The Influence of Novelty and Engagement on Neural Dynamics

Zachary McNulty ZACHARY_MCNULTY@BERKELEY.EDU
Department of Mathematics

Abstract

The visual cortex is one of the most commonly studied regions of the brain due to its diverse array of neural dynamics. Since vision is one of the primary ways we learn from and navigate this constantly changing natural world, it stands to reason these encodings may be structured in such a way to foster these processes: learning and prediction. In this paper, we study optical physiology data collected from mice trained to perform a simple classification task with visual stimuli. In doing so, we aim to study the role novelty and expectation play in the neural dynamics.

1. Introduction and Background

The visual cortex is one of the most commonly studied and intriguing regions in the human and animal brain. Acting as a bridge between visual stimuli and response, the visual cortex processes environmental information for decision making. The complexity of the environments it has to process as well as the crucial role vision plays in navigating such environments requires a high degree of precision when it comes to neural computation. This necessitates efficient ways of encoding, interpreting, and decoding such stimuli. The mechanisms by which the visual cortex accomplishes these tasks are still far from understood. More recently, some studies have suggested these encodings may facilitate the prediction of future events in an ever-changing natural environment and how feedback mechanisms naturally foster the development of low-dimensional representations of the world in these encodings [Rao and Ballard (1999), Recanatesi et al. (2021)]. As such, developing a deeper understanding of the computation and encoding properties of this area of the brain could help understand the role of neural feedback mechanisms and how encodings may be structured to facilitate specific real-world tasks.

In order to try to understand some of the ways in which the visual cortex encodes the complexities of the natural world, the Allen Institute has developed large-scale, intensively quality-controlled, and publicly-accessible datasets of brain activity across several different regions and on a variety of different tasks. In this paper, we focus on the Visual Behavior Optical Physiology dataset which aims to evaluate the influence of experience, expectation, and task engagement on neural coding and dynamics in the mouse brain [de Vries (2020)]. In these experiments, mice were presented with a sequence of natural images, where the same image was shown several times consecutively, and trained to respond when there was a change in the presented image[1]. Throughout this process, data surrounding the task-performance and neural activity were collected. The latter was measured using optical physiology, specifically 2-photon calcium-imaging. This technique detects calcium release in the brain via fluorescent indicator molecules which give off light upon binding to calcium ions [Grienberger and Konnerth (2012)]. Calcium ions play a crucial role in the transmission of electrical signals from neuron to neuron and hence high extracellular calcium levels are indicative of high levels of neural activity. In the collected optical physiology data, these high calcium ion levels manifest themselves in the form of high levels of fluorescence. Instead of collecting a continuous stream of images from the calcium imaging, the datasets provides two summary statistics: the max projection and average projection. These represent the maximum and average fluorescence levels of a given pixel attained throughout the session respectively. The basic behavior task setup and an example of the 2-photon calcium-image recorded are displayed in Figure (1).

In this project we attempt to infer properties of the testing environment purely from the neural dynamics through the calcium imaging data. Specifically, we were interested in the influence of novelty and task engagement. Since the mice underwent several recording sessions and were trained and tested on a few different image sets, novelty was measured as a binary variable: had the mouse seen this particular image set before. On the other hand, to facilitate training and encourage task participation, on some sessions (but not others)

1. More details on the project's data-collection process, task-training methodology, and quality-control procedures are discussed in the project's whitepaper.

