**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.