UNIVERSITY OF CALIFORNIA, MERCED

A Spiking Neuron Model of Classical Conditioning Phenomena

A dissertation submitted in partial satisfaction of the requirements

for the degree Doctor of Philosophy

in

Cognitive and Information Sciences

by

Jeffrey Joseph Rodny

Committee in charge:

Professor David C. Noelle, Chair
Professor Christopher T. Kello
Professor Jeffrey Yoshimi

2017

The Thesis of Jeffrey Joseph Rodny is approved, and it is acceptable

in quality and form for publication on microfilm and electronically:

Christopher T. Kello

Jeffrey Yoshimi

David C. Noelle, Chair

University of California, Merced

2017

Dedicated to my Shawnee,

Thank you Shawna Banducci for everything you do. This dissertation is as much your accomplishment as it is mine.

I also dedicate this work to my family and friends:

to my brother, Kyle Rodny, and my parents, Bill and Christine Rodny, I would not have accomplished anything without your bottomless guidance and support;

to my friends, old and new, I never thought such a support structure could make home feel so close;

and to my graduate advisor, David C. Noelle, nothing I say here will come close to being able to thank you for your endless understanding and support throughout my graduate career. Thank you.

Table of Contents

LIST OF FIGURES

ACKNOWLEDGEMENTS

I need to acknowledge and thank the incredible work of William Benjamin St. Clair, who generously helped me understand the complexities of the Izhikevich neuron code, and who let me build my models on top of his reimplementation of the Izhikevich neuron code. This dissertation would have been drastically different without your help. Thank you Ben.

I want to thank my friends and family, for helping me stay the course, believing in me, and supporting me. Without a support structure like that I'd never be able to do anything near to what I've accomplished here. I am so incredibly lucky. Thank you.

Also I want to give a big thank you to Shawna Banducci, for significantly-othering like nobody's business. Your support of me through the past 5 years has created a foundation I've unknowingly realized I needed. You've supported me throughout this in so many ways and you've made sure I've never had to worry about anything but graduate school. I've been so lucky to have your love and support. Thank you.

I also want to thank my committee, Chris Kello, Jeff Yoshimi, and David Noelle. Your support and guidance throughout my graduate career helped me to have confidence in myself and to think more critically about the important balance between what is important to model and what can be abstracted away.

Thank you to all of the current and past members of the Cognitive and Information Sciences graduate group at UC Merced for being professional yet warm and friendly throughout my time at UC Merced. Thank you for laughing at the jokes I put in my presentation slides, as well as for asking the questions that needed to be asked when my research needed another perspective.

I also want to thank all of my previous mentors throughout my academic career, starting in high school. Thank you, Gerry Navarra, for cultivating in me a desire to understand how the brain works. Without your AP Psychology class, I would not have had the curiosity for Psychology that pushed me towards Cognitive Science. I would have focused solely on computer programming, and never discovered anything interesting about how the brain works. It all started with you.

Thank you to Whit Tabor, for your guidance with my research as an undergraduate at UConn; your guidance and patience with me while I learned how to first program neural networks in MATLAB was indelible. Thank you for allowing me to research what I was interested in. I now understand how rare that can be for undergraduates.

And finally thank you to my advisor, David Noelle. You're too kind, Dave, you really are. Every year at the Cognitive Science Society meetings, when they describe the qualifications for receiving the David Rumelhart prize, and they talk about David Rumelhart's academic accomplishments and his unending patience and kindness for his students, I always think of you. Every year I secretly hoped you'd get it, even though I never actually nominated you. If I regret anything, it's never nominating you. You work too hard and give too much of your time for your students. Take a break! You deserve it.

CURRICULUM VITAE

Jeffrey J. Rodny
jrodny@ucmerced.edu
https://www.linkedin.com/in/jeff-rodny/

Education

2017, Ph.D.	University of California Merced, Cognitive and Information Sciences Graduate Research Advisor: David C. Noelle
2011, B.S.E.	University of Connecticut, Computer Science and Engineering (Honors) Minor: Math
2011, B.S.	University of Connecticut, Cognitive Science (Honors) Minor: Psychology Undergraduate Research Advisor: Whitney Tabor

Peer-Reviewed Journal Publications

2016	**Rodny, J. J.**, Shea, T. M., & Kello, C. T. (2016). Transient localist representations in critical branching networks. Language, Cognition and Neuroscience, 1-12.
2014	**Rodny, J. J.**, & Noelle, D. C. (2014). Modeling the actor-critic architecture by combining recent work in reservoir computing and temporal difference learning in complex environments. Proceedings of the 13th Neural Computation and Psychology Workshop. *Progress in Neural Processing*, 21, 237-248
2012	Kello, C. T., **Rodny, J.**, Warlaumont, A. S., & Noelle, D. C. (2012). Plasticity, learning, and complexity in spiking networks. *Critical Reviews in Biomedical Engineering*, 40(6).

Oral and Poster Presentations

2016	**Rodny, J. J.**, Noelle. D. C. (2016). The computational role of dopamine, basal ganglia, and hippocampus in extinction and spontaneous recovery. Program No. 850.12. Neuroscience 2016 Abstracts. San Diego, CA: Society for Neuroscience, 2016. Online.
2015	**Rodny, J. J.**, Noelle. D. C. (2015). Modeling the Role of Hippocampus in Extinction and Spontaneous Recovery. Member Abstract. Poster presented at the 35th Annual Conference of the Cognitive Science Society. Pasadena, CA: Cognitive Science Society.
2014	**Rodny, J.**, Kello, C. T. (2014). Learning and Variability in Spiking Neural Networks. Oral presentation at the Proceedings of the 36th Annual Conference of the Cognitive Science Society. Quebec City, Quebec, Canada: Cognitive Science Society.
2013	**Rodny, J. J.**, Noelle. D. C. (2013). Approximating the Value Function in the Actor-Critic Architecture using the Temporal Dynamics of Spiking Neural Networks. Member Abstract. Poster presented at the 35th Annual Conference of the Cognitive Science Society. Berlin, Germany: Cognitive Science Society.

Undergraduate Research

2010 Tabor, W., Cho, P., Kukona, A., Coppola, J., Halle, J., Jennett, P., Lucas, J., Madruga, M., McCloskey, B., Noccioli, E., **Rodny, J.**, Szkudlarek, E., and Wantroba, R., (2010 April) Self-Organization in Language and Society. Poster presented at the UConn Language Fest Spring 2010, Storrs, CT.

2009 Tek, S., Jaffery, G., Piotroski, J., **Rodny, J.**, Fein, D., & Naigles, L. (2009, May) The Shape Bias: Investigations of Word Learning with Children with Autism. Poster presented at the International Meeting for Autism Research, Chicago, IL.

Teaching Experience

2017, Spring Teaching Assistant, COGS 123: *Computational Cognitive Neuroscience*, Instructor: David C. Noelle

2016, Fall Teaching Assistant, COGS 125: *Machine Learning*, Instructor: David C. Noelle

2016, Spring Teaching Assistant, PHIL 002: *Philosophy of Ethics*, Instructor: David Jennings

2015, Fall Teaching Assistant, COGS 140: *Perception and Action*, Instructor: Ramesh Balasubramaniam

2015, Spring Teaching Assistant, COGS 130: *Cognitive Neuroscience*, Instructor: Anne Warlaumont

2014, Fall Teaching Assistant, CSE 175: *Introduction to Artificial Intelligence*, Instructor: David C. Noelle

2014, Spring Teaching Assistant, COGS 001: *Introduction to Cognitive Science*, Instructor: Ben Pageler

2013, Fall Teaching Assistant, COGS 175: *Spatial Cognition*, Instructor: Marcus Perlman

2013, Spring Teaching Assistant, COGS 001: *Introduction to Cognitive Science*, Instructor: Ben Pageler

2012, Fall Teaching Assistant, COGS 001: *Introduction to Cognitive Science*, Instructor: Michael Spivey

2012, Spring Teaching Assistant, COGS 001: *Introduction to Cognitive Science*, Instructor: Ben Pageler

2011, Fall Teaching Assistant, COGS 128: *Cognitive Engineering*, Instructor: Jeanne Milostan

ABSTRACT

This dissertation focuses on the biological structures that allow animals to exhibit classical conditioning. The project presents a computational neuroscience model combining insights from two influential accounts of learning. The first, Chorley & Seth (2011), offers a biologically realistic model of dopamine activity and association learning. The second, Redish et al. (2007), is an abstract model of internal state representation, perhaps residing in the hippocampus, that accounts for many classical conditioning phenomena. Combining these two models produces a biologically realistic explanation of both association learning and unlearning. The proposed model exhibits classical conditioning phenomena while demonstrating how spiking neurons in the brain could implement reward prediction. Specifically, this dissertation project makes the following original contributions: 1) a replication of the model of Chorley and Seth (2011), 2) the presentation of a new model, based on Chorley and Seth (2011) but incorporating a simple spiking neuron model of the hippocampus inspired by Redish et al, (2007), 3) a demonstration that the hippocampal model continues to exhibit association learning, 4) and now exhibits extinction, 5) reinstatement, 6) spontaneous recovery, 7) and blocking, 8) as well as a demonstration that the hippocampal model mirrors the results of relevant lesion studies, 9) and, finally, an analysis of how the hippocampus model represents context. While previous models of learning and unlearning based on temporal difference methods are powerful and account for some classical conditioning phenomena, this dissertation suggests that, contrary to temporal difference methods, learning and unlearning associations are not the result of a single mechanism.

CHAPTER 1: INTRODUCTION

This dissertation focuses on processes for obtaining reward and the faculties required to support reinforcement learning. It pertains to faculties all mammals have, irrespective to the type of input the environment may give (e.g. auditory, visual, tactile), that help to associate unemotional environmental input with positive or negative valence, resulting in taking actions to seek or avoid that input. Environmental input here refers to anything the animal can sense, from a ringing bell, or a ripe berry, to a dollar bill, a bowl of food, or even abstract things like safety, trust, friendship, and frustration. An animal can pass an emotional judgment on all of these things, and from there it can take actions to seek or avoid these environmental inputs.

This kind of learning has been understood by modeling *associations*; abstract numeric strengths of how much an animal likes or dislikes an environmental input (Rescorla & Wagner, 1972). These mathematical learning models made great progress in explaining a variety of learning phenomena, like blocking and extinction (explained in more detail further on in this proposal). These models were limited, however, to only reproducing classical conditioning, which is the passive learning of associations in the environment. Twenty years later, a group of scientists studying macaque monkey brains found a surprising neural firing pattern in the basal ganglia that seemed to involve learning to predict rewarding stimuli (Ljungberg, Apicella, & Schultz, 1992; Romo & Schultz, 1990; Schultz & Romo, 1990). Around the same time, a group of computer scientists put forth a formal model of reinforcement learning that built upon the passive learning in the Rescorla and Wagner (1972) model. This computational model allowed the learning agent to interact with its environment, and, critically, it involved the learning of predictions of future reward (Barto, 1995; Barto, Sutton, & Anderson, 1983; Sutton, 1988). A seminal insight arose when the activity of certain cells in the macaque limbic system - cells that deliver the neurotransmitter dopamine – were shown to encode for changes in expected future reward – a key variable in the abstract computational models of reinforcement learning (Montague, Dayan, & Sejnowski, 1996). This gave rise to an account of learning in which associations were updated based on the difference between a prediction of reward when a stimulus was presented and the actual presence or absence of reward moments later (Sutton, 1988). This signal of the difference over time in expected and actual reward was called the *temporal difference (TD) error*.

While this work made huge progress in marrying neuroscience with computational models of action selection, the approach could not explain all aspects of reinforcement learning, and it did have a few drawbacks. For example, in environments with clear goals (e.g. find a goal location, produce a specific action sequence, etc.) it was proven that this model will lead to an efficient solution. However, large problems (e.g., involving many inputs or allowing for many different actions) can require an intractable amount of time to solve through this form of reinforcement learning. Real-world problems that involve noisy environmental input, including cases in which sensory states are never exactly revisited, require value function approximators – learned functions from inputs to expected reward. Unfortunately, the use of such value function approximators violates the assumptions of the proofs of learning convergence, making those proofs irrelevant for large real-world problems. There are some ways to circumvent the problems of tractability, but using those methods also removes any guarantee of learning a correct solution (Boyan & Moore, 1995; Sutton, 1996). The model also did not accurately characterize the behaviors of animals, such as the phenomena of spontaneous recovery and gradual extinction (which will later be explained in detail). There were also some critics of the TD error approach who claimed

that it did not accurately model the processes that occurred in the mammalian brain. These critics had been frustrated by the lack of a reinforcement learning account implemented using models of spiking neurons similar to those found in the brain. While some work has been done to implement the temporal difference error approach in biologically plausible neuron models, it hasn't been able to rely solely on spiking neurons to implement the learning process. For example, most models required the abstract subtraction of two neural activation values, separated in time, representing rewards and predictions of rewards (Barto, 1995; Castro, Volkinshtein, & Meir, 2009; Florian, 2005, 2007; Potjans, Morrison, & Diesmann, 2009; Rao & Sejnowski, 2001; Roberts, Santiago, & Lafferriere, 2008; Rusu & Florian, 2009). However, none of these approaches relied solely on the function of networks of spiking neurons. They all required at least some abstract non-neural mathematics to be done in support of the modeled neural processes.

