space, are countered by both the likelihood and prior (**Fig. 3E**). Lastly, for illustration purposes, we can increase the prior probability of the reconstruction relative to the optimal by making it spatially or chromatically more uniform (**Fig. 3F**), but doing so drastically decreases the likelihood.

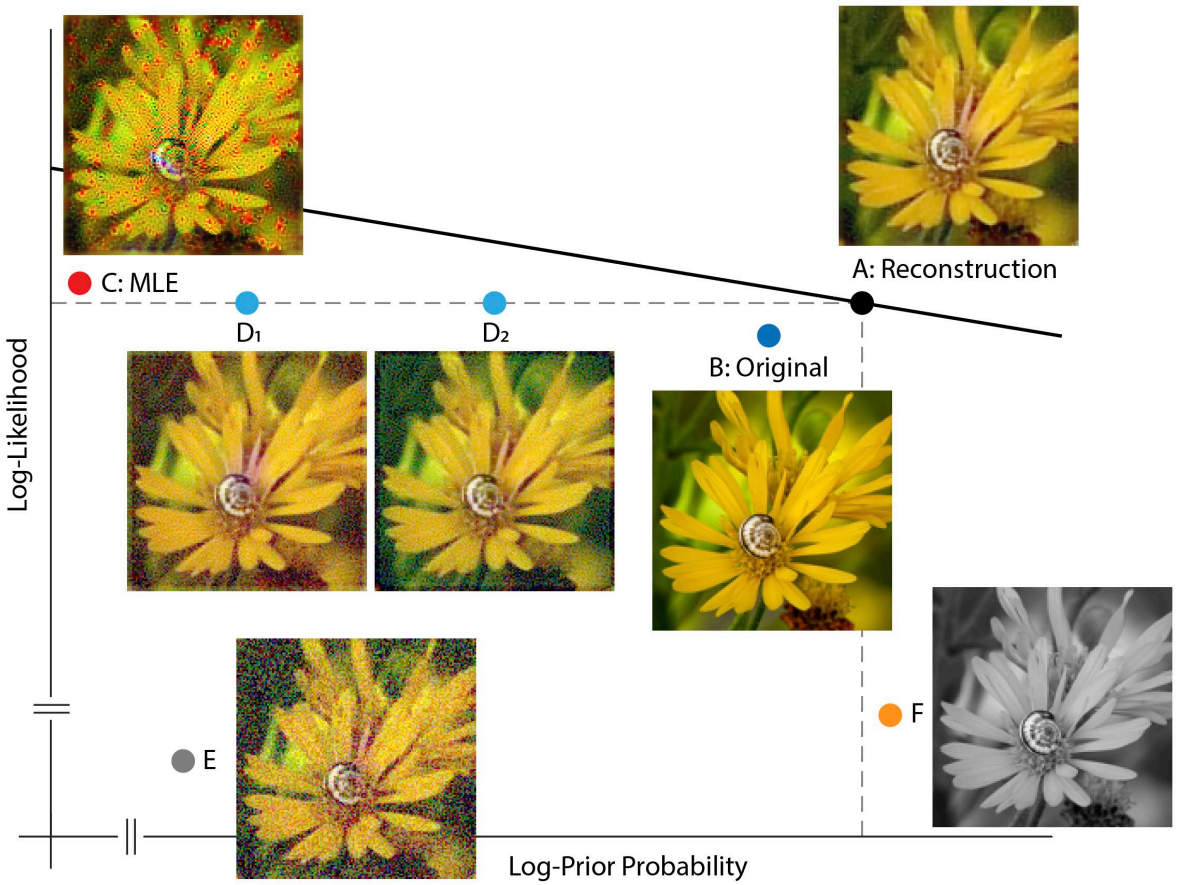

Figure 3: Solution Space of Image Reconstruction. Given a particular instance of cone excitations, we can evaluate the (log-)prior probability (x-axis) and (log-)likelihood value (y-axis) for arbitrary images. Here, a few representative images are shown together with their corresponding location in a log-prior, log-likelihood coordinate system. **(A)** The optimal reconstruction obtained via the reconstruction algorithm. The solid line shows $\gamma x + y = c$, with the value of c evaluated at the optimal reconstruction and with the value of γ matched to that obtained through cross-validation. **(B)** Original input image (ground truth). **(C)** The reconstruction generated by maximum likelihood estimation (MLE, set $\gamma = 0$). **(D)** Optimal reconstruction, corrupted by patterns randomly sampled from the null space of the likelihood render matrix (see *Methods*). **(E)** Optimal reconstruction, corrupted by white noise in RGB space. **(F)** Grayscale version of the optimal reconstruction.

Optimal Allocation of Retinal Photoreceptors:
Within the Bayesian reconstruction framework, the goal of the early visual system can be characterized as minimizing the average error in reconstruction across the set of natural images. In this context, we can ask how to choose various elements of early vision, subject to constraints, to minimize the expected reconstruction error under the natural image prior (Levin, Freeman, and Durand 2008; Manning and Brainard 2009). More formally, we seek the "design" parameters θ of an early visual system:

$$\theta = argmin_{\theta} \quad E_{p(x)} \left(E_{p(m|x;\theta)} \, L(\hat{x}(m;\theta), x) \right),$$

where $\hat{x}(m;\theta) = argmax_x \quad p(m|x;\theta)p(x)$. Here x represents individual samples of natural images, m represents instances of cone excitation (i.e., sensory measurements), and $p(m|x;\theta)$ is our model of early vision (i.e., likelihood function). The particular features under consideration of the modeled visual system are indicated explicitly by the parameter vector θ. The MAP image reconstruction is indicated by $\hat{x}(m;\theta)$, and $L(\cdot,\cdot)$ is a loss function that assesses reconstruction error. In practice, the expectations are approximated by taking the average over large samples of natural images and cone excitations.[1]

One intriguing design problem is the allocation of cone photoreceptor types: The maximum number of photoreceptors (cones) per unit area is bounded due to biological constraints. How should the visual system assign this limited resource across the three different types of cones? It has been observed in human subjects that there is a relatively sparse population of S cones, while large individual variability exists in the L/M cone ratio (Hofer et al. 2005). Previous research has used information-theoretical measures combined with approximations to address this question (Garrigan et al. 2010). Here, we empirically evaluated a loss function (i.e., we used root sum of squares distance in the RGB pixel space as well as the S-CIELAB space) within our model of early vision, while systematically changing the allocation of retinal cone types (**Fig. 4**).

Interestingly, we found that large variations (nearly a 10-fold range) in the assignment of L and M cones have little impact on the average reconstruction error (**Fig. 4A**). Only when the proportion of L or M cones becomes very low is there a substantial increase in reconstruction error, as the modeled visual system approaches dichromacy. On the other hand, the average reconstruction error as a function of the proportion of S cones shows a clear optimum at a small S-cone proportion (~10%; **Fig. 4B**).

Our results are in agreement with previous analysis in showing that the empirically observed allocation of retinal photoreceptor type is consistent with the principle of optimal design, which maximizes information transfer (Garrigan et al. 2010). The indifference to L/M ratio can be explained by the large spatial and chromatic correlations present in natural images, together with the high overlap in L- and M-cone spectral sensitivities. This leads to a high correlation in the excitations of neighboring L and M cones in response to natural images, allowing cones of one type to be substituted for cones of the other type with little effect on reconstruction error (see the next section for additional analysis on this point). Additional analysis (**Fig. S1**) revealed that the sensitivity to S cone proportion is due to a combination of two main factors: 1) chromatic aberrations, which blur the retinal image at short wavelengths and reduce the value of dense spatial sampling at these wavelengths; and 2) S cones mainly contribute to the estimation of pixel values in the B-pixel plane, whereas L and M cone contribute to both the R- and G-pixel planes (see **Fig. S1**). This makes L and M cones more informative than S cones, given the particular loss functions we employ to evaluate reconstruction error. To further validate our conclusion, we have also replicated our analysis with a

[1] For simplicity in the development here, we did not include the parameter γ that we incorporated into our reconstruction algorithm in the equations above. It was included, however, in the actual computations that investigated the effect of the parameters of early vision on reconstruction performance.

dataset of hyperspectral (as opposed to RGB) images (Nascimento, Ferreira, and Foster 2002; Chakrabarti and Zickler 2011), with a loss function applied directly to the whole spectrum, and have obtained similar results (**Fig. S2**, also see *Methods*).

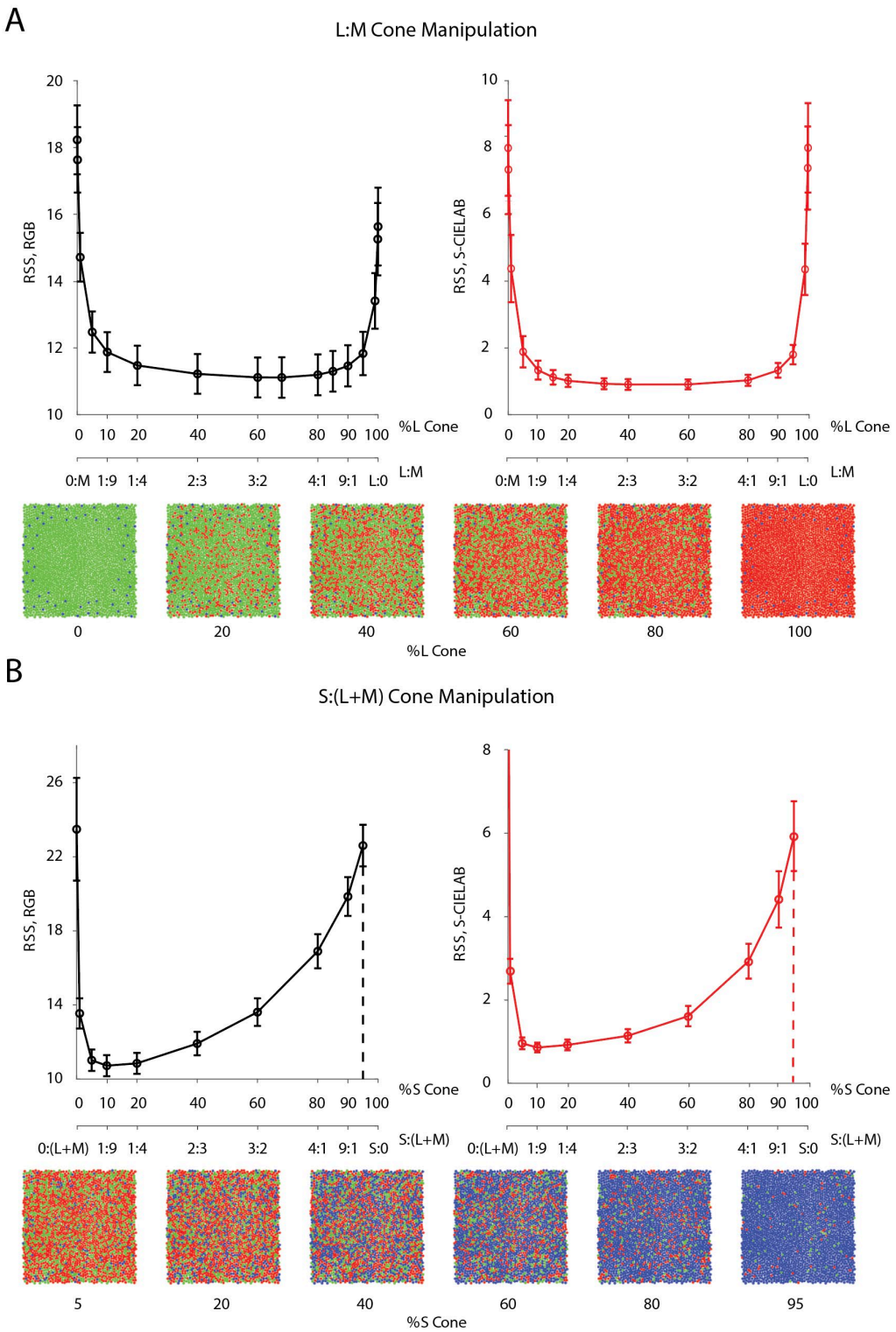

Figure 4: Effect of the Allocation of Retinal Cone Types on Reconstruction. Average image reconstruction error from a 1-deg foveal mosaic on a set of natural images from the evaluation set, computed as root sum of squares (RSS) distance in the RGB pixel space (y-axis, left panels) and the S-CIELAB space (y-axis, right panels), as a function of different allocations of retinal photoreceptor (cone) types in the mosaic. **(A)** Average (over evaluation images) reconstruction error as a function of %L cone (top x-axis), or L:M cone ratio (bottom x-axis). Example mosaics with different %L values are shown below the plot. Error bars indicate +/- 1 SEM. **(B)** Average reconstruction error as a function of %S cone (top x-axis), or S:(L+M) cone ratio (bottom x-axis). Example mosaics with different %S values are shown below the plot. Error bars indicate +/- 1 SEM across sampled images. See **Fig. S2** for a replication of the same analysis with hyperspectral images.

To further study the role of statistical regularities in the optimal allocation of photoreceptor type, we repeated the analysis above of evaluating the optimal design objective as a function of L-cone proportion, but on different sets of synthetic image datasets for which the spatial and chromatic correlations in the images were manipulated explicitly (see *Methods*). The dependence of the average reconstruction error on the L-cone proportion decreases as the chromatic correlation in the signal increases (**Fig. 5**). A decrease of spatial correlation has little impact on the overall pattern itself, but decreases the overall magnitude of reconstruction error (**Fig. 5**; to highlight the shape of the curves, the scale of the y-axis is different across rows and columns). This analysis highlights the importance of considering visual system design in context of the statistical properties (prior distribution) of natural images, as it shows that the conclusions drawn can vary with these properties. Natural images are thought to have both high spatial and high chromatic correlation (Webster and Mollon 1997; Nascimento et al. 2002; Garrigan et al. 2010), making the results shown in **Fig. 5** consistent with those in **Fig. 4**.