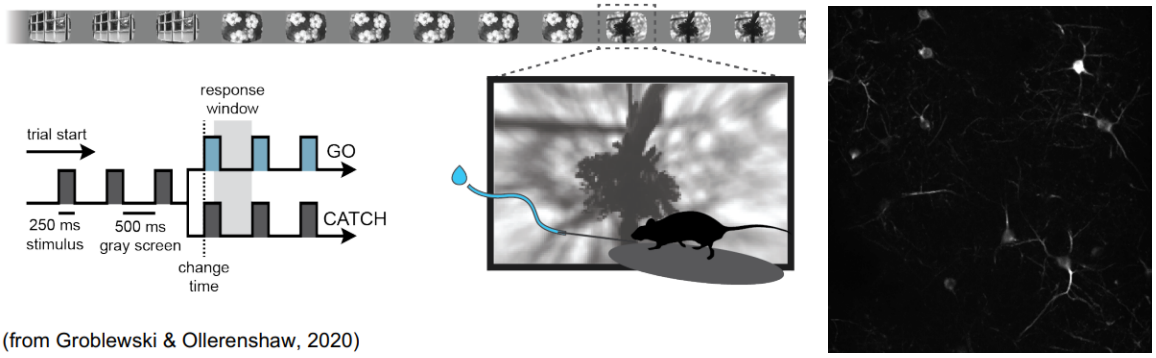

Figure 1: (left; from project's whitepaper) Behavior task setup. Top row shows the continuous stream of stimulus presentations (250ms per stimulus), with inter stimulus interval gray screen (500ms) that are displayed to the mouse during a behavior session. Lower left shows the two trial types in the task, "go" trials where the stimulus identity changes and the mouse must lick within a 750ms response window to earn a water reward, and "catch" trials where the image identity doesn't change and false alarm licking behavior is quantified. (right) Example of Ca^+ imaging of neural activity during behavior task.

the mice were given rewards upon successfully performing the task. Again, we model this as a binary variable measuring whether or not rewards were provided in a given session. We call trials with rewards **active** and those without **passive**.

2. Methodology

2.1 Hierarchical Bayes

One problem that often arises in data science is having a large number of overall datapoints in a population, but very few datapoints associated to each individual. Using such a small sample to estimate individual parameters is inherently noisy and unreliable. Hierarchical Bayesian Inference leverages the fact that the statistics of individuals are likely governed by parameters which themselves share some distribution across the entire population. Thus estimation of the individual parameters can be further informed by using statistics from all individuals in the population. Such a technique has already been deployed to study a variety of neurocomputational mechanisms [Ahn et al. (2017)]. For an example, consider:

$$\theta_i \overset{iid}{\sim} \textbf{Beta}(\alpha, \beta) \quad X_{ij} \mid \theta_i \overset{ind}{\sim} \textbf{Binomial}(k_i, \theta_i) \quad i = 1, ..., n \;\; j = 1, ..., m$$

where $\{X_{ij}\}$ are our individual statistics, with m samples from each of the n individuals, and $\{\theta_i\}_{i=1}^n$ are our individual parameter governed by some unknown hyperparameters α, β. Since Beta is the conjugate prior of Binomial and $\sum_j X_{ij} | \theta_i \sim \textbf{Bin}(mk_i, \theta_i)$ is a sufficient statistic for θ_i, we have posterior:

$$\theta_i \mid \{X_{ij}\}_{j=1}^m \sim \textbf{Beta}\left(\alpha + \sum_{j=1}^m X_{ij},\ \beta + mk_i - \sum_{j=1}^m X_{ij}\right) \qquad \mathbb{E}\left[\theta_i \mid \{X_{ij}\}_{j=1}^m\right] = \frac{\alpha + \sum_{j=1}^m X_{ij}}{\alpha + \beta + mk_i}$$

where the above posterior mean is the Bayes Estimator under MSE. Hence if we can estimate α, β using all the X_{ij}, we have a natural predictor for each θ_i : the posterior mean. One way of doing this is to assume they have a known prior distribution $(\alpha, \beta) \sim P$ in which case standard Markov Chain Monte Carlo (MCMC) methods such as the Gibbs Sampler or the Metropolis–Hastings algorithm can be used to sample from the posterior for α, β [Carpenter et al. (2017)]. Another option, known as Empirical Bayes, is to simply estimate α, β from the data using the MLE or some other estimator. We chose the latter approach. If $S_i := \sum_j X_{ij}$ the joint distribution has log-density[2]:

$$\log(\mathbb{P}(S_1, ..., S_n)) = \sum_{i=1}^{n} \log\left(\mathbb{P}\left(\textbf{BetaBinomial}(\alpha, \beta, mk_i) = S_i\right)\right)$$

which we maximized using the differential evolution algorithm to find the MLE [Bilal et al. (2020)]. In this way, the data from all individuals can be leveraged to improve our estimate of the individual parameters θ_i by helping estimate the population hyperparameters (α, β).