In 2011, Chorley and Seth published a paper demonstrating how a network composed entirely of spiking neurons could model the dopamine signal that was found by Schultz and colleagues (Ljungberg et al., 1992; Romo & Schultz, 1990; Schultz & Romo, 1990). This was the first model of the temporal difference error in the brain that used only biologically realistic spiking neurons, making use of a specific pattern of connectivity. Instead of having an abstract variable for the dopamine signal, as others were forced to do, this model calculated it through the spikes arising from environmental input traveling through the model. This model was able to hold a predicted expected future reward value in the form of spiking activity and pass that value to the rest of the model.

As a preliminary modeling effort and a part of this dissertation work, the model described in Chorley and Seth (2011) was replicated and analyzed (described in more detail later in this dissertation). Analyzing the results of the replicated model revealed that it lacked an important capability; the Chorley and Seth model did not have the ability to extinguish learned associations. While the model could learn associations between neutral stimuli and reward, it did not "unlearn" these associations when reward was no longer delivered. This deficit had further ramifications. Since the model did not extinguish associations, it could not show spontaneous recovery, gradual extinction, and other learning phenomena related to extinction.

This observation prompted a review of the neuroscientific literature on extinction in mammals. Through this review, it was discovered that the hippocampus is crucial to the extinction process. Importantly, the model in Chorley and Seth (2011) did not model the hippocampus. Similarly, rats with bilateral lesions of the hippocampus have trouble extinguishing learned associations. When placed in a T maze, these rats learned that food is down one arm of the "T", but when the food was instead placed in the other arm, they could not extinguish the previously learned association. They continued to behave as if they thought the food was down the original arm (Kimble & Kimble, 1970). In 1986, Weikart and Berger performed similar hippocampal lesions on rabbits and trained them on an eye-blink conditioning task with two stimuli, a bell and a light. One stimulus predicted an air puff to the eye, and the other did not. Air puffs were aversive, and they caused the rabbits to blink. The rabbits with hippocampus lesions could learn as well as controls that one of the two stimuli strongly predicted an air puff to the eye, and that the other could be ignored. Learning was seen in a tendency to blink when presented with the predictive stimulus but not when presented with the other. Halfway through the experiment, however, the stimuli were switched, and the learned stimulus no longer predicted the air puff. Instead, the ignored stimulus began to strongly predict the air puff.

Control animals quickly learned to ignore the newly non-predictive stimuli, responding instead to the newly predictive stimulus as a predictor of the air puff. In contrast, the hippocampally lesioned animals continued to produce a blink response to the originally predictive stimulus while also successfully acquiring an association with the newly predictive stimulus. Thus, the lesioned animals came to respond to both stimuli (Weikart & Berger, 1986). This, and other similar studies, suggested that for an animal to extinguish a previously learned association it needed a hippocampus. This implied that, for the Chorley and Seth (2011) model to properly extinguish learned associations, it also needed a hippocampus.

Redish *et al.* (2007) proposed a novel way to think about how learning and extinction occur. While the Redish model was largely agnostic concerning the neural basis of its proposed mechanisms, there was one key aspect of the model that reflected interesting properties of the hippocampus. We will revisit this model later in this document. For now, it is important to note that a central claim of the Redish model was that, in order for a learning system to show both extinction and spontaneous recovery, it needed to be able to create new internal representations of the current state of the environment. This theory suggested that the learning of extinction depends critically on an animal's *internal representation of the environmental state*, and how the context cues and environmental state change throughout the extinction process. According to this account, the internal state represented by the animal *during the learning of an association* does not change much throughout association learning trials, but the internal state represented by the animal *during the extinction of an association* changes substantially over time. This change in context representation is based on the expanding amount of time since the previously associated unconditioned stimulus (US - reflecting an external reward signal) last occurred. This theory posited that entrance into the experimental environment invokes a specific state representation in the rat's brain, and, with each presentation of an association learning trial, the state stays largely the same. In contrast, on each trial during which the US is expected but remains absent, the context changes slightly, so that the learning occurring throughout extinction trials involves associations within slightly different internally represented contexts. The Redish model thus correctly modeled extinction, as well as the related phenomena of spontaneous recovery and renewal (Redish, Jensen, Johnson, & Kurth-Nelson, 2007).

The research project described in this dissertation is about adding a spiking model of the hippocampus to the Chorley & Seth model, producing a spiking neural network model that exhibits extinction and other related phenomena. Specifically, this dissertation project offers the following original contributions:

- a replication of the model of Chorley & Seth
- the presentation of a new model, based on Chorley & Seth but incorporating a simple spiking neuron model of the hippocampus
- a demonstration that the hippocampal model continues to exhibit association learning
- a demonstration that the hippocampal model exhibits extinction
- a demonstration that the hippocampal model exhibits reinstatement
- a demonstration that the hippocampal model exhibits spontaneous recovery
- a demonstration that the hippocampal model exhibits blocking
- a demonstration that the hippocampal model mirrors the results of lesion studies when the hippocampus component of the model is removed
- an analysis of how the hippocampus model represents context

The newly proposed model offers a viable alternative to abstract TD accounts – an alternative that is more aligned with established biological findings.

CHAPTER 2: BACKGROUND

Learning

This section reviews relevant current research into the nature of reward learning. There are many learning processes in the mammalian brain, from the many levels of visual learning in the occipital, temporal, and parietal lobes to learning associated with action selection in the basal ganglia - a kind of learning that is directly related to reward. Here I will cover those brain areas most closely associated with reward learning; brain structures known to have a relation to reinforcement learning and its immediately supporting structures. The focus is on mechanisms for behavioral change so as to obtain reward and the learning facilities required to support this reinforcement learning. This includes learning to select actions so as to achieve reward, sensitive to association learning and motor control behavior. Not all learning processes depend directly on the delivery of an external reward, and reinforcement learning depends on many learning mechanisms in the brain that lack such a reward-based orientation. Let us consider the fundamental nature of learning.

The American Psychological Association defines learning as “a process based on experience that results in a relatively permanent change in behavior or behavioral potential” (Gerrig & Zimbardo, 2010). In a very abstract sense, learning is the process of gathering information about one’s environment and using it to modify how one will interact with the environment in the future. In this abstract sense, there are three categories of timescales of learning, in which any living thing (from a virus to a human) modifies the relationship between it and its environment.

The largest timescale is over many lifespans along the line of heredity of an organism. Processes at this time scale are very slow in how they modify the relationship between the living thing and its environment; these are the processes of biological evolution. The organism, viewed as a species, modifies its relationship with its environment by changing how its descendants interact with the environment. This timescale is on the order of tens of thousands of years for humans, and hours for viruses, and the resulting modifications are stored in the organism’s DNA.

The next timescale is over a single lifespan of a living thing, and this is what we generally refer to as learning. At this timescale, the organism modifies its relationship to its environment over its lifespan. This is not done through modifying its descendants, but through modifying its behavior, or the organizations of its movements. Here, it modifies this relationship on the order of minutes, days, months, or years for humans, and the modifications are stored in the organism’s body.

The shortest timescale is over fractions of a lifespan of a living thing. This is the timescale of action and movement. Here, the organism modifies its relationship to its environment very quickly, by changing its orientation or spatial location, modifying how it will interact with its environment in the next few seconds for humans or milliseconds for viruses, and this modification is stored in the organism’s immediate spatial location and orientation.

Notice that I did not provide an example for viruses at the intermediate timescale of learning. This is because (to the author’s knowledge) it has not yet been shown that viruses change their patterns of behavior based on information from their environment at a higher level than just movement. It may be that not all living things implement all of these organism-environmental modifiers, and there may be others not mentioned here (such as culture or religion, which would occur on a timescale between evolution and learning), however it is interesting to note these relationships, nonetheless.

The purpose of these modifications of the relationship between an organism and its environment is a long-debated issue, but here we simply argue that they are made for the organism to exist longer and reproduce more. While we will not fully cover that argument here, the basic explanation is that, whether or not there is a purpose for modifying an organism-environment relationship, it just so happens that those organisms that exist longer and reproduce more exist in larger numbers. In contrast, those organisms that don't exist very long and do not reproduce very much do not exist in very large numbers, and they eventually go extinct. Therefore, while the purpose of the ability to change an organism's relationship with its environment is unclear, those that do so in ways that keep them and their species existing in larger numbers are the ones that mostly make up things that exist, and thus mostly make up the things that we can study and learn from.

Let us now focus on reward learning. This is the modification of an organism's pattern of behavior so as to increase the likelihood and magnitude of external reward. Typically these modifications support efforts by the organism to exist longer and to reproduce more. I will review the biological evidence of the importance of different brain areas to reward learning, summarize some current research and models of reward learning, and explain what we can learn from them.

Classical Conditioning Phenomena

There are some important terms to know that are necessary to understanding classical conditioning phenomena. An *unconditioned stimulus* (US) is a stimulus that automatically elicits a response from the subject, such as a reward (e.g., food). A *conditioned stimulus* (CS) is a stimulus that is neutral initially to the subject, such as a bell or a light. While a CS is initially neutral, over many association trials with a US where the CS precedes the US, the CS can come to elicit the same response that the US elicits.

In many learning experiments, associations between stimuli and reward are not only created but are also perturbed to better understand the learning process. One such perturbation process involves the unlearning of a previously learned association. After many association trials of a stimulus and a reward, the animal expects the reward to occur after the stimulus is presented. When the animal is then repeatedly presented with the stimulus without reward, the association between the stimulus and the reward is slowly unlearned. This is called *extinction*. Another interesting phenomenon occurs when an animal learns an association, has that association extinguished, and then is removed from the experimental setting, only to be returned after some time has passed. After returning to the original setting, and upon presentation of the stimulus, the animal responds as if the association has not been completely extinguished. This is called *spontaneous recovery* (Pavlov, 1928). In other experiments, multiple stimuli are presented to the animal and the specific stimuli that predict reward are changed after an initial association is acquired. Here, a stimulus that previously predicted reward stops doing so, and a previous stimulus that did not predict reward starts to predict reward. This learning perturbation process is called *reversal*. During reversal, the loss of the association between the previously predictive stimulus and reward can be seen as a kind of extinction.

We can learn more about how learning works by modifying the way stimuli are presented to the subject. If, after association and extinction trials, the subject is presented with just the reward, and no other stimuli, the subject usually shows a *reinstatement* of the previously extinguished association. It's important to note that this

phenomenon is similar to spontaneous recovery in that, for both, the subject learns, extinguishes, and then shows the association again. However, the difference is that in spontaneous recovery, the subject is presented with a large amount of time outside of the experimental environment and then returned to the original context it learned in, at which time the subject is again presented with the conditioned stimulus. In reinstatement, immediately after extinction occurs, and in the same context, the subject is presented with the rewarding stimulus, and this results in a brief reappearance of the association with the conditioned stimulus. Another similar phenomenon is *renewal*, where, after learning and extinction occur, if the subject is placed in a new context and presented with the conditioned stimulus, the subject sometimes shows renewed responding, as if the extinction did not occur.

There are other classical conditioning phenomena involving the presentation of multiple conditioned stimuli. For example, in *blocking*, if one stimulus is already associated with the US and another is not, the presentation of those two stimuli at the same time, followed by the US, blocks the learning of an association between the new stimulus and the US. Here, because the first conditioned stimulus already provides the animal with all of the information needed to predict the upcoming US, there is no need to form an association involving the second conditioned stimulus, as such a new association would provide no additional predictive power. We know that blocking occurs because, afterward all of the learning trials, when the subject is just shown the second CS, the subject treats the stimulus as neutral, ignoring it (Kamin, 1969). Another phenomenon involving multiple stimuli is *overshadowing*. This is where one stimulus is more salient, or more powerful. In this case, when both stimuli are presented before the US, an association is formed only between the US and the more powerful, more salient stimulus. You can also get *second order conditioning*, where one stimulus is paired with a reward, and then another stimulus is paired with the first stimulus. The subject treats the initially learned conditioned stimulus as rewarding, allowing learning even when the subsequent stimulus never precedes the US during training (Holland & Rescorla, 1975).

Each of these learning phenomena is important in order to understand what the brain does when it learns and unlearns associations. Damage can be done to specific parts of the brain to understand which classical conditioning phenomena are affected. Knowing which classical conditioning phenomena are affected by damaging which brain areas can help us better understand how the brain works.

Neuroscience

The Basal Ganglia

The basal ganglia are brain structures located below the folds of the cerebral cortex in the human brain. Homologues of these structures can be found in most vertebrate animals and even some arthropods (Strausfeld & Hirth, 2013). The basal ganglia take on multiple functional roles, including action selection, habit formation, memory, and learning associations (Packard & Knowlton, 2002).

