Trichromatic Reconstruction from the Interleaved Cone Mosaic: Bayesian Model and The Color Appearance Of Small Spots

David H. Brainard
Department of Psychology, University of Pennsylvania, Philadelphia, PA
brainard@psych.upenn.edu
To Whom Correspondence Should Be Addressed

David R. Williams
Center for Visual Science, University of Rochester, Rochester, NY
david@cvs.rochester.edu

Heidi Hofer
College of Optometry, University of Houston, Houston, TX
hhofer@optometry.uh.edu

Please do not quote without permission. Comments welcome.
Submitted to Journal of Vision, August, 2007
Abstract

Observers use a wide range of color names, including white, to describe monochromatic flashes with a retinal size comparable to that of a single cone. We model such data as a consequence of information loss arising from trichromatic sampling. The model starts with the simulated responses of the individual L-, M-, and S-cones actually present in the cone mosaic and uses these to estimate the L-, M-, and S-cone signals that were present at every image location. We incorporate the optics and mosaic topography of individual observers, as well as the spatio-chromatic statistics of natural images. We simulated the experiment of Hofer et al. (2005), and predicted the color name on each simulated trial from the average chromaticity of the spot reconstructed by our model. The main features of the data across observers emerged naturally as a consequence of the measured individual variation in the relative numbers and arrangement of L, M, and S-cones. The model’s output is also consistent with the appearance of larger spots and of sinusoidal contrast modulations. Finally, the model makes testable predictions for future experiments that study how color naming varies with the fine structure of the retinal mosaic.
Introduction

Trichromatic sampling

Human color vision is mediated by three classes of retinal cones, the L, M, and S cones. That there are only three classes means that color vision is trichromatic: two lights with distinct spectra that produce the same photopigment isomerization rates in all three classes will be indistinguishable to the visual system. Such lights are referred to as metamers. Metamerism is a special case of aliasing, where two physically distinct images produce the same responses in an extended array of photoreceptors.

Figure 1. Retinal cones are arranged in an interleaved mosaic.

The figure shows the locations and class of cones in an individual human retina. The cone locations were determined by acquiring fundus images through an adaptive optics system that corrected for the optical aberrations in the eye of an individual observer. Cone types were determined using microspectrophotometry in conjunction with the adaptive optics system. Figure reproduced from Roorda and Williams (Roorda & Williams, 1999). L cones shown in red, M cones in green, and S cones in blue.

Standard treatments of human trichromacy (e.g. Wyszecki & Stiles, 1982; Kaiser & Boynton, 1996; Wandell, 1995; Brainard, 1995) generally neglect the fine spatial structure of the mosaic, and this is justified as long as the spatial scale over which stimuli can vary is large compared to average
spacing between cones of different classes. These treatments of trichromacy break down, however, at fine spatial scales. This is because retinal cones are arranged in an interleaved fashion, with at most one cone sensing light at any particular location (see Figure 1.) Thus two patterns that differ in both their spatial and spectral properties can none-the-less be aliases of one another (Williams, Sekiguchi, Haake, Brainard, & Packer, 1991; Brainard & Williams, 1992; Brainard, 1994). This idea is illustrated in Figure 2.

**Figure 2. Aliasing can occur with interleaved mosaics.** Left panel. The figure illustrates how two physically distinct signals that vary spatially and spectrally can produce the same responses in an interleaved mosaic of cones. For simplicity, the example is for a single spatial dimension and a retina containing only L and M cones. The full spatio-spectral image is thus described by the L- and M-cone isomerization rates at each spatial location (L and M signals). Two distinct images are represented, one by the solid line in both plots and one by the dashed line in both plots. The solid line shows an image consisting of a 3 cycle per image isochromatic grating in which L and M signals vary together. The dashed line shows a 1 cycle per image red/green grating in which L and M signals vary in counterphase. These images are sensed by a mosaic consisting of 8 cones, 4 L cones and 4 M cones, arranged in a regular interleaved fashion. Locations of

---

1 Standard treatments do, however, incorporate differences in color matching that occur between small (e.g. 2°) and large (e.g. 10°) fields because of changes in the density of inert pigments between the fovea and periphery (see for example Wyszecki & Stiles, 1982).
cones are shown as triangles at the top of each plot. The two images have the same L-cone signal at each location where there is an L cone, and the same M-cone signal at each location where there is an M cone. Thus the two images are aliases, as they produce the same responses in each retinal cone receptor. Note that this aliasing occurs even though both patterns have spatial frequencies below the 4 cycle per image Nyquist rate for the mosaic as a whole. Panel adopted from Brainard (1994). Right panel. Aliasing of the sort shown in the left panels occurs in digital color imaging. Here the high-spatial frequency intensity variation of the jacket is rendered by the camera as its lower spatial frequency red/green alias. Image courtesy of J. Kraft; the face of the photographic subject has been pixilated to preserve anonymity.

The computational problem posed by interleaved mosaic design is illustrated by Figure 3. The visual system has available the isomerization rates of one cone class (L, M, or S cone) at each location. To provide a representation of a spatially varying trichromatic signal at the resolution of the mosaic, the visual system must estimate the isomerization rates of the two missing cone classes at each cone location.

**Figure 3. The estimation problem posed by interleaved trichromatic sampling.** The horizontal dimension of the figure represents spatial variation in one dimension. Once we take trichromacy into account, spectral variation in the stimulus is described by the isomerization rates of the L, M, and S cones. This variation is represented by the vertical dimension. A trichromatic reconstruction of the stimulus at the spatial resolution of the mosaic
requires specification of three isomerization rates at each location. That is, a number must be assigned to each box in the figure. The data available to make the estimates are the isomerization rates of the actual cones present. These are represented by the colored circles (red for L cones, green for M cones, blue for S cones). The figure makes clear that the estimation of the full spatio-chromatic image is underdetermined. In our more general treatment, we consider two spatial dimensions and also allow for image locations where there is no cone at all.

How the visual system copes with the estimation loss and solves this computational problem is largely unknown. In this paper, we present a model of the process and compare its predictions with psychophysical data and observations.

_Psychophysical data_

Historically, it has been difficult to identify empirical phenomena that result from interleaved trichromatic sampling and thus allow study of the computations related to this sampling. Although aliasing of intensity variation into chromatic variation is a routine artifact in digital imaging (see Figure 2), demonstrations of parallel phenomena in humans have been elusive. Williams et al. (1991) argued that an effect known as Brewster’s colors, wherein colored patterns can be seen on top of high-contrast luminance gratings, represent one such example. They suggested that the subtle nature of the effect resulted from clever processing by the visual
system in the way that signals from separate cones are combined to produce the overall percept. None-the-less, the aliasing artifacts were subtle and difficult to exploit for systematic modeling. Williams et al. (1991) were able to argue that the visual system is unlikely to interpolate signals from each cone submosaic separately.

Filling-in at the tritanopic area (e.g. Williams, MacLeod, & Hayhoe, 1981) provides another example of spatio-chromatic aliasing. Brainard and Williams (1992) showed that the perceived S-cone component of the stimulus at the tritanopic area is influenced by L- and M-cone signals, a direct empirical verification that signals from different cone classes can interact in the reconstruction algorithm used by the visual system. Brainard and Williams (1992) did not find evidence for interaction between S-cone signals and L- and M-cone signals for extended grating stimuli that were above the resolution limit for the S cones alone.

Recently, Hofer et al. (2005) reported experiments where observers named the color appearance of small flashed retinal spots. By using adaptive optics to correct the aberrations in individual observers, they were able to present spots whose retinal size was similar in scale to the acceptance aperture of single cones. Moreover, they were able to measure the spatial arrangement and types of the cones in the same general retinal region as they presented the flashed spots. Their experiment provides an opportunity to study the estimation algorithm employed by the visual
system, since a report of the color name of a flash that stimulates only one cone implies estimates of the signals seen by the two cone types missing at the spot location.

*Bayesian model of trichromatic reconstruction*

Here we report a quantitative model of the experiment of Hofer et al. (2005). The model is based on a Bayesian algorithm that solves the general estimation problem illustrated by Figure 3. We implemented the algorithm using measurements of individual observer cone mosaics and optics, and used simulation to evaluate the algorithm’s performance for the small spot experiment. This allowed us to compare the human data to the algorithm’s predictions, in aggregate. We find that the model provides a good first-order account of the data, and that it also accounts for other color appearance phenomena. Moreover, the model makes clear predictions for the results of future experiments.
Methods

Measurements of small spot colors

The experimental procedures and results are described in detail by Hofer et al. (2005). Briefly, an adaptive optics system was used to measure and correct for aberrations in the optics of individual observers. This enabled resolution of individual cones in acquired fundus images. Microspectrophotometry performed using the adaptive optics system allowed determination of the type (L, M, or S) cones in localized retinal regions at ~1° retinal eccentricity. Such regions were characterized for 5 observers, and ~12' by ~12' subregions of characterized mosaic for each observer are shown schematically in Figure 4. We used the subregions shown in the computations presented here.

