**Specific aims**

The retina is composed of multiple pathways, with each pathway processing specific visual information. These pathways do not process all visual information equally: some information is better represented by neurons than other, supposedly because it is more adaptive. While we know the physiologically behind such phenomenon, our understanding of why evolution designed the retina to better process certain types of information, is currently lacking. To truly understand what the retina does, we need mathematical theories that can explain as much of the retinal physiology with as few assumptions as possible. The efficient coding hypothesis is one of the most successful theories in this vein, which states that sensory systems should remove redundancies in their inputs to optimize the information they process. This hypothesis has been especially successful in the retina, where it can explain many features of retinal encoding such as center-surround receptive fields and ON-OFF pathways1-5. However, most of these models make simplifying assumptions that are biologically unrealistic, such as an infinite number of neurons and the possibility of negative firing rates. Because of these simplifying assumptions, it is currently unclear what predictions the efficient coding hypothesis makes about retinal pathways. To fill this gap in knowledge, I will use machine learning model to build a more flexible model that can make efficient coding predictions about the structure of neurons into different pathways. More specifically, I will provide efficient coding predictions about how the retina should process chromatic inputs (Aim 1), and about how the retina should process motion (Aim 2). By comparing these results with experimental data, we will know whether the efficient coding hypothesis is sufficient to explain how the retina encodes natural scenes.

**Aim 1: Expand efficient coding models to encompass chromatic information**

Encoding colors starts at the level of cone photoreceptors in the fovea of the retina. which come in three types –Long (L), Medium (M), and Short (S), roughly encoding red, green and blue stimuli, respectively6. This color information is ultimately encoded by RGCs, which are separated into different pathways. In macaques, midget cells represent roughly 80% of RGCs and encode the discrepancy between L and M cones, which tend to be highly correlated in natural images. Parasol cells (~10%) are separated into two pathways (ON and OFF) that sum L, M and S inputs. Bistratified cells (~10%) integrate ON inputs from S cones and OFF inputs from L and M cones. Bistratified cells are unique in that they asymmetric, and are the only major cell type to have an ON but not an OFF pathway. My model will try to explain two major findings: 1) Why this asymmetry in bistratified cells occurs, and 2) why midget cells represent most neurons despite encoding events that rarely occur in natural images. Completion of this aim will inform us whether efficient coding can successfully explain how the retina processes color in different pathways.

**Aim 2: Expand efficient coding models to explain motion-selectivity in RGCs**

Visual scenes are typically in motion, either because of objects moving or optic flow from our own movements. The encoding of visual motion starts as early as the retina, with several subtypes of retinal ganglion cells (RGCs) having stronger responses to one direction of motion than to its opposite7, 8. The major type of motion encoding RGCs are ON-OFF direction-selective ganglion cells (DSGC)9, which are divided into four subtypes that respond preferentially to each of the four cardinal directions8. There are currently no efficient coding predictions about these ON-OFF DSGCs. To fill this gap in knowledge, I will build an efficient coding model that jointly estimates receptive fields in space and time – a requirement to encode directional motion. I will then test whether we can replicate ON-OFF DSGCs that only encode the four cardinal directions8, and also compare my findings to experimental data from the Field Lab at UCLA. Completion of this aim will result in testable predictions for efficient coding of motion in natural images, which will help validate or invalidate the efficient coding hypothesis.

**Significance**

There are three explanatory processes that underlie scientific explanation: Descriptive (what), mechanistic (how) and normative (why) explanations10. The retina is fairly well-understood both from descriptive and mechanistic perspectives11, 12. However, normative models that explain the purpose of the system are still lacking. Efficient coding is one of the most successful theories that can explain how the retina is organized, but what it can explain is still relatively sparse compared to what is mechanistically known. The current work will try to expand efficient coding theory to replicate how the retina processes color and motion.