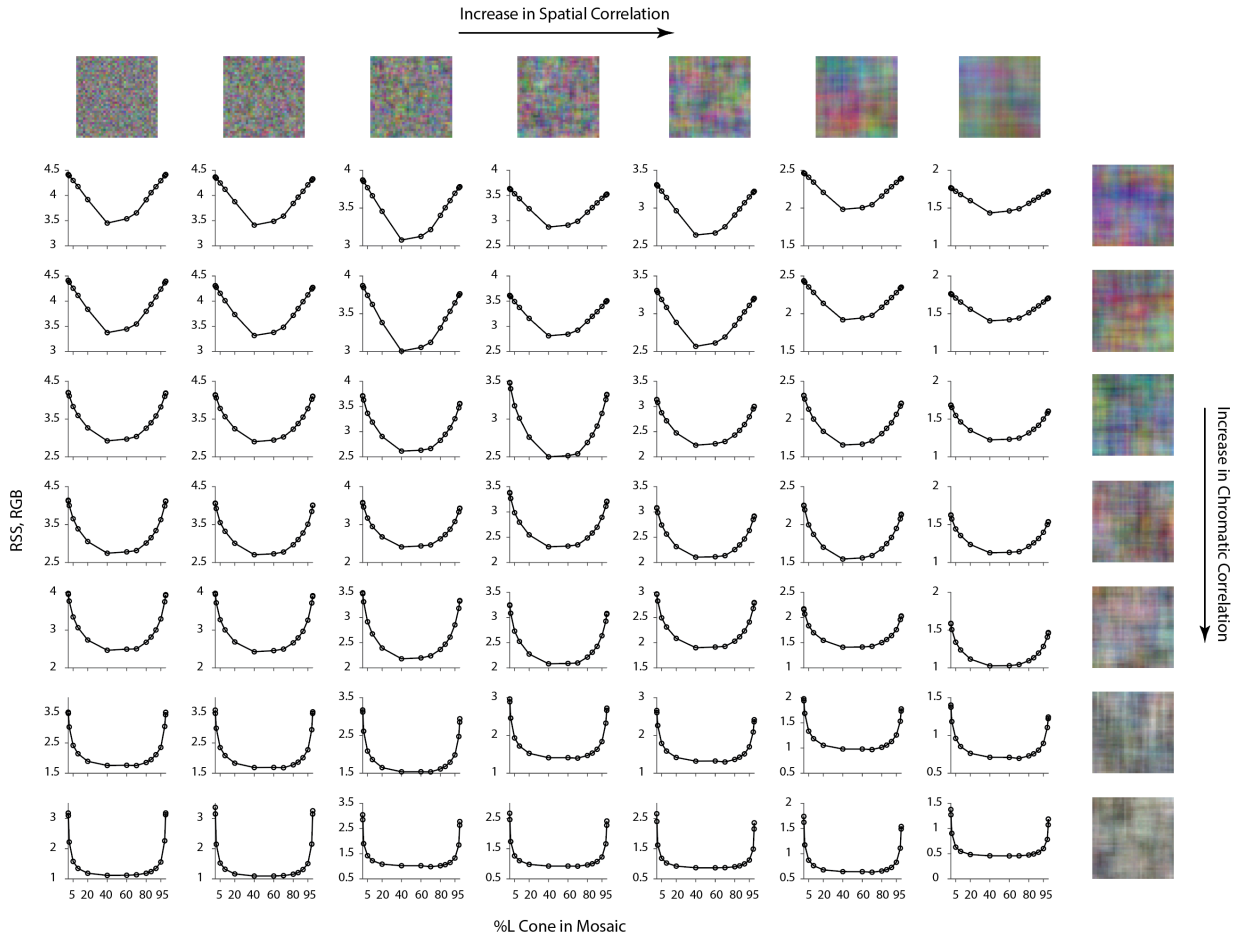

Figure 5: Effect of Spatial and Chromatic Correlation on the Optimal Allocation of Photoreceptors. Average image reconstruction error from a half-degree square foveal mosaic on different sets of synthetic images, computed as root sum of squares (RSS) distance in the RGB pixel space, as a function of %L cone (L:M cone ratio) of the mosaic (i.e., similar to **Fig. 4A**, left column). Within each subfigure, synthetic images were sampled from a Gaussian distribution with changing spatial and chromatic correlation, as indicated by example images on the top row and rightmost column, and reconstruction was performed with the corresponding Gaussian prior (see *Methods*). The overall RSS is reduced compared to **Fig. 4** due to the smaller image size used and the fact that the images corresponded to a different prior, as well as that the prior now exactly described the images. In addition, error bars were negligible due to the large image sample size used.

Visualization of Color Deficiency with Image Reconstruction:
In addition to quantification, the reconstruction framework also provides a method for visualizing the effect of information loss in the early visual pathways. We know that extreme values of L:M cone ratio create essentially dichromatic retinal mosaics, and from the analysis above we observed that these lead to high reconstruction error. To understand the nature of this error, we can directly visualize the reconstructed images.

Fig. 6A shows reconstructions of a set of example images with different dichromatic retinal mosaics. While the spatial structure of the original images is largely retained in

the reconstructions, each type of dichromacy creates a distinct pattern of color confusions and shifts in the reconstructed color.

One might speculate as to whether the reconstructions predict color appearance as experienced by dichromats. To approach this, we compare the reconstructions with the results of methods developed based on reports of the color percepts experience by dichromats (Brettel, Viénot, and Mollon 1997). Our reconstruction produces images that are qualitatively similar to the output of that method (**Fig. 6B**). We also show (**Fig. 6C**) the results of an approach based not on reconstructing images from cone responses but rather on reconstructing spectra from cone excitations (Jiang, Farrell, and Wandell 2016). This method too produces similar results. We also note that our algorithm takes into account the spatial structure of the image, while the previously proposed methods operate pixel-wise.

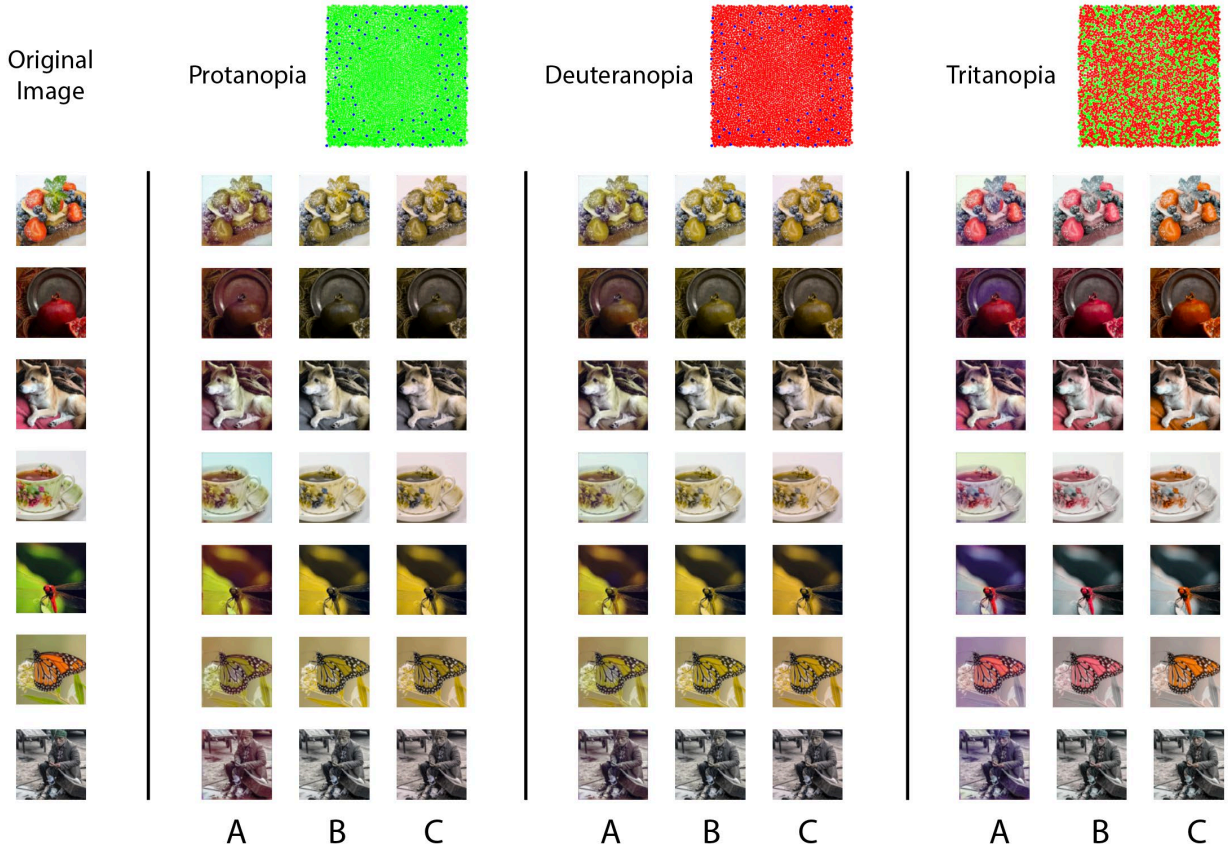

Figure 6: Visualization of Effect of Dichromacy. Reconstructions of a set of example images in the evaluation set from different types of 1-degree foveal dichromatic retinal mosaics (protanopia, deuteranopia, tritanopia) together with other previously proposed methods for predicting color appearance for dichromats. **(A)** Our method; **(B)** Brettel, Viénot, and Mollon (1997); **(C)** Jiang, Farrell, and Wandell (2016).

Anomalous trichromacy is another form of color deficiency that is commonly found in human observers. For example, in deuteranomaly, the spectral sensitivity of the M cones is shifted towards that of the L cones (**Fig. 7B**). Since the three cone spectral

sensitivity functions are linearly independent of each other, in the absence of noise we should be able to obtain a trichromatic reconstruction from the excitations of the deuteranomalous mosaic. However, in the presence of noise, we expect that the high degree of overlap between M and L spectral sensitivities will result in a lower signal-to-noise ratio (SNR) in the difference between M- and L-cone excitations, compared to that of a normal trichromatic observer, and thus to worse reconstructions. We performed image reconstructions for a normal trichromatic (with a peak spectral sensitivity of M cone at 530 nm) and a deuteranomalous (with a peak spectral sensitivity of M cone at 550 nm) 1-deg foveal mosaic at different overall light intensity levels (**Fig. 7**). Due to the nature of Poisson noise, the higher the light intensity, the higher the SNR of the cone excitations. At high light intensities, the reconstructions are similar for the normal and deuteranomalous mosaics. At lower intensities, however, the deuteranomalous reconstruction lacks chromatic content still present in the normal reconstruction. A loss of chromatic content is also seen for the reconstruction from the normal mosaic at the lowest light level. This observation may be related to the fact that biological visual systems that operate at low light levels are typically monochromatic (e.g., the human rod system; see Manning and Brainard 2009 for a related treatment).

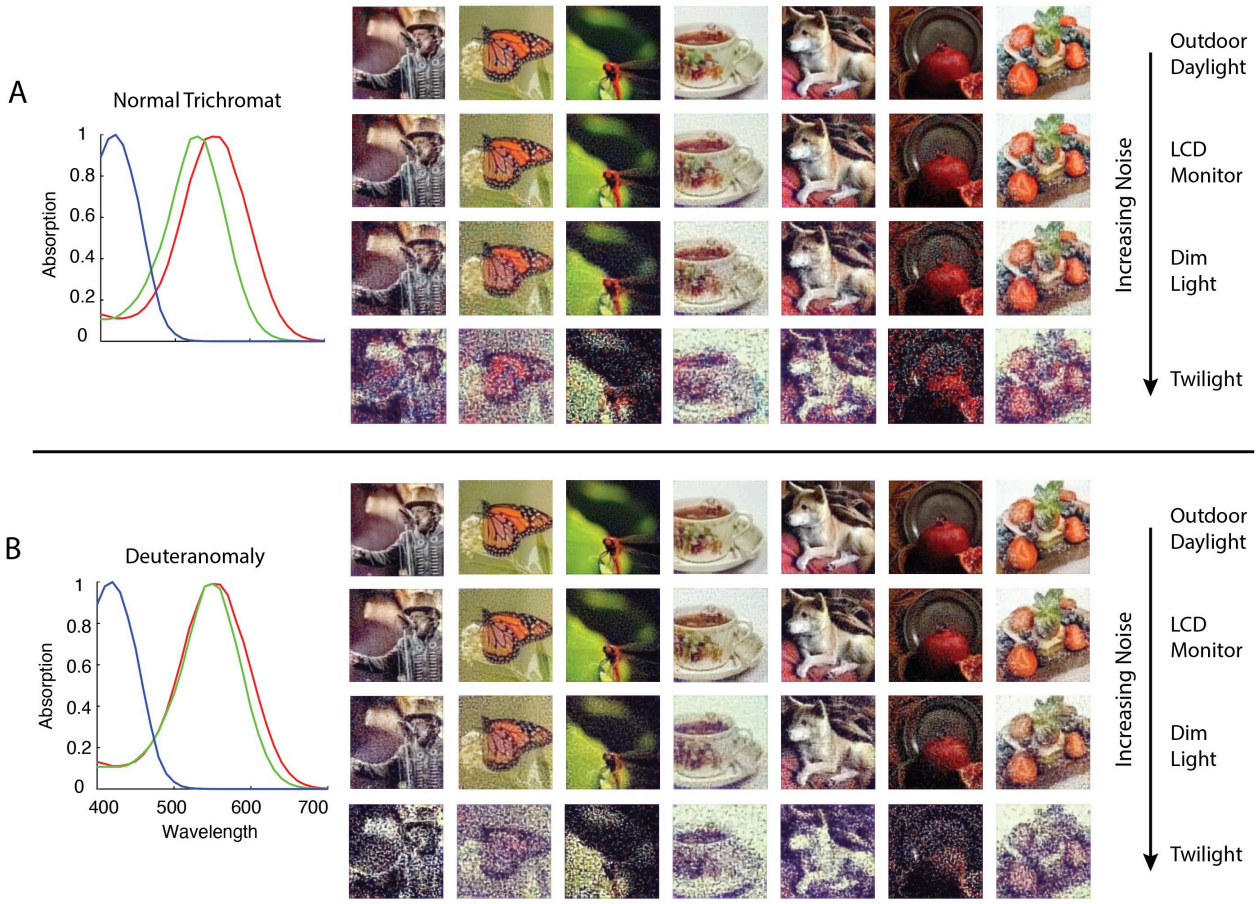

Figure 7: Comparison of Normal and Deuteranomalous Observer at Light Intensities. Image reconstructions for a set of example images in the evaluation set from 1-degree, foveal **(A)** normal trichromatic and **(B)** deuteranomalous trichromatic mosaics at five different overall light intensity levels that lead to different Poisson signal-to-noise ratios in the cone excitations. The average excitations (photo-isomerizations) per cone per 50 ms integration time is chosen to be approximately 10^4 for *Outdoor Daylight*, 10^3 for *LCD Monitor*, 10^2 for *Dim Light*, and 10^1 for *Twilight* (Lewis and Li 2006; Stockman and Sharpe 2006).