The adaptive optics system was used to present small monochromatic spots (~0.3 minute fwhm retinal size) at the retinal region where the mosaic had been characterized. The location of each spot within the region varied from trial-to-trial because of fixational eye movements. Spots were presented at near threshold intensities, determined individually for each observer, and observers named the color of each seen spot. Observers chose the color name from the set red, orange, yellow, yellow-green, green, blue-green, blue, purple, or white. A few spots were also judged Unnamable by observers, an interesting feature of the data that we note but do not dwell on further. Not all observers employed all available color names. The
bottom panel of Figure 4 show the color naming performance of all 5 observers averaged over 500 nm, 550 nm, and 600 nm spots. Several features of the data are noteworthy. First, observers required a wide range of color names to describe their experience; all observers employed at least 5 color names across the experimental conditions studied. Second, all observers described a non-trivial percentage of spots as white. Third, there was substantial individual variation in the naming data. For example, the percentage of spots named white was smallest for observers with roughly equal numbers of L and M cones and largest for observers with either L-cone dominated or M-cone dominated mosaics.

**Figure 4. Basic data from Hofer et al. (2005).** Top panels. Schematics of retinal mosaics used for the 5 observers. These are subsets of the full regions characterized for each observer. Cones identified as L cones are colored red, those as M cones green, and those as S cones blue. LM ratios of mosaics used: HS 1:3.1; YY 1.2:1; AP 1.3:1; MD 1.6:1, BS, 14.7:1. Bottom panel. The bar plot shows the color naming data for each observer. Each bar represents performance for one observer. The proportion of the bar depicted in each color provides the percent of color names used by that observer. From bottom to top of each bar, the possible color names are red, orange, yellow, yellow-green, green, blue-green, blue, purple, and white. Data averaged over 500, 550, and 600 nm spots and correspond to 50% seeing for each observer. Data replotted from Hofer et al. (2005).
Model Structure

Overview. Figure 5 shows the steps in the model calculation. In the treatment here, various numerical quantities are given symbolic names. Table 1 provides a summary of the notation and typical values.

Figure 5. Overview of model structure. The first step was to simulate the presentation of small monochromatic spots. For each presentation, we then calculated the retinal image that would be produced by the spot, and from this the array of cone responses. A model of spot detection was implemented, so that color names were evaluated for spots predicted as seen. The cone responses for seen spots were used as the input to the Bayesian reconstruction algorithm, which produced a reconstructed image for each seen spot. The chromaticity of the reconstructed spot was used to predict its color name.

Stimulus representation. On every trial of the simulation, an $N_{\text{pixel}}$ by $N_{\text{pixel}}$ image representing a small monochromatic spot was created. The image corresponded to $K_{\text{image}}$ by $K_{\text{image}}$ minutes of arc, and the size of the simulated spot was that of the physical stimulus used in the experiment ($K_{\text{spot}}$ minutes in diameter.) The spot location $x_{\text{spot}}$, $y_{\text{spot}}$ was chosen at random within the image, subject to the requirement that its center be at least one cone aperture diameter away from the edge of the image. The wavelength $\lambda_{\text{spot}}$ of the simulated spots matched those used in the experiments. We specified spot intensities ($I_{\text{spot}}$) specified in arbitrary units
and linked these to the physical spot intensities used in the experiments through a psychophysical model of spot detection (see below).

**Calculation of retinal image.** The retinal image is formed when the stimulus is blurred by the eye’s optics. We computed the retinal image using point spread functions calculated for each individual observer at the spot wavelength from measurements of the residual optical aberration of that subject acquired during adaptive optics correction (Hofer, Singer, & Williams, 2005). The adaptive optics point spread function was represented at the same pixel resolution as the image. For convenience in computing cone isomerization rates, we also incorporated blurring by the cone aperture at this stage. We assumed a Gaussian aperture for each cone, with a full width at half max of 61.5% of the nominal cone diameter ($K_{coneaperture}$, in minutes of arc).

**Calculation of cone responses.** Measurements of cone location and cone type for each observer were used to sample the retinal image. These mosaic properties (cone locations and cone types) were determined by imaging the retina of each observer with correction by adaptive optics, in conjunction with retinal densitometry (Roorda & Williams, 1999; Hofer, Carroll, Neitz, Neitz, & Williams, 2005; Hofer, Singer, & Williams, 2005). For a small number of cones, the densitometry data were ambiguous as to cone type. Such cones were randomly assigned as either L or M cones according to the measured proportion of these two cone types in that observer’s
retina. We used $I_{\text{spot}}$, $\lambda_{\text{spot}}$, and the Smith-Pokorny (Smith & Pokorny, 1975; DeMarco, Pokorny, & Smith, 1992) estimates of cone spectral sensitivity, with the function for each cone normalized to have a quantal sensitivity of one at its wavelength of maximum sensitivity, to compute the mean isomerization rate $u_{\text{mean}}$ of each of the $N_{\text{cones}}$ cones contained within the $K_{\text{image}}$ by $K_{\text{image}}$ minute mosaic. As noted above, intensities in the simulations were specified in arbitrary units, so we simply took the mean isomerization rates directly as the inner product of the spot spectrum and the normalized spectral sensitivities; coupling of intensity to the experiment was achieved through the detection model.

The mean isomerization rates were used to compute actual simulated isomerizations for each cone through a draw from a Poisson distribution with mean $u_{\text{mean}}$. We refer to the result as the cone responses. In the calculations below, it is convenient to represent the mean cone responses as an $N_{\text{cones}}$-dimensional column vector $y_{\text{meanmosaic}}$, and the responses with noise added as $y_{\text{noisymosaic}}$.

*Detection model.* In the experiments, observers made a yes-no judgment as to whether they saw each flash and named the color of flashes seen. Experiments were run at near threshold intensities, so we include a model of detection in the simulations. The goal of this model was not to provide a precise account of the detection of small spots, which is beyond
the scope of this paper. Rather the model yoked simulated and physical spot intensities, and provided an indication of which of our simulated flashes would be seen, so as to select flashes whose names should be predicted. Only the appearance of seen flashes was modeled.

For each flash, we identified the cone with the largest response. We then summed the response of that cone and its $N_{\text{neighbors}}$ nearest neighbors in the mosaic. If this pooled response $y_{\text{pooled}}$ exceeded a criterion level $y_{\text{criterion}}$, the flash was classified as seen. This is, in essence, the pooled detection model used by Hofer et al. (Hofer, Singer, & Williams, 2005). Supplemental Figures S1 and S2 compare frequency of seeing curves from the detection model to experimental data. The agreement is good when the relation between model intensity and physical intensity is scaled independently for each observer and flash wavelength to optimize the quality of the fit (Supplemental Figure S1). The detection model does not capture the within observer experimental variation in frequency of seeing curves with flash wavelength (Supplemental Figure S2), nor between observer variation. This is not surprising, given that the detection model does not incorporate photoreceptor dark noise or limitations imposed by post-receptoral processing. To predict color naming performance below, we set the relation between model intensity and physical intensity separately for each observer and wavelength.
**Trichromatic reconstruction.** The cone responses $y_{\text{noisymosaic}}$ for seen flashes were used as input to a Bayesian algorithm that reconstructs a trichromatic image. The output of the algorithm is an $N_{\text{pixel}} \times N_{\text{pixel}} \times 3$ color image. The three image planes represent L-, M-, and S-cone components of an idealized retinal image estimated by the algorithm. By this, we mean the image that would have been imaged on the retina in the absence of blur by the eye’s optics.

We can represent estimated image by a column vector $x_{\text{estimate}}$ with $N_{\text{pixel}} \times N_{\text{pixel}} \times 3$ entries. The first $N_{\text{pixel}} \times N_{\text{pixel}}$ entries represent the L-cone plane in rasterized order, the second $N_{\text{pixel}} \times N_{\text{pixel}}$ entries represent the M-cone plane, and the last $N_{\text{pixel}} \times N_{\text{pixel}}$ entries represent the S-cone plane. For purposes of this overview the algorithm may be thought of as a procedure that takes observations $y_{\text{noisymosaic}}$ and produces an image estimate $x_{\text{estimate}}$. The algorithm itself, which forms the core of our model, is described in more detail below.

**Mapping to color names.** We used the estimated image to assign a color name to each seen flash. First, we extracted and averaged the L-, M-, and S-cone values from the reconstructed image in the neighborhood of the reconstructed flash. To find this neighborhood, we identified the location in that had the largest value of the quantity $\sqrt{L^2 + M^2 + S^2}$. Here where the symbols $L, M,$ and $S$ represent the reconstructed values for each cone class.
at a pixel of the reconstructed image. We then computed the center of mass
of the quantity $\sqrt{L^2 + M^2 + S^2}$ over pixels whose value of $\sqrt{L^2 + M^2 + S^2}$ was at
least 30% of the maximum and that were within one spot diameter of the
location with the largest value. We took the resulting center of mass as the
center of the reconstructed spot, and took the region of the reconstructed
spot as one spot diameter around this center. We excluded from this
analysis reconstructed spots that were close to the border of the simulated
image. We then found the CIE u‘v’ chromaticity corresponding to the
average LMS triplet.

