A COMPUTATIONAL FRAMEWORK FOR MODELING EMERGENCE OF COLOR VISION IN THE HUMAN BRAIN

Atsunobu Kotani & Ren Ng
Department of Electrical Engineering and Computer Sciences
University of California, Berkeley
{akotani,ren}@berkeley.edu

ABSTRACT

It is a mystery how the brain decodes color vision purely from the optic nerve signals it receives, with a core inferential challenge being how it disentangles internal perception with the correct color dimensionality from the unknown encoding properties of the eye. In this paper, we introduce a computational framework for modeling this emergence of human color vision by simulating both the eye and the cortex. Existing research often overlooks how the cortex develops color vision or represents color space internally, assuming that the color dimensionality is known a priori; however, we argue that the visual cortex has the capability and the challenge of inferring the color dimensionality purely from fluctuations in the optic nerve signals. To validate our theory, we introduce a simulation engine for biological eyes based on established vision science and generate optic nerve signals resulting from looking at natural images. Further, we propose a bio-plausible model of cortical learning based on self-supervised prediction of optic nerve signal fluctuations under natural eye motions. We show that this model naturally learns to generate color vision by disentangling retinal invariants from the sensory signals. When the retina contains N types of color photoreceptors, our simulation shows that N-dimensional color vision naturally emerges, verified through formal colorimetry. Using this framework, we also present the first simulation work that successfully boosts the color dimensionality, as observed in gene therapy on squirrel monkeys, and demonstrates the possibility of enhancing human color vision from 3D to 4D.

1 INTRODUCTION

"Color is the place where our brain and the universe meet." – Paul Klee

We experience colors in everyday life so effortlessly that it is easy to take the underlying neural computations for granted. In fact, the sensory signals exiting our eye, called optic nerve signals (ONS), are nothing like our color vision (see Fig. 1 and Supplementary Video 0:30 [1]). For example, ONS are spatially warped, akin to an image taken with a fish-eye lens, due to varying densities of photoreceptor cells in the retina (Curcio et al., 1990). ONS does not come in color either – colors of the scene are spectrally sampled by different types of color sensitive cells (cone cells) in the retina, appearing as a layer of spatial noise in the ONS. Furthermore, other processes, such as lateral inhibition and action potentials, render the image structure barely recognizable, in gradient domain where spatiotemporal "edges" dominate. Now, the question is: *how are we still seeing colors?*

Specifically, this paper introduces a novel computational framework for modeling the emergence of human color vision by simulating the eye and the cortex. For the eye, we present a biophysically accurate implementation of a textbook scientific model of the retinal neural circuitry. For the brain, we hypothesize a low-level, self-supervised learning mechanism in the cortex that operates purely on the optic nerve signal stream. For color representation in the brain, we propose modeling color in the brain as a high-dimensional vector, rather than assuming any specific color dimensionality and show that the correct color dimensionality emerges naturally through the proposed learning.

For eye simulation, our goal is to create a computational engine that takes in any spectral image of the world and outputs the corresponding optic nerve signal stream, based on established vision science.

[1] Code, video, dataset, and tutorials are available on our project website: https://matisse.eecs.berkeley.edu.

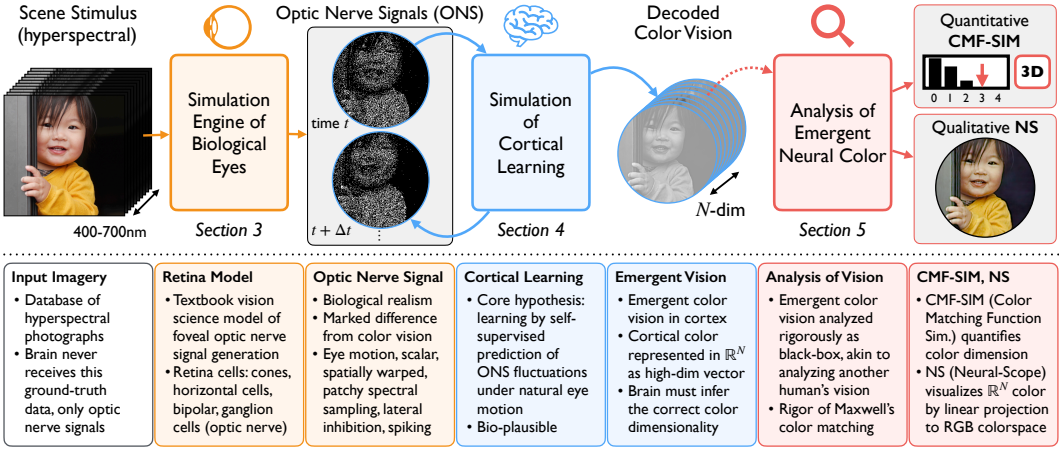

Figure 1: Overview of our proposed framework for modeling the emergence of human color vision. Our simulation engine of biological eyes converts a scene stimulus (hyperspectral image) to a stream of optic nerve signals (Section 3 & Video 0:30). We simulate cortical learning purely from these optic nerve signals (Section 4) and show the emergence of color vision. We show how to analyze the emergent neural color quantitatively with Color Matching Function test Simulator ($CMF\text{-}SIM$) and qualitatively with Neural Scope (NS) (Section 5).

Our model captures how scene form, color, and motion become spatiotemporally entangled in the optic nerve signal stream. Our simulation is based on the "textbook" model of vision science (Rodieck, 1998) for midget, private-line visual pathways, detailed in Section 3, and comprising fixational eye motion, spectral encoding by cone cells, foveation, and lateral inhibition.

For brain simulation, we hypothesize that the cortex could disentangle the optic nerve signal from the invariant retinal properties to generate color vision – purely through a self-supervised learning process that aims to predict the constant fluctuations in cellular-level activations of the optic nerve signal during small eye movements. The neural conditions for such self-supervised learning are biologically plausible in the sense that the cortex continuously receives optic nerve signals under the tiny gaze movements of fixational eye drift.The theoretical intuition for why such learning might succeed is that eye motion repeatedly draws a static scene image across the invariant spectral and spatial sampling properties of the retina, potentially enabling the retinal properties and scene images to be mutually filtered out of the optic nerve signal where they are entangled. We show that this simple learning mechanism succeeds at discovering color vision with the correct dimensionality.

But first, a somewhat esoteric yet technically critical feature of the modeling framework needs discussion: color representation in the brain. Existing research often overlooks how the cortex develops color vision or represents color space internally, assuming that the color dimensionality is known a priori, e.g. RGB. We argue that the cortex has the capability and the task of inferring color dimensionality, purely from fluctuations in the optic nerve signals. Therefore, we propose representing color in the cortex as a high-dimensional vector in $\mathbb{R}^N$ and find that the correct color space and color dimensionality emerge naturally as geometric properties of the hypothesized learning mechanism. We show how to formally quantify and visualize the emergent color space (Section 5).

Remarkably, the hypothesized learning mechanism results in clear color vision of the correct color dimensionality. When the retina contains 1, 2, 3 or 4 types of color photoreceptors (cones), correct color dimensionality emerges: respectively, 1D mono-, 2D di-, 3D tri- or 4D tetra-chromatic vision. In fact, it is an esoteric but well-known fact in vision science that the three types of 2D dichromacy result in highly specific color-spaces (blue-yellow, blue-orange, and teal-pink), which we see emerge naturally. Even more, the simulation presents a model of how color dimensionality boosting occurred in the squirrel monkeys (Mancuso et al., 2009). We simulate injection of the third cone pigment virus into the dichromat retina, which results in boosting from dichromacy to trichromacy. Intriguingly, the model also shows the possibility that normal human trichromacy could be boosted to tetrachromacy.

In sum, our proposed framework formulates the emergence of human color vision in a computational manner. The simulation engine of realistic optic nerve signals generates training data to this problem,

and an intentional and challenging constraint is that the cortical model must strictly learn only from the optic nerve signal with no auxiliary information. This paper presents the first, simple and yet complete existence proof that such cortical inference is possible, and we show that this learning simulation is consistent with various surprising and unexplained vision science phenomena.

2 Related Work

2.1 Vision Science on Optic Nerve Signal Encoding

In the "textbook" model, the retina transforms light into electrical signals through three primary functions: color sampling via cone cell spectral response functions, lateral inhibition via horizontal neural connections, and spatial sampling via cell positioning (Fig. 2). Under daylight, this process begins when light activates cone cells, which are of three types for most humans (Young, 1801; 1802; Von Helmholtz, 1867), each sensitive to different wavelengths (Stockman & Sharpe, 2000). These signals are then modulated by horizontal cells that enhance visual contrast through lateral inhibition (Rodieck, 1965; Dacey et al., 1996; Verweij et al., 2003). The signals continue to bipolar and retinal ganglion cells (RGCs), with a direct connection in the fovea via the midget private-line pathway (Dowling & Boycott, 1966; McMahon et al., 2000; Wool et al., 2018), vital for high-resolution color vision. Notably, the cone cell density varies, reaching its peak in the fovea (Osterberg, 1935; Curcio et al., 1990). The axons of the ganglion cells bundle to form optic nerve signals, maintaining their spatial arrangement, which results in a retinotopy in the visual cortex (Holmes, 1918; Tootell et al., 1982; Dougherty et al., 2003). Fixational eye movements cause dynamic photoreceptor activation at all times by constantly shifting the gaze, even when focusing on static objects (Rucci & Victor, 2015; Young & Smithson, 2021; Martinez-Conde et al., 2013).