Mechanistic explanations of retinal processing

****Retinal processing of visual information follows a well-known structure11, 12: First, photoreceptors transform light from the outside world into electrical activity. They then send this information to bipolar cells, and bipolar cells send this information to retinal ganglion cells (RGCs). These RGCs are the output layer of the retina. Their axons form the optic nerve and sends information to the thalamus, which then transmits it to the primary visual cortex. RGCs are separated into two different pathways (ON and OFF), and each neuron within a pathway processes a small region of visual space — its receptive field. These receptive fields form ‘mosaics’ (one per RGC type) that tile visual space. The receptive fields of RGCs have a center-surround organization: ON RGCs encode light in the center and dark in the surround, and vice-versa for OFF RGCs.

The mechanistic explanation of retinal processing has been pushed a step further by computational models of RGCs. These models usually involve estimating a function that approximates the computations performed by a single recorded neuron. The most popular class of RGC model is linear-nonlinear filters, which take a weighted linear summation of each image pixel, followed with an output nonlinearity (e.g. a softplus function). The weights are optimized to minimize discrepancies between the neuron’s recorded responses to images and the model responses to the same images. In the retina, linear-nonlinear models tend to converge to difference-of-gaussian receptive field and manage to explain most of the variance in a neuron’s responses. Linear-nonlinear models can be outperformed by more complex models with additional nonlinearities, which can explain almost all of a neuron’s average responses to images13. Interestingly, these more complex models are also purely feedforward, suggesting that feedback connections are not necessary to accurately model RGC responses.

Normative explanations of retinal processing

Even though linear-nonlinear models provide us with simple functions that approximate what computations RGCs perform, these models do not explain *why* receptive fields in the retina are the way they are. To answer this question, we need to replicate RGC receptive fields not by predicting recorded responses, but instead from simple assumptions and normative principles. These computational theories typically involve three main components: The inputs (e.g. natural images or gaussian mixtures), the model (e.g. linear-nonlinear), and the normative principles.

The efficient coding hypothesis is the most prevalent theory in the retina. This theory states that early sensory systems form an efficient neural code that decorrelates its inputs to maximize information while minimizing the number of spikes used. Early theoretical work on efficient coding explained how the center-surround organization of RGCs arises from decorrelation, both for achromatic and for color inputs1, 2. They managed to do so with very simple assumptions, such as gaussian inputs, a linear model that allows negative firing rates, and an infinite number of neurons. However, these simple assumptions did not manage to account for other properties of retinal processing, such as the segregation of neurons across different cell types. The retina has different RGC types, with neurons within a cell type tiling the entire retina to form a mosaic, with each neuron type processing a specific type of visual information.

Recent work from our lab leveraged machine learning to make efficient coding predictions with natural image inputs, a limited number of neurons and linear-nonlinear models. These new assumptions allowed us to replicate mosaics across different cell types. This model was also able to replicate the anti-alignment between ON and OFF mosaics, and suggests that this anti-alignment occurs due to the high output noise of RGCs. Previous work from the lab has also showed that it is efficient for neurons to encode either high spatial or high temporal frequencies, but not both, similar to midget and parasol cells.

This new type of efficient coding model raises the possibility of asking whether other properties of RGC types are efficient. The inputs to RGCs are much more complex than static achromatic images – RGCs receive input images from multiple color channels, and visual scenes are usually in motion. While efficient coding can predict how the retina should process achromatic stimuli, its predictions for color and motion processing – two crucial aspects of natural stimuli – are still lacking. My project is going to test whether we can use efficient coding to explain (1) How different RGC types process chromatic information and (2) How different RGC types process motion.

**Innovation**

**Technical innovation:** To complete either aims, I will need to develop new machine learning techniques to train efficient coding models with multiple correlated channels (cones or latencies), which implies increasing the number of parameters by multiple folds. I will solve this overparameterizing problem by designing new methods to parametrize receptive fields across color channels and latencies. By doing so, we will pave the way for future research to solve efficient coding problems with very larger number of parameters.