Each region in the chromaticity diagram was assigned a color name. The
same naming boundaries were used for all observers, and the exact
boundaries were determined by numerical search to optimize how well the
overall model accounted for the data. This search minimized the sum of
squared differences between actual and predicted naming percentages over
all observers and wavelengths. The parameters adjusted described the
location and size of a circle that defined the region of chromaticity space
corresponding to white, and the angles of the linear boundaries in the
chromaticity space that radiated from the center of the white circle and
which separated the color categories red, orange, yellow, yellow-green,
green, blue-green, blue, and purple. To perform the numerical search, we
used soft categorical boundaries defined by a cumulative normal distribution
function. The location of the lines separating color categories was
constrained so that the final boundaries did not stray excessively far from typical nominal locations. Figure 6 shows the naming boundaries obtained for the best fit with the best choice of algorithm prior parameters (see below.)

By simulating repeated flash presentations and comparing their reconstructed chromaticities to the color naming boundaries, we produced a histogram of predicted color naming percentages for each observer and flash wavelength. These predicted percentages were computed using hard categorical boundaries, rather than the soft boundaries used to facilitate the numerical search.

**Figure 6. Mapping of estimated spot chromaticity to color**

*names.* The chromaticity diagram is divided into regions corresponding to different color names. The central circular region corresponds to white. The surrounding regions, divided by radial lines in the chromaticity diagram, correspond (anti-clockwise) to red, orange, yellow, yellow-green, green, blue-green, blue, and purple as indicated by the colored dots shown in the figure. The naming boundaries used were held constant across observers and wavelengths. They were determined by numerical search to optimize the overall agreement between predictions and data as described in the text. The solid black line shows the spectrum locus from 400 nm to 700 nm. The black circles plotted on the spectrum locus mark the chromaticities of monochromatic lights in 50 nm increments (i.e. 400 nm, 450 nm, ..., 700 nm).
Bayesian algorithm

The core of the model is the algorithm that maps cone responses $y_{\text{noisy mosaic}}$ to trichromatic image estimates $x_{\text{estimate}}$. We adopted a Bayesian algorithm for trichromatic reconstruction (sometimes called demosaicing) developed by Brainard in the context of digital color image processing (Brainard, 1994). The implementation was modified for use with the spatially irregular sampling of the human cone mosaic, and we describe the algorithm below.

Bayesian estimation. Bayesian estimation is based on using probabilities to express i) the relation between the quantity to be estimated and the observed data and ii) a priori constraints on the quantity to be estimated.

The first probability is called the likelihood and may be written as $p(y | x)$. Here $x$ represents an idealized retinal image as described above and $y$ represents observed cone responses. The likelihood tells us the probability of any vector of cone responses $y$ occurring, given that the image is $x$.

The second probability is called the prior and may be written as $p(x)$. This tells us the a priori probability of any image $x$ occurring.

Given the likelihood and the prior, Bayes rule allows us to express the posterior probability $p(x | y)$ as $p(x | y) = C p(y | x) p(x)$ where $C$ is a normalizing constant. The posterior tells us the probability of any image $x$ given the observed data $y$. It is then possible to choose an estimate of $x_{\text{estimate}}$ of $x$ from the posterior, for example as the mean or maximum of the
posterior. Figure 7 illustrates Bayesian estimate for a very simple example image reconstruction problem.

**Figure 7. Bayesian estimation applied to simple example image reconstruction.** The figure illustrates the basic principles of Bayesian estimation for a highly simplified version of the trichromatic estimation problem considered in this paper. The image to be estimated consists of \( \mathbf{x} = \begin{bmatrix} L & M \end{bmatrix}^T \), the L and M cone responses at a single image location. The observations are the noisy response of a single L cone at that location, \( \mathbf{y} = \begin{bmatrix} L \end{bmatrix} + n \) where \( n \) represents additive noise. The top left panel shows a prior distribution \( p(\mathbf{x}) \) over the (L, M) image space. This is modeled as a bivariate normal distribution that expresses a correlation between L and M cone responses. Such correlations are typical of natural images (Burton & Moorehead, 1987; Ruderman, Cronin, & Chiao, 1998). The top right panel illustrates the likelihood \( p(\mathbf{y} | \mathbf{x}) \) for the case \( \mathbf{y} = 3 \). The likelihood of observing 3 is relatively large for values of \( \mathbf{x} \) whose first component is near 3, independent of the second component. Thus in this case the likelihood plots as a ridge in the (L, M) space. The posterior is obtained by multiplying the likelihood with the prior, and then normalizing. The bottom panel shows the posterior for the case \( \mathbf{y} = 3 \). The combination of likelihood and prior constrain the solution more than either factor alone. Here the mean and maximum of the posterior coincide and provide a reasonable estimate of \( \mathbf{x} \). Functions shown in the figure plot probabilities in arbitrary units.
The substance of a Bayesian algorithm is captured by the formulation of the likelihood, the prior, and the rule used to choose \( x_{\text{estimate}} \) from the posterior.

**Likelihood.** For our application, the likelihood is essentially a model of the image formation process. Above we described how we simulated cone responses produced by monochromatic flashes. Let \( f(\lambda_{\text{spot}}, I_{\text{spot}}, x_{\text{spot}}, y_{\text{spot}}, K_{\text{spot}}) \) be the function that returns \( y_{\text{meanmosaic}} \) as a function of the simulated spot properties. The function \( f() \) simulates the eye’s optics, blurring by photoreceptor sampling, sampling by the trichromatic mosaic, and absorption of light by the L-, M-, and S-cone photopigments. To implement \( f() \), we used estimates of the eye’s optics under normal viewing, measurements of the location and types of the cones in a patch of each observer’s retina (measured as described above), and the Smith-Pokorny estimates of cone spectral sensitivity. When viewed as a function that maps images \( x \) to mean observations \( y_{\text{meanmosaic}} \), \( f() \) is linear in \( x \).

We incorporated normal optics into the algorithm’s construction, rather than adaptive optics, on the assumption that the processing applied by the visual system is not so plastic as to adapt significantly to the adaptive optics correction applied during the experiment.

For observers YY, AP, and MD, estimates of the normal optics for each cone class were obtained from wavefront sensor measurements of the eye's
optics prior to adaptive optics correction. Aberrations were measured at 820 nm over a 6.8 mm pupil and represented by 10 orders of Zernike coefficients. These were then used to compute the optical point spread function for monochromatic lights across the visible spectrum for a 3mm pupil using standard estimates of the eye’s lateral chromatic aberration. For each cone class, the monochromatic point spread functions were weighted by that class’s spectral sensitivity and averaged. Measurements were not available for observers HS and BS. For these observers, we used the circularly symmetric average of the point spread functions obtained for YY, AP, and MD. To convert the deterministic imaging model expressed by $f(\cdot)$ into a likelihood, we assume that the actual cone responses are obtained from the mean cone responses through the addition of zero-mean normally distributed additive noise. That is, we write $y_{\text{noisy mosaic}} = N(y_{\text{mean mosaic}}, K_{\text{noise}})$, where $K_{\text{noise}}$ is an $N_{\text{cones}} \times N_{\text{cones}}$ diagonal matrix with each entry equal to the variance of the noise for the corresponding cone in the mosaic. The variance of the noise for each cone was taken as the mean response of that cone to the prior mean image, which may be thought of as implementing a normal approximation to Poisson photon noise. The reason for using the normal approximation is that it allows for a closed-form solution for the Bayesian estimate.
Prior. We take \( p(x) \) as a multivariate normal distribution, so that
\[
p(x) \sim N(u_x, K_x)
\]
In typical applications, we compute with \( N_{\text{pixel}} = 100 \), which means that the dimensionality of \( x \) is \(~ 30,000 \). This is too large to allow an arbitrary choice of \( K_x \), so that it is necessary to impose some additional structure on the form of the prior.

First, we assumed that the prior was separable in space and color. This assumption means that the probability distribution over L-, M-, and S-cone isomerization rates at each pixel do not depend on the pixel’s location, and that the probability distribution over spatial structure in each L-, M-, and S-cone plane is the same. This assumption allows us to write
\[
u_x = u_{LMS} \otimes u_{\text{space}}
\]
and
\[
K_x = K_{LMS} \otimes K_{\text{space}},
\]
where \( \otimes \) represents the Kronecker product.

Measurements of natural images indicate that this separability assumption is reasonable (Burton & Moorehead, 1987; Ruderman, Cronin, & Chiao, 1998; Parraga, Brelstaff, Troscianko, & Moorehead, 1998; deviations from separability for some image types reported in Parraga, Troscianko, & Tolhurst, 2002).