Recordings of real optic nerve signals exist, but cannot be used for our cortical simulations because they are orders of magnitude too low resolution (from only a few thousand cells) and only in response to grayscale rather than color imagery (Litke et al., 2004; Brackbill et al., 2020; Marre et al., 2017; Liu et al., 2022). To overcome these data limitations, we present a simulation engine for the midget private-line pathway in the fovea in Section 3.

2.2 Color Representation in Computational Neuroscience

It is an open question how to meaningfully model neural representations of color in simulations of visual perception. Most computational neuroscience sidesteps this issue, "hard-coding" a dimensionality of 3, representing and constraining cortical color to tristimulus values, such as RGB (Parthasarathy et al., 2017; Botella-Soler et al., 2018; Brackbill et al., 2020; Kim et al., 2021; Zhang et al., 2022; Wu et al., 2022), LMS (Young, 1802; Von Helmholtz, 1867) or cone-opponent space (Derrington et al., 1984; MacLeod & Boynton, 1979). Constraining cortical models to such a ceiling of three for dimensionality clearly conflicts with the study of a functional tetrachromat observer with 4D color vision (Jordan et al., 2010; Rezeanu et al., 2021). Instead, we model the brain with the capability and the challenge of deducing the inherent color dimensionality encoded in the optic nerve signals.

2.3 Theory on How Vision Emerges in the Brain

In computational vision science modeling, it is often overlooked that the cortex relies solely on a stream of optic nerve signals to discover color vision, with no access to a teacher signal or perceptual ground truth. Rather, various efforts have been made to reconstruct visual stimuli from neural responses by giving the cortical model access to ground truth stimulus image (Naselaris et al., 2009; Nishimoto et al., 2011; Parthasarathy et al., 2017; Botella-Soler et al., 2018; Brackbill et al., 2020; Kim et al., 2021), but the neural reality is that the brain never has direct access to the visual scene. This characteristic makes the human visual system a quintessential example of self-supervised learning, which is a growing field in computer vision (De Sa, 1993; Chen et al., 2020; He et al., 2022). In this work, we propose a learning principle in the cortex which aims to predict the cellular-level fluctuations in activation that occur during small eye movements, which is associated with the idea of temporal prediction (Palmer et al., 2015; Lotter et al., 2016; Singer et al., 2018; 2023b), as well as the broader concept of predictive coding (Rao & Ballard, 1999; Srinivasan et al., 1982). It is also closely linked to the sensorimotor contingency theory from cognitive science (O'Regan & Noë, 2001) and

slow feature analysis (Hinton, 1990; Földiák, 1991) which suggests that the brain learns to anticipate the sensory outcomes of motor actions (e.g. eye movements) and filters out invariant (slow) features.

2.4 Theory on How Brains Infer Color Dimensionality

Previous studies has investigated inference of invariant retinal properties from sensory signals. For instance, research has demonstrated the ability to deduce cone cell types (Wachtler et al., 2007; Brainard et al., 2008; Benson et al., 2014) and the positions of photoreceptors (Maloney & Ahumada, 1989; Brainard et al., 2008) via statistical methods from sensory signals. Brainard et al. (2008) hinted at analyzing sensory inputs from different time points during fixational drift as a way to reveal retinal features, and in this paper we provide a computational realization of this idea with a specific learning mechanism that achieves complete disentanglement of retinal invariants from optic nerve signals.

2.5 Measurement of Color Dimensionality in Human Color Perception

In this paper, we need to rigorously measure the color dimensionality of the emergent color from cortical simulation. To do so, we adapt Maxwell's famous color-matching experiments (Maxwell, 1856), which laid the foundation of colorimetry that remains at the heart of all color reproduction technology today. Color matching experiments confirmed the trichromatic theory (Young, 1802; Von Helmholtz, 1867; Grassmann, 1853) in which the 3-dimensional nature of human color vision has its basis in the three different cone types in the human retina. Jacobs's recent review (Jacobs, 2018) of color dimensionality in animal vision, however, reminds the community that the dimensionality of color vision does not automatically equal the number of cone types, and that the most rigorous way to measure it remains Maxwellian color matching – as we follow in this work.

2.6 Complexity of Human Color Vision

This paper focuses on color dimensionality because it is the foundational characteristic of an observer's color experience, but many layers of further perceptual complexity exist atop that foundation. Examples include chromatic adaptation (color constancy) (Von Kries, 1902; Land, 1977; Foster, 2011), perceptual nonuniformity across colorspace (CIE, 1976; MacAdam, 1942), complex contextual interactions (Fairchild, 2013), and even surprising flood-fill features (Pinna, 1987; Pinna et al., 2001). Parts of these perceptual phenomena are scientifically mapped to neural correlates, such as nonlinear photoreceptor responses (Krauskopf & Karl, 1992; Angueyra et al., 2022), or parvocellular pathway and neural processes spanning V1, V2, and V4 (Livingstone & Hubel, 1987; Zeki, 1978; Li et al., 2014; Liu et al., 2020; Angueyra et al., 2022). This paper leaves these additional layers of perceptual complexity as future modeling work.

3 Simulation Engine for Biological Eyes and Optic Nerve Signals

We model the primary functions of the human retina based on the known science. The inputs to our retina model are hyperspectral images (Fig.2.A; details in Appendix A.1). A model of fixational eye drift (Rucci & Victor, 2015; Young & Smithson, 2021; Martinez-Conde et al., 2013) generates a sequence of frames that sample different parts of the image. Each frame is projected on the retina, stimulating a randomized array of cone cells, according to known cell density variation as a function of eccentricity (distance from the fovea) (Curcio et al., 1990) and known statistics of different cone cell types (Carroll et al., 2000) (Fig. 2.B). Each cone cell converts the scene light into photoreceptor activations based on cone type spectral response functions (Stockman et al., 1999; Stockman & Sharpe, 2000) (Fig. 2.C), followed by lateral inhibition from horizontal cells (Wool et al., 2018; Rodieck, 1965) (Fig. 2.D). These signals are transformed into spike trains (Fig. 2.E), then bundled into optic nerve signals (ONS), resulting in spatial distortion (Fig. 2.F), turning a color scene stimulus into a noisy, spatially distorted ONS. Additional details are in Appendix A and Video 1:35.

One observation to make is that there are three invariant properties of the simulated retina (i.e. retinal invariants): cell positions, cell types, and horizontal cell connections. These properties are held constant during generation of ONS for a particular eye, but we use the engine with different values for these properties to generate ONS datasets for a diverse set of eyes. For example, we create datasets with different cone types, including monochromatic (L, M, S), dichromatic (LM, LS,

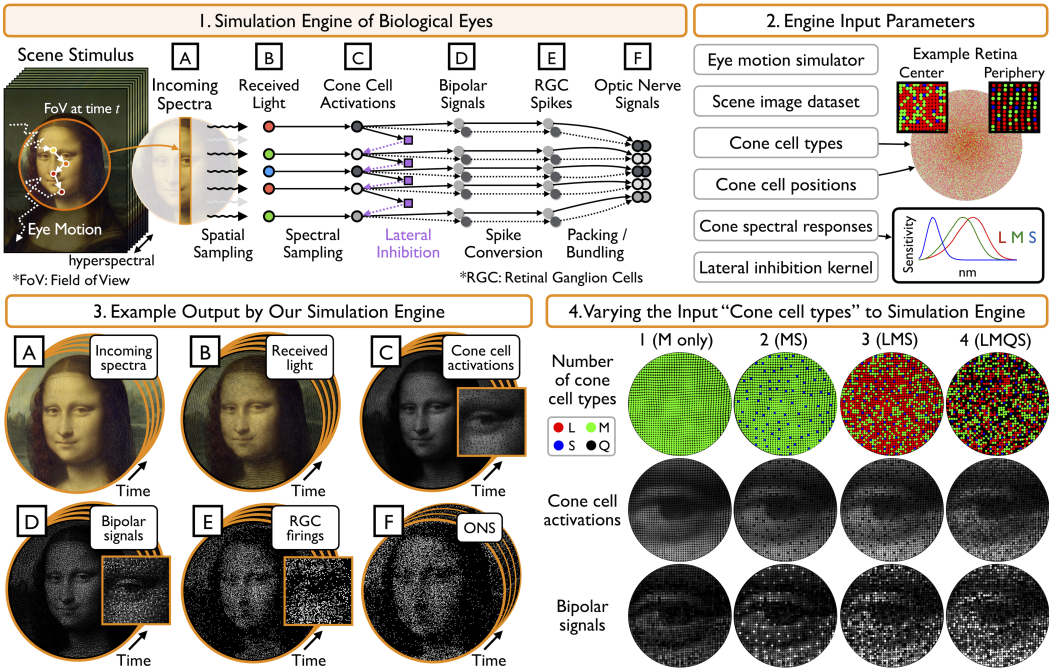

Figure 2: Overview of our simulation engine of biological eyes. 1. This engine takes a scene stimulus as an input, processes it through a "textbook" model of eye motion and retinal neural circuitry, to generate a stream of optic nerve signals. 2. This engine accepts custom eye and retina parameters. 3. It allows visualization of neural signals in steps (A-F), illustrating the progressive entanglement of scene imagery with retinal properties. 4. Visualization of changing one of the input parameters, the number of cone cell types – showing that the signals become noisier as the number increases.

MS), trichromatic (LMS) and tetrachromatic (LMSQ) configurations (Fig. 2.4). The generated ONS becomes spatially noisier as the number of cone types increases, and we study how the cortex infers the inherent color dimensionality of these eyes purely from the differences in their respective ONS.