Effect of Physiological Optics and Mosaic Spatial Sampling:
So far, our visualizations have focused on chromatic information loss due to the limited number of cone types. However, the imperfection in the physiological optics, combined with the spatial sampling of retinal mosaic, also introduce significant loss of information. Furthermore, the interleaved nature of the mosaic means that color and pattern are entangled at the very initial stage of visual processing (Brainard 2019). To highlight these effects, we reconstructed natural images from 1-deg patches of mosaics at different retinal eccentricities across the visual field, with both 1) changes in optical aberrations (Polans et al. 2015) and 2) increase in size and decrease in density of the photoreceptors (Curcio et al. 1990) simulated accurately. The degradation in the quality of the reconstructed images can be clearly observed as we move from the fovea to the periphery (**Fig. 8**; See **Fig. S3** for an enlarged view of the mosaic and optics). For some retinal locations, the elongated point-spread function (PSF) also introduces a salient directional blur (**Fig. 8E, F**). For a simple quantification of the average reconstruction error as a function of visual eccentricity, see **Fig. S4**.

The consequences of irregular spatial sampling by the cone mosaic have been previously studied with the framework of signal processing (Snyder, Laughlin, and Stavenga 1977; Yellott 1983). Our results highlight that optimizing early vision depends in rich ways on the interplay between the cone sampling and the optics, which determines the information available at each retinal location, as well as how various aspects of reconstruction performance are assessed by the choice of the loss function. To emphasize this richness, note that in **Fig. 8**, we observe a trade-off across visual eccentricity between spatial and chromatic vision. In the image of the dragonfly, for example, the reconstructed colors are desaturated at intermediate eccentricities (e.g., **Fig. 8C, D**), compared with the fovea (**Fig. 8A**) and more eccentric locations (**Fig. 8E, F**). At the same time, the reconstructions generally become more blurred with increasing eccentricity. While less information (i.e., more eccentric location) does lead to lower quality reconstructions (**Fig. S4**), exactly which aspects of the reconstructions are incorrect can vary in subtle ways with the details of the sampling and optics. In an additional set of analysis, we systematically varied the size of the PSF for a fixed retinal mosaic. This revealed that (**Fig. S5**): 1) A larger PSF does lead to better estimate of chromatic content, albeit eventually at the cost of spatial content. 2) In general, an appropriate amount of optical blur is required to achieve the best overall image reconstruction performance, presumably due to its prevention of aliasing. We will treat spatial aliasing further in the next section.

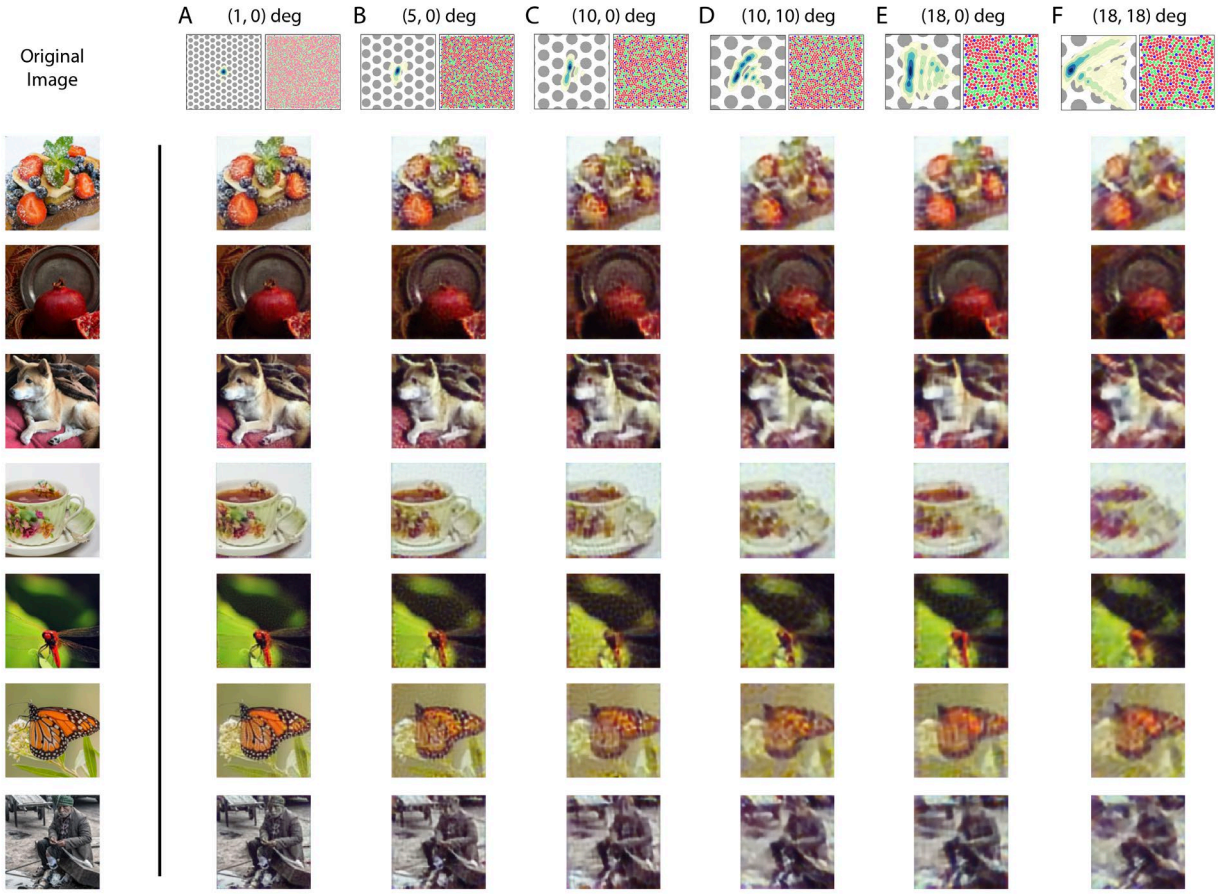

Figure 8: Image Reconstruction with Optics/Mosaic at Different Retinal Eccentricities. Image reconstructions for a set of example images in the evaluation set from 1-degree patches of mosaic at different retinal eccentricities. The coordinates at the top of each column indicate the horizontal and vertical eccentricity of the patch used for that column. The image at the top left of each column shows a contour plot of the point-spread function relative to an expanded view of the cone mosaic used for that column, while the image at the top right of each column shows the full 1-degree mosaic (see **Fig. S3** for an enlarged view of the mosaic and optics).

As we have alluded to above, the retinal mosaic and physiological optics can also interact in other important ways: Both in humans and other species, it has been noted that the optical cut-off of the eye is reasonably matched to the spacing of the photoreceptors (i.e., the mosaic Nyquist frequency), enabling good spatial resolution while minimizing spatial aliasing due to discrete sampling (Williams 1985; Snyder, Bossomaier, and Hughes 1986; Land and Nilsson 2012). In contrast to our work, these analyses did not take into account the fact that the cone mosaic interleaves multiple spectral classes of cones (Williams et al. 1991; Brainard 2015), and here we revisit classic experiments on spatial aliasing for a trichromatic mosaic using our reconstruction framework.

Experimentally, it has been demonstrated that with instruments that *bypass* the physiological optics and present high contrast grating stimuli directly on the retina, human subjects can detect spatial frequencies up to 200 cyc/deg (Williams 1985). For

foveal viewing, subjects also report having a percept resembling pattern of “two-dimensional noise” and/or “zebra stripes” when viewing those high spatial frequency stimuli (Williams 1985). For peripheral viewing, high frequency vertical gratings can be perceived as horizontal (and vice-versa; Coletta and Williams 1987). We explored these effects within our framework as follows: We reconstructed a set of vertical chromatic grating stimuli from the cone excitations of a foveal and a peripheral mosaic. To simulate the interferometric experimental condition in Williams (1985), we used diffraction-limited optics with no lateral chromatic aberration (LCA), allowing high-frequency stimuli to reach the cone mosaic directly. For gratings that are above the typical optical cut-off frequency, we obtained reconstructions that 1) are quite distinct from a uniform field, which would allow them to be reliably detected in a discrimination protocol; and 2) lack the coherent vertical structure of the original stimulus (**Fig. 9**). Concretely, the reconstructions recapitulate the “zebra stripe” percept reported at approximately 120 cyc/deg in the fovea (**Fig. 9A**); as well as the orientation-reversal effect at an appropriate spatial frequency in the periphery (**Fig. 9B**). Both corroborate pervious theoretical analysis and psychophysical measurements (Williams 1985; Coletta and Williams 1987), but now taking the trichromatic nature of the mosaic into account. On the other hand, with full optical aberration, the reconstructed images became mostly uniform at these high spatial frequencies (**Fig. S6**). Since our method accounts for trichromacy, we have also made the prediction that for achromatic grating stimuli viewed under similar diffraction-limited conditions, while the spatial aliasing pattern will be comparable, additional chromatic aliasing should be visible (**Fig. S7**).

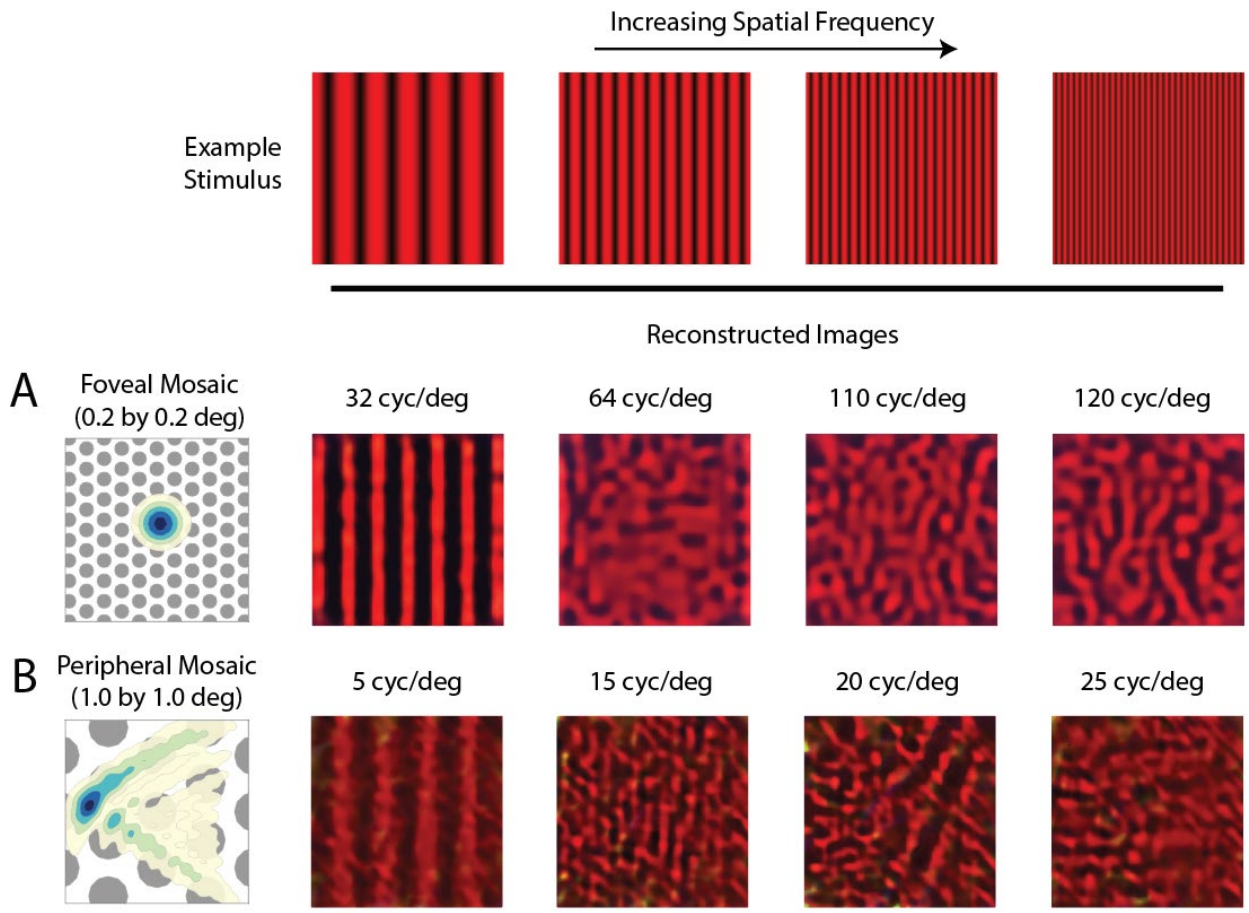

Figure 9: Reconstruction of Chromatic Grating Stimuli without Optical Aberrations. Image reconstruction of chromatic grating stimuli with increasing spatial frequency from **(A)** a 0.2-deg foveal mosaic and **(B)** a 1-deg peripheral mosaic at (18, 18) degree retinal eccentricity, using diffraction-limited optics without LCA. The leftmost images show an expanded view of the cone mosaic relative to a contour plot of a typical point-spread function at that eccentricity. Images were modulations of the red channel of the simulated monitor, to mimic the 633 nm laser used in the interferometric experiments. The exact frequency of the stimuli being used for each condition is as denoted in the figure. For a more extended comparison between reconstructions with and without optical aberrations, see **Fig. S6, S7**.

Contrast Sensitivity Function:

Our framework can also be adapted to perform ideal observer analysis for psychophysical discrimination (threshold) tasks, which have been commonly used previously to evaluate the information available in early vision. Here we use the reconstructed images to make discrimination-type decisions. This is potentially important since even the early visual representation (e.g., retinal circuits), on which the downstream decisions must be based, is inevitably shaped by the regularities of our visual environment (Borghuis et al. 2008; Karklin and Simoncelli 2011). Our method provides on way to extend ideal observer analysis to incorporate these statistical regularities.

Concretely, we predicted and compared the diffraction-limited spatial contrast sensitivity function (CSF) for gratings with a half-degree spatial extent (see *Methods*). First, we applied the classic signal-known-exactly ideal observer to the Poisson distributed excitations of the simulated cone mosaic. We computed CSFs for both achromatic (L+M) and chromatic (L-M) grating modulations, with matched root mean square (RMS) cone contrasts. As expected, the ideal observer at the cone excitations produces nearly identical CSFs for the contrast-matched L+M and L-M modulations; also, as expected, these fall off with spatial frequency, primarily because of optical blur (**Fig. 10A**).

Next, we reconstructed images from the cone excitations produced by the grating stimuli. A template-matching observer based on the noise-free reconstructions was then applied to the noisy reconstructions (see *Methods*). The image-reconstruction observer shows significant interactions between spatial frequency and chromatic direction. Sensitivity in the L+M direction is relatively constant with spatial frequency. Sensitivity in the L-M direction starts out higher than L+M at low spatial frequencies, but drops significantly and is lower than L+M at high spatial frequencies (**Fig. 10B**). We attribute these effects to the role of the image prior in the reconstructions, which leads to selective enhancement/attenuation of different image components.