**Conceptual innovation:** Most of the efficient coding research in vision involves a single input channel that is encoded by a large number of neurons. However, neurons in the retina have multiple correlated input channels, such as different colors and latencies. This project is conceptually innovative because I consider efficient coding models with multiple correlated channels. By doing so, we will learn how efficient coding models handle correlated channels, and whether this solution is similar to the computations RGCs perform.

**Approach**

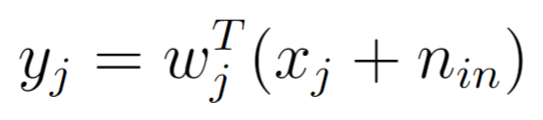
**Previous work**

Our lab previously built efficient coding models to explain the organization of RGCs. why retinal ON and OFF mosaics are anti-aligned. We also used a similar model to explain why RGCs encode either high spatial and low temporal frequencies, or low spatial and high temporal frequencies. Such models not only replicate the center-surround organization, but also their spatial arrangement. Model RGCs are separated into different ON and OFF pathways that process light and dark information, respectively. Each neuron processes a small region of visual space, and each pathway has neurons that are spatially organized to form a ‘mosaic’ that tiles the entire visual scene. Our lab found that whether efficient coding predicted that these ON and OFF mosaics should be aligned or anti-aligned depends on the noise, with mosaics going from aligned to anti-aligned as both the input and output noise levels increase.

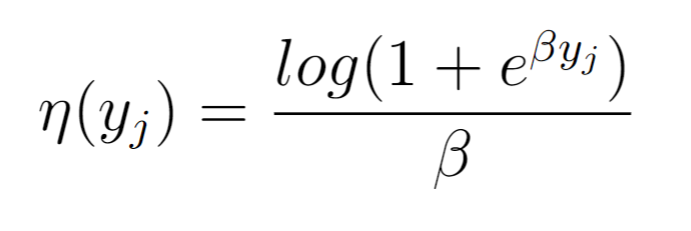
However, such models have previously been either limited (1) to one single input channel3, 14 or (2) by making strong assumptions about how receptive fields should change across channels15. The goal of this proposed project is to extend such models to natural stimuli that have multiple correlated channels, which will allow us to study how RGCs process color (Aim 1) and motion (Aim 2).

**Efficient coding model**

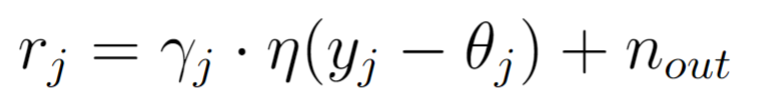
The efficient coding model we use will be similar to other work from our lab14, 15. The inputs will consist of D pixels patches of Natural Images X corrupted by gaussian input noise ~ N(0, . The model RGC j takes as input a linear combination of these natural images:



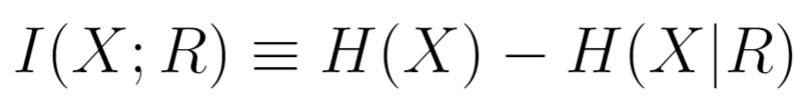
Where wj are unit-norm kernels (||wj|| = 1) and represent how much each photoreceptor contributes to the response of model RGC j. yj is then passed through a softplus nonlinearity:



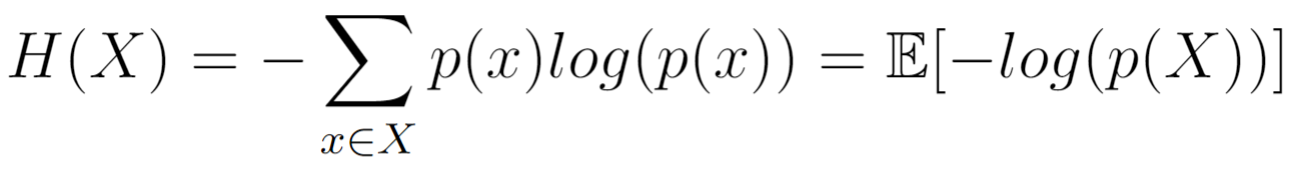
And further corrupted by output noise to produce firing rate with threshold and gain :