The basal ganglia, within the limbic system, have been found to be very important in the formation of associations between related stimuli. The substantia nigra pars compacta (SNc), specifically, is adapted to using stimuli in the environment to predict changes in expected reward. Here, reward is a term used for a class of stimuli that, through evolutionary pressures, has shaped the brain to treat them as intrinsically desirable. Over an evolutionary timescale, this class of stimuli has proven to extend the existence of an organism and increase its ability to reproduce. Thus, organisms have evolved to develop behaviors that lead to the acquisition of such reward. Through

evolution, these organisms have evolved to perceive the reception of reward when they, for example, eat food with nutrients used to maintain the organism's body. The basal ganglia learn which stimuli in the environment reliably predict reward. Neurons in the SNc actually fire for the arrival of unexpected reward in the same way that they fire for stimuli that predict reward (Romo & Schultz, 1990; Schultz & Romo, 1990).

Dopamine is an important neurotransmitter released primarily by neurons found in two nuclei in the midbrain. The substantia nigra pars compacta (SNc) projects dopaminergic connections to the striatum. Also, just ventrally to the basal ganglia, the ventral tegmental area (VTA) projects dopaminergic connections to the frontal lobe, nucleus accumbens, amygdala, and hippocampus.

Corbett and Wise (1980) sunk electrodes into the midbrain of rats looking for an area that encodes for reward. When the rats pressed a lever, the electrode would stimulate a specific area of the brain. They found that rats would self-stimulate in this way most energetically when the electrode was touching the dopamine fiber bundles that originate in the VTA and enervated the frontal lobes (Corbett & Wise, 1980). Bozarth and Wise (1981) injected morphine into the VTA of rats, where this dopaminergic fiber bundle originates, and found that rats will self-administer morphine in the same way (Bozarth & Wise, 1981). Both of these studies suggested that this dopaminergic fiber bundle, and dopamine in general, are involved in the representation of reward. While this early research suggested that dopamine explicitly encodes reward, more recent research has offered an alternative interpretation.

In a collection of studies (Ljungberg et al., 1992; Romo & Schultz, 1990; Schultz & Romo, 1990), dopaminergic neurons in the substantia nigra pars compacta (SNc) were recorded via electrodes sunk into the brains of macaque monkeys. The electrodes recorded a low level of activity, consisting of a slow base rate of electrical action potentials (or spikes), but this activity changed as the monkeys were presented with stimuli. The monkeys were kept in a semi-hungry state and presented with a box that the monkeys could reach inside in order to collect a rewarding food morsel. The researchers found three interesting results. The first result was that when the monkeys received the food morsel, the electrodes sensed an increased firing rate of the dopamine (DA) neurons in the SNc. The DA neurons' rapid firing did not correlate with touching anything else inside of the box, and it did not correlate with muscular contractions used to make similar reaching movements. The second result was that, over the course of many trials, the DA neurons in the SNc reduced their rate of firing for the onset of food and began to fire rapidly to the opening of the latch over the box. The latch consistently opened immediately before the monkeys could collect the food morsel. The SNc neurons began to fire less for the reward, and more for the stimuli that predicted reward. Indeed, the rate of firing upon the onset of food eventually returned to the base firing rate of the DA cells. The third result that the researchers found was that, after many trials, if the latch over the box opened and there wasn't food inside, the neurons in the SNc fired at a rate even lower than their base rate of firing upon a reach into the box. This research suggests that the SNc is important for indicating the unexpected presence of reward and for learning stimuli that predict future reward (Ljungberg et al., 1992).

In 1996, Montague, Dayan, and Sejnowski created a model of how an animal might use this dopamine signal to learn to select actions that lead to reward. They proposed that the dopamine activity in the SNc indicates prediction error about reward; dopamine activity above baseline indicates that the actual reward is greater than the amount of reward predicted. Conversely, dopamine activity below baseline indicates the actual reward is less than that predicted. The model can use the fluctuations in

dopamine levels to improve reward predictions so as to more accurately match the environment. While this publication was mainly a computational modeling paper (which will be described in more detail later), it is important to note that a dopamine signal like that seen in the SNc of the macaque monkey can be used to help it to learn to act so as to maximize its future reward (Montague et al., 1996).

While the substantia nigra pars compacta (SNc) in the basal ganglia is important for learning, so is the striatum. The striatum is a major recipient of incoming synapses from dopaminergic neurons in the SNc. The importance of the striatum for action selection is demonstrated by Kravitz *et al.* (2010). Kravitz and colleagues optogenetically stimulated the SNc and striatum in rats and observed their behaviors. Optogenetics is the technique of adding light-receptive proteins to specific neurons. These proteins can depolarize or hyperpolarize a neuron depending on the color of light shone on them. In this paper, Kravitz and colleagues targeted specific neurons in the striatum: neurons with either D1 or D2 receptors, two different kinds of neural receptors that are sensitive to dopamine (DA). D1 receptors depolarize the cell, driving it toward firing, while D2 receptors hyperpolarize the cell, inhibiting it and reducing the likelihood that it will fire. Within the striatum, there are neurons for which D1 receptors dominate on their dendrites and other, distinct, neurons for which D2 receptors dominate on their dendrites. Thus, D1 neurons fire more frequently when DA is present and D2 neurons fire more frequently when DA is absent. In the Kravitz paper, these researchers were able to excite D1 and D2 neurons in the striatum separately, and observe the results. What they found was that activating the D2 neurons in the striatum led to an increase in substantia nigra pars reticulata (SNr) activity, while activation of D1 neurons led to inhibition of the SNr. The SNr is a portion of the substantia nigra that is separate from the SNc; the SNr indirectly inhibits the thalamus. Thus, D1 neurons in striatum inhibit an inhibitor, effectively exciting the thalamus, while D2 neurons in striatum excite the SNr, which inhibits the thalamus. Furthermore, in an open environment, they found that rats that had their D1 neurons activated showed increased movement and exploratory activity, while similar rats with their D2 neurons activated exhibited increased freezing behavior, refusing to move or explore (Kravitz et al., 2010).

This research is consistent with some computational models of the striatum. In these models, separate neurons driven by D1 and D2 receptors in the striatum are characterized as being ‘Go’ and ‘No Go’ neurons, respectively. Activation of the ‘Go’ neurons corresponds to choosing an action, while activation of ‘No Go’ neurons corresponds to suppression of an action. In these models, cortical connections to the striatum determine the action that is selected. Thus, the pattern of connections from cortex to striatum can determine which stimuli prompt which actions. When reward is predicted with DA firing, “Go” cells activated by cortex are strengthened, and the connections between the active cells become stronger, making it more likely that the current stimulus will trigger the currently considered action in the future. With low DA, the stimulus is associated with the “No Go” cells for the considered action. This decreases the likelihood of taking the considered action, and it results in action selection learning in the basal ganglia (Frank, 2005; O’Reilly & Frank, 2006). Models of this kind account for association learning, but there are other learning effects that are not clearly covered by this account, such as some of the reversal learning effects seen in patients with Parkinson’s disease.

In patients with Parkinson’s disease, the dopaminergic neurons in the SNc die, and initiating actions becomes hard to do. Patients with Parkinson’s disease have great trouble walking, and they usually experience rigidity and hand tremors. It is not

immediately clear how these motor symptoms relate to the fact that the dopaminergic neurons in the SNc spike for rewarding stimuli and stimuli predictive of rewarding stimuli. Do Parkinson's disease patients have trouble predicting rewarding stimuli?

Shohamy, Myers, Hopkins, Sage, and Gluck (2009) showed that patients with Parkinson's disease can use preceding stimuli to predict rewarding stimuli; however, they do so differently than control patients. In Shahomy *et al.*, patients with Alzheimer's disease (a disease where neurons die throughout the brain, but focally in the hippocampus), along with patients with lesions to the hippocampus, were compared with patients with Parkinson's disease, as well as age-matched control participants. The task was a slot-machine-like game, where three different images appeared on a computer screen and coins were guaranteed to come out for every trial. The participants were asked to guess which of two colors the coins coming out of the machine would be, given the images on the computer screen. The color of the coins was determined by the pattern of images shown. The relationship between the displayed pattern of images and the color of the coins was deterministic, so memorizing all combinations of patterns would allow the participant to achieve 100% performance on predicting the color of the coins. However, simply using one of the three images to guess the color of the coins provided 80% accurate performance. Results showed that most participants opted to use this single-image prediction strategy. The relationship of the pattern of images to the color of the coins was set up such that the errors a participant made could be used to determine which particular image the participant was using to guess the color of the coins. Recall that each of the three images individually was correlated with the correct answer, such that a guess based on only one of the images could produce 80% accuracy, leaving 20% error trials. The researchers could use these error trials to determine which image each participant was using to make their predictions.

Halfway through the experiment, the color of the coins associated with each image was reversed. The participants were not told what was happening, and the images they were previously using to predict one color, now predicted the opposite color 80% of the time. The researchers found several different results. The participants with Alzheimer's disease or hippocampal damage perseverated on the image for which they originally learned the association, and they never learned the new coin color association. The controls learned that the color had switched after a few of trials, and they used the same image they were using before to guess the opposite color coin. The patients with Parkinson's disease also learned the new coin color association as easily as controls. However, the puzzling result is that the patients with Parkinson's disease found it easier to learn a new association of a different image to the color, rather than simply reversing the learned association between the old image and the previous color. In other words, these patients chose to learn a new association when given the option to reverse what they learned based on one image or learn a new association involving a different image (Shohamy, Myers, Hopkins, Sage, & Gluck, 2009). Theoretical and computational models have yet to thoroughly explain this phenomenon.

This finding in conjunction with the data from Ljungberg, *et al*. (1992) suggests that the SNc is important for creating and modifying associations. Neural activity in this nucleus is clearly implicated in the ability to learn an association, as well as the ability to unlearn an association that is no longer valid. Consistent with this, Parkinson's patients have trouble unlearning an association.

While these results concerning the substantia nigra pars compacta are intriguing, there are further results that speak to the roles played by nearby systems in the midbrain. Within the dorsal portion of the striatum are the caudate nucleus and the

putamen. These structures of the striatum are similar, but they vary anatomically. It has been found that Parkinson's patients seem to lose neurons in the posterior putamen as well as the SNc, and the posterior putamen is associated with habitual behaviors, while the rostromedial putamen is predominantly associated with goal-directed behavior. This loss of neurons in patients with Parkinson's disease in their posterior putamen adds to our understanding of these brain structures, as these patients lose the ability to walk, with the orchestration of actions required to walk being a habitual behavior (Redgrave et al., 2010).

The Hippocampus

While the basal ganglia are involved with learning associations, the hippocampus is involved with memory and space. In many learning experiments, the memory of the participant is measured by perturbing their learning and extinction processes. This section discusses what happens to some of the previously covered classical conditioning phenomena when the hippocampus is damaged.

In 1957, Scoville and Milner found that, when performing lobotomies and other brain surgeries to alleviate patients of debilitating epilepsy or intractable psychosis, medial temporal lobe lesions led to severe memory impairment, while working memory and IQ remained intact. While these lesions sometimes alleviated the symptoms of patients with epilepsy or psychosis, these patients suffered from severe memory deficits when the surgeries were located near the medial temporal lobe, primarily when damage occurred to the hippocampus. The hippocampus is located medially in the temporal lobe, near the insular cortex. Scoville and Milner described these patients, such as patient Henry Molaison (H.M), in the following way:

> [Patient H.M.] *could no longer recognize the hospital staff nor find his way to the bathroom … he did not remember the death of a favorite uncle three years previously, nor anything of the period in the hospital, yet could recall some trivial events that had occurred just before his admission to the hospital. His early memories were apparently vivid and intact … he will read the same magazines over and over without finding their contents familiar. This patient has even eaten luncheon in front of one of us (B. M.) without being able to name, a mere half hour later, a single item of food he had eaten; in fact, he could not remember having eaten luncheon at all …* (Scoville & Milner, 1957)

Patients with these medial temporal lobe lesions suffered from severe memory loss. Most patients suffered from anterograde amnesia; they had trouble forming new memories after the surgeries. Some even experienced additional partial retrograde amnesia; they were unable to remember some events that happened up to three years before the surgeries. For most patients however, memories formed three or more years before the surgeries were left relatively intact.

Scoville and Milner concluded that the hippocampus was important for memory. They found that for complete bilateral hippocampal lesions, patients were unable to form new memories from the time of the surgery onward. They also found that the extent of lesioning to the hippocampus predicted the amount of memory impairment. Patients with

lesions near the hippocampus such as the amygdala and uncus showed little to no memory impairment, as well as patients with only unilateral lesions of the hippocampus (Scoville & Milner, 1957). This research profoundly shaped our understanding of learning, and it provided evidence that the hippocampus is important for learning and memory.

In 1970, Kimble and Kimble performed hippocampectomies on rats and ran them in a T-maze, with food on one end of the T-maze arm, signaled by a light. Three groups of rats underwent training: rats with hippocampectomies, rats with lesions in neocortex overlying the hippocampus, and unoperated controls. All three groups learned the association between the light and food. Halfway through the experiment, extinction trials began. The food was removed so that neither arm of the T-maze had food, but the light continued to appear. The researchers measured the amount of time that the rats spent at the top of the T-maze, considering which arm to travel down. They argued that incremental changes in this delay time suggested incremental changes in the animal's association between the light and the food, eventually learning that the light was no longer predictive of a food reward. The time that the control and neocortically lesioned rats spent at the top of the T-maze increased over trials once the extinction trials (removal of the previously present association) had begun, whereas the rats with hippocampectomies continued to run towards the light, not learning that the light no longer predicted food. These rats did not explore the other arm of the T-maze and continued to run toward the arm with the light throughout the entire experiment. The authors concluded that hippocampal lesions had negative effects on the abilities of rats to learn extinction. This shows that the rats without hippocampi are able to learn associations; however, they fail to extinguish a learned association (Kimble & Kimble, 1970). While the hippocampus seems to play an important role in learning new memories, it does not seem necessary to learn an association between a stimulus and a reward. It does seem to be required to unlearn such a stimulus-reward association, however.

