

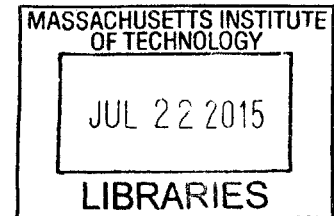
Generation and tuning of learned sensorimotor behavior by multiple neural circuit architectures

by

Michael Lynn

B.Sc. Biology
University of Ottawa, 2012

ARCHIVES



SUBMITTED TO THE DEPARTMENT OF BRAIN AND COGNITIVE SCIENCES
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE IN BRAIN AND COGNITIVE SCIENCES
AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

JUNE 2015

© 2015 Michael Lynn. All rights reserved.

The author hereby grants to MIT permission to reproduce and to distribute
publicly paper and electronic copies of this thesis document in whole or in part in
any medium now known or hereafter created.

Signature of Author: Signature redacted
Brain and Cognitive Sciences
May 8, 2015

Certified by: Signature redacted
Matthew A. Wilson
Professor of Neuroscience
Thesis supervisor

Accepted by: Signature redacted
Matthew A. Wilson
Professor of Neuroscience
Graduate Officer

Generation and tuning of learned sensorimotor behavior by multiple neural circuit architectures

by

Michael Lynn

Submitted to the Department of Brain and Cognitive Sciences on May 8,
2015 in partial fulfillment of the requirements for the degree of Master of
Science in Brain and Cognitive Sciences

ABSTRACT

Organisms have a remarkable ability to respond to complex sensory inputs with intricate, tuned motor patterns. How does the brain organize and tune these motor responses, and are certain circuit architectures, or connectivity patterns, optimally suited for certain sensorimotor applications? This thesis presents progress towards this particular problem in three sub-projects. The first section re-analyzes a large data set of single-unit recordings in zebra finch area HVC during singing. While HVC is known to be essential for proper expression of adult vocalization, its circuit architecture is contentious. Evidence is presented against the recently postulated gesture-trajectory extrema hypothesis for the organization of area HVC. Instead, the data suggest that the synaptic chain model of HVC organization is a better fit for the data, where chains of RA-projecting HVC neurons are synaptically connected to walk the bird through each time-step of the song. The second section examines how optimal sensorimotor estimation using a Bayesian inference framework could be implemented in a cerebellar circuit. Two novel behavioral paradigms are developed to assess how rats might tune their motor output to the statistics of the sensory inputs, and whether their behavior might be consistent with the use of a Bayesian inference paradigm. While neither behavior generated stable behavior, evidence indicates that rats may use a spinal circuit to rapidly and dynamically adjust motor output. The third section addresses the formation of habitual behaviors in a cortico-striatal network using rats. Stress and depression are known to significantly alter decision-making abilities, but the neural substrate of this is poorly understood. Towards this goal, rats are trained on a panel of decision-making tasks in a forced-choice T-maze, and it is shown that a chronic stress procedure produces a dramatic shift in behavior in a subset of these tasks but not the rest. This behavioral shift is reversed by optogenetic stimulation of prelimbic input to striatum, pinpointing a circuit element which may control stress-induced behavioral changes. Furthermore, a circuit hypothesis is presented to explain why sensitivity to changing reward values diminishes with overtraining.

Thesis supervisor: Matthew A. Wilson
Title: Professor of Neuroscience

1: Evidence against a GTE model of HVC dynamics during songbird vocalization

Introduction

Mature songbirds are able to produce a remarkably stereotyped, accurate song which varies from individual to individual and is based on imitating a tutor's song. How is this stereotypy produced by the brain? The premotor nucleus HVC has been shown to be essential for proper adult vocalization and song stereotypy: lesions in HVC severely degrade the song (Nottebohm et al, 1976). One model of song production (Fee et al, 2004) has a series of synaptically connected HVC neurons activate in sequence and walk the bird through each time-step of the song. Here, each HVC neuron bursts at the same time-point during each song rendition; and activates a set of RA motor neurons which control the muscles which should be active at that particular time-point in the song. By synaptically connecting chains of these neurons together, nucleus HVC can represent a temporally ordered sequence of complicated motor actions which is replayed virtually automatically. This model has received experimental support: First, cooling HVC slows the song uniformly, while cooling other regions, which may also be involved in encoding time, does not slow the song (Long & Fee, 2008). Notably, cooling HVC does not slow certain timescales of the song more than others (eg syllables). This is good evidence that a synaptic chain in HVC controls song timing exclusively, without the involvement of other regions. A second piece of evidence is that intracellular recordings from HVC neurons show activity consistent with a synaptic chain with no depolarization ramp before bursting (Long et al, 2010).

However, recent work has proposed an alternate model where biomechanical elements of the singing bird dictate HVC dynamics. Amador et al (2013) deconstructed the auditory features of bird's songs, extracting information about the bird's air sac pressure and labial tension. They then identified points in the song where these two parameters reached extreme points, called gesture-trajectory extrema (GTEs). In their model, HVC neurons would preferentially burst at these GTE points, representing vocal gestures rather than representing time as in the clock model. The authors presented a minimal data set of 6 HVC neurons, which seemed to support their hypothesis. Notably, Amador et al reported the precise statistics of how HVC bursts were distributed around GTE events (a normal distribution with standard deviation = 4 ms and mean < 1 ms.) To compare these two models I compiled and analyzed a larger data set comprised of 40 HVC projector neurons.

I will first compare the distribution of inter-burst intervals in the extended HVC data set with predictions from the clock model (predicting a homogenous poisson distribution) and from the GTE model (predicting an inhomogenous poisson process), using Monte Carlo simulations to arrive at empirically calculated 95% confidence intervals. I will also use point-process modeling to compare the two models with the extended data set in a principled way, constructing Kolmogorov-Smirnov plots. Finally, I will analyze the critical prediction of the clock model that HVC neuron bursts should cover time in a reasonably uniform manner, by looking at the probability mass function of the two models and of the data for different bin sizes. In all cases we find a clear agreement between the extended data set and predictions from the clock model, and the GTE model's predictions are not consistent with the data set.

Extracting burst times

The large data set consists of the activity of 40 HVC projector neurons. Each neuron was recorded over a number of song iterations, and generally produced spikes at the same time during each song iteration. Spikes which were not produced reliably across trials were

eliminated, and the remaining spikes were convolved with a gaussian with a standard deviation of 1.5ms, with the burst times being the center of mass of each cluster of gaussian-convolved spikes. Burst times were rescaled based on syllable length for each rendition. Fig. 1 shows the 66-burst set extracted from the data.

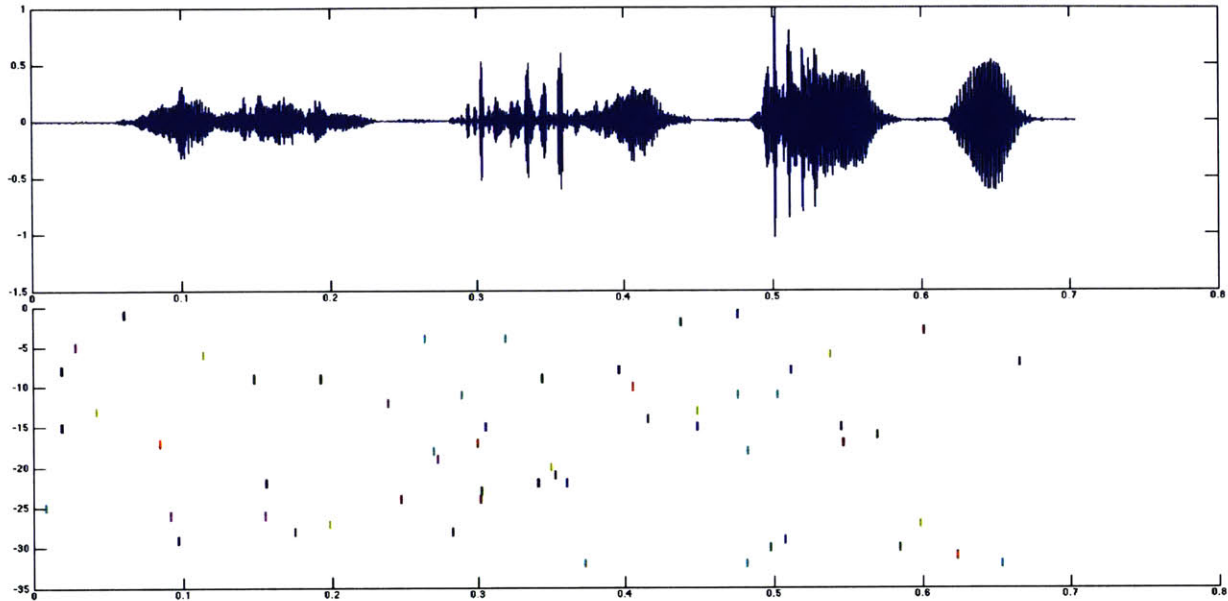


Figure 1: Extracted HVC bursts for 40 neurons (bottom) during a stereotyped song rendition (top).

Inter-burst interval distributions: Clock model and GTE model compared with empirical data

The clock model and the GTE model make different predictions about the statistics of HVC activity during song production. The clock model predicts that HVC bursts will not be centered around any song features, which is consistent with a homogenous poisson distribution. Other distributions could also be consistent with this theory, as long as the distribution covers all times so that a syn-fire chain can connect the neurons. Here, we will focus on a homogeneous poisson process to explain the clock model, as it is the most parsimonious explanation. The GTE model, on the other hand, requires that HVC bursts are centered on gesture extrema, making the testable prediction that HVC burst statistics would be represented by an inhomogenous poisson distribution with λ peaking around GTE events. In this section, I examine the inter-burst interval distribution from the data, and compare it to what would be predicted from a homogenous and inhomogenous poisson distribution.

Fig. 2 shows the distribution of inter-burst intervals for the 66-burst data set (blue), with an interval width of 5ms. If we consider the data as arising from a homogenous poisson process, $\lambda = \text{total events} / \text{total time} = 93.8$. Figure 2 also shows the theoretical curve of inter-burst intervals arising from a homogeneous poisson distribution with this λ (red), which appears consistent with the data.

To more rigorously compare the data with a theoretical homogeneous poisson process, we can calculate confidence intervals for the poisson process using Monte Carlo simulations. For each of 1000 trials, burst times were drawn from a homogenous poisson process with $\lambda = 93.8$, over the song interval [0,0.703]. For each inter-burst interval value, I then used the 1000 trials to find the empirical 95% confidence intervals, by calculating the upper and lower bounds that 95% of the simulated data fit into for each inter-burst interval value (figure 2, red stars). The data's inter-burst interval values are within the 95% confidence intervals in each case, meaning that we cannot reject the null hypothesis that the data is significantly different than the homogenous poisson distribution ($p > 0.05$). This provides evidence for the clock model of HVC function.