4 Simulating Cortical Learning and Emergence of Color Vision

Our cortical model is structured to learn three functions in a pipeline (Fig. 3.1): 1. Φ that decodes the optic nerve signal at time t into its internal percept; 2. Ω that translates this percept according to eye motions inferred from the signal over a short time interval, dt; and 3. Ψ that re-encodes the translated percept back into a predicted optic nerve signal at time $t + dt$, which is compared against the real signal received at that time. Mathematically, given the optic nerve signal O_t at time t;

$$\hat{O}_{t+dt} = \Psi(\Omega(\Phi(O_t))) \tag{1}$$

where $\hat{O}_{t+dt}$ is the predicted ONS at time $t + dt$. Here the task of decoder Φ is analogous to inverting the retinal processes to transform O_t into the visual percept image V_t (i.e. $V_t \leftarrow \Phi(O_t)$). Likewise, the re-encoder function Ψ resembles the retinal processes, as it aims to reproduce an optic nerve signal from the visual percept (i.e. $O_t \leftarrow \Psi(V_t)$). Therefore, Φ and Ψ are pseudo-inverses.

The learning objective is to minimize the prediction error $E_{\text{prediction}}$, the difference between predicted and real optic nerve images at time $t + dt$, such that:

$$E_{\text{prediction}} = \|O_{t+dt} - \hat{O}_{t+dt}\|_2^2 = \|O_{t+dt} - \Psi(\Omega(\Phi(O_t)))\|_2^2 \tag{2}$$

where O_{t+dt} is the real observed ONS at time $t + dt$.

Three main ideas drove the evolution of our selected model features. First, we reasoned that it would be a big step towards successful decoding and re-encoding if the cortex could infer the key encoding properties of each cone cell. We gave the cortical model sets of learnable parameters, which we

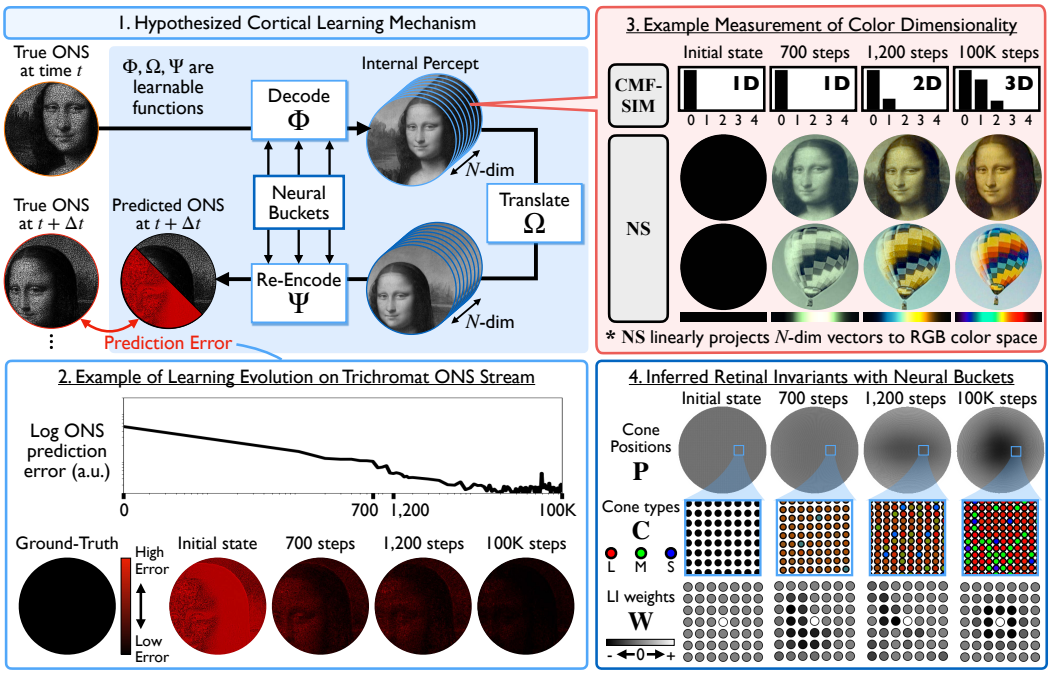

Figure 3: Overview of our hypothesized cortical learning mechanism and exclusive study of the learning behavior of the cortical model with trichromatic retina. 1. Given the stream of optic nerve signals as the only input data, the cortical model aims to predict the next ONS from the current one with 3 learnable functions, decoder Φ, translation operator Ω and re-encoder Ψ. 2. Prediction error decreases as learning progresses, converging after 100K learning steps. 3. During learning, the color dimensionality of the internal percepts transition from 1D, 2D to 3D, formally measured by *CMF-SIM* and visualized by *NS* (Fig. 4 & Section 5). 4. The cortical model infers the retinal invariant properties during learning: cell positions P (higher density in fovea), cone cell types C, and lateral inhibition weights W (center-surround receptive field).

call "neural buckets", in which to store and update guesses of these properties during learning. The buckets contain the following information for each cone cell: 2D position in visual space (P), cone spectral identity (C), and lateral inhibition weights (W) to neighboring cones.

The second main idea was the observation that decoder Φ and encoder Ψ are mathematically factorizable into a pipeline of subfunctions, such that:

$$\Phi = \Phi_{\mathrm{P}} \circ \Phi_{\mathrm{C}} \circ \Phi_{\mathrm{W}}$$
$$\Psi = \Psi_{\mathrm{W}} \circ \Psi_{\mathrm{C}} \circ \Psi_{\mathrm{P}}.$$

Each sub-function is an operator conditioned on its corresponding neural bucket, cell positions P, cone spectral types C, and lateral inhibition weights to neighboring cells W. In case of decoder Φ, it first executes inversion of lateral inhibition using Φ_{W} in order to estimate the activations of each photoreceptor associated with an optic nerve axon; second, projects scalar cone activations into $\mathbb{R}^N$ by Φ_{C} using inferred spectral identities in C (and interpolating color across space – the third main idea below); and finally inverting the spatial distortion of foveation via Φ_{P}. The re-encoding function is a pipeline of analogous subfunctions in reverse order: re-applying spatial warping with Ψ_{P}; re-projecting color into scalar photoreceptor activations with Ψ_{C}; and re-applying lateral inhibition with Ψ_{W}. Further implementation details are provided in Appendix B.

The third main idea was that, in order to accurately re-encode after translation, the model needs to learn to interpolate color information spatially, because there is only one cone type at each point on the retina. This need to interpolate is analogous to the situation in cameras with image sensors (Bayer, 1976; Fossum, 1997; Kimmel, 1999) that physically sample only one of the R,G,B channels at each pixel, and fundamentally require demosaicking algorithms to interpolate full R,G,B values at all pixels. The required interpolation function in the cortical model is more complex because the spectral

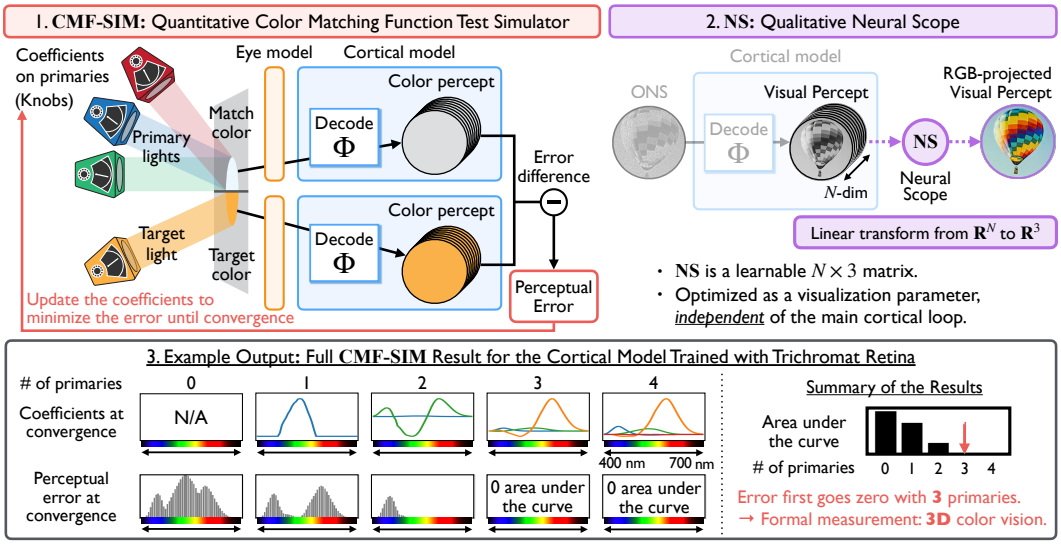

Figure 4: Overview of our measurement methods of emergent color dimensionality, Color Matching Function Test Simulator (*CMF-SIM*) & Neural Scope (*NS*). 1. *CMF-SIM* determines the minimum number of primary colors needed to match any target color by iteratively updating coefficients to minimize perceptual error. 2. *NS* visualizes visual percepts independently of the cortical learning loop, optimized as a learnable $N \times 3$ matrix to minimize projection error to the target RGB image. 3. Example *CMF-SIM* output for a trichromat retina-trained cortical model shows matching errors for 400–700nm spectral light converging to zero with three primaries, confirming 3D color vision.

sampling pattern is random, so learning this function is entangled with correctly resolving C. We enable and force the cortical model to learn the color interpolation function by representing it as a convolutional neural network with neural bucket parameters D (Appendix B.2.2).