In this paper, we use the above Hierarchical Bayes model to estimate the true classification accuracy of each individual mouse in this dataset. We chose this approximation over the simple sample mean because some mice had a relatively low number of trials, especially when it came to the catch trials and false alarms where some had less than 10 such trials. Specifically, we aim to estimate the true and false positive rates (TPR, FPR). The hierarchical model was run separately for each of the four pairings familiar/novel with TPR/FPR. It was not possible to perform this analysis for the active/passive pairing because in the passive trials the mice are not performing the classification task. The main purpose of these estimates is that they give us some indicator of how correlated the neural dynamics may be to the task at hand: poor performance might suggest changes in the presented stimulus have little casual relationship to any observed changes in the neural dynamics. In which case, any hope of predicting properties of the stimulus from the neural dynamics or vice versa might be doomed to fail. In general, they give us a crude baseline for comparison of what performance levels we might expect our models to achieve.

2.2 Downsizing Sparse Images

As with all image processing, we aim to compress the 512 x 512 images produced through calcium imaging by removing irrelevant or redundant information. First we attempt to exploit the sparsity of the calcium images. As is apparent in Figure (2) below, these black and white images are almost entirely black. The brightness of each pixel is encoded as a number in $[0, 1]$, so simply zeroing out values below some threshold can drastically reduce the number of nonzero pixels. As Figure (2) demonstrates, even relatively small thresholds are incredibly effective, and visibly it is not apparent if any information is lost. By introducing this sparsity, we can limit overparameterization in our classification models and utilize efficient algorithms designed for sparse matrices. Since this dataset is quite large, this reduction is necessary to make the classification problem computationally feasible. We used a threshold of 0.1 for this project which reduced the image to around 5-10% of its

2. Because the gamma function can get quite large, for computational purposes it may be better to minimize the log of the joint density as the log gamma function grows much slower.

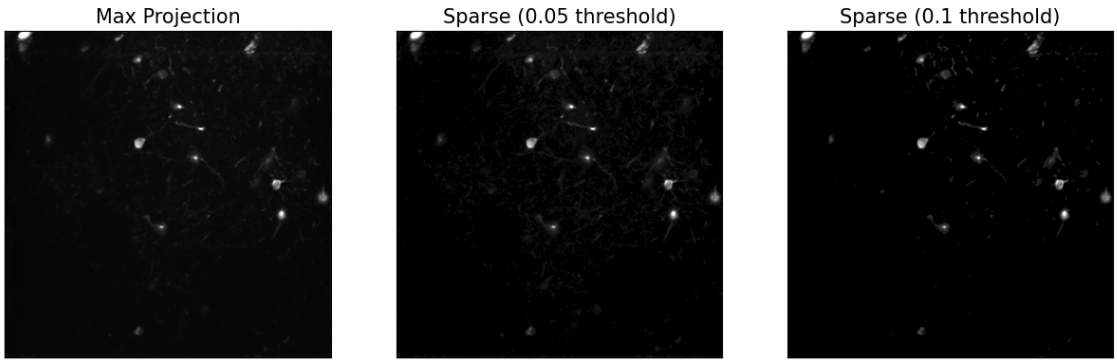

Figure 2: Example of thresholding used to reduce image size. Using the 0.05 and 0.1 thresholds we reduced the pixel count to 29% and 3.3% of the original size respectively.

original size.

Following this procedure, we ran a (truncated) SVD algorithm to compute some of the first few principal components of this image space. In particular, we noticed keeping only the first 750 principal components still captured 97% of the variance in the samples. Thus these two techniques have reduced our feature size from $262,144$ features to just 750, and the latter is certainly an ample number of parameters for our purposes. We also tried running these classification algorithms on the original sparsified images, but their size limited our ability to do proper hyperparameter tuning and to use more complicated models.