The GTE model, on the other hand, predicts that the inter-burst interval distribution from HVC neurons will be consistent with an inhomogenous poisson process, with a time-varying λ which peaks around GTE times. This section uses data reported in the paper originally hypothesizing the GTE model (Amador et al, 2013) to estimate how λ changes around GTE events, generating a large number of simulations each with their own time-varying λ and associated burst train.

Amador et al (2013) pinpointed 39 GTEs in the song associated with their data set. For each of 1000 simulations, I place 39 GTEs randomly on the interval [0,0.703], which is the song length. To generate a simulation's time-varying λ , I convolve each GTE with a gaussian (standard deviation 4ms, mean 0ms, as reported in Amador et al (2013).) and then normalize the area under this λ 's curve such that it is identical to the area under the time-invariant λ 's curve shown earlier ($\lambda = 93.8$). This allows a direct comparison between the inhomogenous and homogenous poisson distribution. There is now a time-varying λ associated with each simulation, which is based on realistic, reported parameters of the GTE hypothesis. For each simulation, a simple thinning procedure was used to generate a spiketrain drawn from λ , from which inter-burst interval distributions were used to calculate 95% confidence intervals similar to in the previous simulation. Figure 2 also compares the data's inter-burst intervals to those of the clock model (homogenous poisson process, red) and of the GTE model (inhomogenous poisson process, green). By examining the 95% confidence intervals of both models' distributions, it is clear that the data is fit best by the clock model, and is inconsistent with a GTE model distribution ($p > 0.05$ for inter-burst intervals between 0ms and 5ms).

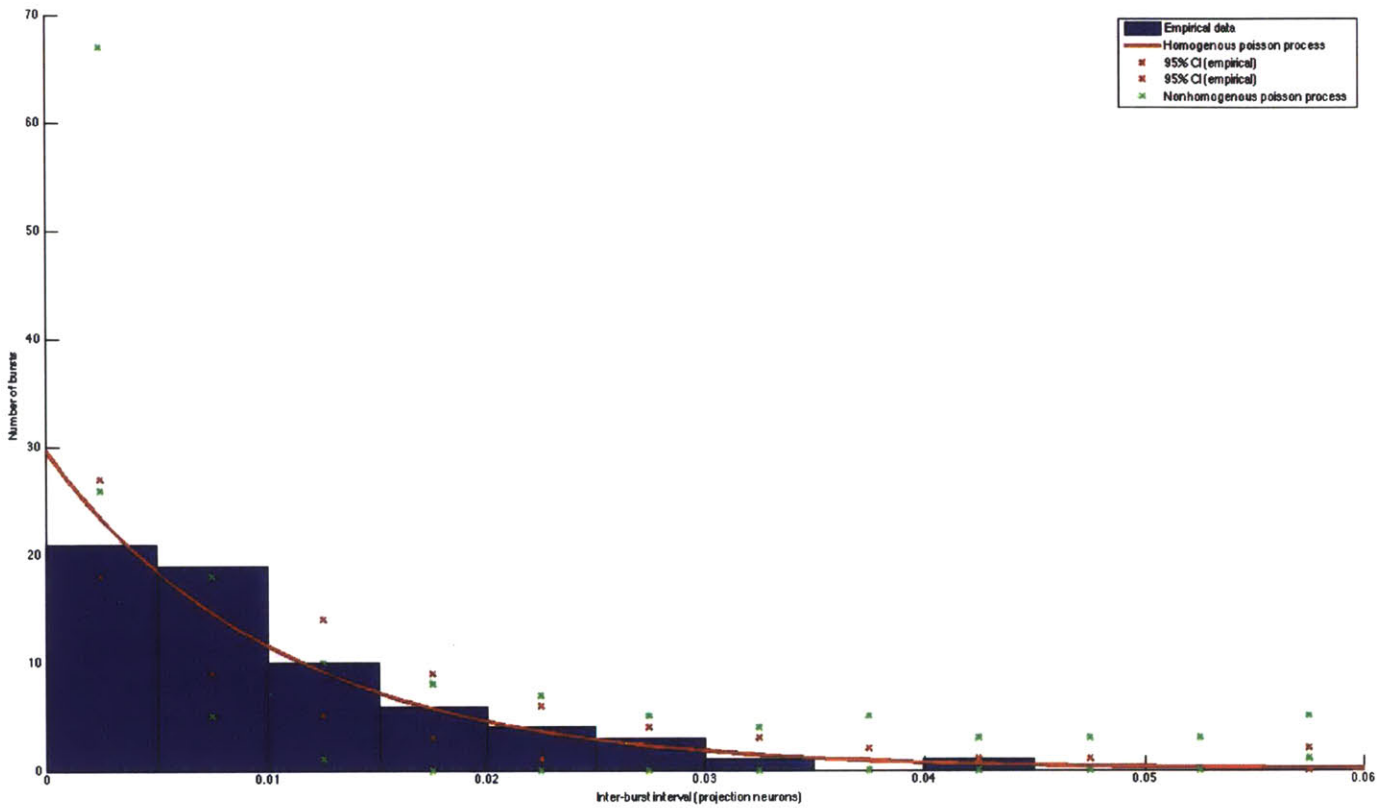


Figure 2: Inter-interval distribution of HVC bursts (blue), compared with the distribution arising from a homogeneous poisson process with 95% confidence intervals calculated from 1000 Monte Carlo simulations (red), and with a distribution arising from a nonhomogeneous poisson process (green).

Kolmogorov-Smirnov plots: the clock and GTE models compared with empirical data

The previous sections predominantly focused on using Monte Carlo simulations to compare the empirically observed inter-burst intervals with those from the two models. However, Kolmogorov-Smirnov plots provide a principled way of comparing models to data within the framework of point processes, with the advantage that they do not rely on simulations. Here, Kolmogorov-Smirnov plots are constructed to compare how well the clock and GTE models fit the data.

To compare the clock model with the data, I first time-rescaled the burst times based on a constant $\lambda = 93.8$, consistent with the clock model's homogeneous poisson distribution. I then placed these time-rescaled burst times on the interval $[0,1]$ by calculating $u(j) = 1 - \exp(-\text{time-rescaled_burst_time}(j))$ and sorted this new u . I calculated theoretical values $b(j) = (j - 1/2)/J$, where J is the total number of bursts. The pairs $(b(j), u(j))$ should fall on the unity line if the model agrees well with the data. I used 95% confidence intervals defined as $CI(j) = b(j) \pm 1.36/(J^{1/2})$. Figure 3 shows the K-S plot of the data compared with the clock model (homogeneous poisson distribution). The clock model fits the data well.

I then compared the GTE model with the data. All of the data's neurons were associated with one bird, and I used real GTE events for that bird's song, generating a time-varying λ as before. Figure 4 shows the K-S plot comparing the data with the GTE model, with 95% confidence intervals. The relationship is outside of the confidence intervals, especially with $CDF < 0.5$. It appears, then, that the GTE model is not consistent with the data ($p < 0.05$).

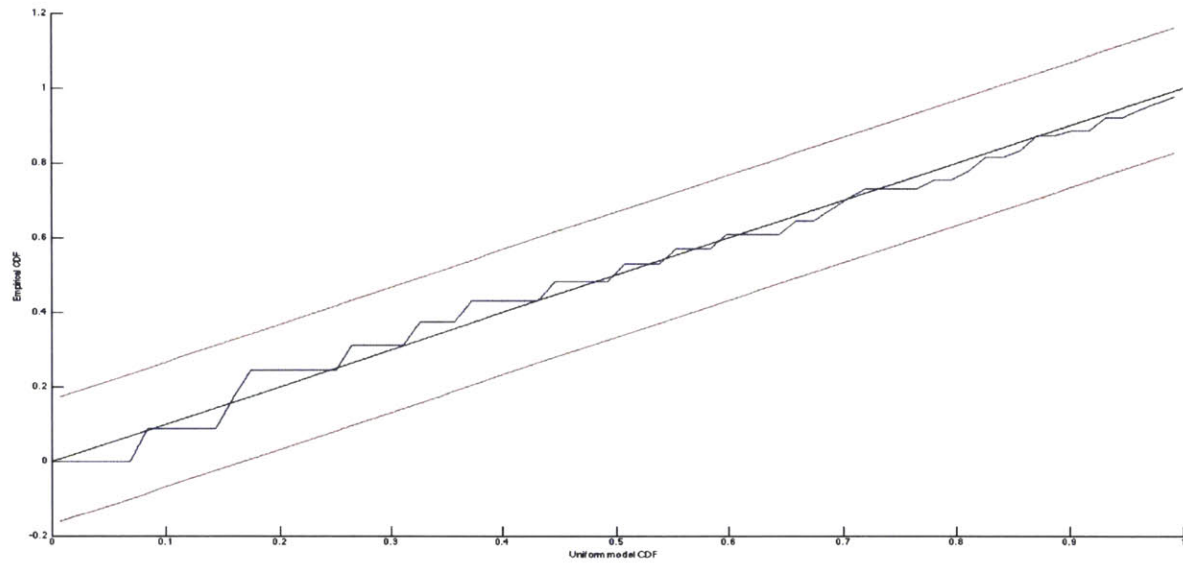


Figure 3: Kolmogorov-Smirnov plot comparing empirical data with the clock model (homogenous poisson distribution). 95% confidence intervals are shown.

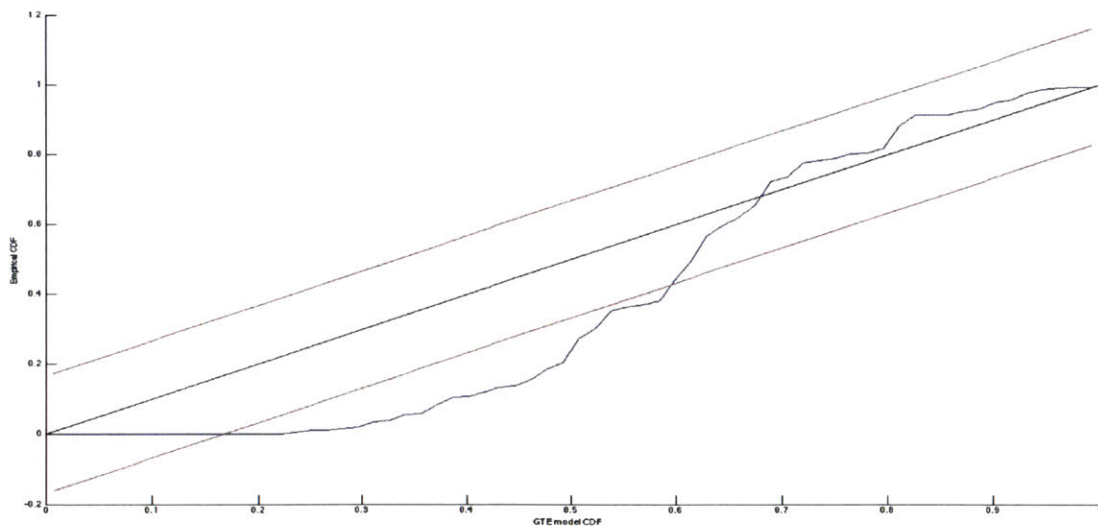


Figure 4: Kolmogorov-Smirnov plot comparing empirical data with the GTE model (inhomogenous poisson distribution). 95% confidence intervals are shown.