An important detail is the handling of eye motion in the simulation. In one experiment we show that the model can learn a subfunction that estimates the eye motion translation between times t and $t + dt$ purely from the optic nerve signals at those times. The main computational challenge here is the spatial warp in optic nerve signals. The uniform translation in stimulus space corresponds to a non-uniform translation in the optic nerve signal space, which makes the prediction of the eye motion dependent on the inference of cell positions P. We find that the cortical model iteratively updates the neural bucket P from imperfect initial eye motion estimate, which helps to improve the prediction of eye motion, and vice versa. This inference converges to the correct eye motion estimate, as the model learns to minimize prediction error (Appendix B.3).

With this pipeline of subfunctions and associated neural buckets, the hypothesized learning is equivalent to parallel numerical optimization of all neural buckets in striving to minimize prediction error. We simulate learning using stochastic gradient descent (Kingma & Ba, 2014).

5 Neural Representation of Color Space and Analysis of Emergent Color Dimensionality

We model color in the cortex as a vector in high-dimensional space, $\mathbb{R}^N$. This decision represents our view that the brain has both the freedom and the challenge of somehow inferring the intrinsic color dimensionality of the visual signals it is receiving along the optic nerve. Specifically, we define a cortical decoding function Φ that takes the optic nerve signal O_t at time t and transforms it into its internal visual percept V_t (Fig. 3), such that:

$$V_t = \Phi(O_t) \tag{3}$$

where each pixel in V_t is a N-dimensional vector. Inside $\mathbb{R}^N$, we assume that cortical color space emerges as a K-dimensional manifold. To measure this intrinsic dimensionality K, we introduce two methods: formal, numerical *CMF-SIM*; and intuitive, visual *NS*.

CMF-SIM (Color Matching Function Test Simulator) is our tool to formally quantify the color dimensionality that has emerged in the $\mathbb{R}^N$ color space of the cortical model (Fig. 4.1). *CMF-SIM* treats the retina model and cortical model as a black-box color observer. Specifically, we limit ourselves to showing the model two patches of scene color at random, distinct locations on the retina to reflect the effect of real-world viewing conditions, obtaining only a scalar score as feedback to indicate the difference in color appearance betweem the two patches (with zero representing a color match). This interface is intentionally limited and identical to color matching experiments with human subjects. Then, resting on the formal, technical bedrock for colorimetry established by Maxwell (1856) and Grassmann (1853), we exhaustively probe to determine the minimum number of color primaries needed to match any test color through linear combination (Fig. 4.1 & Appendix C). For example, Maxwell found that most humans require 3 primaries and are formally trichromatic, but red-green colorblind persons require only 2 primaries and are dichromatic. We model diverse color observers and measure the color dimensionality of their emergent vision.

NS (Neural-Scope) is our tool to display emergent colors in $\mathbb{R}^N$ by projecting them into RGB color space (Fig. 4.2). *NS* is a learnable $N \times 3$ matrix mapping the emergent cortical color manifold in $\mathbb{R}^N$ linearly to conventional RGB color, enabling visual inspection of emergent color vision. To compute *NS*, we first project our hyperspectral image dataset in two ways: (1) using the retina and learned cortical model to map hyperspectral images into the $\mathbb{R}^N$ cortical colorspace, and (2) using conventional color processing to convert hyperspectral data to RGB via CIEXYZ (CIE, 1931). *NS* is then determined as the least squares transform from the former data in $\mathbb{R}^N$ to the latter in RGB. Importantly, *NS* is independently optimized as a visualization parameter, separate from the main cortical loop, ensuring that the cortex processes ONS without any exposure to input images in either hyperspectral or RGB form. *NS* provides striking color visualizations and complementary intuition visually, which are fully consistent with formal *CMF-SIM* results. For example, *NS* allows us to compute and contrast the color palettes of the three different types of human dichromacy (Fig. 5).

In sum, we model cortical color as vectors in $\mathbb{R}^N$, allow color space to emerge naturally through the hypothesized learning mechanism, and measure the emergent color space's dimensionality quantitatively with *CMF-SIM* and analyze it visually with *NS*.

6 Simulation Results - Emergence of Color Vision

Figure 3.2 begins with a model of typical human retina containing L, M and S cones, and illustrates the time-varying behavior of the cortical model as it learns color vision. The visualized prediction error decreases as the training progresses, and the internal percept converges to 3-D color vision both formally (*CMF-SIM*) and visually (*NS* imagery), shown in Figure 3.3. Notably, the visual timeline highlights that the cortical model learns color vision one dimension at a time: achieving monochromacy at 700 learning steps, dichromacy at 1,200 steps, and converging to trichromacy after 10^6 steps. At convergence, the cortical model has accurately inferred all retinal properties: spectral identity of each cell, cell positions and lateral inhibition neighbor weights (Fig. 3.4). For a cortical model trained with a trichromat retina, *CMF-SIM* results closely align with human psychophysical data (Stiles & Burch, 1955) (Appendix G.1) and the model consistently demonstrates 3D color vision across different noise initializations (Appendix G.2), highlighting our model's validity and robustness.

The remainder of this section shows *CMF-SIM* and *NS* results at convergence for a diversity of simulation scenarios. Fig. 4.2 dissects the *CMF-SIM* analysis for the case where the retina contains 3 cone types. The graphs show color matching function results using optimal sets of color primaries from 0 to 4 primaries, along with the residual perceptual error as a function of wavelength. The area under the curve (AUC) for each error graph suggests that the error falls to near-zero only with at least 3 primaries – this formally proves that the color dimensionality is 3.

Figure 5 presents results of the hypothesized cortical learning in a diversity of color observers where the retina contains different numbers of cone cell types. This shows that the model learns K-dimensional color vision when the retina contains K cone types. That is, when the retina contains 1, 2, 3 or 4 cell types, the cortical model converges on mono, di, tri, or tetrachromat color vision, formally quantified with *CMF-SIM* (further analysis of tetrachromat models in Appendix F). And qualitatively, we observe that the *NS* images are grayscale for $K = 1$, colorblind with only shades of blue and yellow for $K = 2$, and full color with all trichromatic hues only with $K = 3$ (Fig. 5.1). Use of *NS* is limited to color dimensionality up to 3, and is to not applicable to $K = 4$ case, but

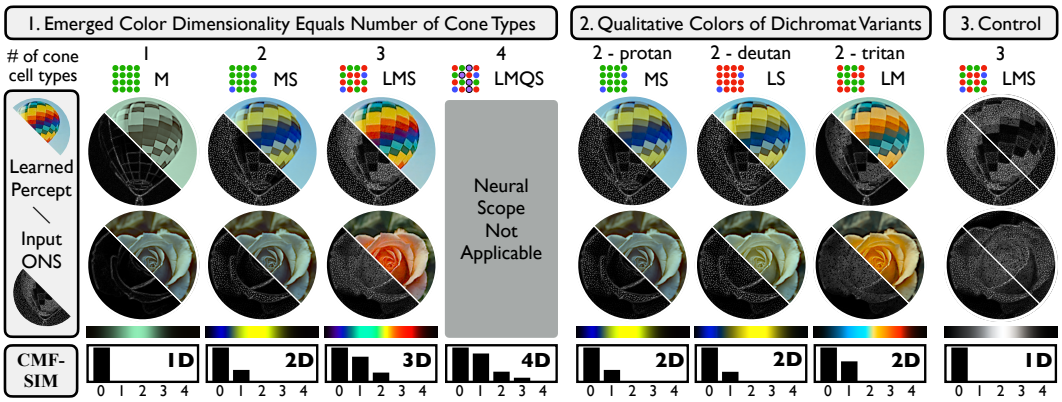

Figure 5: Results of simulating emergence of color vision from various retinas, with analysis of learned color dimensionality using qualitative visualization (NS) and formal methods ($CMF\text{-}SIM$). 1. Cortical models trained with dataset generated with retinas containing 1, 2, 3, 4 cone types result in mono-, di-, tri, tetrachromatic color vision, respectively. 2. Qualitative color of dichromat variants is consistent with known vision science on color vision deficiency. 3. Control experiment with a trichromat retina, but with cortical learning deliberately removed: $CMF\text{-}SIM$ measures color as 1-D, highlighting that cortical learning is necessary for emergence of correct color dimensionality.

$CMF\text{-}SIM$ formally confirms that 4-dimensional color emerges there. All variants of retinas with two cone types (protanopia, deuteranopia and tritanopia), converge to 2D color vision. But more striking, NS reveals hue shifts among these models that are consistent with color vision deficiencies studies (Brettel et al., 1997) (yellow-blue hues for deuteranopia / protanopia (Judd, 1948; Graham & Hsia, 1959) and teal-pink hues for tritanopia (Alpern et al., 1983)) (Fig. 5.2).

In a control experiment, we verify that cortical learning is essential for color vision by comparing two scenarios: (1) a baseline where optic nerve signals directly form percepts, with all cortical learning deliberately removed (Fig. 5.3), and (2) the proposed model including cortical learning (Fig. 5.1, LMS case). The control baseline experiment fails standard color-matching tests, reducing color dimensionality to 1-D despite 3 cone types, demonstrating that cortical learning is indeed necessary for emergence of correct color dimensionality (details in Appendix D). In contrast, the proposed learning-based model results in correct 3-D color, as expected.