In 1986, Weikart and Berger performed similar hippocampectomies on rabbits and presented them with an eye-blink conditioning task. In this experiment, before training, one group of rabbits had hippocampectomies, while another group had lesions in the overlying parietal neocortex, and a third group of rabbits received sham surgeries, with no brain damage, acting as the control group. Both a light and a bell were presented intermittently to the rabbits, with one stimulus signaling an imminent puff of air to the rabbit's eye. Rabbits naturally blink to puffs of air to their eyes. After learning, stimuli that predicted imminent air puffs caused rabbits to blink their eyes in preparation. Thus, a blink following the stimulus that predicted the puff of air (even if no puff of air was delivered) indicated that the rabbit was learning the association between the light or bell and the puff of air. The rabbits from all conditions learned the association of the predictive stimuli to the puff of air. Halfway through the experiment, the predictive and unpredictive stimuli were reversed, resulting in the previously unpredictive stimuli now predicting the onset of the air puff and the previously predictive stimuli no longer predicting the onset of the puff of air. The neocortical lesion and control rabbits stopped blinking to the previously predictive stimulus and began blinking to the stimulus that now predicted the air puff. However, the researchers found that the rabbits with hippocampal lesions continued to blink to the first stimulus. Surprisingly, these rabbits also acquired the association between the second stimulus and a subsequent puff of air. These rabbits with hippocampal lesions could learn both associations, but they did not unlearn the first association when it no longer was useful for predicting the air puff. The authors

concluded that the hippocampus is important in order to properly learn tasks that require the reversal of an association (Weikart & Berger, 1986).

In 1997, Packard and Teather trained rats in one of two tasks and injected an antagonist of N-Methyl-D-aspartate (NMDA, a neurotransmitter implicated in synaptic change) into the hippocampus and dorsal striatum to understand these structures' roles in the rats' spatial learning. It had been previously shown that inhibiting NMDA receptors in the brain inhibits long-term-potentiation, and inhibiting NMDA receptors in the hippocampus impairs learning and memory (Collingridge & Bliss, 1987). The two tasks in the Packard and Teather study involved the Morris water maze task, where the rat is placed in a tub of clouded water, with a small hidden platform just below the water's surface that the rat can rest on (Morris, 1984). Since these animals find swimming aversive, they intrinsically seek out the platform in order to escape their current situation. In the first experiment, the rats were placed in the bucket at different locations, while the platform was always found in the same location. This was called the hidden platform task. In the second task, the rats were placed in the same location each time, while visual cues on the sides of the bucket indicated the location of the platform. This was called the visually cued platform task. In the hidden platform task, 2-amino-phosphonopentanoic acid (AP5, an NMDA antagonist) injected into the hippocampus impaired learning performance, while AP5 injected into the dorsal striatum did not. In contrast, in the visually cued platform task, AP5 injected into the hippocampus had no effect on learning, while AP5 injected into the dorsal striatum impaired learning performance. This research suggests that both the hippocampus and the dorsal striatum use NMDA receptors when learning, and that effectively inactivating those NMDA receptors with NMDA antagonists impairs their ability to learn. Crucially, this research shows that these two structures learn different aspects of a task without requiring the other to function. Namely, the hippocampus is recruited for spatial tasks while the basal ganglia is recruited for cued response tasks (Packard & Teather, 1998)

Packard and McGaugh (1996) performed similar experiments on rats in a cross-maze (a modular T-maze with two possible, opposite, starting positions). They injected lidocaine (an anesthetic which inhibits activation) into the hippocampus and the caudate nucleus within the basal ganglia while training rats to consistently approach an arm of the T-maze baited with food. Importantly, the animals were trained using only one of the two starting positions. After eight and sixteen days of training, the rats were put in the opposite starting position, and the researchers observed which arm the rats entered. Rats that entered the baited arm of the maze on these test trials learned the spatial location of the baited arm (place learning), while rats that entered the non-baited arm learned the individual action to take at the location of choice (response learning). After eight days, control rats showed *spatial learning*, however after sixteen days, the response became over-learned, and they sometimes exhibited *response learning*. Rats with hippocampally injected lidocaine showed no strategy preference after eight days, but after sixteen days showed *response learning* from over-learning. Rats with lidocaine injected into their caudate nucleus showed *spatial learning* after both eight and sixteen days of training. This research suggests that when the hippocampus is chemically inhibited, *spatial learning* is impaired, and when the caudate nucleus is chemically inhibited, *response learning* is impaired (Packard & McGaugh, 1996). This is further evidence of the functional differences in learning between the hippocampus and the basal ganglia; the hippocampus learns spatial associations, while basal ganglia learn state-response associations.

These studies can give us a glimpse into how the hippocampus works. It is not necessary for forming associations; however, it is necessary for unlearning associations. Without a hippocampus, rats are unable to explore other options to learn new associations as in the T-maze (Kimble & Kimble, 1970). The Weikart and Berger (1986) work suggests that animals can learn other associations, but they have great difficulty extinguishing the first association that they learned. The work of Packard & Teather (1998) and Packard and McGaugh (1996) suggest that an animal without a hippocampus can learn state-action associations using the basal ganglia, and those with lesions to the basal ganglia can learn spatial associations using the hippocampus. When combined with the other studies discussed here, this provides evidence that the basal ganglia can learn associations but cannot extinguish them without the additional support of the hippocampus. Overall, these studies show that the hippocampus may not be needed for forming associations, but it is crucial for unlearning those associations.

These studies examined the functional properties of the hippocampus, as a whole, but there has been some work examining the differential properties of distinct regions of the hippocampus. In 1993, Moser, Moser, and Anderson selectively lesioned portions of the rat hippocampus and found that spatial memory was partially affected. Using rats, they lesioned either the dorsal or ventral portion of the hippocampus and put them in the visually cued Morris water maze with spatial cues that allow the rat to orient itself (Morris, 1984). After a few trials of placing a healthy rat in the tub, it can find the platform quickly and regularly. In Moser *et al*., both sets of hippocampally lesioned rats were able to learn the location of the platform; however, the rats with dorsal lesions to their hippocampus took significantly longer to learn the location of the platform. Rats with ventral lesions to the hippocampus showed no significant difference compared to sham surgery rats without lesions and healthy control rats. The authors concluded from this research that the dorsal portion of the rat hippocampus is important for spatial learning (Moser, Moser, & Andersen, 1993).

Related work has also been done in humans. In 2000, Maguire *et al.* measured the size of the posterior and anterior hippocampi in New York City taxi drivers. They found a positive correlation between length of time spent as a taxi driver and the size of the posterior hippocampus. While it's possible that people with larger posterior hippocampi are drawn to being taxi drivers for longer lengths of time, it seems more plausible that spending time as a taxi driver increases the size of your posterior hippocampus. This increase in posterior hippocampus led to an equal decrease of anterior hippocampus (Maguire et al., 2000). This is consistent with the work done with rats.

Taking these two papers on spatial learning into consideration, it seems that the dorsal hippocampus in rats and posterior hippocampus in humans are important for spatial learning and memory, while the ventral/anterior hippocampus is comparatively more important for other types of learning and memory. It is important to make the distinction between learning an association and spatial memory, as both are affected by the hippocampus. Lesions to the hippocampus do not hinder the learning of an association (Kimble & Kimble, 1970; Quirk & Mueller, 2008; Tracy, Jarrard, & Davidson, 2001; Weikart & Berger, 1986), but hippocampal lesions do interfere with extinction. In contrast, lesions to the dorsal/posterior hippocampus will decrease spatial understanding as shown in Moser *et al.* (1993) and suggested in Maguire *et al.* (2000).

In 2012, Liu, *et al.* studied rats that had been conditioned to associate the sound of a tone with a light foot shock. Learning was indicated by an innate fear response - freezing - upon the sound of the tone. Using optogenetic methods, they were able to

identify hippocampal neurons that become active during the learned freezing response, and they were able to activate those neurons at a later time, when no tone was presented. The way they did this was by marking neurons that were active during the fear response with a protein, and using this protein marker to add specific light-sensitive proteins to those marked neurons. Later, when the animal was moved to a new environment, the researchers could shine a light on the hippocampus, thereby selectively activating those neurons that were active during the learned fear response via the light-sensitive proteins. The researchers found that the rats froze in place significantly more often when the specific fear-conditioning-related hippocampal neurons were active. This is a clear demonstration that the activation of hippocampal cells can give rise to a learned response. The authors conjectured that, upon this hippocampal stimulation, the rats re-experienced episodes involving both the tone and the shock, during which the association was learned (Liu et al., 2012).

This collection of studies provides a wealth of information concerning the function of the hippocampus: 1) the hippocampus provides explicit memory formation for introspection as we see from Scoville and Milner (1957), 2) lesions to the hippocampus destroy the ability to explore alternative actions in a familiar environment as we see from the rats in Kimble and Kimble (1970), 3) the hippocampus is not necessary for learning associations but appears to be necessary for extinction as we see in both Weikart and Berger (1986) and Kimble and Kimble (1970), 4) the hippocampus and basal ganglia learn tasks in different ways and can do so without the other as seen in Packard and Teather (1998) and Packard and McGaugh (1996), 5) the dorsal/posterior portion of the hippocampus is related to spatial memory as shown in Moser *et al.* (1993) and Maguire *et al.* (2000), and 6) that neurons in the hippocampus encode some fear conditioning memories as shown in Liu *et al.* (2012).

Modeling Learning

Since the earliest days of scientific psychology, there have been efforts to mathematically and computationally specify the mechanisms of learning. These *reinforcement learning* (RL) paradigms have captured some learning phenomenon quite robustly. Pavlov (1928), Skinner (1938), and Thorndike (1933) all discussed reinforcement learning, but Thorndike (1911) first discussed this in terms of the Law of Effect: animal responses that are closely followed by a satisfactory environmental state will be more likely to occur again (Hull, 1943; Pavlov, 1928; Skinner, 1938; Thorndike, 1898; Thorndike & Jelliffe, 1912). Some accounts of learning involved the simple strengthening of associations with the co-occurrence of stimuli (Hebb, 1949). While these models aligned well with observations of neural synaptic plasticity, they could not account for many aspects of learning. The Rescorla-Wagner model of learning, which strengthens associations in proportion to the difference between expectations and the actual appearance of stimuli (such as rewards), was among the first models to provide an explanation for some learning phenomena seen in experimental data, such as blocking and extinction (Rescorla & Wagner, 1972).

When the Rescorla-Wagner model is used to build an association between a neutral stimulus and reward, it can be seen as a form of temporal difference (TD) learning, updating the association based on the difference between a prediction of reward when the stimulus is presented and the actual presence or absence of reward a moment later (Sutton, 1988). This difference in expected and actual reward is called the temporal difference (TD) error. Building on the Rescorla-Wagner model, Barto, Sutton, and Anderson (1983) developed the *adaptive actor-critic* architecture, in which learning

the expected future reward in each environmental state, and learning the action that should be taken in each state, are divided into two separate systems. The *adaptive critic* system learns how good a state of the environment is, and the *actor* system learns the best action for each state (Barto et al., 1983). Described briefly, the adaptive critic produces a prediction of expected future reward, given features of the current environmental state, and this prediction is used to calculate the TD error. The adaptive critic learns by updating reward associations with features of the environment in proportion to the TD error. Similarly, a positive TD error increases the associations between current environmental features and the action selected by the actor system, while a negative TD error decreases the likelihood that the chosen action will be selected again, in that state, in the future. The *actor-critic* division of labor in this architecture allows the model to select actions that maximize the amount of environmental reward. Sutton (1988) proved that the *actor-critic* architecture would always converge to the optimal solution given a discrete environment and sufficient time, and this proof was extended in generality by Dayan (1992). An actor-critic system learning from TD error would always find the best way to receive the most reward, given enough experience (Dayan, 1992; Sutton, 1988). There is increasing evidence that TD learning using the *actor-critic* architecture provides a good description of dopamine-based learning in the basal ganglia. Phasic firing of DA neurons mirrors the temporal difference error - the difference between expected and actual rewards - and DA can affect synaptic plasticity in the manner required by TD learning (Barto, 1995; Montague et al., 1996).

The TD learning approach, using the *actor-critic* architecture, has been found to be quite powerful. For example, Boyan (1992) and Tesauro (1992) used the *actor-critic* architecture to play backgammon at the champion level. Importantly, these applications of TD learning used a mathematical value function approximator (VFA) to implement the adaptive critic, rather than implementing the adaptive critic as a look-up table, mapping environmental states to expected future reward values. One benefit of using a VFA is that it supports generalization from previously encountered environmental states to novel ones, estimating future reward for a novel state based on its similarity to experienced states. Using a VFA can also speed up learning in larger environments, such as backgammon, so that not every environmental state needs to be visited individually in order to learn a general policy for selecting actions (Boyan, 1992; Tesauro, 1992).