Probability mass functions: the clock and GTE models compared with empirical data

The critical difference between the clock model and the GTE model is that the clock model requires that area HVC continuously represents time during the song. A homogenous poisson distribution would be especially beneficial for this, since it would guarantee that as the number of neurons increased, the song would almost certainly be uniformly covered with HVC neuron bursts. In this section I test this crucial tenet of the clock model (uniform song coverage) by comparing probability mass functions between real data and Monte Carlo simulations of the clock and GTE model. An oversampling procedure is used to reduce the variability which may arise if the bin's start and end location are placed such that it does not see any bursts. Briefly, the data was binned, and then the bin start points were shifted by 5ms increments and binned again, repeating until the total number of shifts possible were reached. This oversampling procedure reduced the variability which may arise if the bin's start and end location are placed such that it does not see any bursts.

I analytically calculated the probability mass function for an idealized homogenous poisson process. For each bin size, $\lambda_{\text{bin}} = \lambda \cdot \text{bin_size}$. Then the probability mass function $\text{pmf}(k) = ((\lambda_{\text{bin}}^k) / k!) \cdot \exp(-\lambda_{\text{bin}})$, with k being the number of projectors active in a bin. I used Monte Carlo simulations to generate a mean probability mass function for an inhomogenous poisson process. I then plotted the probability mass functions from the two models and from the empirical data, with bin sizes of 10ms and 50ms, although similar results are obtained with other bin sizes. Fig. 5 and 6 present these probability mass functions. The clock model (homogenous poisson process) fits the data much better than the GTE model in all cases.

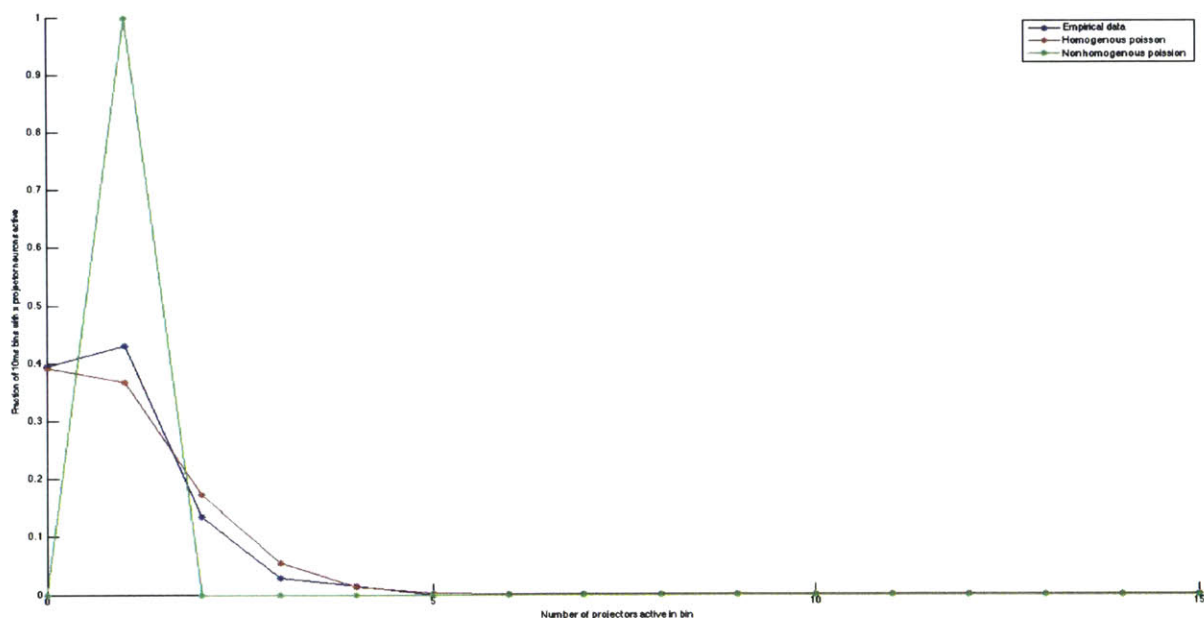


Figure 5: Probability mass function of the data, the clock model, and the GTE model, with a 10ms bin size. The clock model's curve is calculated analytically, and the GTE model's curve is calculated by Monte Carlo simulation.

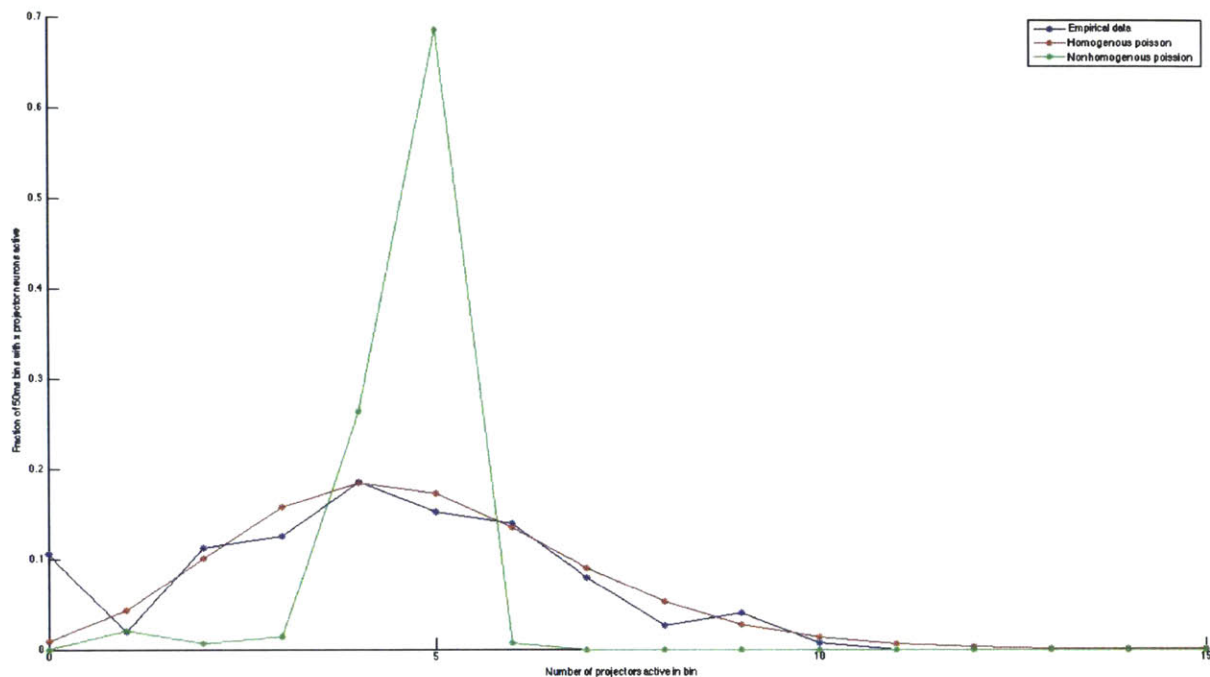


Figure 6: Probability mass function of the data, the clock model, and the GTE model, with a 50ms bin size. The clock model's curve is calculated analytically, and the GTE model's curve is calculated by Monte Carlo simulation.

Discussion and conclusion

The clock model is an elegant way of explaining how premotor area HVC could control stereotyped song production, with neurons connected in a syn-fire chain to form a continuous representation of time through a sequence of sparse bursts propagating through the chain. While the GTE model appears to incorporate elements of biological significance (for instance, having 'important' song components represented preferentially by HVC neurons), this model does not provide a clear explanation for how stereotyped songs could be produced virtually automatically under the control of HVC. Fortunately, both models make opposing predictions about the statistics of HVC bursts during song production, and here I compile and analyze a larger set of HVC neural data to find major evidence against the GTE model of HVC dynamics.

The first section compares inter-burst interval distribution predictions from both models with distributions from the real data. Here, the GTE model explicitly requires an inhomogenous poisson distribution for HVC burst times: bursts must be concentrated around gesture extrema. This has clearly been shown to be inconsistent with the data, especially in the interval of 0-0.05 seconds where the GTE model would predict more inter-burst intervals than are in the real data set. The clock model does not require a homogenous poisson distribution for HVC burst times (as long as the song coverage is continuous, the model is not ruled out), but a homogenous poisson distribution does indeed seem to fit the data well, even using 95% confidence intervals generated by repeated Monte Carlo simulations. One idea is of course a hybrid model where HVC neurons are connected in a syn-fire chain, but proportionally more neurons are active at time-slices with complex vocalizations than at time-slices with simpler vocalization. However, this is not supported by the data either, since the 0-0.05 second inter-burst interval bin in the

real data set has slightly *fewer* intervals than in the homogenous distribution simulation (although within the margin of error), rather than slightly more as predicted by the hybrid model. This under-sampling of short inter-burst intervals could have many causes, including variability due to a small n .

In the second section, Kolmogorov-Smirnov plots are used to compare the difference between cumulative distribution functions of the real inter-burst intervals with predictions from both models. This metric is useful because it provides a principled way of comparison with confidence intervals. Here, we can conclude that while the clock model is in good agreement with the data, the GTE model does not agree, especially in short inter-burst intervals, with 95% confidence. Finally, in the third section, we compare quantitative predictions about the probability mass distributions for both the clock and the GTE model, finding that the data fits with the clock model much better. In this section, the difference between the models is dramatically illustrated: the GTE model predicts that in a given time-bin, there will be a less variable number of projector neurons active, but the clock model predicts that the number of projectors active will be slightly more variable.

This section presents evidence against the GTE model of HVC singing-related dynamics. However, Amador et al (2013) raise an intriguing point: Is it necessary for song structure to be reflected in HVC burst structure, given that HVC controls stereotyped song? Amador et al (2013) clearly think this point to be biologically obvious. However, we can consider a simple alternate explanation whereby HVC *only* encodes time through a syn-fire chain of projector neurons (identical numbers of neurons are active at each point in the chain), but that song complexity is instead dictated by the number of synapses onto downstream RA neurons. Here, HVC projector neurons active at times of gesture extrema may project to comparatively more RA neurons than HVC projectors active at times of silence.