2.3 Novelty and Engagement Classification

After the above preprocessing of our data, we built a classifier to take the PCA coordinates of our neural images and predict if the images observed during a given recording session were novel versus familiar and a separate classifier to predict passive versus active, our measure of engagement. After standardizing the entries for each PCA coordinate, we trained both a logistic regression model and a kernalized Support Vector Regression (SVR) model, an extension of SVM, to do the binary classification. We chose the former due to the convenient probabilistic interpretation of its predictions as the chance a given session was familiar rather than novel. We chose the latter because of its ability to introduce nonlinearities into the features. As we see in Figure (4), the datapoints are highly clustered in principal component space with no clear division, so these nonlinearities may be essential to the success of the classifier.

3. Results

3.1 Hierarchical Bayes

Figure (3) and Table (1) list the results of the Hierarchical Bayes analysis. As we can see in Figure (3), the TPRs had a much wider variance as well as a higher mean in comparison to the FPRs which were centered closer to zero. Moreover, in both cases the distributions

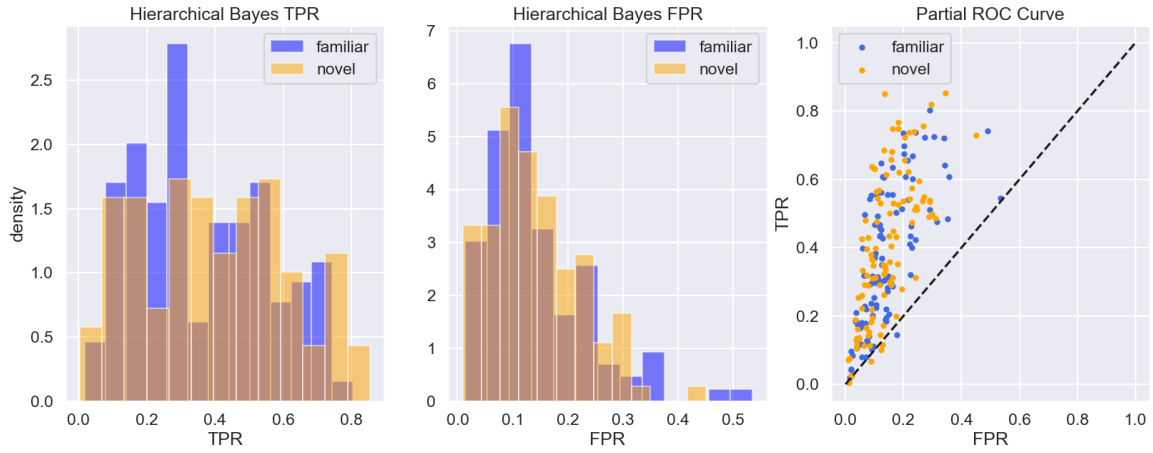

Figure 3: The left two figures show the approximated distributions of TPR/FPR for individual mice in both familiar novel trials. These were estimated using the Beta-Binomial hierarchical Bayesian model. The right figure plots the estimated ROC curve for these models, where the dotted black line is the line for $FPR = TPR$.

for novel and familiar seem nearly identical. This observation is supported by the MLE estimates for (α, β) in our model found in Table (1). In this table we see the parameters for the two TPR distributions are in both cases roughly the same. Similarly, for the two FPR distributions we see α is quite small, while β is much larger. This discrepancy in the scale parameters α, β is what produces the bias towards zero in the two FPR distributions.

In Figure (3) there is a clear trend in performance above the $TPR = FPR$ line. This seems to suggest the mice were at least reasonably proficient at performing the classification task. Still, there are many datapoints near the origin, suggesting the mice may have tended towards inaction in the case of uncertainty.

