David St-Amand

**Specific Aims**

One of the primary goals in neuroscience is to figure out simple principles that explain how systems are organized. Barlow (1961) proposed one of the most successful theories in neuroscience, which states that sensory neurons should optimize the amount of information they process, while also keeping their firing rates to a minimum. This *efficient coding hypothesis* provides us with a mathematical framework to understand how neurons should encode information, which can then be verified experimentally. Efficient coding has been successful at explaining how neuronal information is processed in multiple brain regions including the retina, the primary auditory cortex and the primary motor cortex. To reduce the overall number of spikes, this hypothesis states that sensory information must encode unexpected events – that is, discrepancies between inputs that are redundant. This strategy becomes less and less efficient as the redundancies between inputs increase, since the redundancies represent a larger fraction of the total information. When the redundancies are too high, efficient coding predicts that neurons should stop encoding discrepancies and instead sum redundant inputs. However, this is not how sensory neurons integrate redundant information; they instead encode discrepancies between their inputs instead of the sum, despite the huge loss of information. My project will reconcile predictions from efficient coding with empirical findings; *My central hypothesis is that encoding discrepancies between inputs that are highly correlated is efficient.*

The efficient coding hypothesis has been especially successful at predicting experimental findings in the retina, which makes this system perfect for studying how neurons should efficiently integrate redundant inputs. Efficient coding successfully explains why retinal ganglion cells (RGCs) have center-surround receptive fields, and why RGCs are organized into different functional types, with each type having neurons tiling the entire retina to form a ‘mosaic’. Recently, my lab used efficient coding to explain whether different mosaics should be spatially aligned or anti-aligned, which depends on whether the internal noise levels of RGCs is low or high (Jun, Field & Pearson, 2021). I will use similar models to study the optimal strategy to integrate redundant inputs in the retina, both across different channels (aim 1) and across time (aim 2).

**Aim 1:** Determine how RGCs should efficiently integrate redundant input channels

Hypothesis: The efficient strategy is to encode discrepancies between redundant channels

Retinal ganglion cells integrate inputs from cone photoreceptors, which are split into three different channels: Long (L), Medium (M) and Short (S) cones. The information in these three channels is mostly redundant, with most (~95%) of the information in natural images being achromatic. However, how the retina works seems to contradict that principle: Most RGCs’ responses are tuned to colors, with each neuron type processing a specific color channel. My project will reconcile these two principles and explain why encoding chromatic information is the optimal efficient coding strategy for natural images. To do so, I will build and train an efficient coding model on chromatic natural images and draw parallels from the model neurons to retinal experimental data. Completion of this aim will allow us to understand why it is efficient for neurons to encode discrepancies between redundant input channels.

**Aim 2:** Determine how the retina should encode spatiotemporal correlations across inputs

Hypothesis: We can replicate how the retina encodes motion from efficient coding principles

Neuronal activity is not only correlated in both space and time, but spatial and temporal correlations also interact with each other. The most prominent example of this phenomenon is motion, where we can predict the future location of a moving object based on its current location and velocity. While it is clear that the efficient coding strategy for RGCs should include encoding motion, what exactly this strategy is – how many neurons should process motion and what should their spatiotemporal receptive fields be– is still unclear*. My working hypothesis is that the efficient coding strategy for encoding motion in natural images will replicate experimental findings about motion encoding in RGCs.* To answer this question, I will extend the previous spatiotemporal efficient coding model from my lab to be spatiotemporally inseparable; that is, the model will be able to learn a receptive field that changes across time, a crucial property to encode motion. Completion of this aim will enlighten us as to whether how the retina encodes motion can be fully explained by the efficient coding hypothesis.

**Significance**

Through many rigorous experiments, neuroscientists have unraveled how different sensory systems encode information. However, we still do not fully understand why the brain utilizes these specific coding strategies. Researchers have long speculated about theories that would explain how different sensory systems are wired, and there is one idea that has survived the erosion of time: sensory systems should remove redundancies in their inputs to optimize the information they process. Such a hypothesis, if proven correct, has the potential to uncover the goals of sensory systems, which in turn would reveal why the sensory circuits we’ve discovered are wired in specific ways.