2: Bayesian inference for automatic motor compensation in rats

Bayesian inference and possible neural circuit mechanisms for implementation

When sampling from the environment, individuals can potentially arrive at inaccurate estimates due to measurement error. In one form of optimal estimation, Bayesian estimation, the measurement error can be combined with knowledge of the statistical distribution characterizing an event, to arrive at a statistically optimal estimate of the true event sample. For example, while estimating the speed of an incoming baseball may be a potentially inaccurate task, if the instantaneous measurement and its typical measurement error is combined with knowledge of the mean and standard deviation of baseball velocities experienced by that pitcher, the sample's estimate can be optimized.

Recent work (Jazayeri & Shadlen, 2010) has shown that the framework of Bayesian estimation can convincingly explain human psychophysical results, where individuals are asked to estimate the length of a single presented time interval. Here, the individuals combine an instantaneous, noisy measurement of the sample time interval with a knowledge of the prior distribution of the time interval in previous trials. In particular, given a fixed, presented time interval to estimate, individuals will consistently give a lower estimate of the presented time interval if the past history of interval distributions is low; and they will conversely give a higher estimate of an identical, presented time interval if the past history of interval distributions has a higher mean.

How would this optimal estimation mechanism be implemented neurally? One population of neurons could represent the sample measurement, with the mean firing rate representing the measured estimate while the standard deviation of the firing rate could represent the standard deviation of the measurement error. Yet another population could represent the prior

experienced distribution of samples in a similar manner, and these two population responses could be combined somehow to arrive at an optimal estimate. This basic idea has been outlined in previous theoretical work (Hoyer et al, 2002) but has never seen a real neural implementation. While primate work is currently underway to address elements of this estimation paradigm, rodents provide an elegant model system where circuit elements can be perturbed in a reliable way, and large-scale population recordings from multiple animals is relatively easy. One specific hypothesis for simple cerebellum-dependent sensorimotor task in rodents is that each cerebellar Purkinje cell is performing a Bayesian computation with inputs sampled from the prior and likelihood. In other words, each mossy fiber gives a sample from the likelihood distribution of the sensory input, and the Purkinje cell's firing pause gives the best-estimate of the true value of the sensory input. Each Purkinje cell's firing pause would then be sampled from the posterior distribution in a Monte Carlo fashion.

This section presents experimental progress towards a behavioral paradigm in rodents which would allow rigorous testing of this Bayesian inference framework, using simple sensorimotor tasks. Two basic tasks are presented. In the first task, rats are placed on a rotating rod and must correctly estimate the timing of rod rotation events, or else they lose balance and fall off the rod. In the second task, rats are placed on a treadmill where the treadmill's speed varies, and they must estimate the timing between velocity changes or else they fall back on the treadmill and hit an electrified grating. Both tasks involve building and programming equipment; and the second task additionally involves the use of a high-speed camera to track foot movement. While both tasks were unsuccessful at the behavioral stage, behavioral data is presented which suggests, intriguingly, that basic sensorimotor tasks may rely on a mechanism involving gait dynamics or a spinal circuit to rapidly adjust estimation during the course of a behavioral choice, rather than relying on a sophisticated brain circuit to make an optimal estimation prior to behavioral execution.

For all behavioral tasks, rats were kept under water restriction with a target of 90% pre-restriction body weight; and were given chocolate milk for correct task performance (typically 1mL per trial). In the first task, I set up a rotating rod controlled by a motor and imaged with a high-speed IR camera. The rat sits on the rod, which rotates at some predictable intervals which the rat must guess. The motor and camera are connected to National Instruments cards and are programmed using LabView. The reward system is custom-made and controlled using MatLab. Rats were tested on this paradigm for approximately a week, but the problem here was that we could not motivate the rat to stay on the rod. Since jumping down was not aversive, the rat could not be trained to move at the correct time to avoid falling. Additionally, the rat's position on the rod was extremely variable, and it appeared to use its tail in a variable fashion, which would affect how well it could react to balance perturbations. Therefore we tried another approach.

In the second behavioral paradigm, rats are placed on a treadmill (PanLab single-lane rat treadmill). The treadmill is controlled via the COM port of the computer, using Matlab, and the camera is still controlled via LabView. In this paradigm, the treadmill's speed is changing in a predictable way and the rat must learn to generate anticipatory motor commands. I developed a reward paradigm whereby rats learn to pair chocolate milk dispensing with an LED going off. We can then use the LED as a conditioned response to guide behavior. The rats were tested on a number of paradigms on the treadmill over the course of 2 months. First, rats learned to run at a constant speed for 30-60 seconds to receive a chocolate milk reward. This was primarily to acclimatize them to the chamber. Second, the rat was exposed to a sinusoidal velocity profile on the treadmill. With the period fixed, we predicted that the rat would go through two major learning phases. Early in learning, the rat would fail to correctly predict the speed profile and would instead react to the observed speed, with the outcome that the rat would oscillate back

and forth in the treadmill axis over time. Later in training, if the rat could correctly predict the speed profile and organize a pre-emptive motor plan for each sinusoid, the outcome would be that the rat would be stationary in the treadmill axis over time, rather than oscillating back and forth. It is likely that this simple sensorimotor paradigm would reflect cerebellum-dependent updating of motor predictions, similar to previous work in electric fish (Bell et al, 1997). If this were the case, the paradigm would provide a powerful platform to ask questions about how the prior over different sinusoidal speed regimes would be represented and updated in the cerebellum.

To characterize the rat's position in the treadmill box quantitatively, I developed a simple computer-vision algorithm which was given a stripe on the rats' bodies, and found the most likely position of this stripe in each frame using a least-squares method on normalized pixel intensities. The algorithm was additionally designed to only search for matches near the previous frame's best template match, to render the results more robust against false matches. Additionally, the pixel-value results were rescaled as true distances from the front of the treadmill. The algorithm was able to extract the positions of 2 rats in a robust fashion. Figure 7 shows a frame from one rat's run depicting equipment setup.

The rats' position appeared to vary in step with the treadmill's sinusoidal velocity profile. Figure 8 shows the given treadmill velocity and extracted rat's position for short segment of a full 30-second run. The rat's position appears to reach its most posterior position with a slight time-lag after the maximal treadmill velocity. Could the rat's velocity, rather than position, be modeled as a sinusoid with the same frequency but slightly different phase lag compared to the treadmill's velocity? This comparison of velocity profiles, rather than raw position data, would allow a more direct comparison to models of how the rat's sensorimotor machinery may be reacting to the input (for example, duplicating the treadmill's velocity with some constant time lag). To assess this, the derivative of the rat's position vector was taken to produce a velocity vector for the rat. A Fourier transform was performed on this velocity vector and all frequencies but the highest-amplitude one were filtered out before transforming the data back into the time domain (Fourier filtering). A short segment of this data is plotted along with the treadmill velocity data in Figure 9. As expected, the major component in the rat's velocity data was a sinusoid with its frequency a multiple of the treadmill velocity sinusoid (twice that of the treadmill velocity). The frequency doubling is likely a behavioral artifact arising from the fact the rat's stride frequency is double that of the treadmill velocity, so the body stripe being imaged has the same doubling of frequency. We can also determine the time lag between the peak treadmill velocity and the decreases throughout the running session due to the slightly higher frequency for the rat velocity sinusoid. Collapsing time lags across the entire running session gives a mean lag of 0.413 ± 0.180 s, with the high standard deviation corresponding to the gradual lowering of lag as the running session proceeds. One interpretation of this result could be that as the session proceeds, the rat improves its predictions and reacts with smaller time delays to the treadmill velocity changes. However, this is a simplified set of data due to the Fourier filtering, and the raw extracted position data shows that even two subsequent treadmill velocity perturbations can give different rat position time-lags (Figure 7). Although the data suggest that learning may occur within a session, it is difficult to draw any strong conclusions due to the behavioral variability.

Can the Fourier-filtered rat velocity data be compared with model predictions for how sensorimotor circuitry is producing outputs? I considered a simple model where the rats replicate the observed treadmill velocity with a constant time delay ($v_{rat}(t) = v_{treadmill}(t - t_{delay})$). However, the time shift between the rat's peak total velocity and the peak treadmill velocity converges to period/4 (0.25 s in this case) as t_{delay} approaches infinity, while a shift of 0.413 s is needed as discussed above. Clearly the simple model is not a good fit.

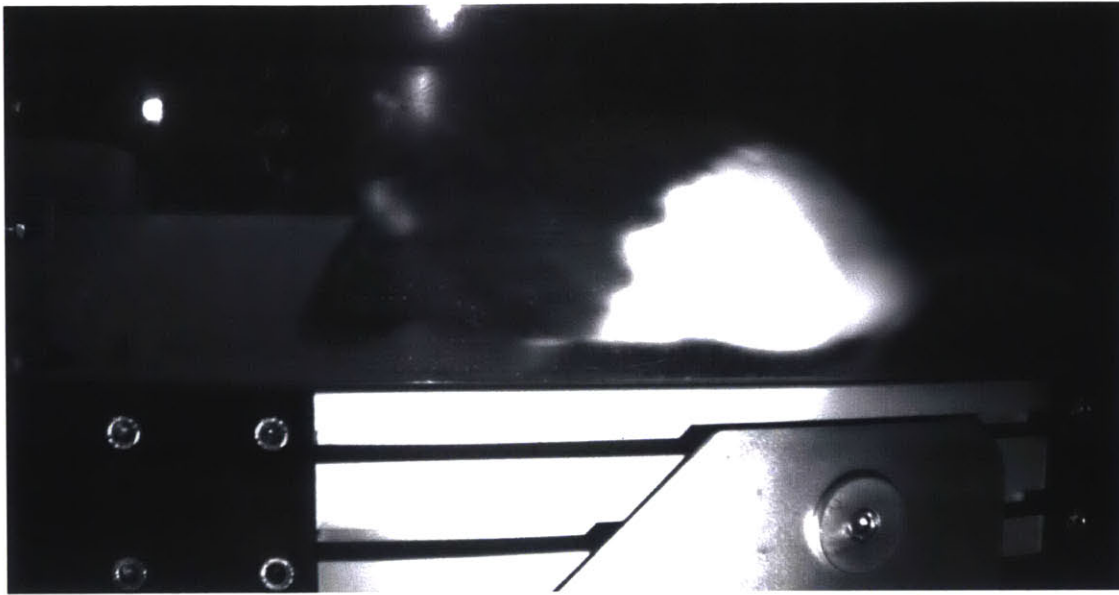


Figure 7: Still frame from a movie of one session, depicting equipment setup and rat placement. The light in the top-left corner indicated maximum treadmill velocity and was used to correctly fit the treadmill velocity profile to the video.