It is intriguing that the CSFs from the reconstruction-based observer show substantially higher sensitivity for L-M than for L+M modulations at low spatial frequencies (with equated RMS cone contrast), but with a more rapid falloff such that the sensitivity for L+M modulations is higher at high spatial frequencies. Both of these features are characteristic of the CSFs of human vision (Mullen 1985; Anderson, Mullen, and Hess 1991; Chaparro et al. 1993; Sekiguchi, Williams, and Brainard 1993). A more comprehensive exploration of this effect and its potential interaction with other decision rule used in the calculation awaits future research.

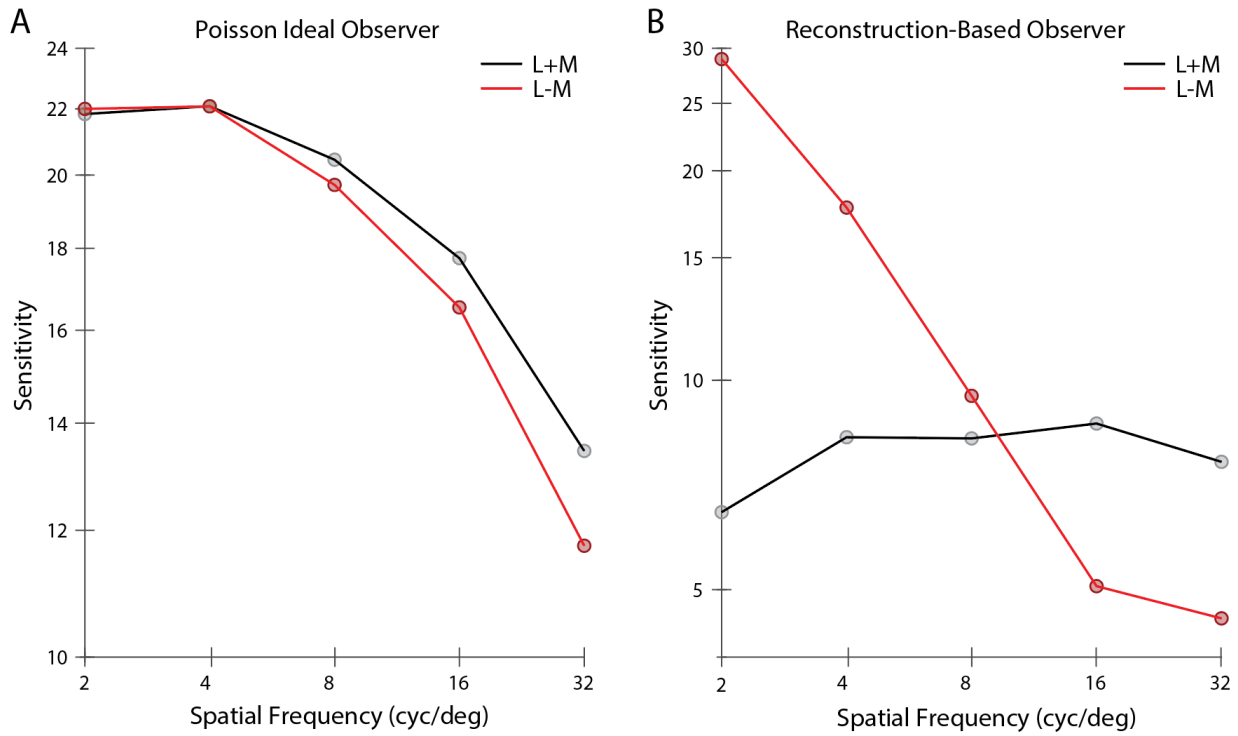

Figure 10: Contrast Sensitivity Function. Contrast sensitivity, defined as the inverse of threshold contrast, for **(A)** a Poisson 2AFC ideal observer, and **(B)** an image reconstruction-based observer (see *Methods*), as a function of the spatial frequency of stimulus in either L+M direction (black) and L-M direction (red) with equal RMS cone contrast.

Discussion:

We developed a Bayesian image reconstruction framework for characterizing early vision, by combining an accurate image-computable forward model of the early visual system, together with the sparse coding model of natural image statistics. Our method enabled both quantification and visualization of information loss due to various factors in early vision, which allowed us to tackle a wide range of questions:

First, we cast retinal mosaic design as a “likelihood design” problem. We found that the large natural variations of L- and M-cone proportion, and the relatively stable but small S-cone proportion, can both be explained as an optimal design that minimizes the expected image reconstruction loss. This is closely related to an alternative formalism, often termed “efficient coding”, which seeks to maximize the amount of information transmission (Barlow 1961; Karklin and Simoncelli 2011; Wei and Stocker 2015; Sims 2018). In both cases, the optimization problem is subject to realistic biological constraints and incorporates natural scene statistics. In fact, previous work (Garrigan et al. 2010) has conducted a similar analysis using an approximation to the information maximization criterion. One main advantage of our work is that we are able to fully simulate a 1-deg mosaic with naturalistic input, as supposed to information-theoretical measures which become intractable as the size of the mosaic and the dimensionality of

the input increase. The fact that the two theories corroborate each other well provides further support for the validity of the result.

Our analysis has implications for understanding retinal mosaic characteristics of different animals. Adult zebrafish, for example, feature a highly regular mosaic with fixed 2:2:1:1 R:G:B:U cone ratios (Engström 1960). Since our analysis has highlighted the importance of prior statistics in determining the optimal design, one might speculate if this regularity results from the particular visual world of zebrafish (i.e., underwater, low signal-to-noise ratio), which demands a more balanced ratio of different cone types to achieve the maximum amount of information transmission. Further study that characterizes in detail the natural scene statistics of zebrafish's environment could help us to better understand this question (Zimmermann et al. 2018; Cai et al. 2020).

Secondly, we applied our framework to cone excitations of retinal mosaics with varying degrees of optical quality, photoreceptor size, density, and cone spectral sensitivity. The reconstructed images reflect accurately the information loss in early vision, including spatial blur due to optical aberration and mosaic sampling, pixel noise due to Poisson variability in the cone excitations, and reduction of chromatic contrast in anomalous trichromacy. Note that although we have mainly focused on the visualization aspect of these effects in our current paper, it is straightforward to perform quantitative analyses by applying appropriate loss functions to the original and reconstructed image as these factors vary. In fact, our reconstruction algorithm could provide a natural "front-end" extension to many image-based perceptual quality metrics, such as spatial CIELAB (Zhang and Wandell 1997), structural similarity (Wang et al. 2004), low-level feature similarity (FSIM; Zhang et al. 2011), or neural network-based approaches (Bosse et al. 2018). Doing so would incorporate early vision factors explicitly into the resulting image quality metrics.

In addition, when SNR is high, we found that we are able to fully recover color information even from an anomalous trichromatic mosaic. As SNR drops, this becomes less feasible. Although our analysis does underestimate the amount of noise in the visual system (i.e., we only consider noise at cone excitations, but see Ala-Laurila et al. 2011 for a detailed treatment of noise in the retinal circuit), this nonetheless suggests that a downstream circuit that properly compensates for the shift in cone spectral sensitivity can, in principle, maintain relatively normal color perception in the low noise regime (Tregillus et al. 2021). This may potentially be related to some reports of less than expected difference in color perception between anomalous trichromats and color normal observers (Bosten 2019; Lindsey, Hutchinson, and Brown 2020).

Thirdly, we speculate that image reconstruction could be a reasonable proxy for modeling percepts in various psychophysical experiments. We found that images reconstructed from dichromatic mosaics resemble results generated by previously proposed methods for visualizing dichromacy, including one that uses explicit knowledge of dichromatic subjects' color appearance reports (Brettel et al. 1997). We have also reproduced the "zebra stripes" and "orientation reversal" aliasing patterns when reconstructing images from cone responses to spatial frequencies above the

mosaic Nyquist limit, similar to what has been documented experimentally in human subjects (Williams 1985; Coletta and Williams 1987); In a similar vein, previous work has used a simpler image reconstruction method to model the color appearance of small spots light stimulus at the resolution of a single cone (Brainard, Williams, and Hofer 2008). We propose to apply our method to a wider range of adaptive optics experiments (e.g., Schmidt et al. 2019; Neitz et al. 2020) designed to probe the early visual system, to fully understand the extent to which image reconstruction can capture subjects' perceptual behavior in these contexts.

Lastly, we have shown that our method can be used in conjunction with analysis of psychophysical discrimination performance, bringing to this analysis the role of statistical regularities of natural images. In our initial exploration, we found that the image-reconstruction based observer exhibits significant interaction between spatial frequency and chromatic direction in its contrast sensitivity function, a behavior distinct from its Poisson ideal observer counterpart, and is more similar to human observer. Future computations will be needed to incorporate realistic simulations of human optics and more complex decision rules to ask whether the reconstruction approach can account for other features of the CSFs of human subjects that are not readily explained by Poisson-limited ideal-observer calculations.

Since our framework is centered on image reconstruction, one may naturally wonder whether we should have applied the more "modern" technique of convolutional neural networks (CNNs), which have become the gold standard for image processing-related tasks (Krizhevsky, Sutskever, and Hinton 2012). We do not doubt that there will likely be a performance gain with CNN, due to its capacity to implicitly represent more complex natural image prior through innate bias and learning (Ulyanov, Vedaldi, and Lempitsky 2018; Kadkhodaie and Simoncelli 2021). However, we believe for our scientific purposes, the Bayesian framework offers an important advantage in its *modularity*, namely, the likelihood and prior are two separate components that can be built independently. This allows us to easily isolate and manipulate one of them (e.g., likelihood) while holding the other constant (e.g., prior), which has been manifested throughout our current paper. In addition, building the likelihood function (i.e., render matrix R, see *Methods*) is a forward process that is computationally very efficient. Performing a similar analysis with the neural network approach (or supervised learning in general) will require re-training of the network with a newly generated dataset (i.e., cone excitation paired with its corresponding images) for *every* single condition in our analysis. Not only is this computationally prohibitive, but it also introduces additional noise due to the training (optimization) process.

The modularity of the Bayesian approach provides a path for extension of the current work. We can employ image priors that capture more of the structure of natural images than the one we used here, for example a Gaussian scale mixture prior (Portilla et al. 2003). Recent work (Kadkhodaie and Simoncelli 2021) develops a method to extract the implicit prior in a denoising deep neural network (an approximation of the gradient of the log-prior, to be precise), allowing this prior to be transferred and combined with different likelihoods. We think this represents a promising direction, and in future work plan to

incorporate more sophisticated priors into our work, to evaluate the robustness of our conclusions to variations in the image prior. In addition to incorporating other priors, we can also apply our method to a likelihood function that represents the computation further along the visual pathways. In this regard, we are particularly eager to incorporate realistic models of retinal ganglion cells into the ISETBio pipeline.

To conclude, we believe our method should be widely applicable to many experiments (e.g., adaptive optics psychophysics) designed for studying early vision, model the effect of changes of various components in early vision (e.g., in clinical conditions), and practical applications (e.g., perceptual quality metric) in which early vision plays an important role.

Methods:

The problem of reconstructing images from neural signals can be considered in the general framework of estimating a signal x, given an (often lower-dimensional and noisy) measurement m. We take a Bayesian approach. Specifically, we model the generative process of measurement as the conditional probability $p(m|x)$ and the prior distribution of the signal as the probability density $p(x)$. We then take the estimate of the signal, $\hat{x}$, as the maximum a posteriori estimate $argmax\ \ p(m|\hat{x})p(\hat{x})$. We next explain in detail how each part of the Bayesian estimate is constructed.

Likelihood Function:
In our particular problem, x is a column vector containing the (vectorized) RGB pixel values of an input image of dimension $N * N * 3$, where N is the linear pixel size of the display. Below we will generalize from RGB images to hyperspectral images. The column vector m contains the excitations of the M cone photoreceptors. The relationship between x and m is modeled by the ISETBio software (Cottaris et al. 2019, 2020; **Fig. 1**). ISETBio simulates in detail the process of displaying an image on the monitor, the wavelength-dependent optical blur of the human eye and spectral transmission through the lens and the macular pigment, as well as the interleaved sampling of the retinal image by the L, M and S cone mosaic. For the majority of simulations presented in our paper, we simulate a 1-deg foveal retina mosaic, which contains approximately 11,000 cone photoreceptors (Curcio et al. 1990). We use a wavelength-dependent point spread function empirically measured in human subjects (Marimont and Wandell 1994), with a pupil size of 3-mm, and an integration time of 50 ms. The input images of size $128 * 128 * 3$ are displayed on a simulated typical CRT monitor (simulated with a 12 bit-depth in each of the RGB channels to avoid quantization artifacts).

Once the RGB pixel values in the original image are linearized, all the processes involved in the relation between x and m, including image formation by the optics of the eye and the relation between retinal irradiance and cone excitations, are well described as linear operations. Furthermore, the instance-to-instance variability in cone excitations is described by a Poisson process acting independently in each cone. Thus $p(m|x)$ is the product of Poisson probability mass functions, one for each cone, with the Poisson mean parameter λ_i for each cone determined by a linear transformation of the input image x. We describe the linear transformation between x and the vector of Poisson mean parameters λ by a matrix R, and thus obtain:

$$p(m|x) = \prod_{i=1}^{M}\ Poisson(m_i \mid \lambda_i = [Rx]_i).$$

We refer to the matrix R as the *render matrix*. This matrix together with the Poisson variability encapsulates the properties of early vision through to the level of the cone excitations. In cases where we parameterize properties of early vision (parameters denoted by θ in the text above), the render matrix is a function of these parameters.

Although ISETBio can compute the relation between the linearized RGB image values at each pixel and the mean excitation of each cone, it does so in a general way that

does not exploit the linearity of the relation. To speed the computations, we use ISETBio to precompute R. Each column of R is a vector of mean cone excitations r_j to a basis image x_j with one entry set to one and the remaining entry set to zero. To determine R, we use ISETBio to compute explicitly each of its columns r_j. We verified that calculating mean cone excitations from an image via Rx yields the same result as applying the ISETBio pipeline directly to the image.

See *Code and Data Availability* for parameters used in the simulation including display specifications (i.e., RGB channel spectra, display gamma function) and cone mosaic setup (i.e., cone spectral sensitivities, lens pigment and macular pigment density and absorption spectra), as well as some of the pre-computed render matrices.

Null Space of Render Matrix:
To understand the information lost between an original RGB image and the mean cone excitations, we can take advantage of the linearity property of the render matrix. Variations in the image space that are within the null space of the (low-rank) render matrix R will have no effect on the likelihood. That is, the cone excitation pattern provides no information to disambiguate between image variants that differ only by vectors within the null space of R. To obtain the null space of R, we used MATLAB function *null*, which computes the singular value decomposition of R. The set of right singular vectors whose associated singular values are 0 form a basis for the null space.