Figure 6 simulates boosting of color dimensionality in adulthood by gene therapy. Previous genetic studies (Jacobs et al., 2007; Mancuso et al., 2009; Zhang et al., 2017) demonstrated that the introduction of an additional class of photopigments in the mammalian cone mosaic, even in adulthood via gene therapy, resulted in a new dimension of chromatic sensory experience. Here, we simulate the experiment performed in squirrel monkeys (Mancuso et al., 2009), finding simulation results consistent with the noted boost in color dimensionality (Fig. 6.1). First, we model the vision of an adult male squirrel monkey with a protanopic retina (M and S cones only) and a cortical model that has converged. Next, we simulate the effects of gene therapy, by modifying the retina model so that a random subset of cones begin to express L opsin. Immediately after this retinal change, NS continues to show blue-yellow dichromatic vision. However, if we allow the cortical model to continue the hypothesized self-supervised learning, vision re-converges to boosted, 3-dimensional color (Fig. 6.1). As an aside, the real gene therapy experiment (Mancuso et al., 2009) resulted in expression of both M and L photopigment in affected cones, but the relative amounts are difficult to ascertain. We additionally model scenarios in which equal M and L expression occurs or a variable amount of L relative to M at each cell, and both scenarios converge to trichromacy as well (Appendix E). These result indicate that the hypothesized self-supervised learning can explain experimental boosting of color dimensionality in adulthood consistent with Mancuso et al. (2009), if the hypothesized learning is assumed to occur continuously even in adulthood.

Figure 6.2 further demonstrates that simulations of boosting color dimensionality succeed even in 3D→4D cases by the addition of a fourth cone between S and M cones. This simulation results present the first theoretical work that highlights the possibility of boosting humans with trichromatic

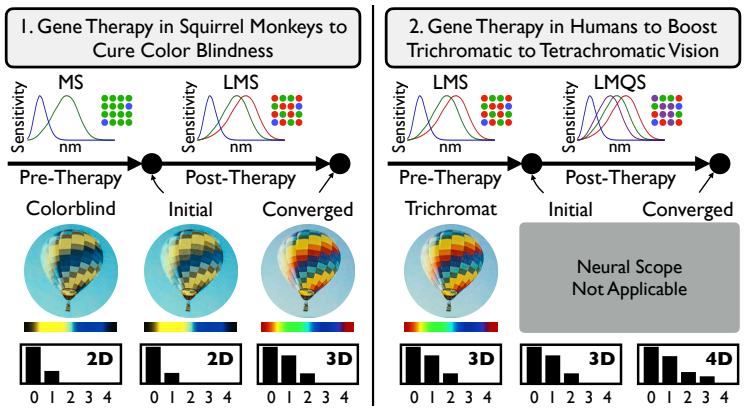

Figure 6: Simulated experiments for boosting color dimensionality via gene therapy. 1. 2D dichromat model is boosted to 3D trichromacy by mutating some M cones to express L opsins, with cortical learning re-converging to 3D color. 2. 3D trichromat model is boosted to 4D tetrachromacy by adding a fourth cone type between M and L cones, with cortical learning re-converging to 4D color.

vision to tetrachromat one by addition of a fourth photopigment. Further details, including the study of human tetrachromat observer model, are described in Appendix F.

7 Conclusion

In this work, we presented a framework for modeling the emergence of color vision in the human brain. We introduced a simulation engine for optic nerve signals, simulated cortical learning purely from such input signals, and measured the emergent color dimensionality qualitatively with NS and quantitatively with $CMF\text{-}SIM$.

We believe that the critical contribution of this work actually lies in the computational formulation of the problem in human perception itself – given access to only a stream of optic nerve signals, we probe how to meaningfully model the emergence of color vision in the cortex via simulated learning. Once the problem is formulated this way, it may not be entirely surprising to machine learning researchers that color with the correct dimensionality can be inferred; however, from the vision science perspective, this new approach is in many ways a foreign way to formulate the problem, because it is more common to think color comes from hardwired neural circuits (Appendix D).

The connection to camera imaging systems is noteworthy as well. The proposed computational framework is akin to a camera processing pipeline attached to an unknown, random color filter array pattern, where even the number of different color filters is a mystery. This is related to a branch of computational imaging called "auto calibration" that jointly solves for the scene and unknown system calibration parameters from measurements. One could imagine a new class of engineered sensing systems that have general processing units along the lines of the learning mechanism in this paper, that enables perceptual inference from a broader range of sensory streams that do not require the precision manufacturing common to many current sensors and cameras.

The learning notion of perception emerging from a process of disentangling it from encoded sensory streams is an interesting intellectual view of how perception may generally emerge in the brain. Any sensory stream going to the brain encodes information from the world entangled with sensing organ characteristics. The self-supervised learning mechanism proposed in this paper might be abstracted into a general neural process of learning to predict fluctuations in sensory stream due to ego perturbation, with perception for that sense emerging neurally as the optimal internal representation and associated decoder/encoder pair that enables accurate prediction.

To conclude, we invite the research community to build on the computational framework proposed in this paper, by improving the visual system components (e.g. even more accurately modeling details in the eye model, or replacing back-propagation in the cortical model with more bio-plausible learning rules (Lillicrap et al., 2020; Hinton, 2022)), or applying the framework to other sensory modalities.

ACKNOWLEDGEMENTS

This project was supported by Air Force Office of Scientific Research's MURI grant (FA9550-20-1-0195 and FA9550-21-1-0230). We are grateful to everyone involved in the MURI project for their support and feedback. Special thanks to Bruno Olshausen, Jessica Lee, and Alexander Belsten for their valuable discussions and insights. We also appreciate Fred Rieke, Lawrence Sincich, and David Brainard for their thoughtful input and engaging conversations.

REFERENCES

Colorimetry. Number 15.2 in CIE Publication. Commission Internationale de l'Éclairage, Vienna, Austria, 2nd edition, 1976.

Larry F Abbott. Lapicque's introduction of the integrate-and-fire model neuron (1907). *Brain research bulletin*, 50(5-6):303–304, 1999.

Russell J Adams, Daphne Maurer, and Margaret Davis. Newborns' discrimination of chromatic from achromatic stimuli. *Journal of Experimental Child Psychology*, 41(2):267–281, 1986.

Eirikur Agustsson and Radu Timofte. Ntire 2017 challenge on single image super-resolution: Dataset and study. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, July 2017.

Albert J Ahumada. Learning receptor positions. *Computational models of visual processing*, pp. 23–34, 1991.

M Alpern, K Kitahara, and D H Krantz. Perception of colour in unilateral tritanopia. *The Journal of Physiology*, 335(1):683–697, 1983.

Alexander G Anderson, Kavitha Ratnam, Austin Roorda, and Bruno A Olshausen. High-acuity vision from retinal image motion. *Journal of vision*, 20(7):34–34, 2020.

Juan M Angueyra, Jacob Baudin, Gregory W Schwartz, and Fred Rieke. Predicting and manipulating cone responses to naturalistic inputs. *Journal of Neuroscience*, 42(7):1254–1274, 2022.

Boaz Arad, Radu Timofte, Rony Yahel, Nimrod Morag, Amir Bernat, Yuanhao Cai, Jing Lin, Zudi Lin, Haoqian Wang, Yulun Zhang, et al. Ntire 2022 spectral recovery challenge and data set. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 863–881, 2022.

Claudia Diaz Araya and Jan M Provis. Evidence of photoreceptor migration during early foveal development: a quantitative analysis of human fetal retinae. *Visual neuroscience*, 8(6):505–514, 1992.

Michael A Arbib. *The handbook of brain theory and neural networks*. MIT press, 2003.

Manel Baradad, Jonas Wulff, Tongzhou Wang, Phillip Isola, and Antonio Torralba. Learning to see by looking at noise. In *Advances in Neural Information Processing Systems*, 2021.

Horace B Barlow et al. Possible principles underlying the transformation of sensory messages. *Sensory communication*, 1(01):217–233, 1961.

Bryce Bayer. Color imaging array. *United States Patent, no. 3971065*, 1976.

Noah C Benson, Jeremy R Manning, and David H Brainard. Unsupervised learning of cone spectral classes from natural images. *PLoS Computational Biology*, 10(6):e1003652, 2014.

Michael J Berry, David K Warland, and Markus Meister. The structure and precision of retinal spike trains. *Proceedings of the National Academy of Sciences*, 94(10):5411–5416, 1997.

Vicente Botella-Soler, Stéphane Deny, Georg Martius, Olivier Marre, and Gašper Tkačik. Nonlinear decoding of a complex movie from the mammalian retina. *PLoS computational biology*, 14(5): e1006057, 2018.

Nora Brackbill, Colleen Rhoades, Alexandra Kling, Nishal P Shah, Alexander Sher, Alan M Litke, and EJ Chichilnisky. Reconstruction of natural images from responses of primate retinal ganglion cells. *Elife*, 9:e58516, 2020.

David Brainard, Haomiao Jiang, Nicolas P Cottaris, Fred Rieke, EJ Chichilnisky, Joyce E Farrell, and Brian A Wandell. Isetbio: Computational tools for modeling early human vision. In *Imaging Systems and Applications*, pp. IT4A–4. Optica Publishing Group, 2015.

David H Brainard. Color and the cone mosaic. *Annual Review of Vision Science*, 1:519–546, 2015.

David H Brainard, David R Williams, and Heidi Hofer. Trichromatic reconstruction from the interleaved cone mosaic: Bayesian model and the color appearance of small spots. *Journal of vision*, 8(5):15–15, 2008.

Hans Brettel, Françoise Viénot, and John D Mollon. Computerized simulation of color appearance for dichromats. *Josa a*, 14(10):2647–2655, 1997.

Gijs Joost Brouwer and David J Heeger. Decoding and reconstructing color from responses in human visual cortex. *Journal of Neuroscience*, 29(44):13992–14003, 2009.

Angela M Brown. Development of visual sensitivity to light and color vision in human infants: A critical review. *Vision Research*, 30(8):1159–1188, 1990.