It is worth noting that proofs of optimality for TD learning do not apply when the adaptive critic is implemented as a VFA, however. Given their experience with VFAs and backgammon, Boyan and Moore (1995) decided to investigate empirically the performance of the *actor-critic* approach when using a VFA. They found that learning can sometimes be robust when using a VFA, but there are relatively simple learning problems for which the introduction of a VFA leads to a divergence in learning. This means that, for some learning tasks, the use of a VFA can result in the adaptive critic never reaching stable estimates of future reward and the actor never reaching a stable mapping from state features to the selection of an action (Boyan & Moore, 1995).

A number of different approaches to addressing this problem have been proposed. Boyan and Moore (1995) proposed a type of off-line learning mechanism in which the learner takes actions in the environment for a while without changing the actor or the critic. These action sequences, along with the corresponding experiences of the learner, are called “roll-outs.” The actor and critic are only updated based on aggregated data from each roll-out. The result is superior learning using a VFA for the adaptive critic. When performing TD learning with a VFA in the standard way, every update to the

adaptive critic can change the learner's expected reward not only for the current environmental state but also for a wide variety of states with similar features. Thus, the function of the adaptive critic can fluctuate wildly as the actor makes even small changes to its action selection policy. Following Boyan's and Moore's roll-out strategy keeps the actor from changing long enough for the learner to converge on a reasonable approximation of the expected reward from different states, reducing the likelihood of divergence[1]. Unfortunately, this strategy involves episodic rather than incremental learning, forcing the learner to remember whole roll-outs before improving the adaptive critic or the actor in any way (Boyan & Moore, 1995).

In response to Boyan and Moore, Sutton (1996) presented an on-line learning version of TD learning with a VFA that was able to learn complex learning tasks. Sutton's innovation was to encode the features of the environmental state using a sparse coarse code - namely the CMAC encoding used in an early model of the cerebellum (Albus, 1975). This encoding restricted changes to the adaptive critic and to the actor so that the responses for only a limited range of environmental states were affected by any one TD learning update. This was shown to support robust learning (Sutton, 1996). One problem with Sutton's solution is that it requires knowledge concerning the similarity structure of the environment in order to design an appropriate coarse code for environmental features. Thus, with this strategy, the representation of environmental features needs to be hard-coded into the model for each specific environment that the learner encounters.

While these models capture many learning phenomena in a high-level and abstract way, they lack biological plausibility or, at least, complete biological implementations. For example, in temporal difference models, predictions of reward are encoded over discrete "time steps", and the mathematical difference between these predictions is the basis for action selection at one specific temporal scale. The TD approach also requires the changing of reward associations from environmental features that are potentially no longer being sensed. For many researchers, whether it is possible for neurons in the brain to implement the needed calculations is not important, as these models are seen as models of behavior rather than brain function. Some work has been done to implement these models using connectionist neurons, *i.e.* continuous activation neurons (Barto, 1995), but even these models require the abstract subtraction of two neural activation values, separated in time, representing rewards and predictions of rewards. Other research has shown that temporal difference algorithms can be implemented using spike-timing-dependent plasticity, which is a biologically based process where the connection strength between two neurons can be strengthened or weakened based on the precise timings of the firing of pre- and post-synaptic neurons (Castro et al., 2009; Florian, 2005, 2007; Potjans et al., 2009; Rao & Sejnowski, 2001; Roberts et al., 2008; Rusu & Florian, 2009).

[1] "Roll outs" are a solution to noisy reward that requires aggregation or a learning algorithm that easily over-learns from single examples. By aggregating the results of multiple runs before updating your VFA, you can avoid over-learning or catastrophic interference (forgetting that results from the learning of new information). There is little biological evidence for a mechanism like "roll outs", but some researchers have suggested that something like this might happen during sleep. Other researchers, however, see sleep as a more complicated process. This account also fails to explain learning that takes place over, say, 30 minute training sessions where the animals can learn tasks without needing to sleep (Sutton, 1996).

Rae and Sejnowski (2001) used spike-timing-dependent plasticity (STDP) in multi-compartment neural models of cortical neurons and found that the neurons started to predict the spike timings of the input, doing so similarly to TD learning. Roberts *et al.* (2008) similarly found another STDP implementation to provide prediction of reward.

Similar follow up studies by Florien (2005, 2007), Di Castro *et al*. (2009), and Potash *et al*. (2009) showed various versions of TD-like reinforcement learning algorithms for spikes emerged within a neural network using a reward signal that modulated STDP, using different flavors of integrate-and-fire neural models. Potash *et al*. (2009), in particular, implemented the actor-critic model of Barto (1995) using specifically connected spiking neurons that employed STDP, proving that an actor-critic model of learning can be implemented with spiking neural networks, producing performance at the same level as Barto's (1995) discrete model.

These models made use of biologically realistic, though somewhat approximate, mathematical characterizations of individual neurons. Hodgkin and Huxley (1952) studied giant squid neurons and their axons and made biophysically accurate models of neurons. These models, however, are computationally intensive to simulate. In 2003, Izhikevich took these models and derived a lower dimensional version that is computationally simple and maintains biological plausibility. Later, Izhikevich (2007) proved that, with spike-timing-dependent plasticity, these neurons can learn associations, purely through the statistics of how often reward and predictive stimuli occur.

In 2011, Chorley and Seth created a spiking model of the striatum and the dopamine (DA) neurons in the substantia nigra pars compacta based on the Izhikevich (2007) model, where all of the learning occurred directly due to the connectivity of the neurons. In this model, all of the calculations were done in spikes. They proved that it was possible to achieve the same DA fluctuations due to reward solely with a spiking neural network where the calculations are done at the neural level, and not at some abstract level outside of the system. The DA neurons in their model gave the same activity levels as in the SNc of monkeys, as previously reported in the literature (Chorley & Seth, 2011).

While this model made many progressive steps forward, it does have some limitations when it comes to modeling the full range of learning phenomena. For example, the model does not account for extinguishing learned associations, and it does not account for spontaneous recovery. These missing pieces of the puzzle can be filled in if we remember that animals without hippocampi also exhibit these limitations – they cannot extinguish, nor can they spontaneously recover associations (as they cannot extinguish them in the first place). A model of the hippocampus might need to be added to the Chorley and Seth (2011) work in order to capture extinction and related phenomena. One approach would be to have the hippocampus learn an internal representation of the current environmental state, with this state representation resulting in better performance than pure TD learning in capturing all of the neuroscience data.

Redish, Jenson, Johnson, and Kurth-Nelson (2007) suggested that a model of the environment is required to exhibit spontaneous recovery. They further gave strong evidence to support their claim that "any mechanism that produces development of a new state in response to repeatedly low *delta* would produce the appropriate extinction with renewal" (Redish et al., 2007). Here, *delta* is the difference between observed and expected reward values - the TD error. Redish and colleagues (2007) claim that, in order

for a learning system to show both extinction and spontaneous recovery, it needs to be able to create new internal representations of the current state of the environment.[2]

This claim by Redish and colleagues (2007) is further solidified by recent work by Gershman *et al.* (2013), where rats were given extinction trials in which the rewarding stimulus was presented less and less frequently in relation to the predicting stimulus. This is called *gradual extinction*, highlighting the difference between this approach and that used in typical extinction paradigms, where reward presentation is suddenly and completely stopped after an association is learned. Rats experiencing gradual extinction showed significantly less spontaneous recovery than rats that were given normal extinction trials (Gershman, Jones, Norman, Monfils, & Niv, 2013). Gershman and colleagues' experimental evidence further strengthens the claim by Redish and colleagues, because the way in which internal representations of the environmental states change during standard extinction versus gradual extinction can explain how reward associations are learned so as to produce spontaneous recovery. Gradual extinction revisits the states where reward recently occurred, using slowly changing internal representations, and weakens the previously learned associations. Normal extinction, in contrast, abruptly changes the internal representation of the environmental state by suddenly removing the presence of any reward in that state, resulting in less weakening of the associations from the features of the original state representation. The larger residual associations in the standard extinction case produce the robust phenomenon of spontaneous recovery.

[2] In Redish and colleagues (2007), the model created new internal representations of the environment when the most recent context cues were sufficiently different from the running average of the previous context cues. This measure of difference could also be affected by strong negative prediction errors, which force the model to reassess the importance of the current cues being used (Redish, 2007). This is further explained in later sections.

CHAPTER 3:
THE FOUNDATIONS OF A NEW MODEL OF ASSOCIATION AND EXTINCTION

Introduction

In this section, I review the computational neuroscience models that this dissertation work builds upon, many of which were mentioned previously. This section goes into much more depth, describing each model in detail. First, I review the work done by Izhikevich (2007) that provided the initial neural network structure that was modified by Chorley and Seth (2011). After that, I review parallel work by a separate modeling group (Redish et al., 2007) which motivates the modifications to Chorley and Seth (2011) that I have made.

Izhikevich Neuron Model

Izhikevich initially created a realistic spiking model of neurons in 2003 by analyzing real neural spiking data and distilling down the equations rendered by Hodgkin and Huxley (1952). Hodgkin and Huxley made scientific history by analyzing giant squid neurons and creating very complex and biologically complete equations of the movement of ions that cause neurons to fire. This made great progress in modeling neurons in a very complete manner, but the equations were very complex and are computationally very expensive to model (Hodgkin & Huxley, 1952). Izhikevich (2003) simplified these equations while maintaining the same biological behavior, allowing for many more neurons to be modeled in the same amount of time (on the order of three orders of magnitude more, x1000). (See Appendix A for details.)

Izhikevich and colleagues made progress understanding and applying this new model (Izhikevich, 2003, 2004, 2006, 2007; Izhikevich & Hoppensteadt, 2009), and in 2007 they proposed a neural network that would produce some of the same neural spiking behaviors observed in dopamine neurons in animals: mainly the shift of dopamine activity to the earliest US-predicting stimuli.

Learning to Predict Using Izhikevich Neurons

In this section, I review some of the important work reported in Izhikevich (2007). This material will help us to understand the reinforcement learning model of Chorley and Seth (2011). An important problem addressed in Izhikevich (2007) was the *distal reward problem* (Hull, 1943), also called the *temporal credit-assignment problem* (Minsky, 1961). It asks, in a general sense: how do we predict reward? This is the problem that, given a reward, and any number of preceding stimuli that occurred before that reward, how do we know which of the stimuli (or temporally separated combinations of the stimuli) are predictive of the reward? How do we learn that one stimulus tells us that reward is coming when there are many different stimuli that occurred before the reward? This problem has been approached from many different angles, with varying degrees of success. In Izhikevich (2007), the problem was approached from the perspective of the neuron: how do neurons know which spike caused reward? And how is dopamine (a neurotransmitter signaling reward prediction error that is diffused in a global manner to many neurons at a time) used in conjunction with this? In what way are the efficacies of certain synapses increased or decreased to support this type of learning?

In order to solve these problems, the paper suggested that, on each synapse between any two neurons, there is an *eligibility trace* - a value that keeps track of how eligible that synapse is for a change in synaptic strength – a weight change. The idea is that if a pre-synaptic neuron fires an action potential that contributes to the activation of a synapse to a post-synaptic neuron, and the post-synaptic neuron fires shortly

thereafter (sending an action potential on to other neurons), then the pre-synaptic neuron *probably* helped cause the post-synaptic neuron to fire. Whether it actually did or not in a particular situation is not important. The general idea is if a post-synaptic neuron fired after a pre-synaptic neuron fired, the pre-synaptic neuron *probably* helped. The eligibility trace keeps track of how often events of this kind occur, and the value increases greatly if the time between pre- and post-synaptic firing is very small, inversely related to the difference in time. (Times less than a few milliseconds have the most effect, while times larger than 50ms have little effect, and times larger than 5000ms have effectively zero effect.)

The inverse is true if the opposite firing order occurs. If the post-synaptic neuron fires *before* the pre-synaptic neuron, then the eligibility trace is *decreased*, to a degree inversely related to the difference in the time between the spikes. This inversion when the opposite spiking order occurs makes sense: If the post-synaptic neuron fired *before* the pre-synaptic neuron, it is not possible that the spike from the pre-synaptic neuron caused the post-synaptic neuron to fire. This leads to a synaptic plasticity profile similar to that shown in Figure 1b. The closer in time the spikes occur, the larger the change in the eligibility trace, and the sign of the change is dependent on whether the pre- or post-synaptic neuron fired first.

The last important aspect of the eligibility trace is that the value decays to zero over time. This means that, in order to have a very large eligibility trace, the pre-synaptic neuron needs to fire before the post-synaptic neuron closely in time a few times in close temporal proximity.[3]

How does the eligibility trace cause the synaptic strength to increase and decrease? The synaptic strength is modified based on the product of the eligibility trace and the current dopamine level, as seen in Figure 1c, where the pre-synaptic neuron fires closely before the post-synaptic neuron fires. Below that, the extracellular dopamine level is depicted as sharply rising a short time later. Directly below the dopamine plot is the eligibility trace, which increases radically because of the closeness in time between the firings of the pre- and post-synaptic neurons. The eligibility trace then slowly decays. In this model, the synaptic strength is modified simply by the product of the extracellular dopamine and the eligibility trace, so we can see when both the dopamine and eligibility trace values are greater than 1, the synaptic strength increases, as shown at the bottom of Figure 1c.