As an illustration, we generate random samples of images from the null space by taking linear combinations of its orthonormal basis vectors, where the weights are sampled independently from a Gaussian distribution with a mean of 0 and a standard deviation of 0.3. As shown in **Fig. 3D**, altering an image by adding to it samples from the null space has no effect on the likelihood.

Prior Distribution:
We also need to specify a prior distribution $p(x)$. The problem of developing statistical models of natural images has been studied extensively using numerous approaches, and remains challenging (Simoncelli 2005). The high-dimensionality and complex structure of natural images makes it difficult to determine a high-dimensional joint distribution that properly captures the various forms of correlation and higher-order dependencies of natural images. Here, we have implemented two relatively simple forms of $p(x)$. We hope to incorporate more complex models of natural image statistics in the future (see *Discussion*).

We will first introduce a simple Gaussian prior $p(x)$ to set up the basic concepts and notations for image prior based on basis vectors. In particular, for the Gaussian prior, we assume $p(x) = \mathcal{N}(x|\mu, \Sigma)$. For convenience, we zero-centered our images when building priors, making $\mu = 0$. The actual mean value of each pixel is added back to each image when computing the likelihood and at the end of the reconstruction procedure. The covariance matrix Σ can be estimated empirically, from a large dataset of natural images. Note that we can write the covariance matrix as its eigen-decomposition: $\Sigma = Q\Lambda Q^{-1}$. Defining $\beta = \Lambda^{-1/2}Q^{-1}x$, we have:

$$p(\beta) = \mathcal{N}(\beta \mid 0, I).$$

This derivation provides a convenient way of expressing our image prior: We can project images onto an appropriate set of basis vectors, and impose a prior distribution on the projected coefficients. In the case above, if we choose the basis vectors as the column vectors of $\Lambda^{-1/2}Q^{-1}$, we obtain an image prior by assuming that the entries of β are each independently distributed as a univariant standard Gaussian (Simoncelli 2005). Such a Gaussian prior can describe the first and second order statistics of natural images, but fails to capture important higher-order structure (Portilla et al. 2003).

Our second model of $p(x)$ emerges from the basis set formulation. Rather than choosing the basis vectors from the eigen-decomposition as above and using a Gaussian distribution over the weights β, we instead choose an over-complete set of basis vectors using independent components analysis, and model the distribution of the entries of weight vector β using the long-tailed distribution Laplace distribution. This leads to a sparse coding model of natural images (Olshausen and Field 1996; Simoncelli and Olshausen 2001). More specifically, we learned a set of K ($K \geq 3N^2$) basis vectors that lead to a sparse representation of our image dataset, through the reconstruction independent component analysis (RICA) algorithm (Le et al. 2011) applied to whitened images, and took these as the columns of the basis matrix E. Our image prior in this case can be written as $p(\beta)$, with $\beta = E^{+}x$. Here E^{+} represents the pseudoinverse of matrix E, and

$$p(\beta) = \prod_{k=1}^{K} \frac{1}{2b}\, exp(-|\beta_k|/b).$$

Note that we further scaled each column of E to equalize the variance across β_k's.

Both methods outlined above can be applied directly to small image patches. They are computationally intractable for larger images, however, since the calculation of basis vectors will involve either an eigen-decomposition of a large covariance matrix, or independent component analysis of a set of high-dimensional image vectors. To address this limitation, we iteratively apply the prior distributions we have constructed above to overlapping small patches of the same size within a large image (Guleryuz 2006).

To illustrate the idea, consider the following example: Assume we have constructed a prior distribution $p(y)$, for small image patches y of size $N_{patch} * N_{patch}$. To model a larger image x of size $pN_{patch} * pN_{patch}$, we could consider viewing x as composed of $p * p$ independent patches of non-overlapping $y's$. Under this assumption, the prior on x could be expressed as the product:

$$p(x) \propto \prod_{j=1}^{p*p} p(y_j),$$

where $y_j's$ describe individual patches of size $N_{patch} * N_{patch}$ within x. The independence assumption is problematic, however, since y_j's are far from independently sampled natural images: they need to be combined into a single coherent

large image. Using this approach to approximate a prior would create block artifacts at the patch boundaries.

The basic idea above, however, can be extended heuristically to solve the block artifact problem by allowing y_j's to overlap with each other. The degree of overlap can be viewed as an additional parameter of the prior, which we refer to here as the stride. Again, for example, consider a large image x of size $pN_{patch} * pN_{patch}$. A stride of 1 will tile through all $(pN_{patch} - N_{patch} + 1) * (pN_{patch} - N_{patch} + 1)$ possible patches of size $N_{patch} * N_{patch}$ within x, yielding a prior distribution of the form:

$$p(x) \propto \prod_{j=1}^{(pN_{patch}-N_{patch}+1)*(pN_{patch}-N_{patch}+1)} p(y_j).$$

Although this form of prior is still an approximation, we have found it to work well in practice, and using it does not lead to visible block artifacts as long as the stride parameter is sufficiently smaller than N_{patch}.

Maximum a Posteriori Estimation:
To reconstruct the image, $\hat{x}$, we find the maximum a posteriori estimate: $\hat{x} = argmax\ \ p(m|\hat{x})p(\hat{x})$. In practice, this optimization is usually expressed in terms of its logarithmic counterpart: $\hat{x} = argmax\ \ [log\ p(m|\hat{x}) + log\ p(\hat{x})]$.

We added an additional parameter γ, which leads to the final formulation of our reconstruction objective:

$$\hat{x} = argmax\ \ [log\ p(m|\hat{x}) + \gamma * log\ p(\hat{x})].$$

For the Poisson likelihood and sparse coding prior, the equation above becomes:

$$\hat{x} = argmax\ \ \left[\sum_{i=1}^{M}\left(-\lambda_i + m_i * log(\lambda_i)\right) + \gamma * \sum_J \sum_K \left|\beta_{jk}\right| + c\right],$$

where $\lambda = R\hat{x}$, $\beta_j = E^+ y_j$, $y_j's$ are individual patches of size $N_{patch} * N_{patch}$ within x. Each β_j is of length K and there are a total of J (overlapping) patches. Lastly, c is a constant that does not depend on $\hat{x}$.

We added the γ parameter due to the approximate nature of our prior, introduced especially by the aggregation over patches. Adding this parameter also provides some level of robustness against misspecification of the prior more generally. For most of the reconstruction results presented in this paper, the value of γ was determined by maximizing reconstruction performance with a cross-validation procedure (see **Fig. 2**). We also found that the optimal γ values are similar regardless of the choice of loss function. Note that the additional flexibility provided by this γ parameter also provides us with a parametric way to manipulate and isolate the relative contribution of the log-likelihood and log-prior terms to the reconstruction (**Fig. 2**).

The optimization problem required to obtain $\hat{x}$ can be solved efficiently using the MATLAB function *fmincon* by providing the analytical gradient to the minimization function:

$$\frac{\partial \log p(m|x)}{\partial x} = \left(-1 + m \circ \frac{1}{\lambda}\right)^T * R,$$
$$\frac{\partial \log p(y)}{\partial y} = sign(\beta)^T * E^+.$$

where $\lambda = Rx, \beta = E^+ y$, $\circ$ denotes element-wise product between two vectors, $\frac{1}{\lambda}$ is the element-wise inverse of vector λ, and

$$sign(\beta_i) = \begin{cases} -1, & \beta_i < 0 \\ 1, & \beta_i > 0 \end{cases}.$$

RGB Image Dataset:
We used the ImageNet ILSVRC (Russakovsky et al. 2015) as our dataset for natural RGB images. Fifty randomly sampled images were reserved as the evaluation set, and the rest of the images were used for learning the prior and for cross-validation. For the sparse prior, we constructed a basis set size of $K = 768$, on image patches of size $16 * 16$ sampled from the training set, and used a stride of 4 when tiling larger images. We randomly sampled 20 patches from each one of the $5{,}000$ images in the training set for learning the prior (ICA analysis), and 500 images for the cross-validation procedure to determine the γ parameter.

In our work, we simulate display of the RGB images on idealized monitor to generate spectral radiance as a linear combination of the monitor's RGB channel spectra. Thus, a prior over the linear RGB pixels values will induce an implicit spatial-spectral prior. To make sure the constraints introduced by RGB images together with the monitor do not influence our results, we also conducted a control analysis using hyperspectral images directly as described in the following section.

Hyperspectral Images:
As a control analyses, we developed priors and reconstructed images directly on small patches of hyperspectral images. The development is essentially the same as above, with the generalization being to increase the number of channels in the images from 3 to N. In addition, since our algorithm treats images as high-dimensional vectors, it can be directly applied to reconstruct hyperspectral images. Here, we used images from Nascimento et al. (2002) and Chakrabarti and Zickler (2011). The dataset of Nascimento et al. (2002) was pre-processed following the instructions provided by the authors, and the images of Chakrabarti and Zickler (2011) were converted to spectral radiance using the camera calibration data provided in that work. We further resampled the combined image dataset with a patch size of $18 * 18$ and 15 uniformly spaced wavelengths between 420 nm and 700 nm for a dataset of ~ 5000 patches. We retained 300 of them as the evaluation set, and the rest for prior learning and cross-validation. The remaining of the analysis (i.e., prior and reconstruction algorithm) followed the same procedures as those used for the RGB images.

See *Code and Data Availability* for the curated RGB and hyperspectral image dataset, and a learned set of basis function for the sparse prior.

Gaussian Prior for Synthetic Images:
We also reconstructed multivariate Gaussian distributed synthetic images with known chromatic and spatial correlations that we can explicitly manipulate (**Fig. 5**). To construct these signals $\mathrm{x} \sim \mathcal{N}(\mu,\ \Sigma)$, where x is RGB image of size $N * N * 3$ ($N = 36$ in our current analysis), we set $\mu = 0.5$, and used a separable Σ along its two spatial and one chromatic dimensions. That is:

$$\Sigma = \Sigma_c \otimes \Sigma_s \otimes \Sigma_s,$$

where Σ_c is the chromatic covariance matrix of size $(3 * 3)$

$$\Sigma_c(i, j) = \sigma_c^2 * \rho_c^{|i-j|},$$

and Σ_s is the spatial covariance matrix of size $(N * N)$

$$\Sigma_s(i, j) = \sigma_s^2 * \rho_s^{|i-j|}.$$

In the covariance matrix constructions, i, j index into entries of Σ_c and Σ_s at i-th row and j-th column. Here $\otimes$ represents the Kronecker product, thus producing the signal covariance matrix Σ of size $(3N^2 * 3N^2)$ (Brainard et al. 2008; Manning and Brainard 2009).

The parameters σ_c^2 and σ_s^2 determine the overall variance of the signal, which are fixed across all simulations, whereas by changing the value of ρ_c and ρ_s, we manipulate the degree of spatial and chromatic correlation presented in the synthetic images (**Fig. 5**).

We introduce an additional simplification for the case of reconstructions with respect to the synthetic Gaussian prior: We approximated the Poisson likelihood function with a Gaussian distribution with fixed variance. Thus, the reconstruction problem can be written as:

$$p(\beta) = \mathcal{N}(\beta \mid 0,\ I),$$
$$p(m|\beta) = \mathcal{N}\big(m;\ RQ\Lambda^{1/2}\beta,\ \sigma^2 I\big).$$

Where R is the render matrix, and $\Sigma = Q\Lambda Q^{-1}$.

The reconstruction problem with Gaussian prior and Gaussian noise matches the ridge regression formulation, and can be solved analytically by the regularized normal equations. Denote the design matrix $D = RQ\Lambda^{1/2}$:

$$\hat{\beta} = (D^T D + \gamma I)^{-1} D^T m,$$
$$\hat{x} = Q\Lambda^{1/2}\hat{\beta}.$$

Note that the γ parameter here is also determined through a cross-validation routine. We adopted this simplification for the simulation results in **Fig. 5**, in order to make it computationally feasible to evaluate the average reconstruction error across a large number of synthetic image datasets.

Variations in Retinal Cone Mosaic:

To simulate a dichromatic observer, we constructed retinal mosaics with only two classes of cones but with similar spatial configuration. To simulate the deuteranomalous observer, we shifted the M cone spectral sensitivity function, setting its peak at 550 nm instead of the typical 530 nm. In both cases, the likelihood function (i.e., render matrix R) was computed using the procedure described above and the same Bayesian algorithm was applied to obtain the reconstructed images.

To simulate retinal mosaics at different eccentricities, we constructed retinal mosaics with the appropriate photoreceptor size, density (Curcio et al. 1990), and physiological optics (Polans et al. 2015), and computed their corresponding render matrices. The same Bayesian algorithm was applied to obtain the reconstructed images.

To simulate the interferometric experimental conditions of Williams (1985), we used diffraction-limited optics without longitudinal chromatic aberration (LCA) for the computation of the cone excitations, but used the likelihood function with normal optics for the reconstruction. This models subjects whose perceptual systems are matched to their normal optics and assumes there is no substantial adaptation within the short time span of the experiment.

Contrast Sensitivity Function:
We compared the spatial Contrast Sensitivity Function (CSF) between a standard, Poisson 2AFC ideal observer, and an image reconstruction-based observer.

We simulated stimuli modulations in two chromatic directions, L+M and L-M, with an equal root mean square (RMS) contrast (i.e., vector length in the L and M cone contrast plane) of maximum value 0.1, at 5 spatial frequencies, $[2, 4, 8, 16, 32]$ cycles per degree. For each chromatic direction/spatial frequency combination, the sensitivity is defined as the inverse of threshold contrast.

We used the QUEST+ procedure (Watson 2017) as implemented in MATLAB by Brainard (mQUESTPlus; https://github.com/BrainardLab/mQUESTPlus) for estimating the simulated threshold efficiently as follows: We initialized the procedure with the contrast in the middle of a pre-defined possible stimulus range. For each contrast, we first generated a null template N, which is the noise-free, average responses of a 0.5 deg foveal mosaic with K number of cones to a uniform background stimulus; and a target template T, which is the noise-free, average cone responses to a grating stimulus at that contrast level. We then simulated 128 individual trials at this contrast. For each trial, two Poisson-noise corrupted cone responses r_1 and r_2 are generated based on N and T, respectively. The simulated observer then identity either r_1 or r_2 as the target based on its decision rule. Based on the observer responses, the QUEST+ procedure chooses the next test contrast according to an information-maximization criterion (Watson 2017). The process is repeated 15 times, for a total of $15 \times 128 = 1920$ trials.

For the Poisson 2AFC observer, we directly compute the likelihood ratio:

$$r = \frac{Poisson(r_1|T)Poisson(r_2|N)}{Poisson(r_2|T)Poisson(r_1|N)}$$

Taking the logarithm of the equation above, the decision rule simplifies to the following:

$$d = \sum_{i=1}^{K}\{(r_1 \circ \log \mathrm{T} + \mathrm{r}_2 \circ \log \mathrm{N}) - (\mathrm{r}_2 \circ \log \mathrm{T} + \mathrm{r}_1 \circ \log \mathrm{N})\}_{\mathrm{i}}$$

where $\circ$ denotes element-wise product between two vectors. The observer chooses r_1 as target when $d > 0$, and r_2 when $d < 0$.