Yoram Burak, Uri Rokni, Markus Meister, and Haim Sompolinsky. Bayesian model of dynamic image stabilization in the visual system. *Proceedings of the National Academy of Sciences*, 107 (45):19525–19530, 2010.

Joseph Carroll, Carrie McMahon, Maureen Neitz, and Jay Neitz. Flicker-photometric electroretinogram estimates of l: M cone photoreceptor ratio in men with photopigment spectra derived from genetics. *JOSA A*, 17(3):499–509, 2000.

Joseph Carroll, Jay Neitz, and Maureen Neitz. Estimates of l: M cone ratio from erg flicker photometry and genetics. *Journal of vision*, 2(8):1–1, 2002.

Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *International conference on machine learning*, pp. 1597–1607. PMLR, 2020.

CIE. Commission internationale de l'eclairage proceedings, 1931.

Bevil R Conway. The organization and operation of inferior temporal cortex. *Annual review of vision science*, 4:381–402, 2018.

Bevil R Conway, Saima Malik-Moraleda, and Edward Gibson. Color appearance and the end of hering's opponent-colors theory. *Trends in Cognitive Sciences*, 2023.

Frans W Cornelissen and Eli Brenner. Is adding a new class of cones to the retina sufficient to cure color-blindness? *Journal of vision*, 15(13):22–22, 2015.

Nicolas P Cottaris, Haomiao Jiang, Xiaomao Ding, Brian A Wandell, and David H Brainard. 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, 2019.

Lisa J Croner and Ehud Kaplan. Receptive fields of p and m ganglion cells across the primate retina. *Vision research*, 35(1):7–24, 1995.

Joanna D Crook, Orin S Packer, John B Troy, and Dennis M Dacey. Synaptic mechanisms of color and luminance coding: rediscovering the xy-cell dichotomy in primate retinal ganglion cells. In *The new visual neurosciences*, pp. 123–144. The MIT Press, Cambridge, MA, 2013.

Christine A Curcio, Kenneth R Sloan, Robert E Kalina, and Anita E Hendrickson. Human photoreceptor topography. *Journal of comparative neurology*, 292(4):497–523, 1990.

Dennis Dacey, Orin S Packer, Lisa Diller, David Brainard, Beth Peterson, and Barry Lee. Center surround receptive field structure of cone bipolar cells in primate retina. *Vision research*, 40(14): 1801–1811, 2000.

Dennis M Dacey. The mosaic of midget ganglion cells in the human retina. *Journal of Neuroscience*, 13(12):5334–5355, 1993.

Dennis M Dacey and Barry B Lee. The'blue-on'opponent pathway in primate retina originates from a distinct bistratified ganglion cell type. *Nature*, 367(6465):731–735, 1994.

Dennis M Dacey, Barry B Lee, Donna K Stafford, Joel Pokorny, and Vivianne C Smith. Horizontal cells of the primate retina: cone specificity without spectral opponency. *Science*, 271(5249): 656–659, 1996.

Virginia De Sa. Learning classification with unlabeled data. *Advances in neural information processing systems*, 6, 1993.

Russell L De Valois and Karen K De Valois. A multi-stage color model. *Vision research*, 33(8): 1053–1065, 1993.

Andrew M Derrington, John Krauskopf, and Peter Lennie. Chromatic mechanisms in lateral geniculate nucleus of macaque. *The Journal of physiology*, 357(1):241–265, 1984.

René Descartes. *Descartes: Selected philosophical writings*. Cambridge University Press, 1988.

Laurent Dinh, Jascha Sohl-Dickstein, and Samy Bengio. Density estimation using real nvp. *arXiv preprint arXiv:1605.08803*, 2016.

Robert F Dougherty, Volker M Koch, Alyssa A Brewer, Bernd Fischer, Jan Modersitzki, and Brian A Wandell. Visual field representations and locations of visual areas v1/2/3 in human visual cortex. *Journal of vision*, 3(10):1–1, 2003.

John E Dowling and Brian Blundell Boycott. Organization of the primate retina: electron microscopy. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, 166(1002):80–111, 1966.

Ch Enroth-Cugell, JG Robson, DE Schweitzer-Tong, and AB Watson. Spatio-temporal interactions in cat retinal ganglion cells showing linear spatial summation. *The Journal of Physiology*, 341(1): 279–307, 1983.

Mark D Fairchild. *Color appearance models*. John Wiley & Sons, 2013.

Greg D Field, Jeffrey L Gauthier, Alexander Sher, Martin Greschner, Timothy A Machado, Lauren H Jepson, Jonathon Shlens, Deborah E Gunning, Keith Mathieson, Wladyslaw Dabrowski, et al. Functional connectivity in the retina at the resolution of photoreceptors. *Nature*, 467(7316): 673–677, 2010.

Peter Földiák. Learning invariance from transformation sequences. *Neural computation*, 3(2): 194–200, 1991.

Eric R Fossum. Cmos image sensors: Electronic camera-on-a-chip. *IEEE transactions on electron devices*, 44(10):1689–1698, 1997.

David H. Foster. Color constancy. *Vision Research*, 51(7):674–700, 2011. ISSN 0042-6989. doi: https://doi.org/10.1016/j.visres.2010.09.006. Vision Research 50th Anniversary Issue: Part 1.

Julian Freedland and Fred Rieke. Systematic reduction of the dimensionality of natural scenes allows accurate predictions of retinal ganglion cell spike outputs. *Proceedings of the National Academy of Sciences*, 119(46):e2121744119, 2022.

Tyler Godat, Nicolas P Cottaris, Sara Patterson, Kendall Kohout, Keith Parkins, Qiang Yang, Jennifer M Strazzeri, Juliette E McGregor, David H Brainard, William H Merigan, et al. In vivo chromatic and spatial tuning of foveolar retinal ganglion cells in macaca fascicularis. *Plos one*, 17 (11):e0278261, 2022.

CH Graham and Yun Hsia. Studies of color blindness: a unilaterally dichromatic subject, 1959.

Hermann Grassmann. Zur theorie der farbenmischung. *Annalen der Physik*, 165(5):69–84, 1853.

John Guild. The colorimetric properties of the spectrum. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 230 (681-693):149–187, 1931.

S Lee Guth. Model for color vision and light adaptation. *JOSA A*, 8(6):976–993, 1991.

Theodros M Haile, Kaitlin S Bohon, Maria C Romero, and Bevil R Conway. Visual stimulus-driven functional organization of macaque prefrontal cortex. *Neuroimage*, 188:427–444, 2019.

H KEFFER Hartline and Floyd Ratliff. Spatial summation of inhibitory influences in the eye of limulus, and the mutual interaction of receptor units. *The Journal of general physiology*, 41(5): 1049–1066, 1958.

Kaiming He, Xinlei Chen, Saining Xie, Yanghao Li, Piotr Dollár, and Ross Girshick. Masked autoencoders are scalable vision learners. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 16000–16009, 2022.

Ewald Hering. *Zur Lehre vom Lichtsinne: sechs Mittheilungen an die Kaiserl. Akademie der Wissenschaften in Wien*. C. Gerold's Sohn, 1878.

Geoffrey Hinton. The forward-forward algorithm: Some preliminary investigations. *arXiv preprint arXiv:2212.13345*, 2022.

Geoffrey E Hinton. Connectionist learning procedures. In *Machine learning*, pp. 555–610. Elsevier, 1990.

Hajime Hirasawa and Akimichi Kaneko. pH Changes in the Invaginating Synaptic Cleft Mediate Feedback from Horizontal Cells to Cone Photoreceptors by Modulating Ca2+ Channels . *Journal of General Physiology*, 122(6):657–671, 11 2003. ISSN 0022-1295.

Heidi Hofer, Joseph Carroll, Jay Neitz, Maureen Neitz, and David R Williams. Organization of the human trichromatic cone mosaic. *Journal of Neuroscience*, 25(42):9669–9679, 2005.

Gordon Holmes. Disturbances of vision by cerebral lesions. *The British journal of ophthalmology*, 2 (7):353, 1918.

David H Hubel and Torsten N Wiesel. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of physiology*, 160(1):106, 1962.

Gerald H Jacobs. Photopigments and the dimensionality of animal color vision. *Neuroscience & Biobehavioral Reviews*, 86:108–130, 2018.

Gerald H Jacobs, Gary A Williams, Hugh Cahill, and Jeremy Nathans. Emergence of novel color vision in mice engineered to express a human cone photopigment. *science*, 315(5819):1723–1725, 2007.

Sandra Jiménez and Jesús Malo. The role of spatial information in disentangling the irradiance–reflectance–transmittance ambiguity. *IEEE transactions on geoscience and remote sensing*, 52(8): 4881–4894, 2013.

Gabriele Jordan, Samir S Deeb, Jenny M Bosten, and John D Mollon. The dimensionality of color vision in carriers of anomalous trichromacy. *Journal of vision*, 10(8):12–12, 2010.

Deane B Judd. Color perceptions of deuteranopic and protanopic observers. *Journal of Research of the National Bureau of Standards*, 41:247–271, 1948.

Deane B Judd. *Response functions for types of vision according to the Müller theory*. US Government Printing Office, 1949.

Young Joon Kim, Nora Brackbill, Eleanor Batty, JinHyung Lee, Catalin Mitelut, William Tong, EJ Chichilnisky, and Liam Paninski. Nonlinear decoding of natural images from large-scale primate retinal ganglion recordings. *Neural Computation*, 33(7):1719–1750, 2021.

Ron Kimmel. Demosaicing: Image reconstruction from color ccd samples. *IEEE Transactions on image processing*, 8(9):1221–1228, 1999.

Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

John Krauskopf and Gegenfurtner Karl. Color discrimination and adaptation. *Vision research*, 32 (11):2165–2175, 1992.

Stephen W Kuffler. Discharge patterns and functional organization of mammalian retina. *Journal of neurophysiology*, 16(1):37–68, 1953.

Rosa Lafer-Sousa and Bevil R Conway. Parallel, multi-stage processing of colors, faces and shapes in macaque inferior temporal cortex. *Nature neuroscience*, 16(12):1870–1878, 2013.

Edwin H Land. The retinex theory of color vision. *Scientific american*, 237(6):108–129, 1977.

Michael F Land and Dan-Eric Nilsson. *Animal eyes*. OUP Oxford, 2012.

L Lapicque. Recherches quantitatives sur l'excitation electrique des nerfs. *J. Physiol. Paris*, 9: 620–635, 1907.

Jessica Lee, Nicholas Jennings, Varun Srivastava, and Ren Ng. Theory of human tetrachromatic color experience and printing. *ACM Trans. Graph*, 43(4), July 2024.

Peter Lennie, P William Haake, and David R Williams. The design of chromatically opponent receptive fields. *Computational models of visual processing*, pp. 71–82, 1991.

Ming Li, Fang Liu, Mikko Juusola, and Shiming Tang. Perceptual color map in macaque visual area v4. *Journal of Neuroscience*, 34(1):202–217, 2014.

Trisha Lian, Kevin J MacKenzie, David H Brainard, Nicolas P Cottaris, and Brian A Wandell. Ray tracing 3d spectral scenes through human optics models. *Journal of vision*, 19(12):23–23, 2019.

Timothy P Lillicrap, Adam Santoro, Luke Marris, Colin J Akerman, and Geoffrey Hinton. Backpropagation and the brain. *Nature Reviews Neuroscience*, 21(6):335–346, 2020.

AM Litke, N Bezayiff, EJ Chichilnisky, W Cunningham, W Dabrowski, AA Grillo, M Grivich, P Grybos, P Hottowy, S Kachiguine, et al. What does the eye tell the brain?: Development of a system for the large-scale recording of retinal output activity. *IEEE Transactions on Nuclear Science*, 51(4):1434–1440, 2004.

Jian K Liu, Dimokratis Karamanlis, and Tim Gollisch. Simple model for encoding natural images by retinal ganglion cells with nonlinear spatial integration. *PLoS computational biology*, 18(3): e1009925, 2022.

Jianfei Liu, HaeWon Jung, Alfredo Dubra, and Johnny Tam. Cone photoreceptor cell segmentation and diameter measurement on adaptive optics images using circularly constrained active contour model. *Investigative ophthalmology & visual science*, 59(11):4639–4652, 2018.

Ye Liu, Ming Li, Xian Zhang, Yiliang Lu, Hongliang Gong, Jiapeng Yin, Zheyuan Chen, Liling Qian, Yupeng Yang, Ian Max Andolina, et al. Hierarchical representation for chromatic processing across macaque v1, v2, and v4. *Neuron*, 108(3):538–550, 2020.

Margaret S Livingstone and David H Hubel. Psychophysical evidence for separate channels for the perception of form, color, movement, and depth. *Journal of Neuroscience*, 7(11):3416–3468, 1987.

William Lotter, Gabriel Kreiman, and David Cox. Deep predictive coding networks for video prediction and unsupervised learning. In *International Conference on Learning Representations*, 2016.

David L MacAdam. Visual sensitivities to color differences in daylight. *Josa*, 32(5):247–274, 1942.

Donald IA MacLeod and Robert M Boynton. Chromaticity diagram showing cone excitation by stimuli of equal luminance. *JOSA*, 69(8):1183–1186, 1979.

Walter Makous. Comment on" emergence of novel color vision in mice engineered to express a human cone photopigment". *Science*, 318(5848):196–196, 2007.

Laurence T Maloney and Albert J Ahumada. Learning by assertion: Two methods for calibrating a linear visual system. *Neural Computation*, 1(3):392–401, 1989.

Katherine Mancuso, William W Hauswirth, Qiuhong Li, Thomas B Connor, James A Kuchenbecker, Matthew C Mauck, Jay Neitz, and Maureen Neitz. Gene therapy for red–green colour blindness in adult primates. *Nature*, 461(7265):784–787, 2009.

Olivier Marre, Gasper Tkacik, Dario Amodei, Elad Schneidman, William Bialek, and Michael Berry. Multi-electrode array recording from salamander retinal ganglion cells. 2017.

Susana Martinez-Conde, Jorge Otero-Millan, and Stephen L Macknik. The impact of microsaccades on vision: towards a unified theory of saccadic function. *Nature Reviews Neuroscience*, 14(2): 83–96, 2013.

James Clerk Maxwell. Theory of the perception of colors. *Trans. R. Scottish Soc. Arts*, 4:394–400, 1856.

J. W. McClurkin and L. M. Optican. Primate striate and prestriate cortical neurons during discrimination. i. simultaneous temporal encoding of information about color and pattern. *Journal of Neurophysiology*, 75(1):481–495, 1996.

J. W. McClurkin, J. A. Zarbock, and L. M Optican. Primate striate and prestriate cortical neurons during discrimination. ii. separable temporal codes for color and pattern. *Journal of Neurophysiology*, 75(1):496–507, 1996.

K McLaren. Xiii—the development of the cie 1976 (l* a* b*) uniform colour space and colour-difference formula. *Journal of the Society of Dyers and Colourists*, 92(9):338–341, 1976.

Matthew J McMahon, Martin JM Lankheet, Peter Lennie, and David R Williams. Fine structure of parvocellular receptive fields in the primate fovea revealed by laser interferometry. *Journal of Neuroscience*, 20(5):2043–2053, 2000.

JD Mollon and JK Bowmaker. The spatial arrangement of cones in the primate fovea. *Nature*, 360 (6405):677–679, 1992.

Thomas Naselaris, Ryan J Prenger, Kendrick N Kay, Michael Oliver, and Jack L Gallant. Bayesian reconstruction of natural images from human brain activity. *Neuron*, 63(6):902–915, 2009.

J Neitz and M Neitz. Evolution of the circuitry for conscious color vision in primates. *Eye*, 31(2): 286–300, 2017.

Jay Neitz and Maureen Neitz. The genetics of normal and defective color vision. *Vision Research*, 51 (7):633–651, 2011.

Isaac Newton. *Opticks: or, A Treatise of the Reflexions, Refractions, Inflexions and Colours of Light*. Courier Corporation, 1704.

Shinji Nishimoto, An T Vu, Thomas Naselaris, Yuval Benjamini, Bin Yu, and Jack L Gallant. Reconstructing visual experiences from brain activity evoked by natural movies. *Current biology*, 21(19):1641–1646, 2011.

J Kevin O'Regan and Alva Noë. A sensorimotor account of vision and visual consciousness. *Behavioral and brain sciences*, 24(5):939–973, 2001.

Gustav A Osterberg. Topography of the layer of rods and cones in the human retina. *Acta ophthalmologica*, 1935.

Orin S Packer and Dennis M Dacey. Receptive field structure of h1 horizontal cells in macaque monkey retina. *Journal of Vision*, 2(4):1–1, 2002.

Stephanie E Palmer, Olivier Marre, Michael J Berry, and William Bialek. Predictive information in a sensory population. *Proceedings of the National Academy of Sciences*, 112(22):6908–6913, 2015.

Nikhil Parthasarathy, Eleanor Batty, William Falcon, Thomas Rutten, Mohit Rajpal, EJ Chichilnisky, and Liam Paninski. Neural networks for efficient bayesian decoding of natural images from retinal neurons. *Advances in Neural Information Processing Systems*, 30, 2017.

Ken Perlin. An image synthesizer. *ACM Siggraph Computer Graphics*, 19(3):287–296, 1985.

Baingio Pinna. Un effetto di colorazione. In *Il laboratorio e la città. XXI Congresso degli Psicologi Italiani*, volume 158. Edizioni SIPs, Societá Italiana di Psiocologia, Milano, 1987.

Baingio Pinna, Gavin Brelstaff, and Lothar Spillmann. Surface color from boundaries: a new 'watercolor'illusion. *Vision research*, 41(20):2669–2676, 2001.

Jonathan R Polimeni, Bruce Fischl, Douglas N Greve, and Lawrence L Wald. Laminar analysis of 7 t bold using an imposed spatial activation pattern in human v1. *Neuroimage*, 52(4):1334–1346, 2010.

Rajeev Ramanath, Wesley E Snyder, Griff L Bilbro, and William A Sander III. Demosaicking methods for bayer color arrays. *Journal of Electronic imaging*, 11(3):306–315, 2002.

Rajesh PN Rao and Dana H Ballard. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature neuroscience*, 2(1):79–87, 1999.

Katja Reinhard and Thomas A Münch. Visual properties of human retinal ganglion cells. *PLoS One*, 16(2):e0246952, 2021.

Ilias Rentzeperis, Andrey R Nikolaev, Daniel C Kiper, and Cees van Leeuwen. Distributed processing of color and form in the visual cortex. *Frontiers in psychology*, 5:932, 2014.

Dragos L Rezeanu, James Kuchenbecker, Rachel Barborek, Marcus Mazzaferri, Maureen Neitz, and Jay Neitz. Explaining the absence of functional tetrachromacy in females with four cone types. *Investigative Ophthalmology & Visual Science*, 62(8):527–527, 2021.

Danilo Rezende and Shakir Mohamed. Variational inference with normalizing flows. In *International conference on machine learning*, pp. 1530–1538. PMLR, 2015.