The above parameters are optimized via Adam to maximize the mutual information between the inputs X and the outputs R, under a firing rate constraint. Mutual information represents the amount of information, in bits, that is transmitted to RGCs from photoreceptors. Mutual information is equivalent to subtracting the entropy of the natural images X from the conditional entropy of X on R:

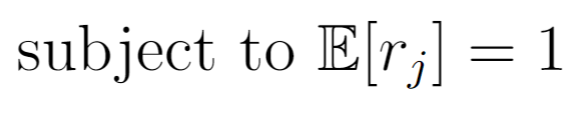


Where the entropy H(X) is defined as:



Previous work has derived a closed-form solution for the mutual information between X and R3, 14, 15. The above parameters are optimized to maximize this mutual information:





Where represents the covariance matrix of the input natural images, and are diagonal matrices that represent the covariance of the input and output noise, respectively, and W is the weight matrix. G is a diagonal matrix that represents the local derivatives of the output responses for a specific set of input images. Since the output nonlinearity is a softmax function, the diagonal of G has binary values (1 if the neuron is firing and 0 if not). This function is maximized using Adam optimization16. To represent the metabolic cost of firing spikes, each neuron is be restricted to have a fixed average firing rate.

In both aims, we will train the model on multiple correlated input channels (color in Aim 1, and time in Aim 2). Having multiple input channels drastically increases the number of parameters, which can make the model significantly harder to optimize. However, preliminary results have shown that these filters consistently converge to difference-of-gaussians even when trained to encode different channels that represent either color or time. We can use this fact to our advantage to facilitate training by reducing the number of parameters. To that effect, we will parametrize the linear filters of each neuron to be difference-of-gaussians:



Where ɑ, b, c, d and z are vectors of length k that represent each of the k different channels. z is the distance from the receptive field center, ɑ is the size of the receptive field center, b is the size of the receptive field surround. d and c are the relative strength of the center and the surround for each color channel, respectively. To allow these values to converge, d is restrained to have an L2-norm of 1. ɑ and b are both restricted to have positive values, and c is restricted to have values ranging from 0 to 1. On top of facilitating training, this parametrization also allows us to directly characterize how each model RGCs integrate either color or time across different channels.



**Aim 1: Expand efficient coding models to encompass chromatic information**

Background and rationale

Color vision starts with cone photoreceptors, which encode three different wavelengths of light. Short cones (S) encode blue, Medium (M) cones encode green, and Long (L) cones encode red. The three main types of RGCs (Parasol, midget and bistratified cells) each integrate this chromatic information in different ways (Figure 2). Parasol cells are achromatic and integrate cones independently of their type. Midget cells integrate L and M cones, and how they do so vary from the fovea to the periphery. In the fovea, midget cells receive excitatory inputs from a single cone in the receptive field center (either L or M), and inhibitory inputs from both L and M cones in the surround. In the periphery, both the center and the surround of midget cells encode a sum of L and M cones, with the center encoding the opposite polarity of the surround (i.e. ON-center and OFF-surround or vice-versa). Bistratified cells encode blue: They only have a receptive field center, which encodes ON S inputs and L + M OFF inputs.While we know a lot about how the retina processes chromatic information, what efficient coding models can explain about color vision in the retina is relatively scarce. For example, mathematically tractable models that can explain why color-opponency is efficient are limited to two cones (L and M) instead of three17. Independent-component analysis can explain why color-opponency with three cones is efficient18, but this results in receptive fields that are qualitatively different and do not have a center-surround organization. We are still missing efficient coding models that can explain in more details how chromatic inputs are processed differently by midget, parasol and bistratified cells. Aim 1 will address this question by building efficient coding models that encode chromatic natural images.