3.2 PCA and Classification

Figure (4) shows the distribution of the first nine principal components which together account for around 32% of the variance in the dataset. As this figure highlights, there is very little separation between these distributions, familiar versus novel and passive versus active, amongst the most significant principal components. Since this might be due to the fact that much of the variance in the data comes from the noise of ambient neural dynamics, we also analyzed the principal components that separated the two distributions as much as possible. We quantified this distance using the Wasserstein metric (after normalizing the features). Nonetheless, we noticed the same clustering as before. This supports the results

	Familiar TPR	**Familar FPR**	**Novel TPR**	**Novel FPR**
α	3.93	1.98	3.86	1.44
β	3.82	13.28	5.20	9.87

Table 1: Hierarchical Bayesian estimates of hyperparameters α, β

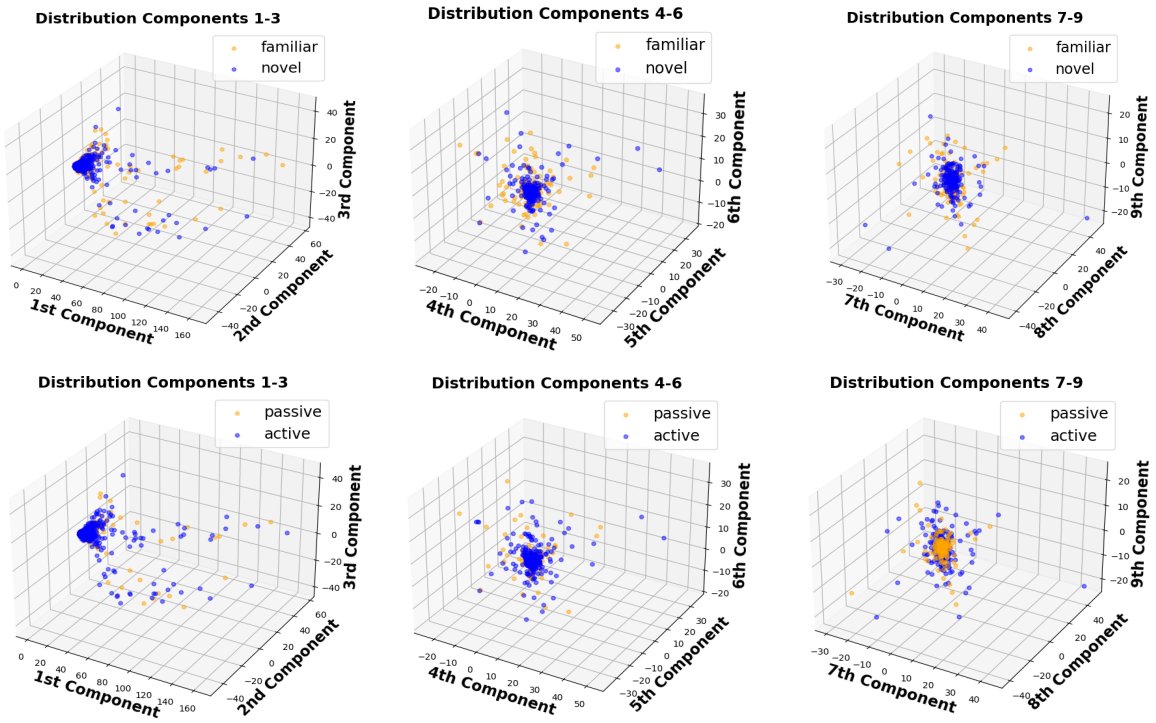

Figure 4: Principal components 1 through 3 (left), 4 through 6 (middle), 7 through 9 (right). The datapoints corresponding to familiar image sessions versus novel sessions were colored orange and blue respectively. The first row shows the distribution of familiar trials versus novel trials in this space, and the second row shows the distribution of passive versus active trials.

of our Hierarchical Bayesian analysis in which we observed the same lack of separation. In line with these observations, our logistic regression and SVR models were not able to definitively form a distinction between these two pairs of classes. In both the familiar versus novel and passive versus active comparisons, the SVR model (with a polynomial kernel) outperformed the logistic model. But in both cases, the model accuracy was only around 60% test accuracy, barely above the rate expected from randomly guessing.