The efficient coding hypothesis has already succeeded at explaining many experimental findings across many senses, such as audition, touch and vision (Lewicki, 2002; Miller et al., 2019). Early studies focused on explaining how the retina work from assumptions that are as mathematically simple as possible, albeit not always biologically realistic. For example, Atick & Redlich (1990) used efficient coding principles to explain why retinal ganglion cells have center-surround receptive fields. However, their mathematical analysis assumed that the retina had an infinite number of neurons, with receptive fields only varying along a single dimension. More recently, researchers have expanded such ideas to scenarios that are more biologically plausible, which allowed them to make more accurate comparisons between the efficient coding hypothesis and experimental data. For example, Karklin & Simoncelli (2011) built an efficient coding model with a limited number of neurons and receptive fields in two dimensions. Using this model, they were able to explain why retinal ganglion cells separate into two different types – ON and OFF – which process light and dark information, respectively. They found that it was efficient for neurons within a type to encode distinct regions of visual space, repelling each other to form a ‘mosaic’ that tiles the entire retina. Such models were further expanded by my lab to explain more details about retinal physiology. We showed how efficient coding can explain why ON and OFF mosaics are anti-aligned (Jun, Field & Pearson, 2021), and why it is efficient for RGCs to encode low spatial with high temporal frequencies, and high spatial with low temporal frequencies (Jun, Field & Pearson, 2022).

Despite this progress, efficient coding is still far from being able to explain everything we know about retinal physiology; more research is required to test whether efficient coding principles can reliably explain how the retina processes information. To answer this question, this project will push the efficient coding hypothesis to its limits and test whether efficient coding can explain the receptive fields of RGCs in complex scenarios. In **Aim 1**, we will investigate whether efficient coding can explain how RGCs integrate Long, Medium and Short cones to process color stimuli. Interestingly, a large majority of RGCs in the fovea integrate long and medium cones to process red-green opponency, despite red-green opponency only representing a small fraction of the information in natural scenes. Instead, preliminary results suggest that a more efficient strategy would be for most RGCs to encode achromatic stimuli, where most of the information lies. By building alternative efficient coding models with different assumptions, we will resolve this discrepancy and discover why RGCs encode red-green opponency. We also aim to replicate other experimental findings about how RGCs process color, such as a small proportion of RGCs process yellow-blue opponency in the ON but not in the OFF pathway. In **Aim 2**, we will explain motion selectivity in the retina from efficient coding principles, which requires inferring receptive fields both across space and time. My lab previously achieved the difficult task of building such a spatiotemporal efficient coding model. However, to do so, they separately estimated the spatial and temporal dynamics of the receptive fields. Such a model is inappropriate to encode motion since a motion-selective RGC has its spatial receptive field change across time. In this project, I will build an efficient coding model that can encode motion and use it to explain why RGCs are motion-selective.

**Innovation**

**Technical innovation:** To complete either aims, we will need to develop new machine learning techniques to train efficient coding models with a high number of parameters. In aim 1, modeling the fovea will require us to increase the number of RGCs to be multiple times that of cones, which in turn increases the number of parameters to optimize by multiple folds. In aim 2, we will invent a new way to estimate spatial receptive fields at multiple latencies. estimating spatial receptive fields at multiple latencies will also increase the number of parameters by many folds. Therefore, the technical innovation of this project is the new implementation of machine learning approaches to solve efficient coding problems with very large numbers of parameters.

**Conceptual innovation:** Most of previous work about efficient coding is mathematically abstract and makes assumptions that are biologically unrealistic. This study will push the efficient coding to its limits and make biologically realistic assumptions to explain the details of what we know about retinal physiology. This innovation will tell us whether current efficient coding models are reliable at explaining how the retina processes information, or whether we need to create better models to replicate current experimental findings.

*References*

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

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

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

Lewicki, M. S. (2002). Efficient coding of natural sounds. *Nature neuroscience*, *5*(4), 356-363.

Miller, L. E., Fabio, C., Ravenda, V., Bahmad, S., Koun, E., Salemme, R., ... & Farne, A. (2019). Somatosensory cortex efficiently processes touch located beyond the body. *Current Biology*, *29*(24), 4276-4283.