Experimental methods and design

To study how to efficiently encode chromatic natural images, we will use the Kyoto Natural Images Dataset19. This dataset consists of 62 images of natural scenes that were gamma-corrected and transformed from RGB to LMS. The transformation matrix from RGB into LMS responses was estimated by minimizing the prediction error of cone spectral sensitivities19. These responses were further transformed with an empirical cone non-linear function20. Each of these images have 1000x1280x3 pixels. We will crop these images into smaller training samples of 16x16x3 pixels.

We will train the efficient coding model described above on the Kyoto Natural Images Dataset. We will analyze the learned parameters of the model, specifically the d parameter, to understand how neurons cluster into different types that process chromatic information differently. We will change the RGC-cones ratio to test whether this ratio changes how midget cells process chromatic information, and compare these results to known receptive fields in the fovea and periphery.

Expected outcomes

I expect model RGCs to have three different receptive field types that are analogous to parasol, midget and bistratified cells. Based on previous research3, 14, 15, each cell type should form a mosaic that tiles the entire visual field. We can categorize which RGC type a model neuron by analyzing the d parameter for L, M and S inputs. Similar to parasol cells, I expect some neurons will have L, M and S inputs that are all positive or all negative, respectively. I expect another group of neurons to have positive S cones and negative L and M cones, similarly to bistratified cells. Finally, there should be another group of neurons that looks similar to midget cells. With a high RGC-cones ratio (>3), these cells should have opponency between L and M cones, and sum L and M cones with lower RGC-cones ratios (~1).



Alternative approaches and hypotheses

Due to the high correlations across color channels, it is possible that receptive fields converge on adding different colors channels instead of subtracting them. If that occurs, we will build a linear version of our model that is mathematically tractable. We will then compare our results to previous literature that found color-opponency to be efficient by making assumptions about the Fourier structure of chromatic natural images17. This mathematically tractable model should give us an intuition about what conditions are sufficient and necessary for different chromatic receptive fields to be efficient.

**Aim 2: Expand efficient coding models to explain direction-selectivity in RGCs**

Background and rationale

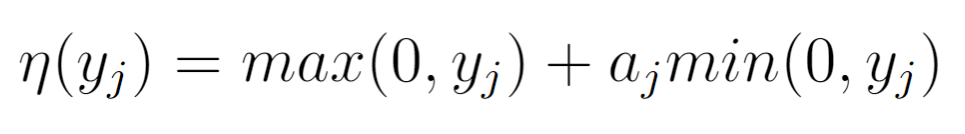
Both the world and ourselves are constantly in movement, which makes motion an important aspect of visual scenes. Consistent with this idea, RGCs respond strongly to moving stimuli. These responses vary as a function of the orientation of motion, and the optimal orientation is usually either pointing towards the fovea or perpendicular to it. This orientation-bias is present in both midget and parasol cells21. However, these neurons can’t individually discriminate the direction of motion: Their responses would be similar if an object is moving left-to-right than right-to-left. Only a subset (~12%) of neurons are direction-selective8. These Direction-Selective Ganglion Cells (DSGCs) are divided into two main types: ON-OFF DSGCs and ON DSGCs. ON-OFF DSGCs are divided into four subtypes that encode each of the four cardinal directions (anterior, posterior, superior and inferior)8. ON-OFF DSGCs respond to the onset of both ON and OFF stimuli, a property similar to complex cells in the primary visual cortex22. These cells were initially discovered in the rabbit retina23, but have also more recently been discovered in the mouse8 and macaque retina24. While ON-OFF DSGCs have relatively small receptive fields and detect local motion, ON DSGCs have larger receptive fields that encode global motion. ON DSGCs are divided into three subtypes that encode the anterior, posterior and inferior directions of motion8. ON DSGCs respond to light (but not dark stimuli), and are divided into two different populations that have either transient or sustained responses8. ON DSGCs are also less numerous than their ON-OFF counterparts. These cells have been discovered in the rabbit and mouse, but have yet to be found in macaques.