We further assumed that the spatial component of the prior was separable in the vertical and horizontal spatial dimensions, and that the vertical and horizontal priors were identical to each other. Thus
\[
u_{\text{space}} = u_{\text{space,1}} \otimes u_{\text{space,1}} \quad \text{and} \quad K_{\text{space}} = K_{\text{space,1}} \otimes K_{\text{space,1}}
\]
where \( u_{\text{space,1}} \) and \( K_{\text{space,1}} \) characterize the properties of the vertical/horizontal priors. We took \( u_{\text{space,1}} \)
to be a constant vector with value $u_{space}$, so that the expected value of the prior was constant across image locations. We computed $K_{space\_1}$ in two steps. First, we chose a single variance $\sigma^2_{space}$ and nearest neighbor correlation $\rho_{space}$ and from these generated the covariance matrix of a first order Markov process along one spatial dimension. The parameter $\rho_{space}$ was expressed in terms of the correlation between locations separated by the mean cone spacing in the mosaic. Measurements of natural images indicate that they are characterized by a significant nearest neighbor correlation (e.g. Pratt, 1978; Field, 1987; Burton & Moorehead, 1987; Ruderman, Cronin, & Chiao, 1998; Parraga, Brelstaff, Troscianko, & Moorehead, 1998; Parraga, Troscianko, & Tolhurst, 2002). 2 Denote by $x_{space\_1}$ a random draw from the normal distribution with mean vector $u_{space\_1}$ and covariance matrix $K_{space\_1}$:

$$p(x_{space\_1}) \sim N(u_{space\_1}, K_{space\_1})$$

In the absence of computational limitations, this is the vertical/horizontal prior we would have used. In practice, however, we needed to reduce the dimensionality of the quantities used in the calculations, so that as second step we approximated this desired spatial prior. We used the singular value decomposition to express $K_{space\_1} = U D V'$ and took the linear model $B_{space\_1}$ to be the first $N_{linmod}$ columns of $U$. We then constructed a mean vector $\bar{u}_{space\_1}$ as the vector that of weights that

---

2 The higher nearest neighbor correlation is often expressed in the spatial frequency domain in terms of the approximately $1/f$ falloff in image amplitude spectrum with spatial frequency.
provided the best least-squares approximation to $\mathbf{u}_{\text{space}_1} = \mathbf{B}_{\text{space}_1} \mathbf{u}_{\text{space}_1}$, and we constructed a diagonal covariance matrix $\mathbf{K}_{\text{space}_1}$ whose diagonal elements were the first $N_{\text{linmod}}$ singular values of $\mathbf{K}_{\text{space}_1}$ (that is, the first $N_{\text{linmod}}$ diagonal entries of $\mathbf{D}$.) We then approximated $p(\mathbf{x}_{\text{space}_1})$ by

$$p(\mathbf{x}_{\text{space}_1}) = \begin{cases} p(\mathbf{w}_{\text{space}_1}), & \text{for } \mathbf{x}_{\text{space}_1} = \mathbf{B}_{\text{space}_1} \mathbf{w}_{\text{space}_1} \\ 0, & \text{otherwise} \end{cases}$$

where $p(\mathbf{w}_{\text{space}_1}) \sim N(\mathbf{u}_{\text{space}_1}, \mathbf{K}_{\text{space}_1})$.

To specify the color component of the prior, we proceeded as follows. We assumed that the cone coordinates $\mathbf{x}_{\text{color}}$ at each location were distributed according to $p(\mathbf{x}_{\text{color}}) \sim N(\mathbf{u}_{\text{color}}, \mathbf{K}_{\text{color}})$. We found $\mathbf{u}_{\text{color}}$ and $\mathbf{K}_{\text{color}}$ by finding the mean vector and covariance matrix of a sample of 10000 cone coordinate vectors that were generated as follows. First we generated a sample of 10000 CIE $u'v'$ chromaticity vectors (first entry $u'$, second entry $v'$) by drawing from bivariate normal distribution with mean $u_{u'v'}$ and diagonal covariance matrix $\mathbf{K}_{u'v'}$. The matrix $\mathbf{K}_{u'v'}$ was diagonal and both of its entries were the same and given by parameter $\sigma_{u'v'}^2$. We took $u_{u'v'}$ to be the chromaticity of CIE standard illuminant D65. We converted each of the 10000 chromaticity vectors to a normalized cone coordinate vector, with the magnitude of these vectors chosen so that the mean of the L and M cone coordinates was unity. We then scaled each of the normalized cone coordinate vectors by an intensity factor. The intensity factor used for each
normalized cone coordinate vector was obtained by an independent draw from a univariate normal distribution, \( p(f_{\text{color}}) \sim N(\mu_{\text{factor}}, \sigma^2_{\text{factor}}) \). We set \( \mu_{\text{factor}} \) so that the mean of the L and M cone coordinates of the set of 10000 scaled cone coordinate vectors was equal to the mean L and M cone coordinate of the ensemble of stimuli being simulated. We set \( \sigma^2_{\text{factor}} = \kappa_{\text{color}} \mu_{\text{factor}} \) where \( \kappa_{\text{color}} \) was a specified constant. The resulting set of 10000 scaled cone coordinate vectors was the set used to determine \( u_{\text{color}} \) and \( K_{\text{color}} \).

Note that this procedure introduces a dependence of the color prior on the properties of the stimulus ensemble. This dependence was introduced as a crude model of adaptation, to ensure that the intensity range of the prior was commensurate with the intensity of the spots being flashed. For the flashed spot simulations, we treated spots of each wavelength as a separate ensemble.

The net effect of our specification of prior \( p(x) \) is that the properties of this prior were controlled by a relatively small number of parameters. Although the dimensionality of \( x \) was typically \( \sim 30,000 \), the prior was specified by parameters \( u_{\text{space}}, \sigma^2_{\text{space}}, \rho_{\text{space}}, N_{\text{linmod}}, u_{\text{u,v}}, \sigma^2_{\text{u,v}}, \) and \( \kappa_{\text{color}} \). In practice, we never varied \( u_{\text{space}}, N_{\text{linmod}}, \) and \( u_{\text{u,v}} \) so that in effect the prior was specified by four parameters. Permitting only a small number of degrees of freedom for the prior is crucial to managing the complexity of computational studies such as the one reported here.
**Posterior and estimate.** Our choice of normal prior with known mean and covariance and additive normal noise with known mean and covariance, combined with the fact that the deterministic component of the likelihood function is linear, implies that the posterior distribution is also normal (Gelman, Carlin, Stern, & Rubin, 2004). In addition, for this case, the mean of the posterior may be computed analytically from the observation vector $y_{noisymosaic}$ using standard formulae (Gelman, Carlin, Stern, & Rubin, 2004; Brainard, 1994; Pratt, 1978, development of discrete Wiener estimator). This posterior mean is also the value of $x$ that maximizes the posterior, and we take it as our estimate $x_{estimate}$. Interestingly, the final estimate may be expressed as an affine transformation of the cone responses:

$$x_{estimate} = I y_{noisymosaic} + i_0$$

where $I$ is an $(N_{pixel} \times N_{pixel} \times 3)$ by $N_{cones}$ matrix and $i_0$ is an $(N_{pixel} \times N_{pixel} \times 3)$ vector.

**Determining prior parameters.** It is well-established that natural images exhibit high-correlations between nearby spatial locations and that they produce high-correlations between the responses of different cone classes at the same spatial location. The exact values appropriate for use in models such as the one we present are less clear, in part because some variation in the exact values can be used to compensate for distortions introduced by the normal assumption incorporated in our priors. To determine the best parameters, we ran simulations for 108 different choices of prior
parameters, obtained as all possible combinations of \( \sigma_{u',v'}^2 = [0.07 \ 0.08 \ 0.09 \ 0.10], \)
\( \sigma_{\text{space}}^2 = [0.25 \ 0.5 \ 1.0], \) \( \rho_{\text{space}} = [0.75 \ 0.85 \ 0.95], \) and \( \kappa_{\text{color}} = [2 \ 4 \ 8]. \) For each combination, we simulated 2000 flashed spots at each wavelength for each observer and extracted the chromaticities of the reconstructed seen spots. We then fit the color boundaries to maximize the agreement between predicted and measured naming percentages. Since the simulation involves a random component (spot locations and added noise on each simulated trial), we were cautious about the possibility that some of the boundary fitting might be explaining random rather than systematic variation. To avoid this, we resimulated the experiment for each of the 108 prior choices and computed the correlation between predicted and named percentages from the corresponding boundary that was fit on the first run. We selected the prior parameters that led to the highest second run correlation without refit of the naming boundaries.

To account for regression to the mean in the quality of the fit, we then repeated the simulate-fit-simulate procedure for this best choice of parameters, and report here the boundaries obtained from the fit to the first simulation run together with predicting naming results from the second simulation run. The final parameters used were \( \sigma_{u',v'}^2 = 0.08 , \sigma_{\text{space}}^2 = 0.25 , \)
\( \rho_{\text{space}} = 0.85 , \) and \( \kappa_{\text{color}} = 8. \) The computed correlations between cone classes corresponding to these parameters were LM: 0.96, LS: 0.69, and MS: 0.74.
Large correlations between cone classes are typical in natural images 
(Burton & Moorehead, 1987; Ruderman, Cronin, & Chiao, 1998; see also 
Nascimento & Ferreira, 2002, Figure 3; also Maloney, 1986; Jaaskelainen, 
Parkkinen, & Toyooka, 1990). The prior parameters and color boundaries 
were held constant across all calculations reported in this paper.