This modification of synaptic weights, using the combination of current dopamine level and relative spike timing, is called dopamine-modulated spike-timing-dependent plasticity (DA-STDP). (See Appendix A for details.)

The Izhikevich (2007) paper made multiple contributions, including existence proofs of some interesting properties of networks of neurons implementing DA-STDP. Perhaps most importantly, the paper described a neural structure that includes dopamine neurons which behave similarly to those measured during experimental

---

[3] In this model, it's important to note that if one pre- post- spike pattern occurs in a short enough time span after another pre- post- spike pattern occurs, while the two pre- post- spike pairs will increase the eligibility trace, the eligibility trace will decrease somewhat due to the inter-spike-pair post- pre- spike pattern. It's also important to note that research is still being done related to triplets of spikes and synapse change due to spike-timing-dependent plasticity. Thus more complex functions for eligibility traces have been investigated (Froemke & Dan, 2002; Pfister & Gerstner, 2006).

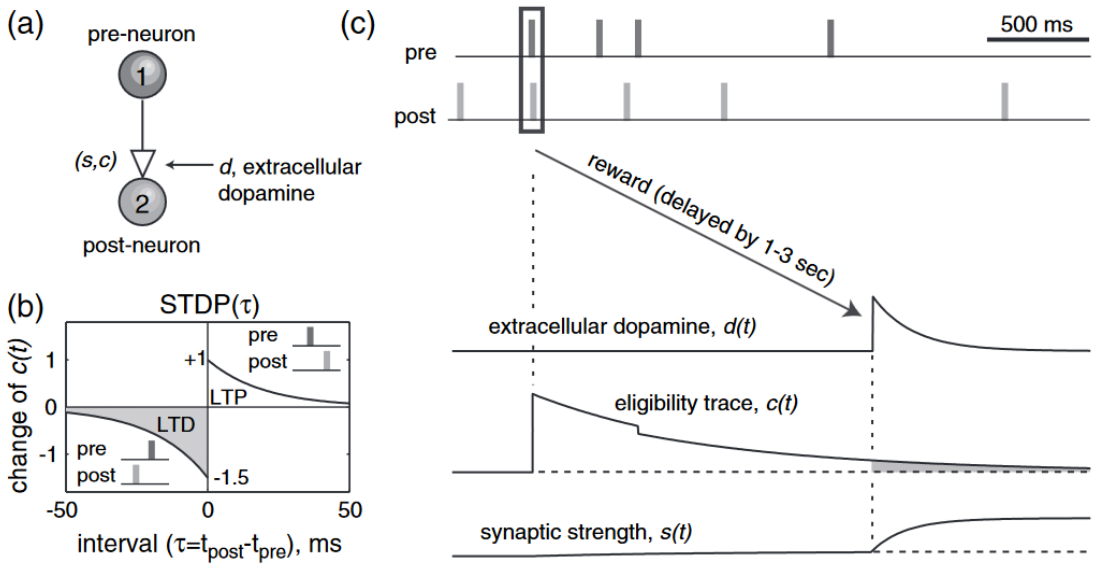


Figure 1. (taken from Izhikevich, 2007). The implementation of spike-timing-dependent plasticity (STDP). **(a)** is a depiction of two connected neurons, one pre-synaptic (1) neuron and one post-synaptic (2) neuron. **(b)** is a graph of the effect of the interval of the spikes between the pre- and post-synaptic neuron (x axis) and the change in eligibility trace that spike interval causes (y axis). Note positive (right of 0, unshaded) intervals mean the pre-synaptic neuron fired before the post-synaptic neuron, which means neuron 1 might have caused neuron 2 to fire, so increase the eligibility trace. While negative (left of 0, shaded) intervals mean the pre-synaptic neuron fired *after* the post-synaptic neuron, which means neuron 1 could not have caused neuron 2 to fire, so decrease the eligibility trace. **(c)** (top) is a depiction of the spike trains of the pre- and post-synaptic neurons, and how it, in addition to extracellular dopamine and the eligibility trace (middle), can affect the synaptic strength (bottom). Notice that the eligibility trace jumps in value when the pre-synaptic neuron fires directly before the post-synaptic neuron, and over time with no other similarly close spike intervals the eligibility trace decays back to zero. Also of interest to note is that the eligibility trace drops a small amount on the pre-synaptic neuron's third spike, because it follows the spike of the post-synaptic neuron closely enough in time to register on the graph in *b* of this figure.

studies. The model involved a neural structure with four groups of neurons (see Figure 2a). One group of neurons represented neurons that either directly or indirectly projected onto midbrain neurons that release dopamine. (While both the ventral tegmental area (VTA) and the substantia nigra pars compacta (SNc) have high concentrations of dopaminergic neurons, the VTA was the focus of this paper.) These neurons, when fired, increased the level of dopamine in the model. Another group of neurons encoded the unconditioned stimulus; these had strong synaptic connections to the dopamine neuron group, as well. The other two groups represented two conditioned stimuli, and they were connected to each of the other three neuron groups with initially weak weights.

In this network structure, when neurons from the first conditioned stimuli (CS) group fired, there was a chance that some of those neurons participated in causing the firing of CS2, US, and VTA neurons, which would increase the eligibility traces on those synapses. When the unconditioned stimulus (US) was presented shortly after, it caused the VTA group to fire, raising the dopamine level temporarily. Those eligibility traces that

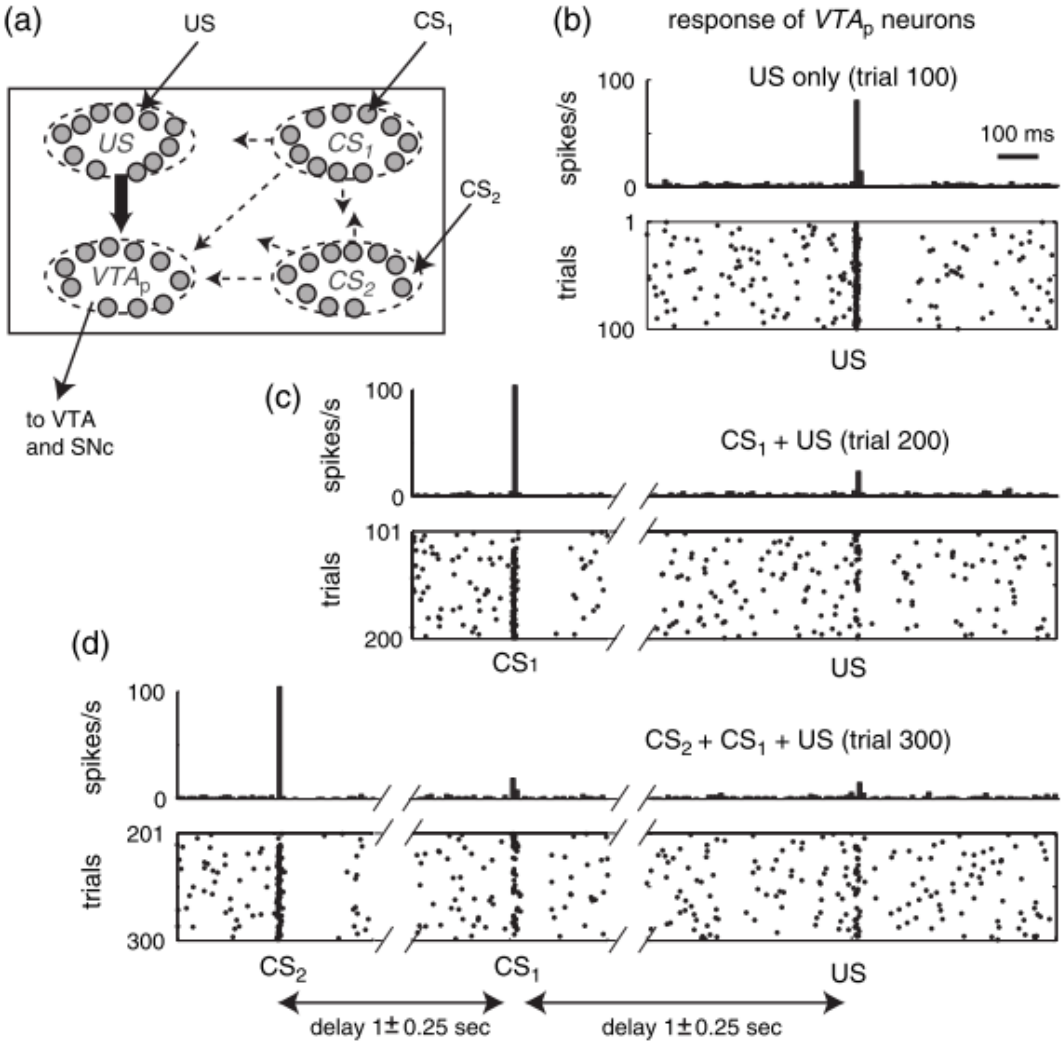


Figure 2: (Taken from Izhikevich, 2007). This is a model that explains the foundational principles of dopamine activity when learning associations between a conditioned stimulus and an unconditioned stimulus. This figure shows the structure and network behavior during association learning. **(a)** shows the structure of the model. There are four interconnected neuron groups, with the ventral tegmental area (VTA) group connected to dopamine (DA) cells, and the unconditioned stimulus (US) group strongly connected to the VTA group. Conditioned stimulus (CS) 1 and 2 are fully connected to the other groups, but with weak strengths. **(b)** After 100 trials of being presented with just the US, DA activity arose for the unpredicted presence of the US. **(c)** For the next 100 trials, CS1 was presented shortly before the US, and the network came to produce DA activity for CS1. Notice that the DA group continues to weakly activate for the US. **(d)** For a further 100 trials, CS2 was presented shortly before CS1 which was presented shortly before the US. Notice that DA activity arose for CS2 which predicted CS1, which in turn predicted the US. Notice, similarly, that the DA activity continued to weakly activate for both predicted stimuli CS1 and US.

were still high enough during dopamine release caused the strengthening of the corresponding synapses, according to DA-STDP. In simulated trials in which CS1 preceded US, the synapses from CS1 to both US and VTA were strengthened because they both caused DA to fire. Since CS2 was not presented during these trials, it produced fewer (noise) spikes that might contribute, through their initial weak weights, to spikes in VTA, so synapses from CS2 were not strengthened in a similar way. As the synapses from CS1 to US and VTA were strengthened, the generation of spikes in CS1 increased the production of DA, via US and VTA. This meant that, over association trials, the firing of conditioned stimulus neurons came to cause VTA neurons to fire.

Thus the model captured the learning of association between a given conditioned stimulus and a rewarding unconditioned stimulus. This was accomplished using a model that solely used a structured spiking neural network, along with eligibility traces. Importantly, there are some candidate biophysical mechanisms for maintaining eligibility traces (Izhikevich, 2007). This model was only an existence proof that something of this sort can occur, and it does have some limitations in its ability to account for some experimental evidence. For example, the dopaminergic neurons in the model continued to fire even for predicted reward, unlike the SNc, which only fires for unpredicted reward and unpredicted stimuli that predict reward. Another phenomenon that this model did not show is the canonical dip in dopamine level when an expected reward does not occur as expected (Ljungberg, Apicella, & Schultz, 1991). This model only showed the transferal of a dopamine burst from the US to the CS and to further CS's beyond that. It did not model extinction or related phenomena.

Overall, this work was instrumental in understanding how to model the changes in synaptic weights that occur over time between two neurons. This work showed how eligibility traces could be implemented in model neurons and argued for the biological plausibility of such synapse-specific variables. It also proposed a number of neural network structures that have helped the modeling community to understand how the dopamine activity in mammalian brains could come to be modeled using a biologically plausible spiking neuron model.

How Chorley and Seth Modeled the Dopamine Response

The model published in Chorley and Seth (2011) built on the neural network structure presented in Izhikevich (2007). A major goal was to better model the response of dopamine (DA) neurons in the substantia nigra pars compacta (SNc) when the animal is presented with unconditioned stimuli (rewarding stimuli), conditioned stimuli (stimuli that predict reward), and erroneous stimuli (stimuli that do not predict reward). Izhikevich (2007) showed how dopamine neurons can learn to fire for stimuli that predict reward, as well as unpredicted reward, but the model continued to fire for rewarding stimuli at a reduced level, even when the stimulus was predicted. This is not normally the case for the dopamine neurons in the SNc in animal subjects (Ljungberg et al., 1992). The Izhikevich (2007) model also did not show the canonical dip in dopamine activity for learned predictable rewarding stimuli that do not appear (Ljungberg et al., 1991). Chorley and Seth (2011) built upon that model to better parallel what is seen in the SNc of mammals. Mainly, this was accomplished by adding a second pathway for stimulus information to enter the model, introducing a way in which the timing of stimuli could be used to predict when dopamine activity would occur in the SNc. This second pathway also provided a mechanism through which an equal and opposite inhibition signal could be sent to the SNc, allowing the original pathway to continue to increase DA activity when predicted reward occurs, but stopping the DA response if it had been predicted.