Previous work from our lab built efficient coding models for spatiotemporal receptive fields. However, these models had constraints that prevented them from having direction-selective neurons15. In Aim 2 I will test the hypothesis that having a subset of direction-selective cells is efficient, due to the predominance of motion in natural movies. To test this hypothesis, I need an efficient coding model that can integrate multiple correlated channels – each channel representing a different latency. The model previously described is a great candidate to solve this problem, as it makes few assumptions about the temporal structure of receptive fields. The contribution of this project will be a theoretical understanding of retinal direction selectivity from efficient coding principles.

Experimental method and design

I will build an efficient coding model to encode natural movies from the Chicago Motion Database25. This model will be conceptually similar to the one from Aim 1 (Figure 3). The main difference is in the input channels, which will be different latencies instead of different colors. At time *t,* the model neurons will receive inputs from the image at time t, t – 1, t – 2, etc. This reflects how RGCs don’t instantaneously fire in response to stimuli but instead integrate their inputs over time26. The weights from each latency will be estimated separately to maximize the mutual information between the inputs and outputs of the model (see Aim 1). The weights will be parametrized by difference-of-gaussians, similarly to Aim 1. We will measure the amount of direction-selectivity in each neuron by computing how much the kernel centers change across latencies.

To encapsulate the responses properties of ON-OFF DSGCs, we will change the output nonlinearity from a softmax function to a Parametric Rectified Linear Unit:



This output non-linearity is conceptually similar to the energy model for complex cells in the primary visual cortex22.

Expected outcomes

Previous work from my lab14, 15 suggests that new RGCs types emerge as we increase the number of neurons. Most of RGCs in the retina are not DSGCs, which is why I expect that the first neurons to emerge will not be DSGCs. Instead, these should have similar spatiotemporal properties to midget and parasol cells15. However, as we increase the number of neurons, I expect that we will eventually find ON-OFF DSGCs, followed by ON DSGCs. ON-OFF DSGCs should have relatively small receptive fields and encode the four cardinal directions. Moreso, the aj parameter from the output nonlinearity should be negative, allowing these neurons to respond to both ON and OFF stimuli. ON DSGCs should have larger receptive fields, and encode motion in the anterior, posterior and inferior direction. The estimated aj should be near zero, allowing these neurons to respond to ON (but not OFF) onset.

The known physiology about DSGCs is fairly detailed. It is possible we will not be able to reproduce all of those experimental findings from efficient coding principles. However, we hope that we will be able to show that ON-OFF DSGCs in the four cardinal directions can be derived from efficient coding principles. Other than that, it will be of interest to see which findings we can replicate from efficient coding principles, and which ones we cannot.

Alternative approaches and hypotheses

Because DSGCs are only a small fraction (~12%) of RGCs, it might require a large RGC-cones ratio to obtain neurons that are directions-selective. If we fail to replicate DSGCs with the computing resources we have, we will remove stationary sections of the movies from our dataset, and train the model exclusively on moving stimuli. We will also try reducing the number of learned parameters; for example, by assuming aj is negative or assuming a, b, c and d do not vary across channels.

**References**

1. Atick, J.J. and A.N. Redlich, *Towards a theory of early visual processing.* Neural computation, 1990. **2**(3): p. 308-320.

2. Atick, J.J. and A.N. Redlich, *What does the retina know about natural scenes?* Neural computation, 1992. **4**(2): p. 196-210.

3. Karklin, Y. and E. Simoncelli, *Efficient coding of natural images with a population of noisy linear-nonlinear neurons.* Advances in neural information processing systems, 2011. **24**.

4. Balasubramanian, V. and M.J. Berry II, *A test of metabolically efficient coding in the retina.* Network: Computation in Neural Systems, 2002. **13**(4): p. 531.

5. Doi, E., et al., *Efficient coding of spatial information in the primate retina.* Journal of Neuroscience, 2012. **32**(46): p. 16256-16264.