**Results**

*Simple example of algorithm performance*

Figure 8 shows the results of the Bayesian reconstruction algorithm for 
two simple cases. The first is shown in the left two panels. The top panels 
depicts a mosaic consisting only of L cones, while the bottom panel shows 
the output of the algorithm when the single L cone at the center of the 
mosaic is stimulated in isolation. We can see the when a single cone is 
stimulated, the algorithm reconstructs that the stimulus was a small spot 
located in the vicinity of that cone. In addition, we note that the color 
appearance of the reconstructed spot is whitish. The intuition behind this 
result is that when the mosaic contains only L cones, its responses provide 
no information about the relative spectrum of the stimulus. Thus the prior 
term dominates that aspect of the reconstructed stimulus and the resulting 
value is determined by the prior. The prior has the mean chromaticity of 
CIE daylight D65 and its effect on the reconstruction is to produce a whitish 
spot.
The right two panels show the reconstruction for a second mosaic. This mosaic differs from the first only in that the central L cone is surrounded by an island of M cones. We ran the algorithm for the same response scenario, where only the central L cone was stimulated. Here the output is also a small spot, but its color appearance is reddish. This change in the relative spectrum of the reconstructed spot occurs despite the fact that in both cases the stimulation of the central L cone is identical and the responses of all the other cones were set to zero. The reason for the change is that the prior incorporates correlations across space, so that information about the stimulus at the central location is carried not just by the cone located there but also by the surrounding cones. Intuitively, the fact that the M cones surrounding the central L cone have zero responses indicates that, had there been an M cone at the location of the central L cone, that M cone would also have had a small response. This, together with the fact that the central L cone itself has a large response leads to a reconstructed stimulus with more power at long wavelengths than middle wavelengths.

**Figure 8. Reconstruction algorithm performance for two artificial mosaics.** The top two panels show two hypothetical mosaics. The left mosaic contains only L cones. The spatial arrangement of the cones in the second mosaic is the same, but here the central L cone has been surrounded by an island of M cones. We used the Bayesian algorithm to reconstruct the stimulus from the responses of these two mosaics, for the case where the only the
central L cone (marked by the white square in each panel) had a non-zero response. That response represented a near threshold intensity level, and was the same for both reconstructions. The reconstructed images are shown below each mosaic. Displayed reconstructions were obtained by mapping LMS planes to the standard sRGB color space (without gamma correction). To suppress ringing away from the central spot, the reconstructed images shown were windowed at each pixel by the quantity \( \sqrt{L^2 + M^2 + S^2} \) and values less than 0 were set to 0. Each displayed images was normalized by a single scale factor so that it occupied the full intensity range of the sRGB color space.

This example illustrates the structural feature of our model that provides the potential to account for the appearance of the small spots flashed in the experiment of Hofer et al., namely that the appearance of a perceived spot resulting from stimulation of a cone of particular type (e.g. an L cone) depends strongly on the properties of the mosaic surrounding that cone.

**Accounting for the experimental data**

As described in the methods section, we simulated the experiment and found the chromaticity of the spots reconstructed by the model. The reconstructed chromaticities for two observers are shown in Figure 9. For each observer, there is a considerable spread in the reconstructed chromaticities. This spread results from the combination of two separate effects incorporated into the model. First, there is added response noise on each trial. Second, the location of the spot varies from trial-to-trial.
There is also between observer variation in the reconstructed chromaticities. For example, fewer reconstructed spots are in the blue and green regions of the chromaticity diagram for BS than for AP, while more are in the red region. This between observer difference in the model predictions occurs because each observer has an individual mosaic, with different relative numbers of L, M and S cones in different arrangements. The change in mosaic means that the type of cone (or cones) mediating detection will vary, so that an observer with an M-rich mosaic will detect a larger fraction of flashes with M cones than an observer with an L-rich retina. In addition, for spots detected by the same class of cone, the average structure of the local neighborhood surrounding the detecting cone (or cones) will vary across observers. Because the model’s predictions are sensitive to the structure of the mosaic surrounding the cone(s) stimulated by the flashed spot, this will also produce individual variation in the chromaticities of the reconstructed spots.

**Figure 9. Chromaticities of reconstructed small spots for two observers.** Each panel shows the reconstructed chromaticities for 0.3 arcminute spots for one observer, for 550 nm flashes. Left panel: Observer AP (LM ratio 1.3:1). Right panel: Observer BS (LM ratio 14.7:1). Color boundaries as in Figure 6.

The variation in reconstructed chromaticities may be converted to variation in predicted color naming. Figure 10 shows the performance of the
model for the white naming data. The solid circles connected by solid lines show the percentage of spots named white against the asymmetry index introduced by Hofer et al. (2005). For all three spot wavelengths, the percentage of flashes named white increases with the asymmetry index. The open circles connected by dashed lines show the model’s predictions in the same format. The model captures the data well. It both predicts the increase in percent white with increasing asymmetry index and the general trend of dependence of percent white on spot wavelength. Note that no free parameters in the model were adjusted specifically to describe wavelength or individual observer differences.

**Figure 10. Prediction of white naming.** Left panel. Solid circles connected by solid lines show the percentage of spots named white against an asymmetry index. The asymmetry index takes on a value of 0 for a mosaic where 50% of the L and M cones are L cones, and a value of 50 for a mosaic where all of the L and M cones are L cones or where all of the L and M cones are M cones. Open circles connected by dashed lines show the model’s predictions in the same format. Blue: 500 nm spots; green: 550 nm spots; red: 600 nm spots. The correlation between predictions and data was 0.86. Right panel. Same predictions and data as in left panel, but averaged over flash wavelength. The correlation between averaged predictions and data was 0.89.

Figures 11 and 12 summarize the prediction performance of the model for the full set of color names (including white.) The model predicts the broad
features of the data, both across observers and across flash wavelength. The overall correlation between observed data and model predictions is 0.82. (If one removes the points named white from the computation correlation, the correlation is 0.79.) There are some features of the data not captured by the model, however. For example, the model does not predict that any flashes will be named yellow, nor does it capture the fine structure of variation across observers in the percentage of non-white color names used.

**Figure 11. Prediction of color names.** Each bar plot shows the percent names for all 5 observers and one wavelength. Top row shows data, while bottom row shows predictions. Each plot is in the same format as the bottom panel of Figure 4. For plots from left to right, wavelengths are 500 nm, 550 nm, and 600 nm respectively. Within each plot, observers are ordered from left to right HS, YY, AP, MD, and BS.

**Figure 12. Summary of prediction of color names.** Each plotted point shows predicted percent named against measured percent name. Data for all 5 observers and all color names (including white). Blue points: 500 nm spots; green points: 550 nm spots; red points: 600 nm spots.

*Model behavior for other stimuli*

The model does a good job predicting the appearance of small-spot colors. More generally the Bayesian algorithm can reconstruct an image
from any set of cone responses, and we thought it important to explore the behavior of this algorithm for stimuli closer to everyday experience than small monochromatic spots viewed under adaptive optics conditions. We investigated the model’s behavior for larger monochromatic spots at suprathreshold intensities, and for sinusoidal modulations along three color directions.

*Larger suprathreshold spots.* Figure 13 shows the reconstructed chromaticities of 8 arcminute spots at 550 nm for two observers, at an intensity level ~0.5 log units above the average (over observers and wavelengths) 50%-seeing threshold intensity for the small monochromatic spots. Three features of the reconstructions are apparent when compared with the corresponding reconstructions for small spots (compare with Figure 8: the variability is greatly reduced within observer, the variability is greatly reduced between observers, and the reconstructed chromaticities are much closer to that of the 550 nm stimulus. These same features hold for the other observers and wavelengths, as illustrated by the naming histograms shown in Figure 14. All observers are predicted to name the stimuli consistently, and the names correspond to what we expect from standard accounts of the appearance of monochromatic lights (e.g. Hunt, 1987, p. 130).

**Figure 13. Reconstructed chromaticities for 8 arcminute spots.**

The figure shows the reconstructed chromaticities for flashed 8
arcminute monochromatic spots, for two observers. The spot intensity was suprathreshold, approximately 0.5 log units above the 50%-seeing threshold for the small (0.3 arcminute) spots. Color naming boundaries shown are those derived from the simulations of the small spot experiments (see Figure 6 above.) Simulations were performed as for the small spots, using the adaptive optics point spread functions to compute the retinal images. Left panel: Observer AP (LM ratio 1.3:1). Right panel: Observer BS (LM ratio 14.7:1).

**Figure 14. Predicted naming for 8 arcminute spots.** Each bar plot shows the percent names for all 5 observers and one wavelength for flashed 8 arcminute monochromatic spots. Each plot is in the same format as the bottom panel of Figure 4. Naming boundaries used to compute color names were those derived from the small (0.3 arcminute) spot simulations. Simulation details as described in the caption of Figure 13 above. Plots from left to right, wavelengths are 500 nm, 550 nm, and 600 nm respectively. Within each plot, observers are ordered from left to right HS, YY, AP, MD, and BS.