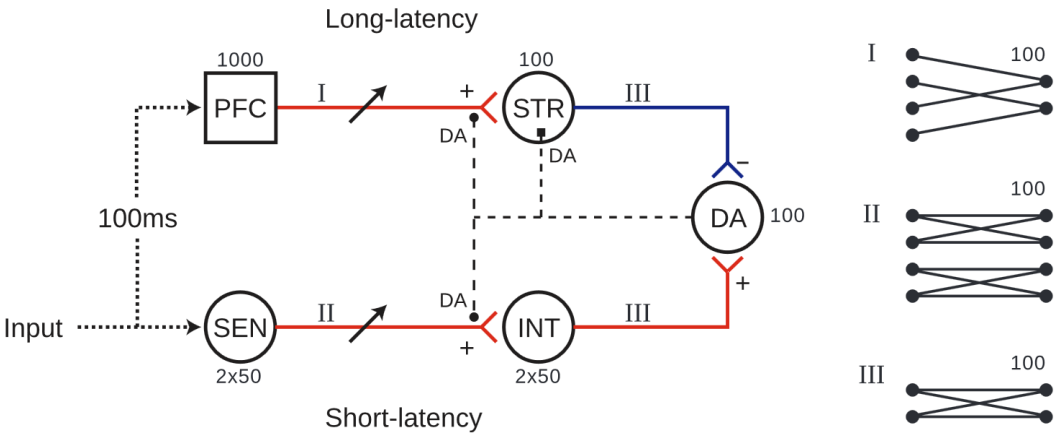


Figure 3: (Taken from Chorley and Seth, 2011) This is the structure of the network as described in Chorley and Seth 2011. There are two pathways in this model. The "*lower pathway*" passes through sensory neurons (SEN), interneurons (INT) and then to dopaminergic neurons (DA). This pathway is analogous to the Izhikevich (2007) model (see Figure 2). The lower pathway knows that the unconditioned stimulus (US - food) is good and fires strongly for it, firing INT and then DA. And it learns that the conditioned stimulus (CS - bell) predicts the food, and thus is also good. Thus after learning occurs, the lower pathway causes brief increases in firing activity in DA for both the unconditioned stimulus and the conditioned stimulus. The "*upper pathway*" passes through the prefrontal cortex (PFC), striatum (STR), and then to dopaminergic neurons (DA). This pathway acts to inhibit those brief increases in DA activity that are predictable. In this model, only the US is predictable (by the presence of the CS), and this "upper pathway" learns to inhibit DA at the right time so that the brief increase of activity that the "lower pathway" causes when the food occurs is met with an equal amount of inhibition, which cancels it out and keeps the DA activity at baseline. This leads to dopaminergic firing activity that replicates the firing activity seen in the substantia nigra of real animal brains. The diagonal arrows through connections symbolize plastic connections, while the dotted line from DA to STR symbolizes dopamine-modulated post-synaptic-facilitation (DA-PSF). The dotted lines from DA to the plastic connections between PFC and STR and between SEN and INT symbolize dopamine-modulated spike-timing-dependent plasticity (DA-STDP). Both of these mechanisms are reviewed in the text. The Roman numerals I, II, and III on the synapses and to the right symbolize the types of connectivity between each neuron group.

This section will further describe that model. See Figure 3 for more details. As a part of this dissertation work, I have replicated the Chorley and Seth model, and will present data from my replication that, to my knowledge, accurately reflect the results reported by Chorley and Seth (2011) (See Figure 4).

The model of Chorley and Seth (2011) was made up of 1400 Izhikevich neurons, divided into 5 groups, connected in a pattern as seen in Figure 3. For simplicity, both the excitatory neurons and the inhibitory neurons were modeled with Izhikevich (2003) regular spiking neurons with appropriate parameters. (See Appendix B.) There were two groups of 50 excitatory neurons that broadly signified sensory neurons that fired in the presence of each of the two stimuli: the unconditioned stimulus (US - the reward) and the conditioned stimulus (CS - the bell)[4]. These two groups of neurons were labeled as

[4] Here, I refer to the conditioned stimulus as "the bell", which initially was not rewarding, but after many trials of the bell preceding the reward, it came to predict the reward. Similarly, I refer to the

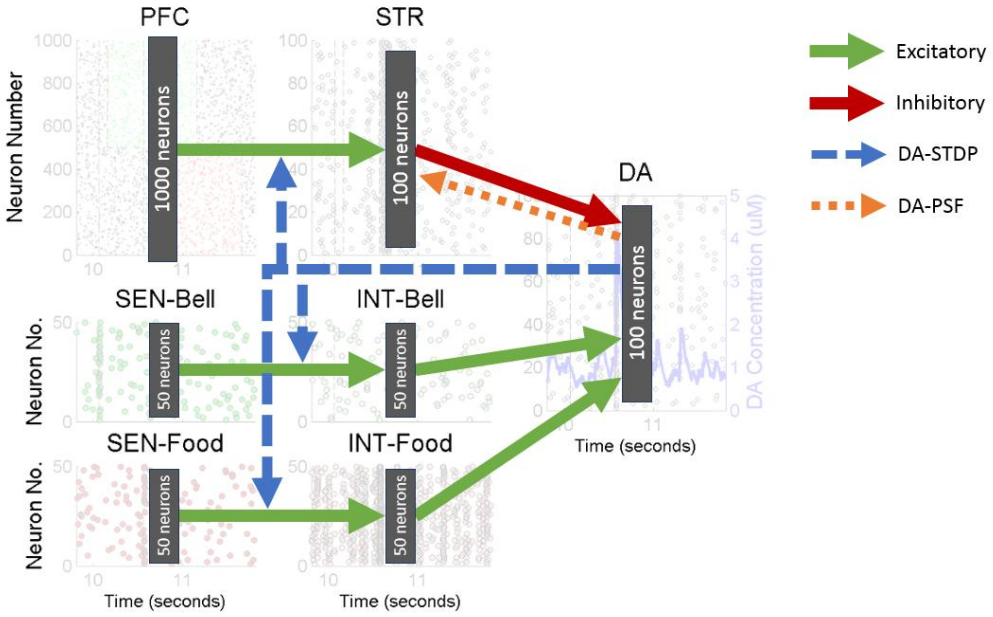


Figure 4: Replication of the Chorley and Seth (2011) model. The only major differences between this figure and Figure 3 is that the sensory and inter-neurons have been split into two different groups to make it easier to understand. Note, however, that the pattern of connectivity in this pathway is identical to that used by Chorley and Seth. The connectivity is as follows. Prefrontal cortex (PFC) has excitatory feed-forward connections to striatum (STR), connected at a 10% connectivity: each PFC neuron is connected to only 10% of the STR neurons. All other connections are all-to-all (100% connectivity) and feed-forward: STR has inhibitory connections to dopaminergic neurons in the substantia nigra pars compacta (DA). The sensory neurons for the bell (SEN-Bell) have excitatory connections to the bell interneurons (INT-Bell), which have excitatory connections to DA. Similarly for the sensory neurons for food (SEN-Food), they have excitatory connections to the food interneurons (INT-Food), and have excitatory connections to DA.

sensory neurons, as they fired when they sensed the presence of a stimulus. All neurons were injected with a small amount of noise, producing a low level of background firing. Firing rates were briefly increased in the appropriate sensory neurons in order to model the presentation of a stimulus. (See Appendix B for details.)

The two sensory neural groups were connected to two corresponding interneuron groups in a feed-forward all-to-all fashion, one interneuron group for each stimulus. (See Appendix B for further details.) Each of those interneuron groups had 50 excitatory neurons. It was these connections between the sensory neuron groups and the interneuron groups that formed the primary locus of learning, capturing the strength of

---

unconditioned stimulus as the reward or “the food”. This is to help the reader more easily intuit the structure of the model. Notice that further on in this document, the conditioned stimulus is sometimes labeled “light” instead of “bell.” There is no meaningful difference between these two from the perspective of the model; conditioning to the stimulus labeled “light” is no different than to the stimulus labeled “bell.” It is only important in later simulations (e.g., on blocking and overshadowing) that there is a need for two separate conditioned stimuli.

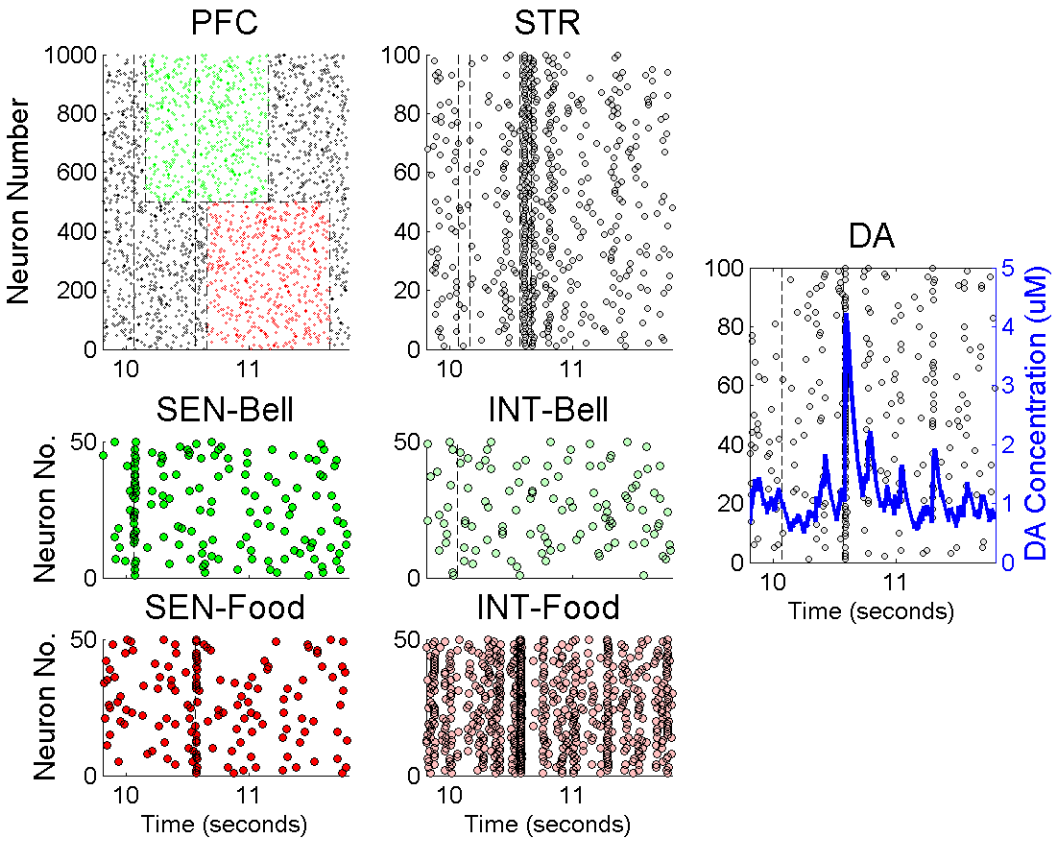


Figure 5: This is a graph of the spiking activity of one trial of a replication of Chorley and Seth (2011) before any training has occurred. Each trial was 3000ms, the presentation of stimuli was 1500ms in total, and the start of the 1500ms of stimuli presentation was randomly chosen within the first 1000ms of a trial. An experiment had between 100 and 400 trials in total. Shown here is a 1500ms portion within one of the first 3000ms trials where the stimuli are presented. The X axis shows time in seconds, and while trial number is not shown here, given that each trial is 3000ms, we can estimate trial number by dividing time in seconds by 3 (thus this is trial number 4 shown here). Notice that the dopaminergic neurons (DA) fired strongly for Food initially, and not for the Bell.

associations between each conditioned stimulus and reward. Thus, these were the connection weights that were learned using dopamine-modulated spike-timing-dependent plasticity (DA-STDP) over conditioning training trials. For the unconditioned stimulus, these connection weights were initialized to their maximal values, capturing the strong association between the US and reward (and corresponding dopamine activity) prior to the onset of conditioning trials. For the conditioned stimulus, these connections were set to their minimal values, corresponding to no initial association with reward. This can be seen in Figure 5. (The illustrated data are from a replication of Chorley and Seth (2011) that is discussed later in this dissertation.) The neuron group SEN-Food briefly increases its firing rate for the presence of food, and because it has strong connections to INT-Food, it causes a similar increase in firing rate in the INT-Food neuron group. In this first trial of an experiment, however, when SEN-Bell increases its activity for the onset of the bell, the connections to INT-Bell are weak, and so there is not a similar increase in firing rate in the INT-Bell neuron group.

The two interneuron groups were then both connected all-to-all to the dopamine (DA) neuron group. This connection was *not* plastic. The synaptic strengths were not learnable, and these strengths were fixed for all neurons at a specific value (6% of the maximum weight value). The DA neuron group represented the dopamine neurons in the substantia nigra pars compacta (SNc). Dopamine neurons in this area have been shown to fire for unpredicted reward and unpredicted stimuli that predict reward (Ljungberg et al., 1992; Romo & Schultz, 1990; Schultz & Romo, 1990).