Austin H Riesen, Robert L Ramsey, and Paul D Wilson. Development of visual acuity in rhesus monkeys deprived of patterned light during early infancy. *Psychonomic Science*, 1(1-12):33–34, 1964.

Mattia Rigotti, Omri Barak, Melissa R Warden, Xiao-Jing Wang, Nathaniel D Daw, Earl K Miller, and Stefano Fusi. The importance of mixed selectivity in complex cognitive tasks. *Nature*, 497 (7451):585–590, 2013.

Robert W Rodieck. Quantitative analysis of cat retinal ganglion cell response to visual stimuli. *Vision research*, 5(12):583–601, 1965.

Robert W Rodieck. The first steps in seeing. 1998.

Martin Rolfs. Microsaccades: small steps on a long way. *Vision research*, 49(20):2415–2441, 2009.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, pp. 234–241. Springer, 2015.

Austin Roorda, Fernando Romero-Borja, William J Donnelly III, Hope Queener, Thomas J Hebert, and Melanie CW Campbell. Adaptive optics scanning laser ophthalmoscopy. *Optics express*, 10 (9):405–412, 2002.

Michele Rucci and Jonathan D Victor. The unsteady eye: an information-processing stage, not a bug. *Trends in neurosciences*, 38(4):195–206, 2015.

Michele Rucci, Ehud Ahissar, and David Burr. Temporal coding of visual space. *Trends in cognitive sciences*, 22(10):883–895, 2018.

Ramkumar Sabesan, Heidi Hofer, and Austin Roorda. Characterizing the human cone photoreceptor mosaic via dynamic photopigment densitometry. *PloS one*, 10(12):e0144891, 2015.

Ramkumar Sabesan, Brian P Schmidt, William S Tuten, and Austin Roorda. The elementary representation of spatial and color vision in the human retina. *Science advances*, 2(9):e1600797, 2016.

Jared M Salisbury and Stephanie E Palmer. Optimal prediction in the retina and natural motion statistics. *Journal of Statistical Physics*, 162(5):1309–1323, 2016.

Elad Schneidman, Michael J Berry, Ronen Segev, and William Bialek. Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature*, 440(7087):1007–1012, 2006.

Manuel Schottdorf and Barry B Lee. A quantitative description of macaque ganglion cell responses to natural scenes: the interplay of time and space. *The Journal of physiology*, 599(12):3169–3193, 2021.

Jérémie Sibille, Carolin Gehr, Jonathan I Benichov, Hymavathy Balasubramanian, Kai Lun Teh, Tatiana Lupashina, Daniela Vallentin, and Jens Kremkow. High-density electrode recordings reveal strong and specific connections between retinal ganglion cells and midbrain neurons. *Nature communications*, 13(1):5218, 2022.

Yosef Singer, Yayoi Teramoto, Ben DB Willmore, Jan WH Schnupp, Andrew J King, and Nicol S Harper. Sensory cortex is optimized for prediction of future input. *elife*, 7:e31557, 2018.

Yosef Singer, Luke Taylor, Ben DB Willmore, Andrew J King, and Nicol S Harper. Hierarchical temporal prediction captures motion processing along the visual pathway. *eLife*, 12:e52599, 2023a.

Yosef Singer, Luke Taylor, Ben DB Willmore, Andrew J King, and Nicol S Harper. Hierarchical temporal prediction captures motion processing along the visual pathway. *Elife*, 12:e52599, 2023b.

Barun Singh, William T Freeman, and D Brainard. Exploiting spatial and spectral image regularities for color constancy. In *Workshop on Statistical and Computational Theories of Vision. Nice, France*. Citeseer, 2003.

Florentina Soto, Jen-Chun Hsiang, Rithwick Rajagopal, Kisha Piggott, George J Harocopos, Steven M Couch, Philip Custer, Josh L Morgan, and Daniel Kerschensteiner. Efficient coding by midget and parasol ganglion cells in the human retina. *Neuron*, 107(4):656–666, 2020.

Mandyam Veerambudi Srinivasan, Simon Barry Laughlin, and Andreas Dubs. Predictive coding: a fresh view of inhibition in the retina. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, 216(1205):427–459, 1982.

WS Stiles and JM Burch. Interim report to the commission internationale de l'eclairage, zurich, 1955, on the national physical laboratory's investigation of colour-matching (1955). *Optica Acta: International Journal of Optics*, 2(4):168–181, 1955.

Vincent Stimper, David Liu, Andrew Campbell, Vincent Berenz, Lukas Ryll, Bernhard Schölkopf, and José Miguel Hernández-Lobato. normflows: A pytorch package for normalizing flows. *Journal of Open Source Software*, 8(86):5361, 2023. doi: 10.21105/joss.05361. URL https://doi.org/10.21105/joss.05361.

Andrew Stockman and Lindsay T Sharpe. The spectral sensitivities of the middle-and long-wavelength-sensitive cones derived from measurements in observers of known genotype. *Vision research*, 40(13):1711–1737, 2000.

Andrew Stockman, Lindsay T. Sharpe, and Clemens Fach. The spectral sensitivity of the human short-wavelength sensitive cones derived from thresholds and color matches. *Vision Research*, 39 (17):2901–2927, 1999. ISSN 0042-6989.

Andrew Stockman, David H Brainard, et al. Color vision mechanisms. *OSA handbook of optics*, 3: 11–1, 2010.

Roger BH Tootell, Martin S Silverman, Eugene Switkes, and Russell L De Valois. Deoxyglucose analysis of retinotopic organization in primate striate cortex. *Science*, 218(4575):902–904, 1982.

Antonio Torralba and Aude Oliva. Statistics of natural image categories. *Network: computation in neural systems*, 14(3):391, 2003.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

Jan Verweij, Eric P Hornstein, and Julie L Schnapf. Surround antagonism in macaque cone photoreceptors. *Journal of Neuroscience*, 23(32):10249–10257, 2003.

Hermann Von Helmholtz. *Handbuch der physiologischen Optik*, volume 9. Voss, 1867.

J Von Kries. Theoretische studien über die umstimmung des sehorgans. *Festschrift der Albrecht-Ludwigs-Universität*, 32:145–158, 1902.

Thomas Wachtler, Eizaburo Doi, Te-Won Lee, and Terrence J Sejnowski. Cone selectivity derived from the responses of the retinal cone mosaic to natural scenes. *Journal of vision*, 7(8):6–6, 2007.

GEORGE WALD. The molecular basis of visual excitation. *Nature*, 219(5156):800–807, 1968.

Yiyi Wang, Nicolas Bensaid, Pavan Tiruveedhula, Jianqiang Ma, Sowmya Ravikumar, and Austin Roorda. Human foveal cone photoreceptor topography and its dependence on eye length. *Elife*, 8: e47148, 2019.

John S Werner, Brennan Marsh-Armstrong, and Kenneth Knoblauch. Adaptive changes in color vision from long-term filter usage in anomalous but not normal trichromacy. *Current Biology*, 30 (15):3011–3015, 2020.

David R Williams, Donald IA MacLeod, and Mary M Hayhoe. Foveal tritanopia. *Vision Research*, 21(9):1341–1356, 1981.

Lance Williams. Pyramidal parametrics. In *Proceedings of the 10th annual conference on Computer graphics and interactive techniques*, pp. 1–11, 1983.

Lauren E Wool, Joanna D Crook, John B Troy, Orin S Packer, Qasim Zaidi, and Dennis M Dacey. Nonselective wiring accounts for red-green opponency in midget ganglion cells of the primate retina. *Journal of Neuroscience*, 38(6):1520–1540, 2018.

William David Wright. A re-determination of the trichromatic coefficients of the spectral colours. *Transactions of the Optical Society*, 30(4):141, 1929.

Eric Wu, Nora Brackbill, Alexander Sher, Alan Litke, Eero Simoncelli, and EJ Chichilnisky. Maximum a posteriori natural scene reconstruction from retinal ganglion cells with deep denoiser priors. *Advances in Neural Information Processing Systems*, 35:27212–27224, 2022.

Günther Wyszecki and Walter Stanley Stiles. *Color science: concepts and methods, quantitative data and formulae*. John wiley & sons, 1982.

Ed Yong. *An immense world: How animal senses reveal the hidden realms around us*. Knopf Canada, 2022.

Laura K Young and Hannah E Smithson. Emulated retinal image capture (erica) to test, train and validate processing of retinal images. *Scientific Reports*, 11(1):11225, 2021.

Thomas Young. Ii. the bakerian lecture. on the mechanism of the eye. *Philosophical Transactions of the Royal Society of London*, (91):23–88, 1801.

Thomas Young. Ii. the bakerian lecture. on the theory of light and colours. *Philosophical transactions of the Royal Society of London*, (92):12–48, 1802.

Semir M Zeki. Functional specialisation in the visual cortex of the rhesus monkey. *Nature*, 274 (5670):423–428, 1978.

Ling-Qi Zhang, Nicolas P Cottaris, and David H Brainard. An image reconstruction framework for characterizing initial visual encoding. *Elife*, 11:e71132, 2022.

Yuxin Zhang, Wen-Tao Deng, Wei Du, Ping Zhu, Jie Li, Fan Xu, Jingfen Sun, Cecilia D Gerstner, Wolfgang Baehr, Sanford L Boye, et al. Gene-based therapy in a mouse model of blue cone monochromacy. *Scientific reports*, 7(1):6690, 2017.

Zhetuo Zhao, Ehud Ahissar, Jonathan D Victor, and Michele Rucci. Inferring visual space from ultra-fine extra-retinal knowledge of gaze position. *Nature communications*, 14(1):269, 2023.