For the image reconstruction-based observer, given the cone responses, it first applies the reconstruction algorithm to obtain the image template $\hat{T}$ and $\hat{N}$ from T and N, and also noisy image instances $\hat{r}_1$ and $\hat{r}_2$ by applying the same algorithm to r_1 and r_2. We then perform a template-matching decision rule as follows:

$$d = \sqrt{||\hat{\mathrm{r}}_1 - \hat{T}||_2^2 + ||\hat{r}_2 - \hat{N}||_2^2} - \sqrt{||\hat{\mathrm{r}}_2 - \hat{T}||_2^2 + ||\hat{r}_1 - \hat{N}||_2^2}$$

where $||\cdot||_2$ represents the L_2 norm of a vector. The observer chooses r_1 as target when $d < 0$, and r_2 when $d > 0$. We choose the template matching procedure for computational convenience. Because the variability in the reconstructed images is not independent across pixels, this procedure is not ideal. Other decision rules will be explored in future work.

Code and Data Availability:
The MATLAB code used for this paper is available at:
https://github.com/isetbio/ISETImagePipeline

In addition, the curated RGB and hyperspectral image datasets, parameters used in the simulation including display and cone mosaic setup, as well as the intermediate results such as the learned sparse prior, likelihood function (i.e., render matrix), are available through: https://tinyurl.com/26r92c8y

References:

Ala-Laurila, Petri, Martin Greschner, E. J. Chichilnisky, and Fred Rieke. 2011. "Cone Photoreceptor Contributions to Noise and Correlations in the Retinal Output." *Nature Neuroscience* 14(10):1309–16. doi: 10.1038/nn.2927.

Anderson, S. J., K. T. Mullen, and R. F. Hess. 1991. "Human Peripheral Spatial Resolution for Achromatic and Chromatic Stimuli: Limits Imposed by Optical and Retinal Factors." *The Journal of Physiology* 442:47–64. doi: 10.1113/jphysiol.1991.sp018781.

Attneave, Fred. 1954. "Some Informational Aspects of Visual Perception." *Psychological Review* 61(3):183.

Banks, Martin S., Wilson S. Geisler, and Patrick J. Bennett. 1987. "The Physical Limits of Grating Visibility." *Vision Research* 27(11):1915–24.

Barlow, Horace B. 1961. "Possible Principles Underlying the Transformation of Sensory Messages." *Sensory Communication* 1(01).

Borghuis, Bart G., Charles P. Ratliff, Robert G. Smith, Peter Sterling, and Vijay Balasubramanian. 2008. "Design of a Neuronal Array." *Journal of Neuroscience* 28(12):3178–89. doi: 10.1523/JNEUROSCI.5259-07.2008.

Bosse, S., D. Maniry, K. Müller, T. Wiegand, and W. Samek. 2018. "Deep Neural Networks for No-Reference and Full-Reference Image Quality Assessment." *IEEE Transactions on Image Processing* 27(1):206–19. doi: 10.1109/TIP.2017.2760518.

Bosten, Jenny. 2019. "The Known Unknowns of Anomalous Trichromacy." *Current Opinion in Behavioral Sciences* 30:228–37.

Bracewell, Ronald Newbold. 1986. *The Fourier Transform and Its Applications*. Vol. 31999. McGraw-Hill New York.

Brainard, David H. 2015. "Color and the Cone Mosaic." *Annual Review of Vision Science* 1:519–46. doi: 10.1146/annurev-vision-082114-035341.

Brainard, David H. 2019. "Color, Pattern, and the Retinal Cone Mosaic." *Current Opinion in Behavioral Sciences* 30:41–47. doi: 10.1016/j.cobeha.2019.05.005.

Brainard, David H., David R. Williams, and Heidi Hofer. 2008. "Trichromatic Reconstruction from the Interleaved Cone Mosaic: Bayesian Model and the Color Appearance of Small Spots." *Journal of Vision* 8(5):15–15. doi: 10.1167/8.5.15.

Brettel, Hans, Françoise Viénot, and John D. Mollon. 1997. "Computerized Simulation of Color Appearance for Dichromats." *JOSA A* 14(10):2647–55. doi: 10.1364/JOSAA.14.002647.

Burge, Johannes. 2020. "Image-Computable Ideal Observers for Tasks with Natural Stimuli." *Annual Review of Vision Science* 6:491–517.

Burge, Johannes, and Wilson S. Geisler. 2011. “Optimal Defocus Estimation in Individual Natural Images.” *Proceedings of the National Academy of Sciences* 108(40):16849–54.

Burge, Johannes, and Wilson S. Geisler. 2014. “Optimal Disparity Estimation in Natural Stereo Images.” *Journal of Vision* 14(2):1–1. doi: 10.1167/14.2.1.

Cai, Lanya T., Venkatesh Krishna, Tim C. Hladnik, Nicholas C. Guilbeault, Scott A. Juntti, Tod R. Thiele, Aristides B. Arrenberg, and Emily A. Cooper. 2020. “Visual Statistics of Aquatic Environments in the Natural Habitats of Zebrafish.” *Journal of Vision* 20(11):433–433. doi: 10.1167/jov.20.11.433.

Chakrabarti, Ayan, and Todd Zickler. 2011. “Statistics of Real-World Hyperspectral Images.” Pp. 193–200 in *CVPR 2011*. IEEE.

Chaparro, A., C. F. Stromeyer, E. P. Huang, R. E. Kronauer, and R. T. Eskew. 1993. “Colour Is What the Eye Sees Best.” *Nature* 361(6410):348–50. doi: 10.1038/361348a0.

Chin, Benjamin M., and Johannes Burge. 2020. “Predicting the Partition of Behavioral Variability in Speed Perception with Naturalistic Stimuli.” *Journal of Neuroscience* 40(4):864–79.

Coletta, Nancy J., and David R. Williams. 1987. “Psychophysical Estimate of Extrafoveal Cone Spacing.” *JOSA A* 4(8):1503–13.

Cottaris, Nicolas P., Haomiao Jiang, Xiaomao Ding, Brian A. Wandell, and David H. Brainard. 2019. “A Computational-Observer Model of Spatial Contrast Sensitivity: Effects of Wave-Front-Based Optics, Cone-Mosaic Structure, and Inference Engine.” *Journal of Vision* 19(4):8–8.

Cottaris, Nicolas P., Brian A. Wandell, Fred Rieke, and David H. Brainard. 2020. “A Computational Observer Model of Spatial Contrast Sensitivity: Effects of Photocurrent Encoding, Fixational Eye Movements, and Inference Engine.” *Journal of Vision* 20(7):17–17. doi: 10.1167/jov.20.7.17.

Curcio, Christine A., Kenneth R. Sloan, Robert E. Kalina, and Anita E. Hendrickson. 1990. “Human Photoreceptor Topography.” *Journal of Comparative Neurology* 292(4):497–523. doi: https://doi.org/10.1002/cne.902920402.

Davila, Karen D., and Wilson S. Geisler. 1991. “The Relative Contributions of Pre-Neural and Neural Factors to Areal Summation in the Fovea.” *Vision Research* 31(7–8):1369–80.

Engström, Kjell. 1960. “Cone Types and Cone Arrangement in the Retina of Some Cyprinids.” *Acta Zoologica* 41(3):277–95. doi: https://doi.org/10.1111/j.1463-6395.1960.tb00481.x.

Field, David J. 1987. “Relations between the Statistics of Natural Images and the Response Properties of Cortical Cells.” *Josa a* 4(12):2379–94.

Garrigan, Patrick, Charles P. Ratliff, Jennifer M. Klein, Peter Sterling, David H. Brainard, and Vijay Balasubramanian. 2010. “Design of a Trichromatic Cone Array.” *PLOS Computational Biology* 6(2):e1000677. doi: 10.1371/journal.pcbi.1000677.

Geisler, Wilson S. 1989. “Sequential Ideal-Observer Analysis of Visual Discriminations.” *Psychological Review* 96(2):267.

Geisler, Wilson S. 2011. “Contributions of Ideal Observer Theory to Vision Research.” *Vision Research* 51(7):771–81.

Geisler, Wilson S. 2018. “Psychometric Functions of Uncertain Template Matching Observers.” *Journal of Vision* 18(2):1–1.

Guleryuz, O. G. 2006. “Nonlinear Approximation Based Image Recovery Using Adaptive Sparse Reconstructions and Iterated Denoising-Part I: Theory.” *IEEE Transactions on Image Processing* 15(3):539–54. doi: 10.1109/TIP.2005.863057.

Hofer, Heidi, Joseph Carroll, Jay Neitz, Maureen Neitz, and David R. Williams. 2005. “Organization of the Human Trichromatic Cone Mosaic.” *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 25(42):9669–79. doi: 10.1523/JNEUROSCI.2414-05.2005.

Jiang, Haomiao, Joyce Farrell, and Brian Wandell. 2016. “A Spectral Estimation Theory for Color Appearance Matching.” *Electronic Imaging* 2016(20):1–4. doi: 10.2352/ISSN.2470-1173.2016.20.COLOR-329.

Kadkhodaie, Zahra, and Eero P. Simoncelli. 2021. “Solving Linear Inverse Problems Using the Prior Implicit in a Denoiser.” *ArXiv:2007.13640 [Cs, Eess, Stat]*.

Karklin, Yan, and Eero P. Simoncelli. 2011. “Efficient Coding of Natural Images with a Population of Noisy Linear-Nonlinear Neurons.” *Advances in Neural Information Processing Systems* 24:999–1007.

Kim, Seha, and Johannes Burge. 2020. “Natural Scene Statistics Predict How Humans Pool Information across Space in Surface Tilt Estimation.” *PLoS Computational Biology* 16(6):e1007947.

Knill, David C., Daniel Kersten, and Alan Yuille. 1996. “Introduction: A Bayesian Formulation of Visual Perception.” *Perception as Bayesian Inference* 1:1–21.

Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. 2012. “Imagenet Classification with Deep Convolutional Neural Networks.” *Advances in Neural Information Processing Systems* 25:1097–1105.

Land, Michael F., and Dan-Eric Nilsson. 2012. *Animal Eyes*. 2nd edition. Oxford ; New York: Oxford University Press.

Le, Quoc, Alexandre Karpenko, Jiquan Ngiam, and Andrew Ng. 2011. “ICA with Reconstruction Cost for Efficient Overcomplete Feature Learning.” *Advances in Neural Information Processing Systems* 24.

Levin, Anat, William T. Freeman, and Frédo Durand. 2008. “Understanding Camera Trade-Offs through a Bayesian Analysis of Light Field Projections.” Pp. 88–101 in *Computer Vision – ECCV 2008*, *Lecture Notes in Computer Science*, edited by D. Forsyth, P. Torr, and A. Zisserman. Berlin, Heidelberg: Springer.

Lewis, Alex, and Zhaoping Li. 2006. “Are Cone Sensitivities Determined by Natural Color Statistics?” *Journal of Vision* 6(3):285–302. doi: 10.1167/6.3.8.

Lindsey, Delwin, Lindsey Hutchinson, and Angela Brown. 2020. “Unique Yellow and Other Special Colors Seen by Deuteranomalous Trichromats.” *Journal of Vision* 20(11):1249–1249. doi: 10.1167/jov.20.11.1249.

Manning, Jeremy R., and David H. Brainard. 2009. “Optimal Design of Photoreceptor Mosaics: Why We Do Not See Color at Night.” *Visual Neuroscience* 26(1):5.

Marimont, David H., and Brian A. Wandell. 1994. “Matching Color Images: The Effects of Axial Chromatic Aberration.” *JOSA A* 11(12):3113–22. doi: 10.1364/JOSAA.11.003113.

Mullen, K. T. 1985. “The Contrast Sensitivity of Human Colour Vision to Red-Green and Blue-Yellow Chromatic Gratings.” *The Journal of Physiology* 359:381–400.

Nascimento, Sérgio MC, Flávio P. Ferreira, and David H. Foster. 2002. “Statistics of Spatial Cone-Excitation Ratios in Natural Scenes.” *JOSA A* 19(8):1484–90.

Naselaris, Thomas, Ryan J. Prenger, Kendrick N. Kay, Michael Oliver, and Jack L. Gallant. 2009. “Bayesian Reconstruction of Natural Images from Human Brain Activity.” *Neuron* 63(6):902–15.

Neitz, Alexandra, Xiaoyun Jiang, James A. Kuchenbecker, Niklas Domdei, Wolf Harmening, Hongyi Yan, Jihyun Yeonan-Kim, Sara S. Patterson, Maureen Neitz, Jay Neitz, Daniel R. Coates, and Ramkumar Sabesan. 2020. “Effect of Cone Spectral Topography on Chromatic Detection Sensitivity.” *JOSA A* 37(4):A244–54. doi: 10.1364/JOSAA.382384.

Olshausen, Bruno A., and David J. Field. 1996. “Emergence of Simple-Cell Receptive Field Properties by Learning a Sparse Code for Natural Images.” *Nature* 381(6583):607–9.

Parthasarathy, Nikhil, Eleanor Batty, William Falcon, Thomas Rutten, Mohit Rajpal, E. J. Chichilnisky, and Liam Paninski. 2017. “Neural Networks for Efficient Bayesian Decoding of Natural Images from Retinal Neurons.” *Advances in Neural Information Processing Systems* 30.

Pelli, Denis G. 1985. “Uncertainty Explains Many Aspects of Visual Contrast Detection and Discrimination.” *JOSA A* 2(9):1508–32.

Polans, James, Bart Jaeken, Ryan P. McNabb, Pablo Artal, and Joseph A. Izatt. 2015. “Wide-Field Optical Model of the Human Eye with Asymmetrically Tilted and Decentered Lens That Reproduces Measured Ocular Aberrations.” *Optica* 2(2):124–34. doi: 10.1364/OPTICA.2.000124.

Portilla, Javier, Vasily Strela, Martin J. Wainwright, and Eero P. Simoncelli. 2003. “Image Denoising Using Scale Mixtures of Gaussians in the Wavelet Domain.” *IEEE Transactions on Image Processing* 12(11):1338–51.

Russakovsky, Olga, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. 2015. “ImageNet Large Scale Visual Recognition Challenge.” *ArXiv:1409.0575 [Cs]*.

Schmidt, Brian P., Alexandra E. Boehm, William S. Tuten, and Austin Roorda. 2019. “Spatial Summation of Individual Cones in Human Color Vision.” *PLOS ONE* 14(7):e0211397. doi: 10.1371/journal.pone.0211397.