*Suprathreshold gratings.* Sinusoidal modulations have been employed extensively as stimuli to probe visual processing of spatial and chromatic information (e.g. deLange, 1958; Campbell & Robson, 1968; Mullen, 1985; Sekiguchi, Williams, & Brainard, 1993b). We applied the Bayesian reconstruction algorithm to such gratings. We studied isochromatic gratings, red/green isoluminant gratings that modulated the L and M cones in counterphase, and gratings that isolated the S cones. Our interest was to verify that the model produced reasonable output for these spatially-
extended stimuli. Since the model is designed to account for appearance and not thresholds, we ran our simulations at fixed contrasts for each color direction. Figure 15 shows simulated and reconstructed gratings corresponding to 6 cycles per degree (cpd). At this spatial frequency, the isochromatic and red/green gratings are reconstructed veridically. There is a small amount of distortion visible in the S-cone grating reconstruction that arises because of aliasing by the sparse S-cone submosaic aliasing (also see below). The reconstructions at lower spatial frequencies, including 0 cpd, appear similarly veridical. Thus the Bayesian model is consistent with the fact that we see the world with little distortion at low-spatial frequencies.

The left two images in Figure 16 show simulated and reconstructed isochromatic gratings corresponding to 24 cpd. Here distortions are clearly visible. Although the stimulus grating is clearly visible in the reconstruction, it is contaminated by red/green noise. This corresponds to the phenomenon of Brewster’s colors, where red/green mottle is apparent against high-contrast high-spatial frequency luminance gratings. We have argued previously that Brewster’s colors arise as a result of trichromatic sampling by the retinal mosaic (Williams, Sekiguchi, Haake, Brainard, & Packer, 1991). In that report, a simple model of the reconstruction process accounted for the spatial pattern of the red/green mottle, but predicted that it should be much more visually salient than it is. That model was based on independent interpolation of the signals from the L-, M-, and S-cone submosaics. The
Bayesian reconstruction algorithm, which jointly processes the responses of the entire mosaic, predicts more subtle effects that are in line with the perceptual phenomenon.

The center two images in Figure 16 show simulated and reconstructed red/green gratings at 24 cpd. The reconstruction distorts spatial structure of the stimulus grating. Although red/green gratings are typically below detection threshold at 24 cpd when viewed with normal optics (Mullen, 1985), they can be detected when presented using interferometric techniques (Sekiguchi, Williams, & Brainard, 1993a) that bypass optical blurring. In this case, the percept at detection threshold is one of spatial noise (Sekiguchi, Williams, & Brainard, 1993a), qualitatively consistent with the Bayesian algorithm’s reconstruction.

The right two images in Figure 16 show simulated and reconstructed S-cone gratings at 15 cpd. As the spatial frequency of S-cone gratings increases, distortions referred to as S-cone aliasing occur because of the sparse sampling of the S-cone submosaic (Williams & Collier, 1983). The splotchy nature of the reconstructed pattern reproduces this phenomenon.

**Figure 15. Reconstructed low-spatial frequency sinusoidal gratings.** The top row shows patches of isochromatic, red/green isoluminant, and S-cone grating modulations that correspond to a spatial frequency of 6 cpd. The bottom row shows the reconstructed stimuli obtained by applying the Bayesian algorithm to the cone

---

36
responses from these gratings. Simulated grating LMS contrasts were 
\[0.75 \, 0.75 \, 0.75], \, [0.062, \, -0.12, \, 0.00] \text{ and } [0, \, 0, \, 0.66]\] for the three 
modulation directions respectively. These contrasts fit within the 
gamut of a typical CRT monitor. Mean stimulus chromaticity was that 
of CIE illuminant D65 (the prior mean) and the mean stimulus 
intensity was chosen to be \(\sim 1.5\) log units above the small-spot 50%- 
seeing threshold intensity. Simulations were performed using the 
mosaic of observer AP, and the retinal image was computed using the 
normal optical PSF (without adaptive optics correction) for this 
observer.

**Figure 16.** Reconstructed high-spatial frequency sinusoidal 
gratings. The top row shows patches of isochromatic, red/green 
isoluminant, and S-cone grating modulations that correspond to a 
spatial frequencies of 24, 24, and 15 cpd respectively. The bottom row 
shows the reconstructed stimuli obtained by applying the Bayesian 
algorithm to the cone responses from these gratings. Other details as 
with Figure 15 above.

Figure 17 summarizes the relation between simulated and reconstructed 
stimuli for modulations between 0 and 60 cpd for all three color directions 
investigated. For each stimulus, we computed the projection of the 
reconstructed image onto the stimulus image, after subtraction of the mean 
image. We then took the ratio of the contrast power of the stimulus and 
projected images. This provides a single number measure of how much of 
the input image makes it through the reconstruction process. We plotted 
this measure, which we refer to as the Projection Contrast Sensitivity,
against spatial frequency for all three color directions. Each function was normalized by its value for 0 cpd. The plot falls off most rapidly for the S-cone gratings. This effect is due primarily to axial chromatic aberration, which reduces the optical quality for stimuli seen by S cones relative to those seen by M and L cones.

More interesting is the fact that the red/green function falls off more rapidly than the isochromatic function. This effect is a fundamental consequence of the fact that L and M cone signals are highly correlated in prior distribution used by the Bayesian algorithm. At any spatial frequency, a luminance grating is more likely \textit{a priori} than a red/green grating. In addition, the high correlation between neighboring pixels in the prior makes low spatial frequency stimuli more likely \textit{a priori} than high spatial frequency stimuli. As the spatial frequency of the stimulus gratings increases, the reconstructed power falls off in part because of optical blurring. But this blurring affects isochromatic and red/green isoluminant gratings to about the same degree. The differential effect seen in the figure occurs because the prior, which favors low frequency interpretations in the face of ambiguity introduced by trichromatic sampling, begins to dominate earlier for red/green gratings than for isochromatic gratings.

Although a firm conclusion would require a much more detailed model of grating detection than we have elaborated to date, Figure 17 does suggest the possibility that the more rapid fall off of the red/green grating contrast
sensitivity function relative to the isochromatic contrast sensitivity function (Mullen, 1985; Sekiguchi, Williams, & Brainard, 1993b) is at least in part a fundamental consequence of the statistics of natural images combined with the mosaiced design of the human retina. A more detailed model would need to specify explicitly the site or sites of noise that limit detection relative to the sites that implement grating reconstruction, and should also incorporate considerations of efficient transmission and representation of information (e.g. Buchsbaum & Gottschalk, 1983; Derrico & Buchsbaum, 1990; Atick, Li, & Redlich, 1992; van Hateren, 1993; Ruderman, Cronin, & Chiao, 1998; Parraga, Brelstaff, Troscianko, & Moorehead, 1998; Lee, Wachtler, & Sejnowski, 2002; Parraga, Troscianko, & Tolhurst, 2002; Doi, Inui, Lee, Wachtler, & Sejnowski, 2003; Caywood, Willmore, & Tolhurst, 2004).

**Figure 17. Projection Contrast Sensitivity.** The plot shows the Projection Contrast Sensitivity (see description in text) as a function of grating spatial frequency for isochromatic (black), red/green isoluminant (red), and S-cone isolating gratings (blue). Values shown are the average of values obtained for gratings in sine and cosine spatial phase.

**Predictions**

The fit of the model to the small spot data described above, as well as its predictions for larger spots and sinusoidal contrast modulations, indicates
that the behavior of the model is at least qualitatively consistent with extant experimental observations. At the same time, the modeling approach would be more compelling if it successfully predicted phenomena that have not yet been studied.

Here we outline two such predictions. First, how should the fraction of white responses to small monochromatic spots vary, within observer, with the structure of the mosaic in the neighborhood of the test flash. Second, how should the percentage of blue and green responses for flashes detected by M cones vary, again within observer, as a function of the distance to the nearest S cone. Testing these predictions requires refinement of the experimental procedures to allow recording of the retinal location of the flashed spots on a trial-by-trial basis. We expect that such experiments will soon be feasible.

*Effect of local mosaic properties.* Since in our simulations we know the location of every simulated flash, we binned the data over flashes that landed in similar local neighborhoods and computed the percentage of flashes named white as a function of the neighborhood properties. Figure 18 shows the results of this calculation. The x-axis shows the number of L cones among the 10 cones nearest to the location of the flashed spots in each bin; the y-axis shows the corresponding predicted percentage of flashes named white. Each individual line in the plot was derived from the data for a different observer and spot wavelength. The individual lines do
not span the entire x-axis because there were not enough flashes in every bin to compute a meaningful prediction; we only plotted data for bins that contained more than 10 seen flashes (from simulations of 2000 flashes).

The overall form of the predictions is clear. The number of flashes named white depends strongly on the number of L cones in the local neighborhood, and is highest for asymmetric local neighborhoods (i.e. those with mostly L or mostly M cones.) The number of flashes named white is less sensitive to between observer differences or to flash wavelengths, beyond the fact that it may be difficult for some observers and wavelengths to aggregate sufficient data to trace out the entire function shown.

**Figure 18. Predicted dependence of percent named white on local mosaic.** The figure shows the predicted percentage of spots named white as a function of the local mosaic in the vicinity of the spots. The x-axis is the number of L cones among the 10 cones nearest to the retinal location of the simulated flashed spot. The y-axis is the corresponding percentage of spots named white. Each individual line in the plot shows data derived from simulations of one wavelength and observer. Blue: 500 nm spots; green: 550 nm spots; red; 600 nm spots.