4. Conclusions

Unfortunately, as Figure (3) illustrates, the distributions of TPRs in our Hierarchical Bayesian analysis were quite low in both cases, averaging close to the 50% mark of random guessing. As we discussed earlier, this somewhat poor performance might have made it difficult to filter out the effects of novelty and engagement from the ambient noise of the mouse's normal neural activity. Nonetheless, the task at hand is not a choice between two actions, but rather a choice between action and no action at all. As such, we may expect the default of nonaction to carry some amount of inertia, giving preference to it in the case of uncertainty. This would of course naturally drag both the TPR and FPR down. For this reason, the Allen Institute suggested another metric: the "d-prime" metric (aka the

sensitivity index) common in signal processing. This metric combines the TPR and FPR measures to give a single measure of classification accuracy that is not as influenced by a preference for one action over the other[3]. Most of the mice achieved an average d' value somewhere in $[-0.5, 2]$, suggesting possibly a slightly more optimistic view of the classification accuracy than our TPR/FPR analysis.

In terms of classification, our above work suggests that the neural dynamics alone, or at least the summary max projection statistic we looked at, are not sufficient for predicting the novelty or passivity of a given trial. In this paper, we only had the chance to explore the influence of two features, novelty and engagement, on the neural dynamics, but the Visual Behavior Optical Physiology dataset includes many more. Data was collected on the running speed of the mouse, its eye movements, the timestamps and orders of the different stimulus presentations, and the pattern of licking and rewards. It is likely that novelty and engagement influence the neural dynamics in a way that is not apparent without controlling for some of these outside, possibly confounding factors. This may suggest some form of casual model could be in order, but given the complexity of neural dynamics we found it hard to reasonably justify any models simple enough to be tractable. Another limitation of this project was that the calcium imaging data was only available as a single summary statistic for each session: the max projection. As such, we could not analyze the neural activity near events of interest such as image transitions or false alarms. Again, this may have caused the ambient neural activity to be quite overwhelming for our simple classification models.

Lastly, it is possible that PCA strips away much of the spatial information in our images. This could hinder the models ability to recognize the distinct clusters of neurons which might be quite informative. Future studies could try to incorporate some kind of convolutional data preprocessing step that does a better job of isolating this spatial information. Instead, a lot of the variance captured by PCA is may just be noise from the ambient neural dynamics.

5. Code

The code (and documentation) used to run the analysis, interact with the Allen Database SDK, and produce the above figures for this paper can be found on GitHub. The README file there contains information on the roles of different files and how I accessed data from the Allen Institute.

3. This metric is discussed further in the project's whitepaper

References

Woo-Young Ahn, Nathaniel Haines, and Lei Zhang. Revealing neurocomputational mechanisms of reinforcement learning and decision-making with the hbayesdm package. *Computational Psychiatry*, 1:24, 10 2017. doi: 10.1162/cpsy_a_00002.

Bilal, Millie Pant, Hira Zaheer, Laura Garcia-Hernandez, and Ajith Abraham. Differential evolution: A review of more than two decades of research. *Engineering Applications of Artificial Intelligence*, 90:103479, 4 2020. ISSN 09521976. doi: 10.1016/j.engappai.2020.103479.

Bob Carpenter, Andrew Gelman, Matthew D. Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Marcus A. Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell. Stan: A probabilistic programming language. *Journal of Statistical Software*, 76, 2017. ISSN 15487660. doi: 10.18637/jss.v076.i01.

Lecoq J.A. Buice M.A. et al. de Vries, S.E.J. A large-scale standardized physiological survey reveals functional organization of the mouse visual cortex. *Nature Neuroscience*, 23:138–151, 1 2020. ISSN 1097-6256. doi: 10.1038/s41593-019-0550-9.

Christine Grienberger and Arthur Konnerth. Imaging calcium in neurons. *Neuron*, 73: 862–885, 3 2012. ISSN 08966273. doi: 10.1016/j.neuron.2012.02.011.

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

Stefano Recanatesi, Matthew Farrell, Guillaume Lajoie, Sophie Deneve, Mattia Rigotti, and Eric Shea-Brown. Predictive learning as a network mechanism for extracting low-dimensional latent space representations. *Nature Communications*, 12:1417, 12 2021. ISSN 2041-1723. doi: 10.1038/s41467-021-21696-1.