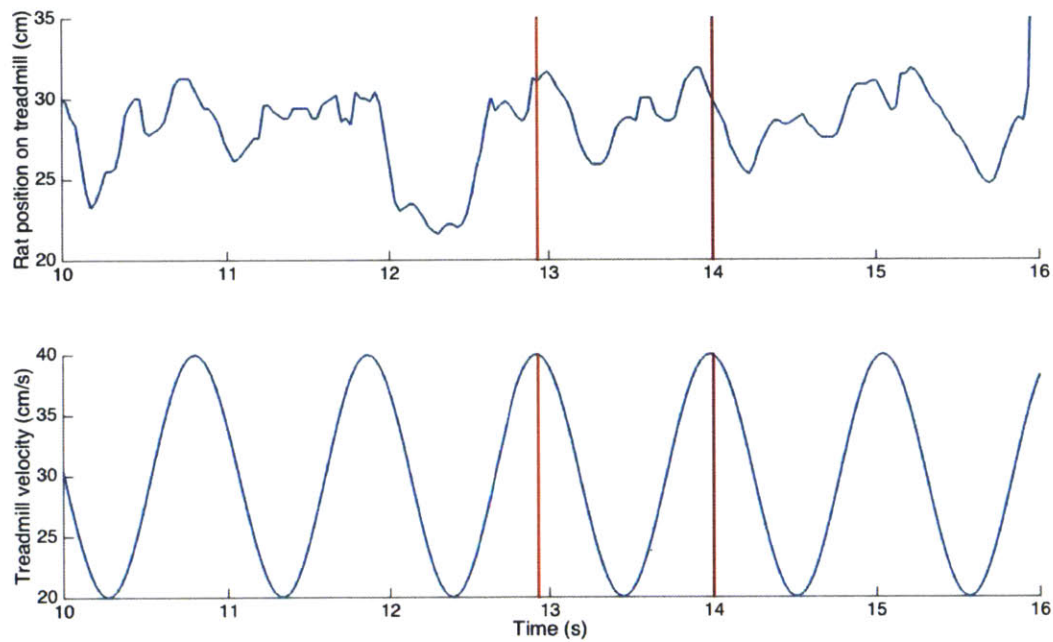


Figure 8: The rat's position on the treadmill, extracted using a computer-vision algorithm, for a short segment of the total 39-second session for one rat. For comparison, the treadmill velocity is depicted at the bottom. The red lines indicate peak treadmill velocities and show behavioral variability in rat position at these points.

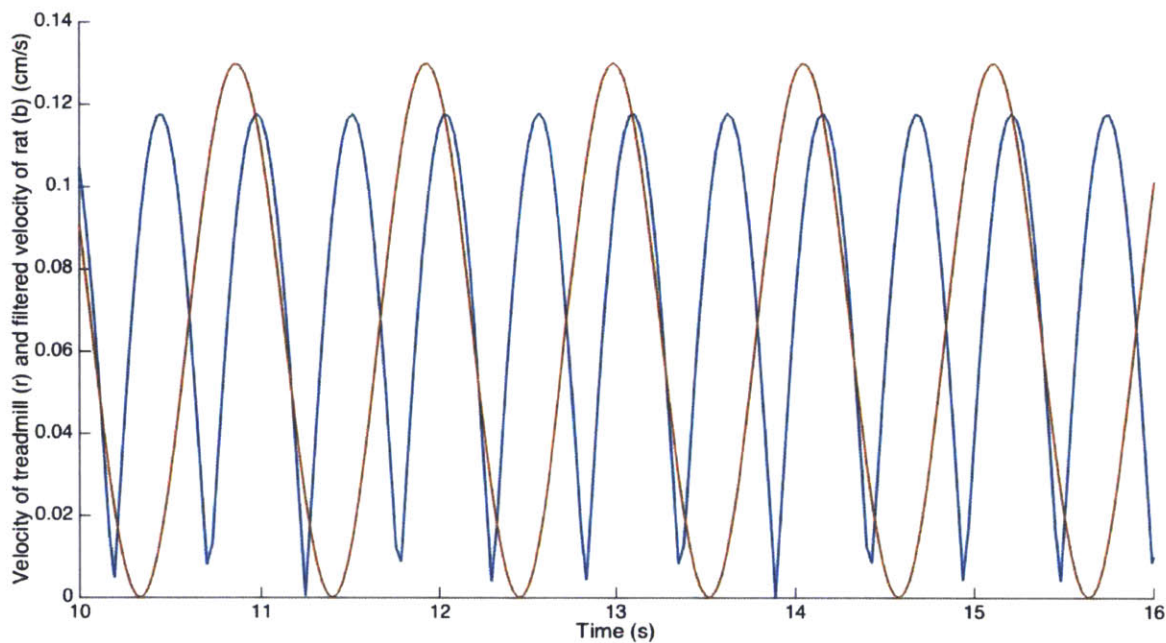


Figure 9: The treadmill's velocity (red) compared with a Fourier-filtered version of the rat's velocity incorporating only the dominant sinusoidal component (blue). A short segment of the entire 39-second run is depicted.

Are the rats truly predicting the treadmill velocity changes, or are they simply reacting to them? To more fully test this, I exposed the rats to square wave velocity functions where the times between square waves were drawn from a Poisson distribution with a mean of 1 second, making it impossible for the rat to predict when the velocity would change. A nonpredictive spinal loop should generate identical reaction times to unpredictable square waves compared with predictable sinusoids, while a predictive cerebellar circuit should produce larger reaction times in the unpredictable square wave test compared with predictable sinusoids. Paradoxically, we saw very short reaction times to square waves (almost instantaneous), which was different than the slightly longer reaction times we saw when giving the rat a sinusoidal speed profile. This was very surprising - if rats were truly predicting in the sinusoidal treadmill velocity task, and shortening their time delays as they learn the task over the course of one trial, then one should expect that unpredictable, random events such as the square waves given should produce *longer* time delays than in the predictive sinusoidal treadmill velocity task. The fact that the opposite happened, and that the rats produced shorter time delays in response to random square waves, suggests that perhaps the rats were not being predictive in the first sinusoid task.

What underlies the vastly different time delay responses to sinusoids and square waves? This finding could be explained in two ways. First, a spinal mechanism could mediate this task, with an input of leg positions and an output of muscle contractions to optimize leg position invariance over time. This spinal mechanism could be more receptive to large accelerations (square wave) than more gradual ones (sinusoid), such that it would be able to quickly correct for square wave perturbations while correcting more slowly for sinusoidal perturbations. Second, the rat could be integrating information from multiple sensory modalities to make a decision, and could be using each sense a different amount on a trial-by-trial basis, which would generate variability in the perceived reaction times. The rats specifically like to

press their whiskers against the front of the treadmill in the square wave task, which seems to lead to very short reaction times; but when they do not have this sensory cue and are further back on the treadmill (sinusoidal wave task), they have longer reaction times. (They are likely using a combination of proprioceptive input and vestibular system input in this case, neither of which are as instantaneous as whiskers).

Discussion and conclusion:

This section explored rodent sensorimotor tasks designed to investigate how noisy sensory inputs are transformed into highly optimal motor outputs, based on prior knowledge of the input statistics. In the first task, rats balanced on a rod and had to make predictive movements as to when the rod would rotate. Here, the rats did not reliably stay on the rod, and their tail usage was unpredictable, generating variable behavior. In the second task, rats were forced to anticipate treadmill velocity changes for sucrose reward. This task generated promising behavior. The rat's position on the treadmill varied as a sinusoidal velocity profile was provided to the treadmill, consistent with the rat making slightly inaccurate measurements about the current speed. Fourier filtering showed that the rat's velocity had a time lag from the treadmill velocity which decreased as the session progressed, possibly indicating learning over the session. However, the rat's behavior was not consistent; and moreover, when Poisson-distributed square waves of velocities were shown to the rat, its motor lag actually *decreased*, contrary to what would be expected from an unpredictable stimulus compared to the more predictable sinusoid. Either the rat is relying on a simple spinal loop to generate rapid motor outputs; or it is integrating sensory input from multiple modalities which vary in contribution over time and lead to varying time lags in the motor output.

A major problem in this section was behavioral consistency. To this end, a similar but more consistent paradigm would be to build a virtual reality machine for mice, where they run on a ball and are presented with a curved screen of visual inputs. For angular movements, the gain between the ball's rotation and the on-screen rotation could be drawn from a distribution, resulting in a different gain value for each trial. If the mice were presented with a curved maze to navigate, then for the first few seconds of running each trial, they would have to integrate limited sensory information about the correspondence between running direction and screen turning with a knowledge of the prior from which gains are drawn. This approach has many strengths. First, it would likely produce far more reliable behavior, since the mice would have a sparser set of sensory cues rather than the combined vestibular, proprioceptive (walls) and auditory cues which fluctuate in their contribution over time in the treadmill task. Second, the motor output would be easily and unambiguously measured as the ball's direction, instead of having to rely on a camera and an algorithm to extract position vectors. Third, the virtual reality machine could be reprogrammed with parameters far more easily than a physical treadmill. However, the main drawback to this paradigm is that the mice find it far harder to produce an angular shift of the ball than to simply run in a straight line; and this angular shift is often accompanied by odd, irregular motor patterns involving all four paws; moreover, angular accuracy is not very high and mice often have to correct angular outputs more than they correct linear acceleration outputs (personal communication, Dr. Chris Harvey). Therefore, it may be unlikely that mice could achieve the high motor precision required for our version of the task.

Finally, in this lab I additionally learned to set up a rodent colony, design a water restriction paradigm, order and build a surgery platform for simple craniotomies and injections, write surgery addenda, and perform perfusions and simple retrobead injections.

3: Prefrontal input to dorsal striatum: A potential role in online biasing of complex decision-making processes

Functional anatomy of dorsal striatum:

Animals can make behavioral choices using a variety of distinct mental frameworks. For example, in what is called goal-directed responding, individuals explicitly associate certain actions with certain outcomes, and have the ability to update outcome values to guide action choice (for instance, if the animals are satiated they will no longer press a lever for sucrose water reward). In a second major type of responding, called habitual responding, a cue reliably drives action production, with no regard for the actual outcome's value, so habitual responders are quicker to execute the action but are less sensitive to changes in reward value, and will often repeat an action long after it is behaviorally optimal. When learning a simple behavioral task such as a T-maze, animals typically start out as goal-directed but shift towards habitual responding over a period of weeks (Thorne et al, 2010). A key question is how this change in mental framework is reflected in the brain, and indeed how the eventual motor output is altered in a reliable manner. Is there a solitary behavioral choice region that shifts its activity during learning, or are multiple regions recruited to compete for behavioral control throughout the course of learning?