*M cones and blue.* In classical color theory, signals from M cones contribute to the sensation of green through a red/green opponent pathway, and to the sensation of yellow through a blue/yellow opponent pathway (Hurvich, 1981; Kaiser & Boynton, 1996). There have, however, been
suggestions that under some conditions signals from M cones can contribute to the sensation of blue, particularly for small spots (Drum, 1989; DeValois & DeValois, 1993; Schirillo & Reeves, 2001, suggestion attributed to B. Drum by the authors; Hofer, Singer, & Williams, 2005, suggestion attributed to D. MacLeod by the authors). Here we examine the predictions of the Bayesian model in this regard for the color appearance of small flashes.

Figure 19 provides intuition. The figure shows the reconstructed spots obtained for two mosaics when only the central M cone is stimulated. On the left is mosaic where the central M cone is surround by a mix of M and L cones. For this mosaic, the reconstructed spot indeed appears to lie on the blue end of the blue-green range. The intuition here is that the absence of an S cone in the neighborhood of the flash means that the visual system has no information about the short wavelength component of the stimulus. Thus the algorithm will fill this in from the prior distribution. In this distribution, M and S signals are correlated, and a large M cone component is thus most likely to be accompanied by a large S cone component. Hence the large blue component in the reconstructed spot. The presence of nearby L cones, together with the spatial correlations incorporated in the prior, bias the reconstruction away from having a substantial long wavelength component.

The mosaic on the right is identical to the one on the left, with the exception that one of the L cones neighboring the central M cone has been replaced by an S cone. Here the reconstruction is considerably greener.
This occurs because the S cone, together with the spatial correlations in the prior, provides evidence that the short wavelength component of the stimulus is small.

**Figure 19. Effect of nearby S cones on appearance of spots seen by M cones.** The top two panels show two hypothetical mosaics. Each has a central M cone (indicated by a white square) surrounded by a mix of L and M cones. The mosaic on the right is identical to that on the left, with the exception that one of the L cones has been replaced by an S cone. The bottom two panels show the reconstructed spots that result from identical stimulation of the central M cone. The effect of adding the S cone is to shift the appearance of the reconstructed spot from blue to green. Image display of reconstructed images was handled as described in the caption for Figure 8.

Figure 20 illustrates the effect for the overall set of simulations that we performed. Seen flashes such that the nearest two cones to the flash location were M cones were selected from all flashes used in the main simulation (5 observers, 3 wavelengths.) For each flash, if there was an S cone within the 100 cones nearest to the flash location, the distance to the nearest S cone was computed. Flashes were grouped according to this distance in 1 arcminute bins. For each bin, the percent of all flashes named Blue and the percent of all flashes named green were then computed. The plot shows that green responses are more prevalent when an S cone is nearby the flash location and decrease as the distance to the nearest S cone.
increases, while the opposite relation holds for Blue responses. Thus, consistent with some of the earlier suggestions cited above, the Bayesian algorithm predicts that whether an M cone contributes to Blueness or greenness varies, and depends on whether or not there is an S cone nearby.

**Figure 20. Effect of nearby S cones on appearance of spots seen by M cones.** Flashes such that the nearest two cones to the flash location were M cones were selected from all flashes used in the main simulation (5 observers, 3 wavelengths.) For each flash, if there was an S cone within the 100 cones nearest to the flash location, the distance to the nearest S cone was computed. Flashes were grouped according to this distance in 1 arcminute bins. For each bin, the percent of all flashes named blue and the percent of all flashes named green were then computed. The plot shows the results (blue line, percent named blue; green line, percent named green.) Data from a bin were plotted if the bin represented at least 10 flashes.

**Discussion**

**Summary**

This paper describes a Bayesian algorithm (Brainard, 1994) for reconstructing full trichromatic images from the responses of interleaved L, M, and S cones. The algorithm is combined with a model of color naming and provides a good account of measurements of the appearance of very small (0.3 arcminute) monochromatic spots presented using adaptive optics (Hofer, Singer, & Williams, 2005). In particular, the model can account for
the fact that small flashed monochromatic spots elicit a wide variety of color names and for the striking individual variation in naming percentages. The latter is driven by corresponding individual variation in the arrangement and relative number of L, M, and S cones in each observer’s mosaic and does not require any free parameters to describe the individual observer differences.

A key emergent feature of the model is that stimulating a single cone of a particular type should yield a different color sensation, depending on the structure of the mosaic surrounding the stimulated cone (see Figures 8 and 19.) A particular consequence of this feature is that the percent of flashes named white are predicted to vary systematically with the LM cone ratio of the mosaic, in a manner consistent with the experimental data.

The model also accounts for other color appearance phenomena. It correctly predicts that within and between observer variation in naming will decrease dramatically for larger (8 arcminute) spots at suprathreshold levels, and it describes distortions of the appearance of fine spatial gratings while at the same time correctly predicting that coarse gratings will be perceived veridically.

Finally, the model makes potentially testable predictions for how the color appearance of very small spots should depend on the fine structure of the cone mosaic.
Quality of model fit/extending the model

It is worth emphasizing that the model does not provide a perfect account of the experimental data. For example, with the parameters we used the model does not predict that any flashes will be named yellow, while observers in fact use yellow to describe the appearance on some trials. It would be possible to improve the model fit by allowing more parameters to describe the color naming boundaries. For example, we could have parameterized the chromaticity region corresponding to white with an ellipse rather than a circle, allowed the boundaries between chromatic color names to curve instead of being straight lines, or have relaxed the constraint on the location of the color boundaries to allow them to deviate further from their typical positions. Each of these changes would be expected to improve the agreement between predictions and data. In addition, we could have changed our fitting criterion to place a high weight on predicting non-zero percentages for each color named at least once for each observer/wavelength combination. This would be expected to lead to a lower overall correlation between predictions and data, but would likely lead to the model predicting the occurrence of at least one instance of each color name actually used. We did not think that any of these data fitting exercises would yield much additional insight: adding parameters can always improve the fit of a model, and adjusting error criteria can generally be used as a technique to alter key features of predictions. Indeed we see the degree of
agreement between model and data as quite remarkable, given the relative simplicity of the approach. For example, the model does not incorporate any individual differences in color naming boundaries, nor the effects of any source of post-isomerization noise on detection and image reconstruction.

In our view, the value of the model as currently implemented is its ability to clarify the fundamental consequences of interleaved trichromatic sampling on visual performance and relate these consequences to experimental data. In this regard, the model does make predictions that can be tested against future experimental data.

Another interesting future direction is to refine the reconstruction algorithm itself. The present algorithm is based on normal priors and additive normal noise. This choice was based on considerations of analytic and computational convenience. More realistic priors would incorporate a positivity constraint on the reconstructed image (L. Paninski, personal communication) and non-normal features of natural images (Simoncelli, 2005). A Poisson noise model would be more realistic than the normal noise (also L. Paninski, personal communication). It will be interesting to see whether and how the reconstruction algorithm’s behavior varies as these more realistic assumptions are added. Our belief is that the basic structural features of the reconstructions will remain unaltered, and that the changes will be reflected in reasonably subtle features of the performance and predictions.
Relation to other work on optimal color processing

A number of authors have considered optimal visual processing of color information. This approach has been effective in accounting for spectral properties of cone photoreceptors (e.g. Regan et al., 2001), the approximate receptive field structure of retinal ganglion cells and cortical units (e.g. Buchsbaum & Gottschalk, 1983; Derrico & Buchsbaum, 1990; Atick, Li, & Redlich, 1992; van Hateren, 1993; Ruderman, Cronin, & Chiao, 1998; Parraga, Brelstaff, Troscianko, & Moorehead, 1998; Lee, Wachtler, & Sejnowski, 2002; Parraga, Troscianko, & Tolhurst, 2002; Doi, Inui, Lee, Wachtler, & Sejnowski, 2003; Caywood, Willmore, & Tolhurst, 2004; Wachtler, Doi, Lee, & Sejnowski, 2007) and the division of information between ON and OFF pathways (von der Twer & MacLeod, 2001; see also Ratliff, 2007), and the shape of ganglion cells’ static non-linearities (von der Twer & MacLeod, 2001). Here we add a variety of color appearance phenomena to the exemplars that may be understood as a consequence of optimal processing. Our work differs from this earlier work in two important ways. First, of the above papers only a few Doi, 2003 #12948; Wachtler, 2007 #12938} explicitly consider the trichromatic sampling of the cone mosaic, while this is a key feature of our model. Second, rather than optimizing signal-to-noise (or information transmitted), our work emphasizes veridicality of the final perceptual representation.
Interestingly, our work also produces a possible rationale for the difference between isochromatic and chromatic spatial contrast sensitivity functions. This is driven by the properties of the prior we used and their interaction with information loss from trichromatic sampling. The contribution of the prior (or equivalently, statistical properties of natural images) also plays an important role in the earlier work cited above that examines optimizing signal-to-noise/information transmission. An interesting direction will be to merge our approach, with its emphasis on front-end sampling and veridical representation, with the earlier work. A recent paper from the first author’s lab (Bjornsdotter Abrams, Hillis, & Brainard, 2007) takes a step in this direction by considering the relation between the adaptation of early visual mechanisms required to optimize color discrimination on the one hand and the adaptation of the same mechanisms required to stabilize color appearance on the other.