6. Crook, J.D., et al., *Horizontal cell feedback without cone type-selective inhibition mediates “red–green” color opponency in midget ganglion cells of the primate retina.* Journal of Neuroscience, 2011. **31**(5): p. 1762-1772.

7. Rasmussen, R. and K. Yonehara, *Contributions of retinal direction selectivity to central visual processing.* Current Biology, 2020. **30**(15): p. R897-R903.

8. Vaney, D.I., B. Sivyer, and W.R. Taylor, *Direction selectivity in the retina: symmetry and asymmetry in structure and function.* Nature Reviews Neuroscience, 2012. **13**(3): p. 194-208.

9. Wei, W., et al., *Development of asymmetric inhibition underlying direction selectivity in the retina.* Nature, 2011. **469**(7330): p. 402-406.

10. Levenstein, D., et al., *On the role of theory and modeling in neuroscience.* Journal of Neuroscience, 2023. **43**(7): p. 1074-1088.

11. Kaneko, A., *Physiology of the retina.* Annual review of neuroscience, 1979. **2**(1): p. 169-191.

12. Dacey, D.M. *Physiology, morphology and spatial densities of identified ganglion cell types in primate retina*. in *Ciba Foundation Symposium 184‐Higher‐Order Processing in the Visual System: Higher‐Order Processing in the Visual System: Ciba Foundation Symposium 184*. 2007. Wiley Online Library.

13. Maheswaranathan, N., et al., *Inferring hidden structure in multilayered neural circuits.* PLoS computational biology, 2018. **14**(8): p. e1006291.

14. Jun, N.Y., G.D. Field, and J. Pearson, *Scene statistics and noise determine the relative arrangement of receptive field mosaics.* Proceedings of the National Academy of Sciences, 2021. **118**(39): p. e2105115118.

15. Jun, N.Y., G. Field, and J. Pearson, *Efficient coding, channel capacity, and the emergence of retinal mosaics.* Advances in neural information processing systems, 2022. **35**: p. 32311-32324.

16. Kingma, D.P. and J. Ba, *Adam: A method for stochastic optimization.* arXiv preprint arXiv:1412.6980, 2014.

17. Atick, J.J., Z. Li, and A.N. Redlich, *Understanding retinal color coding from first principles.* Neural computation, 1992. **4**(4): p. 559-572.

18. Lee, T.-W., T. Wachtler, and T.J. Sejnowski, *Color opponency is an efficient representation of spectral properties in natural scenes.* Vision Research, 2002. **42**(17): p. 2095-2103.

19. Doi, E., et al., *Spatiochromatic receptive field properties derived from information-theoretic analyses of cone mosaic responses to natural scenes.* Neural computation, 2003. **15**(2): p. 397-417.

20. Baylor, D., B. Nunn, and J. Schnapf, *Spectral sensitivity of cones of the monkey Macaca fascicularis.* The Journal of Physiology, 1987. **390**(1): p. 145-160.

21. Antinucci, P. and R. Hindges, *Orientation-selective retinal circuits in vertebrates.* Frontiers in neural circuits, 2018. **12**: p. 11.

22. Mechler, F. and D.L. Ringach, *On the classification of simple and complex cells.* Vision research, 2002. **42**(8): p. 1017-1033.

23. Barlow, H. and W.R. Levick, *The mechanism of directionally selective units in rabbit's retina.* The Journal of physiology, 1965. **178**(3): p. 477.

24. Kim, Y.J., et al., *Origins of direction selectivity in the primate retina.* Nature communications, 2022. **13**(1): p. 2862.

25. Salisbury, J.M. and S.E. Palmer, *Optimal prediction in the retina and natural motion statistics.* Journal of Statistical Physics, 2016. **162**(5): p. 1309-1323.

26. Yasui, S., W. Davis, and K.-I. Naka, *Spatio-temporal receptive field measurement of retinal neurons by random pattern stimulation and cross correlation.* IEEE Transactions on Biomedical Engineering, 1979(5): p. 263-272.