Portions of the striatum appear to play crucial, separable roles in controlling the two types of behavioral responses. The dorsomedial striatum (DMS) receives convergent input from parts of prefrontal cortex including prelimbic (PL), infralimbic (IL) and anterior cingulate cortex (ACC), while the dorsolateral input primarily receives input from the somatosensory and motor cortex (Pan et al, 2010); both send output back to these regions through the direct and indirect outflow pathways of the basal ganglia. DMS appears to be essential for volitional, goal-directed behaviors present early in training: if DMS is lesioned, even animals in the early stages of training become largely insensitive to any reward value alterations - for instance, if the underlying reward amount is decreased, DMS-lesioned animals will continue to perform a lever press previously associated with that reward (Yin et al, 2005). Surprisingly, lesions of the other dorsal striatum subdivision, DLS, appear to have no effect on animals in early stages of training - they continue to respond in a goal-directed fashion to cues and can alter their actions flexibly based on the underlying goal values. However, DLS lesions prevent animals from transitioning to habitual responses, where they would respond quickly and automatically but would not be as sensitive to changes in the underlying reward value (Yin et al, 2004). In other words, DLS lesions produce animals which never lose the ability to change their behavioral response based on changing reward values, but that cannot respond as automatically and quickly to tasks. The functional separability of these two areas has been confirmed in more recent work using inactivation methods (Gremel & Costa, 2013). The emerging story appears to be that while DMS biases prefrontal networks to produce volitional, goal-directed behavior early in training, DLS plays an essential role in biasing motor networks to respond automatically to cues with the appropriate action later in training.

This is borne out by recent electrophysiological studies, testing the prediction that DMS should be active early in training while DLS is active later in training, if the functional interpretations of lesion studies are to be believed. Indeed, Thorne et al (2010) found that DMS neurons produce spikes early in learning, and these spikes are concentrated during cue presentation and at choice points (both midway through the task) - DMS neurons largely stop spiking later in learning, when the animal has been tested to be responding habitually. In contrast, DLS neurons, while active early during training, are mainly active at late stages of training when the animal is insensitive to goal value changes and thus is responding habitually.

Here, DLS neurons form task-bracketing activity rather than midrun activity as with DMS neurons. The task-bracketing activity of DLS neurons appears to be a robust phenomenon (Kubota et al, 2009; Root et al, 2010).

Taken together, these results could be interpreted as supporting a model whereby as animals learn a forced-choice task, there is continual competition between two groups of brain regions: one supporting model-based, flexible responding (including prefrontal regions and DMS) which dominates early in training; and another group supporting rapid, habitual responses (including motor cortex and DLS) which dominates later in training. One computational model suggests these two systems might compete for control of behavior, based on Bayesian uncertainty in the ability to predict outcomes (Daw et al, 2005). An obvious question is how these two groups of brain regions could mechanistically compete to drive different motor outputs based on a fixed sensory input - this has been the subject of intense investigation with no clear answer. A simple model could be that early in training, DMS receives information from limbic centers like PL about perceived goal value and satiety state, and provides a premotor bias to particular channels of ACC incorporating to bias one of the choices over the other. ACC would then convey this choice preference to M1, activating the correct set of muscles for that choice. This concept of a premotor bias is similar to that proposed in birds between striatum and cortex (Fee & Goldberg, 2011). Later in training, PL activity might diminish, decreasing DMS activity and decreasing the bias ACC provides to motor cortex. It is still unclear how DLS would fit into this story, and specifically how lesions in DLS would prevent habitual responding.

Building on this work, this section will address two related questions. First, focusing on the DMS portion of striatum, can a chronic stress procedure change the behavioral choice output of rats, when having to make a forced choice between different amounts of costs and benefits? And could this change in behavioral output be reflected by either a change in PL firing dynamics (encoding choice elements), or by a change in DMS medium spiny neuron dynamics? In other words, during chronic stress is there a change in prefrontal input, or a change in the synaptic strength between prefrontal networks and DMS MSNs? And finally, if the prefrontal-DMS synapses are truly controlling behavioral choice, can we reverse the decision-making shift by chronic optogenetic stimulation of prefrontal-DMS connections? I will investigate this behaviorally and using optogenetic tools, although I also learned electrophysiological recording techniques (there was no access to this data at the time of writing).

The second question relates to how the DMS-DLS divide represents choice throughout learning, and specifically provides a circuit-level hypothesis as to how changing reward values early in learning results in rapid behavioral flexibility, but changing reward values later in learning does not result in the same speed of flexibility. The orbitofrontal cortex is a portion of cortex which projects specifically to dorsomedial striatum (Gremel & Costa, 2013) and tracks perceived reward value well (van Duuren et al, 2007). Here, the hypothesis is that upon a change in the reward value (for instance, reversal of contingencies), orbitofrontal cortex updates its representation quickly and sends a global stop signal to those medium spiny neurons in DMS representing the currently favored behavioral choice (stopping direct pathway outflow to ACC and thus premotor bias), allowing the previously disfavored choice to dominate through ACC premotor bias. I will propose experiments integrating electrophysiology with optogenetics and behavior to test this.

How does stress affect a prefrontal-striatal network during cost-benefit decision making:

As discussed above, the dorsomedial striatum receives convergent input from many parts of prefrontal cortex including prelimbic cortex, and likely integrates information about the animal's state (eg thirst) with outcome values to guide an appropriate premotor bias. Dorsomedial striatum is specifically involved in goal-directed, or non-habitual, responding. How

does this network assess possible behavioral choices which each involve a balance of rewards with costs?

Previous work (Graybiel lab, unpublished) has started addressing this important question. Rats are trained on a forced-choice T-maze task where the two arms each have various combinations of rewards (chocolate milk in different dilutions) and costs (an aversive light at different brightnesses). The task layout is shown in Figure 10. Notably, two versions of the task lead to a simple decision for the rat: In both the benefit-benefit and cost-cost versions, the rat must simply pick the greatest amount of reward or the least cost. However, in the cost-benefit version, the rat must assess the *relative* weighting of costs and benefits, presumably based on internal factors such as satiety or mood. This likely involves the prefrontal-DMS network discussed earlier as it incorporates limbic inputs. The animals appeared to make stable decisions in all three versions of the task. Notably, optogenetic inactivation of PL projectors to DMS only shifted choice in the cost-benefit version (towards the high cost option) and did not affect the other versions, indicating PL may encode task elements unique to this conflict version but not the others. Consistent with this, antidromically identified striatal-projecting PL neurons reached peak spiking early in the cost-benefit version of the task, before the decision, but reached peak spiking later in the other versions of the task after the decision-point - indicating that they may be guiding the decision only in the cost-benefit version of the task but not the others. From this, we can conclude that a network involving medial prefrontal cortex and dorsomedial striatum appears to control complex choices involving weighting the relative value of costs and benefits (based on internal state and perceived value); but does not appear to be involved in simpler non-comparative decision-making.

If PL is providing information about internal state to DMS to guide decision-making, then how would the system and behavioral output change if the animal's internal state is perturbed? To test this, rats were placed in a chronic stress procedure involving hours of complete immobilization every day. I then exposed them to the same T-maze task described above. It appears, as expected, that exposure to chronic stress alters the responses to the cost-benefit version of the task (incorporating elements of internal state) while leaving responses to the other versions intact. Specifically, chronically stressed animals choose the arm with higher reward but higher cost, rather than the arm with lower reward and lower cost as in unstressed animals. Figure 11 describes these results collected from x animals. The interpretation here is that chronic stress has altered the animals' relative weighting of rewards such as chocolate milk compared to aversive costs such as light. They appear to be weighting reward acquisition highly, in a manner more independent of associated costs compared to before - possibly due to a habituation to negative valences due to chronic stress exposure. How would the prefrontal-DMS circuit track this change in behavioral choice? The model most consistent with the previous results described above would be that PL provides information to DMS about the relative weighting of costs and benefits (based on the animal's internal state), that changes in this PL input alone are sufficient to alter behavior, and that upon chronic stress exposure, PL has altered its input to DMS and this is responsible for the altered behavior. This behavioral paradigm also has enormous clinical implications, since the altered decision-making is often a phenotype associated with human conditions such as depression and suicide, where the weighting of rewards and costs is altered. It would have great relevance to pinpoint the pathway involved in this altered decision-making, and to revert the decision-making back to normal parameters.

As a first step, we reasoned that if the PL-DMS pathway is sufficient to coordinate an altered behavior in the cost-benefit paradigm, then optogenetic activation of PL terminals in DMS should alone be enough to revert the animal's behavior back to a more normal weighting of costs and benefits. An excitatory channelrhodopsin, C1V1, was packaged in an AAV5 vector

and put under control of the CaMKIIa promoter, and 0.2uL of virus was bilaterally injected in PL in x rats. After 1 month of chronic stress (immobilization in plastic wraps for 6-8 hours daily), optic fibers were implanted in dorsomedial striatum. During the cost-benefit T-maze task, PL terminals in DMS were optogenetically stimulated from gate opening to reaching of the reward, and behavioral results were tabulated. Figure 12 summarizes these results for 6 animals. Optogenetic stimulation of the PL-DMS pathway reverted decision-making towards the low-cost, low-reward option, while optogenetic stimulation in control rats produced an even greater bias towards the low-cost, low-reward option. These results are critical because, to our knowledge, they represent the first instance of stress altering cost-benefit decision weighting in a task in a repeatable way; and they represent the first time that optogenetic stimulation of a solitary pathway has reversed this decision-making deficit - which may have broad clinical implications for depression and suicide.

Recall that previously, PL projector neurons were shown to have pre-choice spiking in the cost-benefit version of the task, but this pre-choice spiking was not as prevalent in other task versions such as benefit-benefit comparison. This implied that PL projectors may encode information about the relative weighting of costs and benefits or internal state factors involved in making optimal decisions. Based on this, we asked whether the behavioral shift produced by chronic stress was reflected by either a change in PL neuron dynamics, or only by a change in DMS dynamics at the choice-point in the task (the latter would imply that information is communicated identically by PL, independent of stress, but that perhaps synaptic weight between PL and DMS was altered). Two rats were implanted with headstages carrying 6 tetrodes implanted in each of left and right PL and DMS, as well as stimulating electrodes in left and right DMS for antidromic identification of PL projectors. Electrodes were lowered over a period of 14 days until high-quality single unit recordings were obtained during task performance. I also learned spike-clustering and antidromic activation techniques. Unfortunately data from this electrophysiological section is unavailable for this report.

How does chronic stress impact behavioral choice? This section looks at complicated decisions involving assessing relative weights of positive and negative valences, finding that chronic stress reliably decreases the weight of negative valences. Importantly, this shift can be reversed by optogenetic stimulation of the PL-DMS pathway, providing further evidence that this elements of this pathway coordinate complicated decision-making. Specifically, the evidence is consistent with a model where PL provides information about the relative importance of rewards and costs, and DMS integrates these with other factors to arrive at an optimal decision which is communicated to ACC to provide a premotor bias. Importantly, the optimal decision given a set of input parameters may be different between animals, and indeed may even vary within one animal based on fluctuating internal states.