**Neural implications**

The model as we have presented is functional and is based on optimality considerations rather than on known facts about mechanisms in the visual pathways. We noted in Methods that the Bayesian reconstruction may be expressed as $x_{\text{estimate}} = I y_{\text{noisy mosaic}} + i_0$. As emphasized by Brainard et al. (2006), in general the neural implementation of a Bayesian calculation need not resemble its implementation on a digital computer. Here, however, examination of the form of the estimator allows straightforward
interpretation of the algorithm in neural terms. If we neglect the $i_0$ term, we see that the reconstructed image is the weighted sum of a set of basis images, one from each column of the matrix $I$, with the weight on each column simply being the magnitude of the response of the corresponding cone. We know from examination of these basis images that each represents a blurred colored blob, as in the lower panels of Figures 8 and 19. Thus we can conceive of a set of neurons, one for each foveal cone, each of which represents a single basis image. A set of second stage neurons could then pool the output of the first to produce a representation of the Bayesian image reconstruction.

A useful insight gained from this perspective is that the Bayesian model can be understood as associating a localized color percept with each individual cone and then producing an overall percept that is the sum of the individual cone percepts. This is similar to the elemental sensation assumption made in earlier work that analyzed the appearance of small spot colors (Hartridge, 1954; Krauskopf, 1964; Krauskopf & Srebro, 1965; Otake, Gowdy, & Cicerone, 1990; Krauskopf, 2000). This assumption is that a fixed sensation (e.g. red, green, or blue) is associated with stimulation of cones of each class (e.g. L, M, and S.) The elemental sensation assumption is difficult to reconcile with the experimental data of Hofer et al. (2005), and the current Bayesian model indicates that it is not compatible with optimal reconstruction from the interleaved cone mosaic. The possibility that the
elemental sensation assumption does not hold is worth bearing in mind when interpreting conclusions drawn via analyses that incorporate it.

The Bayesian model does suggest a generalization of the elemental sensation assumption, where the percept associated with each cone depends not only on its class but also on the fine structure of the local mosaic in which it is situated. If this more general form of the assumption holds, in which stimulation of any given cone always contributes the same sensation to the overall percept, then we should not expect variability in the color percepts resulting from trials that resulted in the identical array of cone responses. In practice, however, testing this prediction would require accounting for trial-to-trial variability arising from uncontrollable noise sources such as photon noise.

**Acknowledgements**

This work was supported by NIH RO1 EY10016, EY0436 and EY0139, as well as by the National Science Foundation Science and Technology Center for Adaptive Optics, managed by the University of California at Santa Cruz under cooperative agreement No. AST-987673. We thank B. Averbeck, W. Makous, L. Paninski, A. Pouget, J. Schirillo, and B. Wandell for useful discussions and comments. M. Kahana and the Penn Psychology Department for provided access to a computer cluster.
References


<table>
<thead>
<tr>
<th>Symbol</th>
<th>Meaning</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>( N_{\text{pixel}} )</td>
<td>Linear size of simulated image, pixels*</td>
<td>100 - 101</td>
</tr>
<tr>
<td>( K_{\text{image}} )</td>
<td>Size of simulated image, minutes*</td>
<td>11.4 - 12.6</td>
</tr>
<tr>
<td>( K_{\text{spot}} )</td>
<td>Simulated spot size, minutes</td>
<td>0.3</td>
</tr>
<tr>
<td>( K_{\text{cone aperture}} )</td>
<td>Cone aperture, minutes</td>
<td>0.9</td>
</tr>
<tr>
<td>( x_{\text{spot}}, y_{\text{spot}} )</td>
<td>Location of a particular simulated spot</td>
<td></td>
</tr>
<tr>
<td>( I_{\text{spot}} )</td>
<td>Simulated spot intensity</td>
<td></td>
</tr>
<tr>
<td>( \lambda_{\text{spot}} )</td>
<td>Simulates spot wavelength, nm**</td>
<td>500, 550, 600</td>
</tr>
<tr>
<td>( N_{\text{cones}} )</td>
<td>Number of cones in image area*</td>
<td>128 - 216</td>
</tr>
<tr>
<td>( N_{\text{neighbors}} )</td>
<td>Number of cone neighbors in detection model</td>
<td>4</td>
</tr>
<tr>
<td>( y_{\text{criterion}} )</td>
<td>Criterion in detection model</td>
<td>10</td>
</tr>
<tr>
<td>( y_{\text{mean mosaic}} )</td>
<td>Vector of mean cone responses</td>
<td></td>
</tr>
<tr>
<td>( y_{\text{noisy mosaic}} )</td>
<td>Vector of noisy cone responses</td>
<td></td>
</tr>
<tr>
<td>( y_{\text{pooled}} )</td>
<td>Pooled cone response in detection model</td>
<td></td>
</tr>
<tr>
<td>( p(x) )</td>
<td>Prior over trichromatic images</td>
<td></td>
</tr>
<tr>
<td>( p(y</td>
<td>x) )</td>
<td>Likelihood of cone responses, given trichromatic image</td>
</tr>
<tr>
<td>( p(x</td>
<td>y) )</td>
<td>Posterior over trichromatic images, given cone responses</td>
</tr>
<tr>
<td>( x_{\text{estimate}} )</td>
<td>Vector representing estimated trichromatic image</td>
<td></td>
</tr>
<tr>
<td>( u_x )</td>
<td>Prior mean vector for trichromatic images ( X )</td>
<td></td>
</tr>
<tr>
<td>( K_x )</td>
<td>Prior covariance matrix for trichromatic images ( X )</td>
<td></td>
</tr>
<tr>
<td>( u_{\text{space}} )</td>
<td>Prior mean of spatial prior at each location</td>
<td>1</td>
</tr>
<tr>
<td>( \sigma_{\text{space}}^2 )</td>
<td>Prior variance of spatial prior at each location</td>
<td>0.25</td>
</tr>
<tr>
<td>( \rho_{\text{space}} )</td>
<td>Nearest neighbor correlation of spatial prior</td>
<td>0.85</td>
</tr>
<tr>
<td>( N_{\text{linmod}} )</td>
<td>Dimensionality used in approximation of spatial prior</td>
<td>20</td>
</tr>
<tr>
<td>( u_{\text{u'v'}} )</td>
<td>Mean of color prior in CIE u'v' chromaticity</td>
<td>( [0.198, 0.472] )'</td>
</tr>
<tr>
<td>( \sigma_{\text{u'v'}}^2 )</td>
<td>CIE u' and v' variance of color prior</td>
<td>0.08</td>
</tr>
<tr>
<td>( K_{\text{color}} )</td>
<td>Factor relating ( \sigma_{\text{factor}}^2 ) and ( \mu_{\text{factor}} )</td>
<td>8</td>
</tr>
</tbody>
</table>

* Range provided represents range across observers.
** Values given represent values used across experimental conditions.

**Table 1: Symbols and typical values**
Supplemental Figure Captions

Supplemental Figure S1. Frequency of seeing for observers and detection model. For each observer and wavelength, the figure shows a comparison between the frequency of seeing curve predicted by our detection model and measured frequency of seeing data. Predictions generated for $N_{\text{neighbors}} = 5$ and $y_{\text{criterion}} = 10$. The quality of the predictions was not highly sensitive variation of these parameters around the chosen values. For each observer and wavelength, a scale factor linking model flash intensity (arbitrary) units to physical flash intensity was chosen to optimize the agreement between predictions and data. Blue: 500 nm; green: 550 nm; red: 600 nm. Observers indicated in the inset of each panel.

Supplemental Figure S2. Frequency of seeing for observers and detection model. For each observer and wavelength, the figure shows a comparison between the frequency of seeing curve predicted by our detection model and measured frequency of seeing data. Predictions generated for $N_{\text{neighbors}} = 5$ and $y_{\text{criterion}} = 10$. The quality of the predictions was not highly sensitive variation of these parameters around the chosen values. For each observer, a single scale factor linking model flash intensity (arbitrary) units to physical flash intensity was chosen to optimize the agreement between predictions and data simultaneously for all three flash wavelengths. Blue: 500 nm; green: 550 nm; red: 600 nm. Observers indicated in the inset of each panel.
Figure 2
Figure 3
Figure 4
Small monochromatic spots

Adaptive optics

Retinal images

Sampling by mosaic

Cone responses

Detection model

Cone responses of seen spots

Reconstruction algorithm

Trichromatic reconstruction of seen spots

Map chromaticities to names

Color names of seen spots

Figure 5
Figure 6
Figure 7

- $p(x)$
- $p(x|y)$
- $p(y|x)$

$L$ and $M$ labels on the diagrams.
Figure 10
Figure 12

The scatter plot shows a linear relationship between the percent predicted and the percent named, with a correlation coefficient of $r = 0.82$. The data points are color-coded, with different colors indicating different categories.
Figure 16
Figure 19
Figure 20
Supplemental Figure S1
Supplemental Figure S2