Sekiguchi, N., D. R. Williams, and D. H. Brainard. 1993. “Efficiency in Detection of Isoluminant and Isochromatic Interference Fringes.” *Journal of the Optical Society of America. A, Optics, Image Science, and Vision* 10(10):2118–33. doi: 10.1364/josaa.10.002118.

Shepard, Roger N. 1987. “Toward a Universal Law of Generalization for Psychological Science.” *Science* 237(4820):1317–23.

Simoncelli, Eero P. 2005. “4.7 Statistical Modeling of Photographic Images.” *Handbook of Video and Image Processing* 9.

Simoncelli, Eero P., and Bruno A. Olshausen. 2001. “Natural Image Statistics and Neural Representation.” *Annual Review of Neuroscience* 24(1):1193–1216.

Sims, Chris R. 2018. “Efficient Coding Explains the Universal Law of Generalization in Human Perception.” *Science* 360(6389):652–56. doi: 10.1126/science.aaq1118.

Singh, Vijay, Nicolas P. Cottaris, Benjamin S. Heasly, David H. Brainard, and Johannes Burge. 2018. “Computational Luminance Constancy from Naturalistic Images.” *Journal of Vision* 18(13):19–19. doi: 10.1167/18.13.19.

Snyder, A. W., T. R. Bossomaier, and A. Hughes. 1986. “Optical Image Quality and the Cone Mosaic.” *Science (New York, N.Y.)* 231(4737):499–501. doi: 10.1126/science.3941914.

Snyder, A. W., S. B. Laughlin, and D. G. Stavenga. 1977. “Information Capacity of Eyes.” *Vision Research* 17(10):1163–75. doi: 10.1016/0042-6989(77)90151-1.

Stockman, Andrew, and Lindsay T. Sharpe. 2006. “Into the Twilight Zone: The Complexities of Mesopic Vision and Luminous Efficiency.” *Ophthalmic & Physiological Optics: The Journal of the British College of Ophthalmic Opticians (Optometrists)* 26(3):225–39. doi: 10.1111/j.1475-1313.2006.00325.x.

Tjan, Bosco S., and Gordon E. Legge. 1998. “The Viewpoint Complexity of an Object-Recognition Task.” *Vision Research* 38(15):2335–50. doi: 10.1016/S0042-6989(97)00255-1.

Tregillus, Katherine E. M., Zoey J. Isherwood, John E. Vanston, Stephen A. Engel, Donald I. A. MacLeod, Ichiro Kuriki, and Michael A. Webster. 2021. “Color Compensation in Anomalous Trichromats Assessed with FMRI.” *Current Biology* 31(5):936-942.e4. doi: 10.1016/j.cub.2020.11.039.

Ulyanov, Dmitry, Andrea Vedaldi, and Victor Lempitsky. 2018. “Deep Image Prior.” Pp. 9446–54 in *Proceedings of the IEEE conference on computer vision and pattern recognition*.

Wang, Zhou, Alan C. Bovik, Hamid R. Sheikh, and Eero P. Simoncelli. 2004. “Image Quality Assessment: From Error Visibility to Structural Similarity.” *IEEE Transactions on Image Processing* 13(4):600–612.

Watson, Andrew B. 2017. “QUEST: A General Multidimensional Bayesian Adaptive Psychometric Method.” *Journal of Vision* 17(3):1–27. doi: 10.1167/17.3.10.

Webster, Michael A., and J. D. Mollon. 1997. “Adaptation and the Color Statistics of Natural Images.” *Vision Research* 37(23):3283–98. doi: 10.1016/S0042-6989(97)00125-9.

Wei, Xue-Xin, and Alan A. Stocker. 2015. “A Bayesian Observer Model Constrained by Efficient Coding Can Explain ‘anti-Bayesian’ Percepts.” *Nature Neuroscience* 18(10):1509–17. doi: 10.1038/nn.4105.

Williams, David R. 1985. “Aliasing in Human Foveal Vision.” *Vision Research* 25(2):195–205. doi: 10.1016/0042-6989(85)90113-0.

Williams, David R., Nobutoshi Sekiguchi, William Haake, David Brainard, and Orin Packer. 1991. “The Cost of Trichromacy for Spatial Vision.” Pp. 11–22 in *From Pigments to Perception: Advances in Understanding Visual Processes*, *NATO ASI Series*, edited by A. Valberg and B. B. Lee. Boston, MA: Springer US.

Yellott, J. I. 1983. “Spectral Consequences of Photoreceptor Sampling in the Rhesus Retina.” *Science* 221(4608):382–85. doi: 10.1126/science.6867716.

Zhang, L., L. Zhang, X. Mou, and D. Zhang. 2011. “FSIM: A Feature Similarity Index for Image Quality Assessment.” *IEEE Transactions on Image Processing* 20(8):2378–86. doi: 10.1109/TIP.2011.2109730.

Zhang, Xuemei, and Brian A. Wandell. 1997. “A Spatial Extension of CIELAB for Digital Color-Image Reproduction.” *Journal of the Society for Information Display* 5(1):61–63.

Zimmermann, Maxime J. Y., Noora E. Nevala, Takeshi Yoshimatsu, Daniel Osorio, Dan-Eric Nilsson, Philipp Berens, and Tom Baden. 2018. “Zebrafish Differentially Process Color across Visual Space to Match Natural Scenes.” *Current Biology* 28(13):2018-2032.e5. doi: 10.1016/j.cub.2018.04.075.

Supplementary Figures:

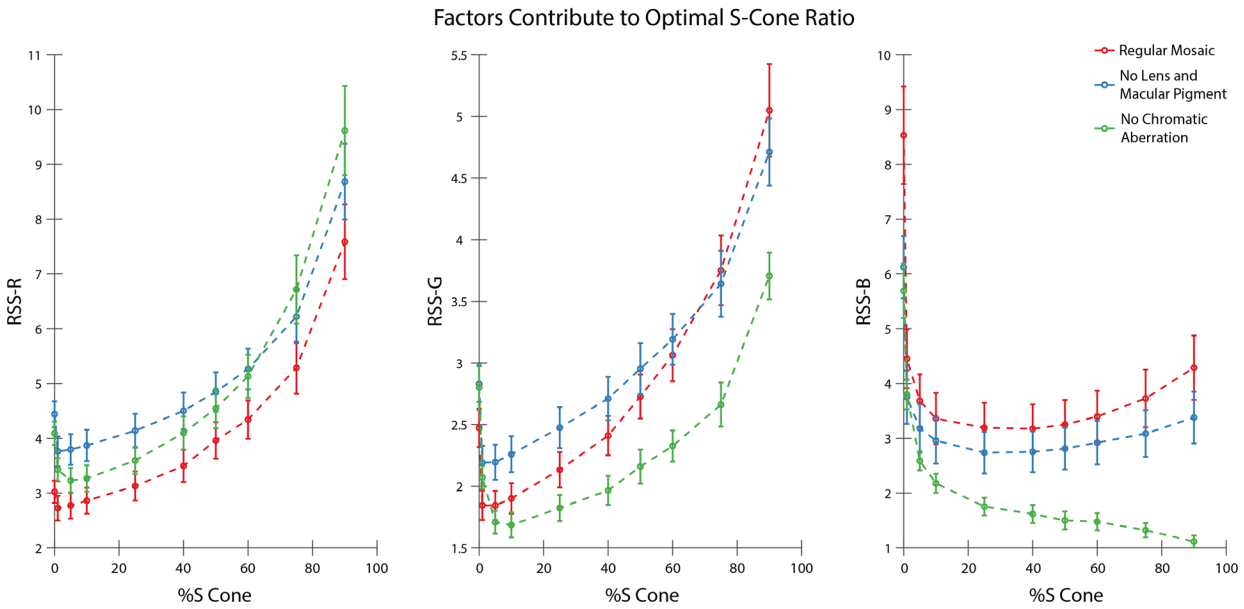

Figure S1: Factors that Contribute to Optimal S Cone Proportion. Average reconstruction error as a function of S-cone proportion, computed as RSS of pixel values for the R- (left), G- (middle), and B-planes (right) respectively. Under typical conditions (red), a low S-cone ratio is optimal for all three planes. Removing lens pigment and macular pigment from the simulations (blue) increases the SNR of the S cones by increasing their average quantum catch, but has little effect on the optimal S-cone proportion for any of the image planes. Correcting chromatic aberration (green) while retaining lens pigment and macular pigment greatly improves the information provided by the S cones for the B-plane, but not for the R- and G- planes. Error bars indicate $\pm$ 1 SEM.

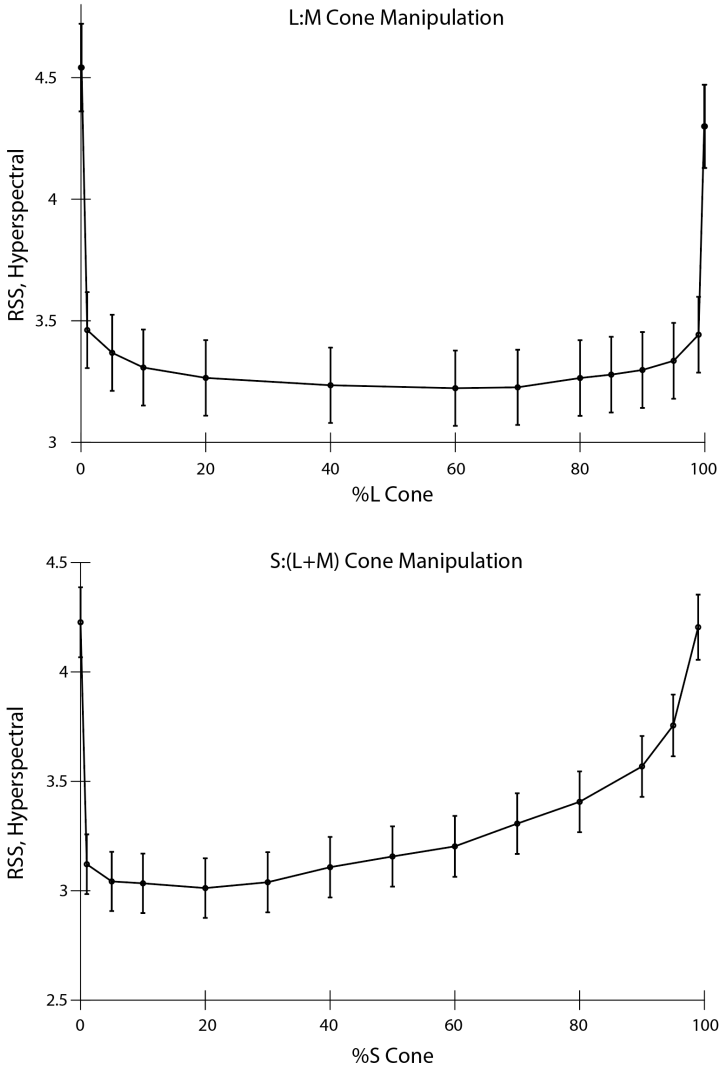

Figure S2: Effect of the Allocation of Retinal Cone Types on Reconstruction of Hyperspectral Images. Average reconstruction error as a function of L-cone proportion (top) and S-cone proportion (bottom), computed as RSS of pixel values over space and wavelength, for a set of evaluation hyperspectral images of size of $18 * 18$ and 15 uniform wavelength sample between 420 nm and 700 nm (see *Methods*). Error bars indicate $\pm$ 1 SEM. The results corroborated our main conclusion obtained with RGB images in **Fig. 4**.

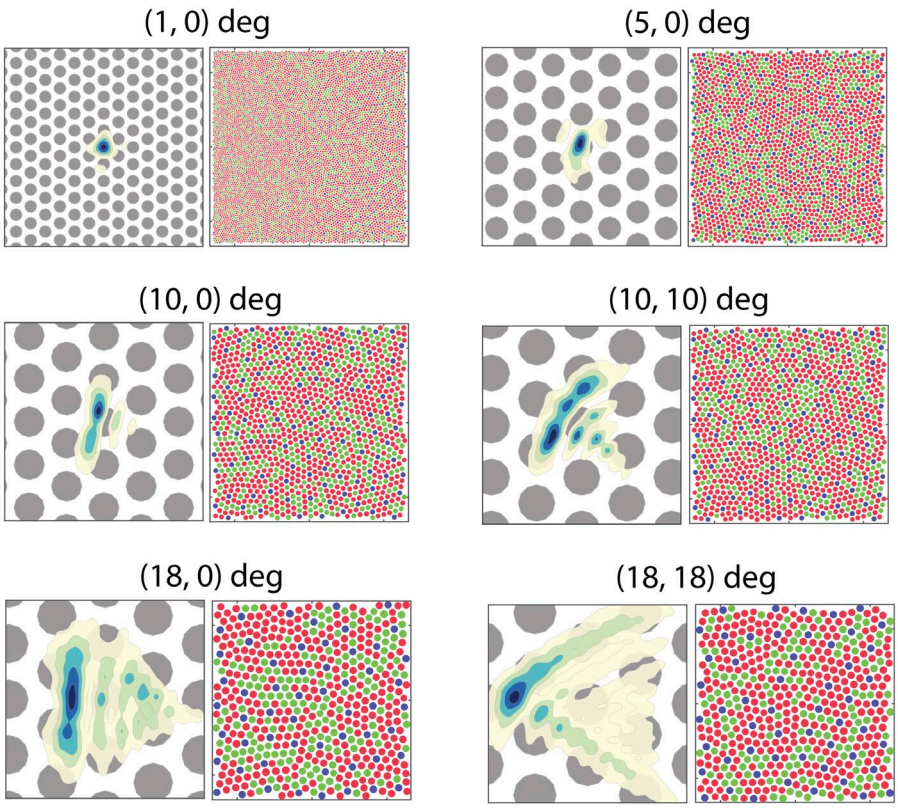

Figure S3: Optics and Cone Mosaic at Different Retinal Eccentricities. Enlarged view of the top panels of **Fig. 8**. The coordinates at the top of each pair indicate the horizontal and vertical eccentricity of the retinal patch. The left image of each pair shows a contour plot of the point-spread function relative to an expanded view of the cone mosaic, while the right image of each pair shows the full 1-degree mosaic used in the simulation.