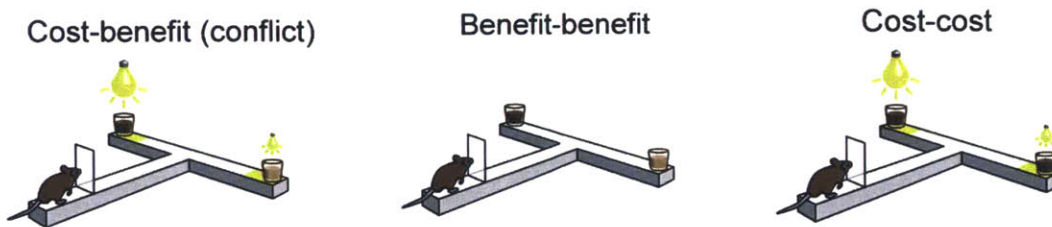


Figure 10: Rats are presented with different combinations of a reward (chocolate milk) and a cost (aversive light). The cost-benefit version of the task is the only version where different optimal decisions can be made based on internal weighting of costs and benefits.

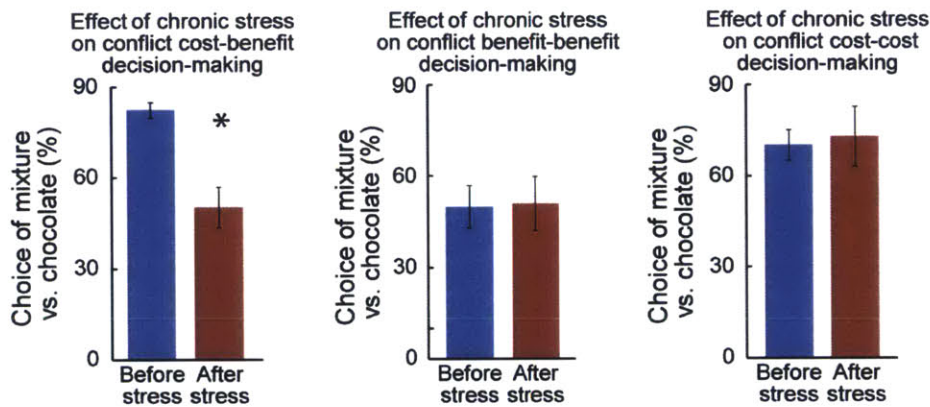


Figure 11: Chronic stress decreases the weight of the cost element in cost-benefit analysis, yet does not alter choice in other task versions. Mixture refers to low-concentration chocolate milk, while chocolate refers to high-concentration chocolate milk. n=6 rats.

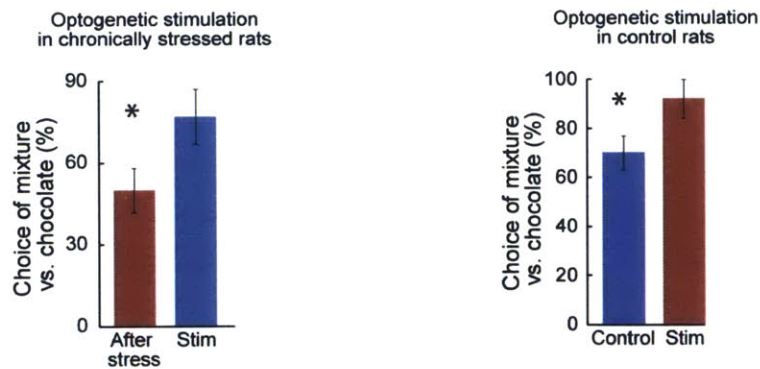


Figure 12: Optogenetic activation of PL terminals in striatum in chronically stressed rats is sufficient to revert decision-making back to a normal weighting of rewards and costs, and is similar to the behavioral shift evoked by optogenetic stimulation of this pathway in control animals. n=6.

A model for behavioral flexibility involving orbitofrontal cortex and striatum

Individuals can quickly learn mappings between environmental cues and the behavioral response which gives maximal reward. For instance, rodents can learn to associate certain odors with pushing a particular lever for sucrose reward; and Americans learn, when approaching a busy street, to always look left for oncoming traffic. However, what if the underlying associations between cues and rewards change? Americans, if taken to Britain, find it difficult to look right instead of left before crossing a busy street. But sometimes, remapping cue-reward associations is done relatively effortlessly, especially in the early phases of learning. This section provides a proposal for how cue-reward associations could be remapped at different phases of learning in an orbitofrontal-striatal network, and is specifically interested in solving the problem of how individuals overtrained on a task, or over-exposed to it, find it harder to adapt to changing contingencies than newly trained individuals.

This problem has been subject to intense investigation, and work has highlighted the important role of the orbitofrontal cortex (OFC) in promoting behavioral flexibility. Specifically, animals with OFC lesions cannot adapt their behavior as rapidly during reward devaluation or reversal learning (Bohn et al, 2003; Ghods-Sharifi et al, 2008; Izquierdo et al, 2004) but these same animals learn initial cue-reward associations unimpaired (Schoenbaum et al, 2003 (1)). This indicates that OFC is necessary to remap cue-reward associations, but is not necessary for initial cue-reward learning. Do OFC neurons encode behaviorally relevant parameters during behavioral tasks? OFC neurons indeed develop cue-evoked activity that codes for reward magnitude and is selective for particular cues, with both properties dependent on basolateral input to OFC (Schoenbaum et al, 2003 (2)). Other work suggests in fact that the OFC may represent animal's current state in an abstract map of the task (Wilson et al, 2014). Given that striatum is known to be important in both goal-directed and habitual behavior, could OFC inputs to striatum change decision-making during periods of altered reward values? Consistent with this general idea, recent work (Burgiere et al, 2013) showed that an OCD mouse model's compulsive behavior can be inhibited if OFC-striatal circuits are optogenetically activated. However, stimulating projections from the more medial portion of OFC to ventromedial striatum produced the opposite behavior: normal mice displayed OCD-like characteristics (Ahmari et al, 2013). These contradictory results highlight the need for more fundamental, mechanistic study of the roles of OFC and striatal circuits during normal behavior.

In line with this, recent experiments (Gremel & Costa, 2013) examined OFC and striatal neuron dynamics during goal-directed and habitual behavior, using a lever-press task with different reward schedules. After the reward was devalued, there was a clear correlation between OFC neurons' firing rates and the change in lever press rate, suggesting that OFC directly controls a behavioral change to reward devaluation. Thus, one model is that after a reward value change, OFC activity changes and this alone is sufficient to alter behavior. This would lead to the prediction that if OFC is optogenetically activated, behavior should *always* be altered identically - since no other factors such as downstream circuit dynamics play a role in the behavioral shift. However, the authors found that optogenetic OFC activation changed lever-press behavior, but *only* in goal-directed, early behavior, and only post-devaluation, not before. Based on this result, OFC is not sufficient to change behavior after devaluation, and other factors such as striatal dynamics or behavioral state must be additionally necessary. The authors also use a task design which prevents direct comparison between OFC dynamics in the goal-directed and habitual state, making it difficult to form a model to directly explain the findings.

Here, I propose a model where a cue arrives, and OFC contains two populations of projector neurons which respond to that cue. One population encodes predicted reward if action one is executed after the cue, and a separate second population encodes predicted reward if

action two is executed after that cue. Both predictions are encoded as firing rates. This reward prediction occurs in all behavioral states, including in overtrained animals, and tracks changes in reward contingencies. OFC projector neurons communicate these predictions to DMS, synapsing onto a local interneuron which places an inhibitory brake on those DMS neurons controlling the *opposite* behavioral choice through output to ACC. In other words, if OFC neurons predict a high reward for action one, then action two will be inhibited, and vice versa. This system has the advantage that if the reward prediction for action one changes to be very low, action two will be disinhibited and thus has a high chance of being behaviorally executed - the DMS inhibitory interneurons synapsing onto these action-two-controlling DMS projector neurons will be less active, since the interneurons receive less excitatory drive from the OFC neurons predicting the low reward for action one. The system can thus quickly respond to changes in reward contingency, in a manner highly consistent with what is known about striatal control of action selection through the classical direct and indirect outflow pathways.

How does this system respond later in training, when it is known that animals do not respond rapidly to reward contingency changes? Recall that early in training, DMS medium spiny neurons are highly active during the decision point and provide input to ACC, which provide a premotor bias to M1, activating or inhibiting the correct or incorrect responses, respectively; this is the basis for the rapid remapping response of different cue inputs as discussed above. However, recall that later in training, DMS is not active to provide premotor bias, and instead DLS starts indicating task-bracketing activity which is associated with habitual control of actions. Late in training, DMS projector neurons are not actively firing to provide input to ACC, and since they are silent, inhibiting them via a local interneuron would have no effect. Therefore, in late-trained animals, while OFC is still actively signaling reward predictions if various actions are taken, its input to DMS has no behavioral effect. This provides a plausible neural mechanism to explain why overtrained animals are generally far less sensitive to changes in reward contingencies, while newly trained animals respond to these same changes in a far more dynamic manner. The model presented here integrates various disparate and sometimes contradictory experimental observations to arrive at an integrated explanation for how an orbitofrontal-striatal network could plausibly control action selection for various environmental cues in a highly dynamic manner. In the next section, I will briefly discuss experiments which will test this model.

Experiment 1: Do the reward-prediction dynamics of OFC projection neurons change throughout training to reflect behavior?

The model relies on the unusual assumption that as reward contingencies change, OFC projector neurons quickly update reward predictions for a given cue-action combination - and that this rapid updating persists even later in training, where animals update their behaviors on a far slower time-scale. In other words, OFC always signals the correct reward prediction, but later in training, striatum is unable to organize an appropriate response, resulting in slow behavioral change. This contrasts with the possibly more parsimonious prediction that as the animals are overtrained and transition to a habitual mode of responding, then OFC projectors actually attenuate their reward-predictions after cues, possibly tracking the strength of the cue-reward association (which decreases later in training). Here, while striatum can always organize a rapid behavioral shift, it is OFC's signaling of correct predictions which attenuates over time and accounts for the resulting slowness of behavioral changes later in training.

Populations of OFC projector neurons have never, to my knowledge, been recorded over longer time-periods such as days or weeks. Therefore, a simple preliminary experiment to test the this model would be to train and overtrain rats on a simple T-maze forced-choice task over a period of weeks, while recording from antidromically identified DMS-projecting OFC neurons.