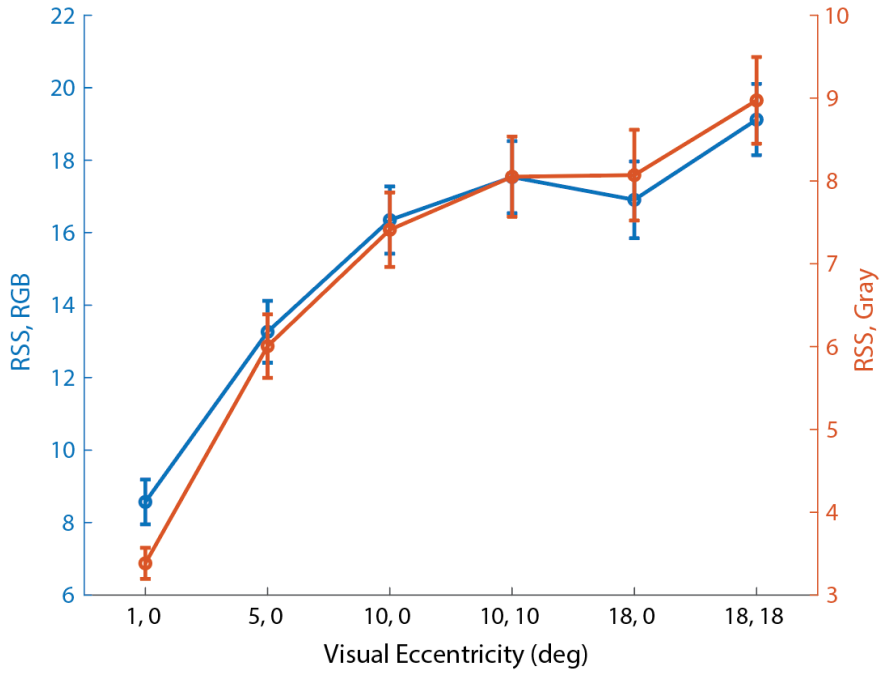

Figure S4: Reconstruction Error at Different Visual Eccentricities. Average image reconstruction error, computed as RSS of pixel values for both the RGB (left y-axis) and corresponding gray scale images (i.e., $R + G + B$, to measure only the spatial reconstruction error, right y-axis), as a function of the visual eccentricity location of a 1-deg retinal mosaic. Error bars indicate +/- 1 SEM.

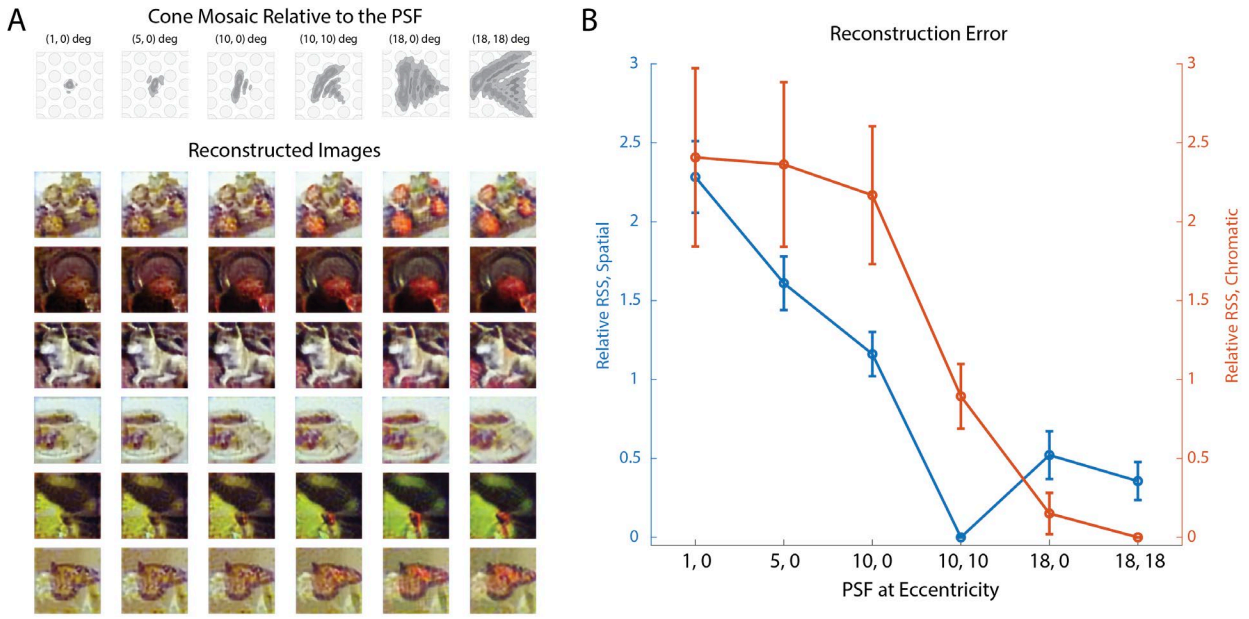

Figure S5: Image Reconstruction with Different Point Spread Functions. (A) Image reconstructions for a set of example images in the evaluation set from 1-degree patches of mosaic at (10, 10) degree eccentricity, but with PSFs sampled from different visual eccentricities as indicated by the top panel. **(B)** The average differential reconstruction error (i.e., difference in RSS compared to the lowest value obtained among the simulations) as a function of the eccentricity of the PSFs used. Error bars represent +/- 1 SEM. To separate the spatial and chromatic error, we perform a PCA analysis on the RGB images. The RSS along the first PC ($0.57R + 0.59G + 0.56B$) corresponds to the spatial error (left axis), while the RSS along the second and third PCs ($0.76R - 0.13G - 0.64B;\ -0.31R + 0.80G - 0.52B$) quantify the chromatic error (right axis). With the range of PSFs in our simulation, the minimal spatial error is obtained with the PSF at (10, 10) deg (i.e., the PSF that matched to the mosaic), and the minimal chromatic error is obtained with the largest PSF, corresponding to (18, 18) deg.

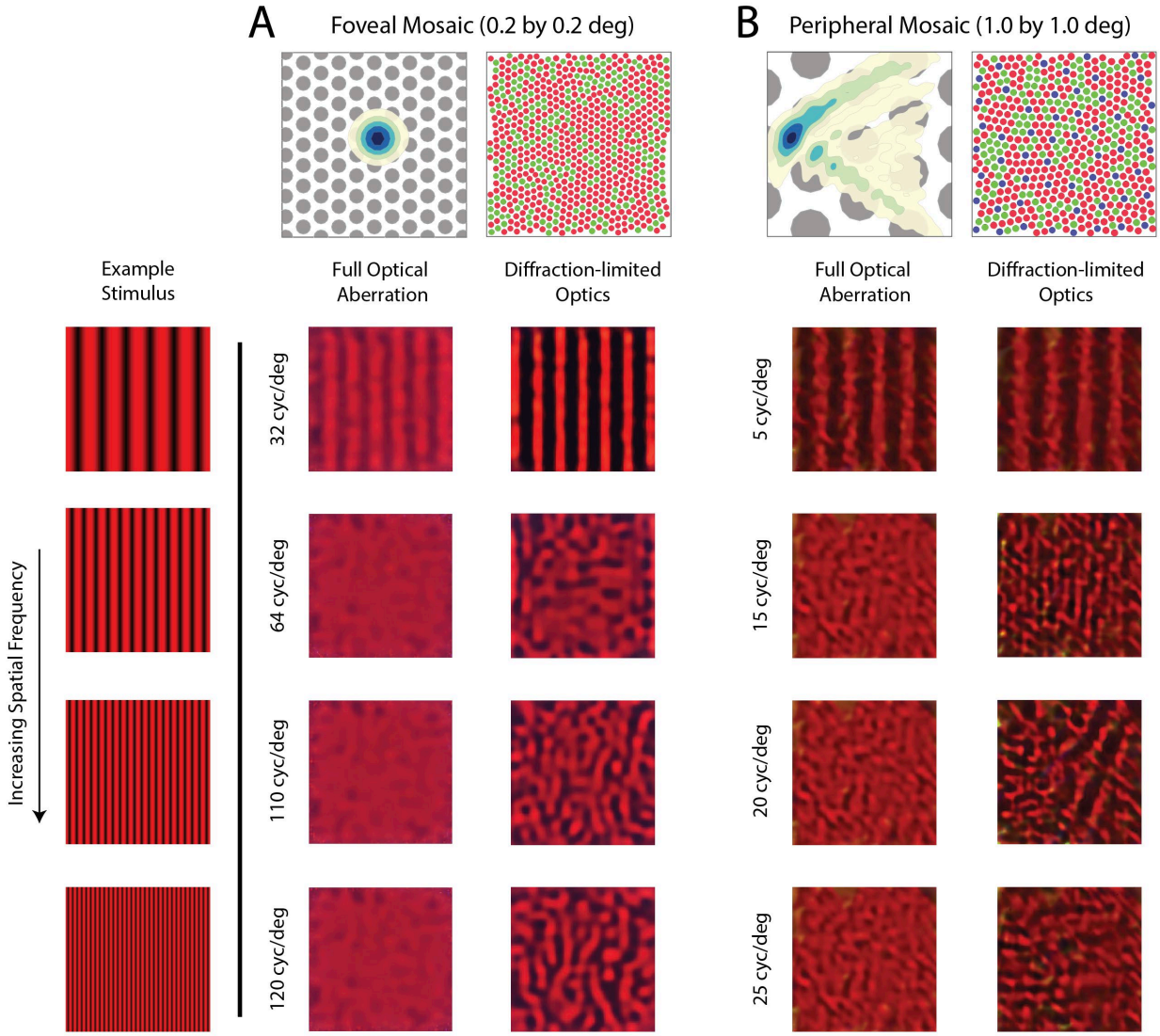

Figure S6: Reconstruction of Chromatic Grating Stimuli with/without Optical Aberrations. Image reconstruction of chromatic grating stimuli with increasing spatial frequency from **(A)** a 0.2-deg foveal mosaic and **(B)** a 1-deg peripheral mosaic at (18, 18) degrees retinal eccentricity with full optical aberrations (left columns) and with diffraction-limited optics (right columns). The top left images show a contour plot of the point-spread function relative to an expanded view of the cone mosaic, while the top right images show the full mosaic. Images were modulations of the red channel of the simulated monitor, to mimic the 633 nm laser used in the interferometric experiments. The exact frequency of the stimuli being used for each condition is as denoted in the figure. Note that the mottle observed in the reconstructions with full optical aberrations at high spatial frequencies match the reconstruction of a uniform field of saturated red stimulus.

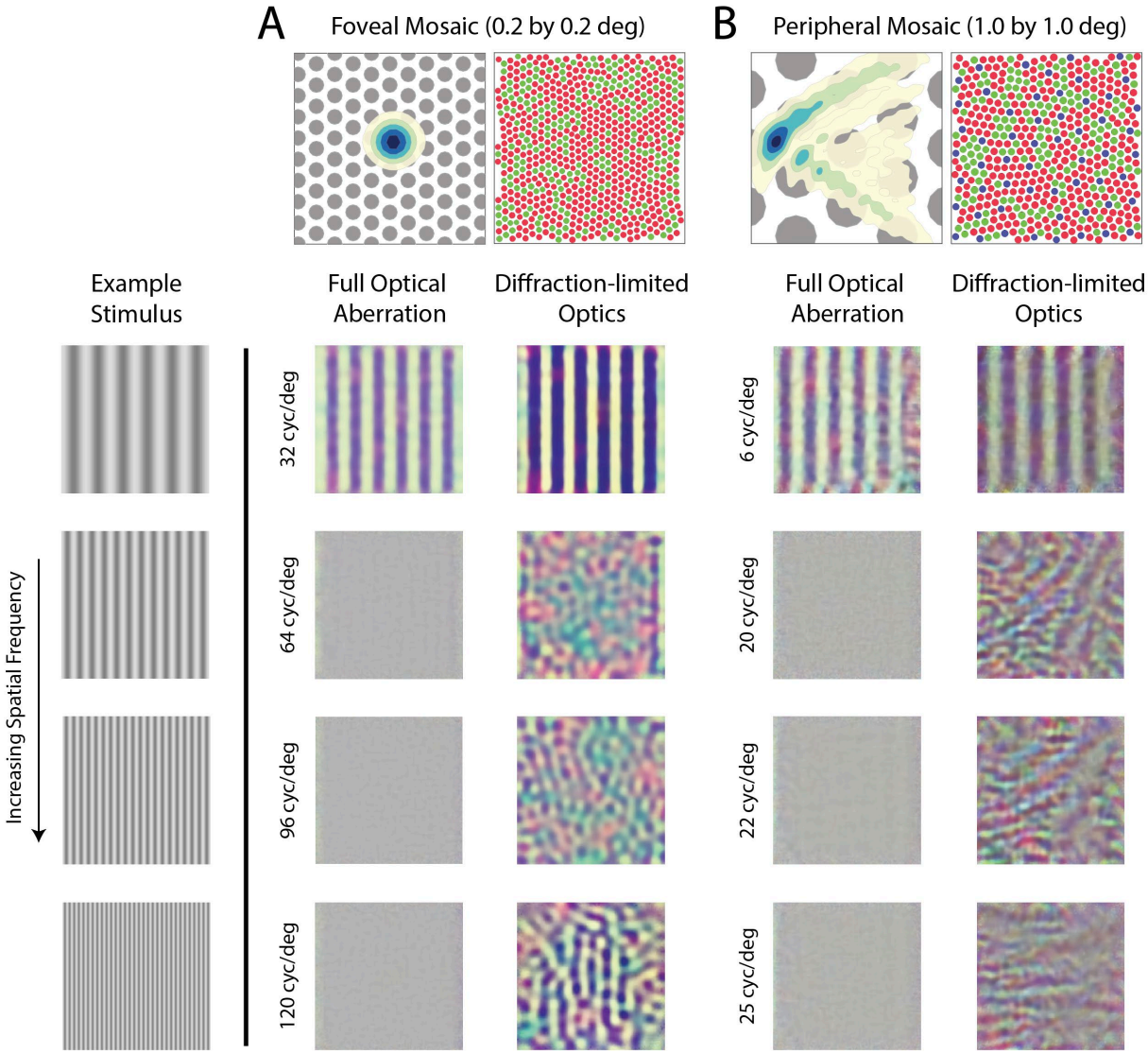

Figure S7: Reconstruction of Achromatic Grating Stimuli with/without Optical Aberrations. Image reconstructions of achromatic grating stimuli with increasing spatial frequency from **(A)** a 0.2-deg foveal mosaic and **(B)** a 1-deg peripheral mosaic at (18, 18) degree retinal eccentricity with full optical aberration (left columns) and with diffraction-limited optics (right columns). The top left images show a contour plot of the point-spread function relative to an expanded view of the cone mosaic, while the top right images show the full mosaic. The exact frequency of the stimuli being used for each condition is as denoted in the figure. The reconstruction shows similar spatial aliasing as in **Fig. 9** and **Fig. S6**, but shows an additional pattern of chromatic aliasing that arises because of the interleaved sampling by a mosaic of different cone types (Williams et al. 1991; Brainard et al. 2008). Whether such chromatic aliasing would actually be observed if a subject viewed achromatic gratings under diffraction-limited conditions is to our knowledge, an open question.