The model proposed above would predict that given a presented cue, one population of OFC projectors would signal a reward prediction if action 1 were taken after this cue (for instance, turning left in the maze) while another population of OFC projectors would signal a reward prediction if action 2 were taken (turning right in the maze). Additionally, the magnitudes of these reward predictions, averaged across the population, would remain relatively constant throughout training and overtraining. If the contingencies were to change (cue 1 should now be mapped onto action 2 rather than action 1), OFC projector neurons would signal this contingency remapping quickly, regardless of whether animals are early-trained or late-trained. Upon cue presentation, the OFC population representing action 2's reward prediction would rapidly increase firing rate, while the OFC population representing action 1's reward prediction should rapidly decrease firing rate. This remapping of population codes should proceed quickly independently of the stage of training. Behaviorally, in contrast, contingency remapping should result in rapid behavioral shifts towards the correct cue in newly trained animals, while resulting in slower behavioral shifts towards the correct cue in overtrained animals. This result would demonstrate that OFC cannot be sufficient to organize a behavioral change, since the timescales of OFC dynamics are always rapid while the behavioral timescales transition from rapid to slow-responding as animals are overtrained.

In contrast, the alternate hypothesis would make different predictions about the timescales of OFC dynamics and behavioral dynamics - they should be correlated. Given a contingency remapping early in learning, DMS-projecting OFC neurons should rapidly adjust reward predictions to reflect this. However, later in learning, the same contingency remapping should result in a much slower adjustment of reward predictions in DMS-projecting OFC neurons. The rate of reward prediction adjustment in all cases should be correlated with the rate of behavioral adjustment to the new, correct reward contingencies. This result would suggest that OFC dynamics alone are sufficient to explain behavioral remapping to new cue-reward associations.

The above experiment is a necessary first step to test whether the model could be plausible, but it only presents correlational rather than causal evidence. Experiment 2 provides a mechanism to causally test the major tenets of the model described above.

Experiment 2: Is optogenetic perturbation of DMS-projecting OFC neurons sufficient to alter behavior?

This experiment provides a causal test of whether changes in activity of DMS-projecting OFC neurons is sufficient to change behavior throughout training. The model outlined above predicts that it will *not* be sufficient in over-trained animals, while an alternate hypothesis predicts that it will be sufficient in both early-trained and overtrained animals. This will be tested by optogenetically activating OFC neurons which project to DMS at the cue point, as rats run a T-maze task. Optogenetic activation will be done in probe trials throughout the course of training, from start to over-training.

Let us first consider the model outlined above. Recall that in OFC, upon presentation of a cue, one population of projectors would signal a reward value prediction if action 1 is taken, while a second population would signal a reward value prediction if action 2 is taken. If action one's reward value prediction is higher than action two's associated prediction, then the downstream DMS medium spiny neurons corresponding to action two will be suppressed by activation of a local inhibitory interneuron synapsing onto only this action's associated DMS neurons. Let us now suppose that in the early stages of training, during the cue presentation the majority of DMS-projecting OFC neurons are optogenetically activated. By the proposed model, this optogenetic perturbation should increase feedforward inhibition on both sets of DMS medium spiny neurons, both corresponding to action 1 and to action 2 - both actions should be

inhibited rather than just one action being inhibited as before the perturbation. Behaviorally, the rat should no longer be biased towards the maze arm giving optimal reward, and would likely choose each arm equally.

Let us consider what the proposed model would predict if late-trained animals, rather than early-trained ones, were to have DMS-projecting OFC neurons optogenetically activated during the cue presentation. As in early-trained animals, OFC projection neurons representing action 1's reward prediction would increase firing, increasing feedforward inhibition onto DMS MSN neurons associated with releasing action 1; and OFC projection neurons representing action 2's reward prediction would also increase firing, increasing feedforward inhibition onto DMS MSN neurons associated with releasing action 2. However, in late-trained animals DMS activity has attenuated to a practically silent state, and instead the DLS portion of striatum seems to be controlling action. Therefore, the increased feedforward inhibition onto both action-1-related DMS neurons and onto action-2-related DMS neurons has no effect, since both of these groups of neurons are already silent. So the proposed model predicts that if OFC neurons projecting to DMS were optogenetically activated late in training, this perturbation should have practically no behavioral effect and rats should continue to favor the maze arm associated with most reward. This is in contrast to the model's prediction for early-training optogenetic perturbation, where the rats should now choose both arms equally rather than favoring the optimal maze arm. Additionally, electrophysiologically, we would predict that in early trained animals any DMS medium spiny neurons active during the cue point would be silenced during optogenetic activation of OFC projector terminals; and that DMS interneurons would be more active. In late-trained animals DMS medium spiny neurons should be silent during cue presentation, and optogenetic activation of OFC projectors should have no effect on this.

Consider the alternative: that OFC dynamics alone were somehow sufficient to organize a behavioral change. Here, we would predict that in both early-trained and over-trained animals, optogenetic activation of OFC projector neurons in DMS would result in rats favoring either arm equally.

In this section, I have outlined a possible model for how an orbitofrontal-striatal circuit can organize behavioral shifts when the reward contingencies of a task change. A major strength of this model is that it provides an explanation for the observation that late-trained animals are less receptive to reward-contingency remapping than early-trained animals, supposing that this difference is due to an alteration in striatal dynamics which occurs during learning. The first experiment provides correlational evidence that OFC reward-prediction signals are not sufficient alone to explain the change in receptivity to reward-contingency remapping which occurs during the course of learning. The second experiment provides more causal evidence that this is the case, using optogenetic manipulations.

References

- Ahmari, S.E., Spellman, T., Douglass, N.L., Kheirbek, M.A., Simpson, H.B., Deisseroth, K., Gordon, J.A. & Hen, R. (2013) Repeated cortico-striatal stimulation generates persistent OCD-like behavior. Science 340:1234-1239.
- Amador, A., Perl, Y.S., Mindlin, G.B. & Margoliash, D. 2013. Elemental gesture dynamics are encoded by song premotor cortical neurons. Nature 495:59-64.
- Bell, C., Bodznick, D., Montgomery, J. & Bastian, J. 1997. The generation and subtraction of sensory expectations within cerebellum-like structures. Brain, Behavior & Evolution 50(Suppl1): 17-31.
- Bohn, I., Gierler, C., Hauber, W. (2003) Orbital prefrontal cortex and guidance of instrumental behavior in rats under reversal conditions. Behav. Brain. Res. 143:49-56.
- Burgiere, E., Monteiro, P., Feng, G. & Graybiel, A.M. (2013) Optogenetic stimulation of lateral orbitofronto-striatal pathway suppresses compulsive behaviors. Science 340:1243-1246.
- Corbit, L.H. & Janak, P.H. (2007) Inactivation of the lateral but not medial dorsal striatum eliminates the excitatory impact of Pavlovian stimuli on instrumental responding. J. Neurosci. 27:13977-13981.
- Daw, N.D., Niv, Y. & Dayan, P. (2005) Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. Nat. Neurosci. 8:1704-711.
- Fee, M.S. & Goldberg, J. 2011. A hypothesis for basal-ganglia-dependent reinforcement learning in the songbird. Neuroscience 198:152-70.
- Fee, M.S., Kozhevnikov, A.A. & Hahnloser, R.H. 2004. Neural mechanisms of vocal sequence generation in the songbird. Ann N Y Acad Sci 1016:153-170.
- Ghods-Sharifi, S., Haluk, D.M. & Floresco, S.B. (2008) Differential effects of inactivation of the orbitofrontal cortex on strategy set-shifting and reversal learning. Neurobiol. Learn. Mem. 89:567-573.
- Gremel, C. & Costa, R.M. (2013) Orbitofrontal and striatal circuits dynamically encode the shift between goal-directed and habitual actions. Nat. Commun. 4:2264.
- Hoyer, P.O., Hyvarinen, A. & Arinen, A.H. 2002. Interpreting neural response variability as Monte Carlo sampling of the posterior. Advances in Neural Information Processing Systems, 15 NIPS.
- Izquierdo, A., Suda, R.K. & Murray, E.A. (2004) Bilateral orbital prefrontal cortex lesions in rhesus monkeys disrupt choices guided by both reward value and reward contingency. J. Neurosci. 24:7540-7548.
- Jazayeri, M. & Shadlen, M.N. 2010. Temporal context calibrates interval timing. Nature Neuroscience 13:1020-1026.
- Kubota, Y., Liu, J., Hu, D., DeCoteau, W.E., Eden, U.T., Smith, A.C. & Graybiel, A.M. (2009) Stable encoding of task structure coexists with flexible coding of task events in sensorimotor striatum. J. Neurophysiol. 102:2142-2160.
- Long, M. & Fee, M. 2008. Using temperature to analyze temporal dynamics in the songbird motor pathway. Nature 456:189-194.
- Long, M., Jin, D.Z. & Fee, M.S. 2010. Support for a synaptic chain model of neuronal sequence generation. Nature 468:394-399.
- Nottebohm, F., Stokes, T.M., & Leonard, C.M. 1976. Central control of song in the canary, *Serinus canarius*. J. Comp. Neurol. 165:457-86.
- Pan, W.X., Mao, T. & Dudman, J.T. 2010. Inputs to the dorsal striatum of the mouse reflect the parallel circuit architecture of the forebrain. Front. Neuroanat. 4:147.

Root, D.H., Tang, C.C., Ma, S., Pawlak, A.P. & West, M.O. (2010) Absence of cue-evoked firing in rat dorsolateral striatum neurons. Behav. Brain Res. 211:23-32.

Schoenbaum, G., Setlow, B., Nugent, S.L., Saddoris, M.P. & Gallagher, M. (2003) Lesions of orbitofrontal cortex and basolateral amygdala complex disrupt acquisition of odor-guided discriminations and reversals. Learn. Memory. 10:129-140.

Schoenbaum, G., Setlow, B., Saddoris, M.P. & Gallagher, M. (2003) Encoding predicted outcome and acquired value in orbitofrontal cortex during cue sampling depends upon input from the basolateral amygdala. Neuron 39:855-867.

Thorn, C.A., Atallah, H., Howe, M & Graybiel, A.M. (2010) Differential dynamics of activity changes in dorsolateral and dorsomedial striatal loops during learning. Neuron 66:781-795.

van Duuren, E., Escamez, F.A.N., Joosten R.N.J.M.A. et al. (2007) Neural coding of reward magnitude in the orbitofrontal cortex of the rat during a five-odor olfactory discrimination task. Learn. Memory. 14:446-456.

Wilson, R.C., Takahashi, Y.K., Schoenbaum, G. & Niv, Y. (2014) Orbitofrontal cortex as a cognitive map of task space. Neuron 81:267-279.

Yin, H. H., Knowlton, B. J. and Balleine, B. W. (2004) Lesions of dorsolateral striatum preserve outcome expectancy but disrupt habit formation in instrumental learning. Eur. J. Neurosci. 19:181-9.

Yin, H. H., Ostlund, S. B., Knowlton, B. J. and Balleine, B. W. (2005) The role of the dorsomedial striatum in instrumental conditioning. Eur. J. Neurosci. 22:513-23.