As described so far, the Chorley and Seth (2011) model was very similar to what was presented in Izhikevich (2007). Both models incorporated sensory neuron groups that fired for conditioned and unconditioned stimuli, and both had a connection pathway from these groups to dopaminergic neurons. In Izhikevich (2007), there was a third conditioned stimulus and no interneurons, but the spiking behavior of the dopaminergic neurons was largely the same across the two models. Using the Izhikevich neurons, along with eligibility traces, the appropriate conditioned stimulus learned to cause the dopaminergic neurons to fire just as strongly as the unconditioned stimulus, correctly learning when a conditioned stimulus was predictive of reward.[5] This learning occurred in the "lower pathway" in the Chorley and Seth (2011) model (see Figure 6). Through the plastic connections between conditioned stimuli neural groups and their corresponding interneurons, the model learned to fire dopaminergic neurons for stimuli that predict reward, essentially learning associations between conditioned stimuli and reward.

Beyond Izhikevich (2007), the Chorley and Seth (2011) model included another neural pathway. The second portion of the Chorley and Seth (2011) model was the "upper pathway" (see Figure 3). This additional pathway learned to inhibit dopaminergic neurons from firing during the presentation of stimuli that were predictable. In this way, only unpredictable rewarding stimuli or unpredictable stimuli that predict reward came to cause a phasic dopaminergic response, as seen in animal studies (Schultz, 1998; Schultz & Romo, 1990). In this upper pathway, there were 1000 excitatory neurons that provided about 1-5Hz of random spiking activity over time. These neurons represented a portion of the prefrontal cortex (PFC), and they were used to maintain information about presented stimuli over time (see Figure 6). When a stimulus was presented to the model, both the lower pathway and the upper pathway were affected, but they were affected in different ways. As previously discussed, the lower pathway briefly increased its firing rate in the neural group associated with the stimulus presented. However, in the PFC portion of the upper pathway, the firing activity was not simply increased, but it was changed to be a stereotyped (i.e., identical whenever triggered) 1000ms time-locked pattern of neural spikes, starting when the stimulus was presented. Of the 1000 PFC neurons, there were 500 neurons that fired in a specific pattern whenever the conditioned stimulus appeared, and there were 500 other neurons that fired in a (different) specific pattern when the unconditioned stimulus appeared. Thus, every time a particular stimulus was presented, the same spatio-temporal pattern of neural spikes was generated across a population of neurons in PFC dedicated to that stimulus. The unconditioned stimulus caused one set of neurons to fire in a specific way over time, and the conditioned stimulus caused another set of neurons to fire in another specific way.[6]

---

[5] It's important to note that given any number of stimuli that do not predict the rewarding unconditioned stimulus, the model can learn the stimuli that best predicts reward, solving the distal reward problem.

[6] If the patterns of spiking in the PFC for the stimuli are sufficiently sparse, there is good reason to think that a common population of neurons could be used to encode all stimuli, rather than

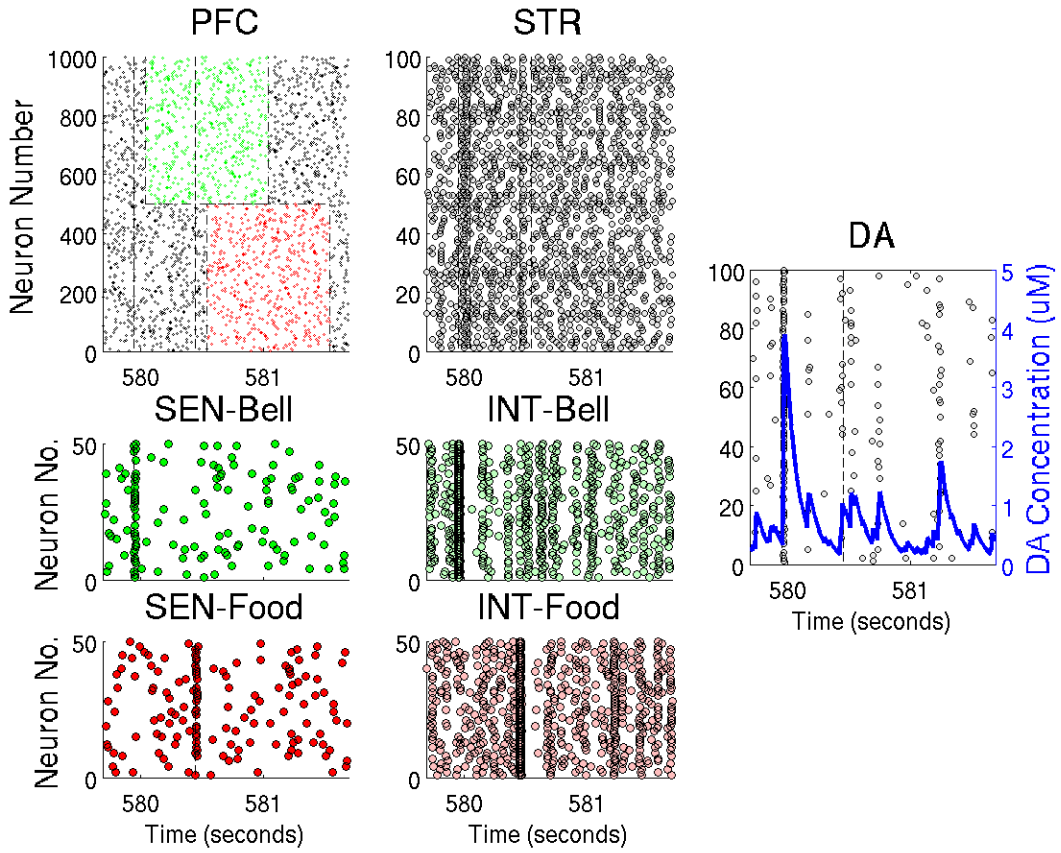


Figure 6: Chorley and Seth (2011) replication, showing representative performance of the model after learning has occurred over 200 trials, each taking about 3 seconds. Notice that increases in activity in both the Bell and Food sensory neuron groups created corresponding increases of activity in both of the inter-neuron groups. This shows that the model continued to respond strongly to the Food, but it had also learned a positive association with the Bell. Also notice that the DA neuron group fired strongly for the Bell after learning, and no longer fired for the Food. This shows that the model learned that the Bell predicted Food.[7]

This mechanism was used to represent a working memory trace of each presented stimulus, maintained in the prefrontal cortex, encoded as a polychronous neuronal group (Izhikevich, 2006). The idea that the brain sometimes encodes stimuli in this way, as an extended temporal pattern of spikes over a population of neurons, has increasing support (Izhikevich & Hoppensteadt, 2009; Martinez & Paugam-Moisy, 2009; St. Clair & Noelle, 2013, 2015). Given that the presentation of a stimulus caused a subset of the cells in this prefrontal cortex (PFC) neuron group to fire in a stereotyped manner, the

---

using segregated groups, with one group for each stimulus (St. Clair & Noelle, 2013). This has not been explicitly tested, however.

[7] It is important to explain the much higher activity in the two inter-neuron groups for bell and food, as well as the STR neuron group. This is caused by the increase in strengths of weights to these neuron groups, and may be a side effect of the exact formula and curve for DA modulated STDP. This is a problem Izhikevich (2007) has dealt with and accordingly increased the amount of long-term depression from reverse spike ordering (a post- and then a pre-synaptic spike) to balance the change in weights over time.

specific PFC neurons that fired at specific points in time after a stimulus was presented were *the same every time*. This meant that the model could reliably learn very specific properties of the representation of a stimulus, such as, after 498ms, neuron 45 in the PFC fires, and after 499ms neuron 348 fires, etc. This is important, and it is the key to how this model learned to inhibit dopaminergic firing for predicted stimuli at the right time.

The 1000 neuron prefrontal cortex (PFC) group was connected, in a feed-forward fashion at 10% connectivity, to a neuron group of 100 inhibitory neurons, labeled striatal (STR) neurons (see Figure 3). This connection between PFC and STR was plastic, changing based on dopamine-modulated STDP. What was learned by these connections was which neurons in PFC predict a phasic burst of activity in the DA neuron group, at some particular (short) delay. In other words, these connections became strong for those PFC neurons that fired when a dopamine spike was about to occur. This synaptic learning arose through a combination of dopamine-modulated STDP (DA-STDP) and dopamine modulated post-synaptic facilitation (DA-PSF).

Dopamine-modulated spike-timing-dependent plasticity (DA-STDP) was explained in the previous section on the Izhikevich (2007) model, as well as in Appendix A. As a quick recap, the timing between a spike at a pre-synaptic neuron and a spike at a corresponding post-synaptic neuron modifies a decaying eligibility trace; a pre-synaptic spike followed closely by a post-synaptic spike greatly increases the eligibility trace, while the opposite order greatly decreases the eligibility trace. Larger times between spikes reduce the strength of eligibility trace change exponentially, with a greater than 50ms difference producing close to no change (5000ms is zero change). The weight of the connection between the pre-synaptic and post-synaptic neurons is increased by the product of the dopamine level and the eligibility trace. If either are below baseline (a negative eligibility trace or a dip in dopamine level) the weight decreases.

Dopamine-modulated post-synaptic facilitation (DA-PSF) involves an increase of the excitability of a neuron when there is an increase of dopamine above its baseline firing level, along with a corresponding decrease in the excitability of the neuron when there is a decrease of dopamine below its baseline firing level. A baseline firing level maintains an even excitability level, around the same value as neurons that are not affected by DA-PSF. The details of this mechanism are described in Appendix B.

It's important to note that DA-PSF was only implemented in the striatal (STR) neuron group in the Chorley and Seth model, and DA-STDP was only implemented in two places: 1) on the connections between the sensory groups and the interneurons and 2) on the connections between the PFC neuron group and the STR neuron group.

In high level terms, what DA-PSF did in the STR neuron group is increase the probability that a spike from the PFC neuron group would cause a spike in the STR neuronal group when there was a phasic dopamine response, so that the eligibility trace would increase and DA-STDP would, therefore, increase the weight of that connection. This is important because there was no increase of spiking activity that the network could rely on for pre-synaptic neurons to "cause" post-synaptic neurons to fire in the upper pathway like there was in the lower pathway. DA-PSF briefly increased the firing rate of the STR neurons in the moments after the US was presented, increasing the probability of the PFC neurons (time-locked in a polychronous neuronal group representing the US) to cause an STR neuron to spike. This increased the weights of the specific neurons that always fired directly before the DA neurons fired for the US.

The striatal (STR) neuron group was connected in a feed-forward all-to-all manner to the dopamine (DA) neuron group. The strengths of the synapses from these

neurons were not plastic, and they were fixed at 6% of the maximum synaptic strength. Because the STR neurons were inhibitory, they could counteract the excitatory input that the DA neurons received from the lower pathway. Recall that the lower pathway learned to activate the DA neurons in the presence of rewards and stimuli that predicted rewards, using spike-timing-dependent eligibility traces and DA-modulated STDP. With the lower pathway activating DA when predictable rewarding stimuli occurred, the upper pathway could learn to strengthen the connections to STR from neurons in the PFC that fire when a phasic dopamine response occurred.

When a phasic dopamine response occurred, DA-PSF increased the probability of a PFC spike causing a spike in STR, thereby increasing the eligibility trace of that connection. At that point, because the DA level was high, the synaptic strength of the PFC to STR connection also increased. Over trials, if the same group of PFC neurons spiked every time before a phasic dopamine response (e.g. those PFC spikes predicted a phasic dopamine response), the PFC to STR connection weight would substantially increase. Note that if it was not the same PFC neurons that reliably fired before the dopamine response then the connections would not increase consistently enough to cause neurons in STR to spike, and this meant that STR would not inhibit the phasic dopamine response. This explains how the model could use the presentation of a previous stimulus - one that occurs regularly before a rewarding stimulus - to inhibit a phasic dopamine response at the time of the subsequent presentation of the rewarding stimulus, matching recordings from SNc. See Figure 6. Notice that there is no increase in dopamine activity at the onset of the Food, but there is one for the onset of the Bell.

Chorley and Seth (2011) found a few interesting results from running their simulations. Their model built upon the work done by Izhikevich (2007), and so some of their results replicated those found in Izhikevich (2007), including: 1) a shift in DA responses from the unconditioned stimulus to the conditioned stimulus, and 2) the model still produced strong DA responses for unpredicted rewarding stimuli. Unlike the Izhikevich (2007) model, their new model produced a measurable dip in the dopamine activity when an expected reward didn't arrive. Chorley and Seth (2011) also tested the robustness of their model in three ways: a) they modified the specific timing delay from the conditioned stimulus to the unconditioned stimulus, showing that there wasn't anything special about the 500ms delay between stimulus and reward that they regularly used - that the model could also learn when the delay was anywhere in the range 200-900ms, b) they modified the specific timing between the conditioned stimulus and the unconditioned stimulus every trial, adding noise and varying the delay each trial by up to 100ms, showing that their results continued to hold in this case, as well, and c) they modified the noise in the PFC spike patterns specific to each stimulus, changing the input of up to 10% of the neurons[8], showing that here, as well, the model maintained similar performance. It's important to note that Chorley and Seth (2011) specifically mentioned in the discussion section of their paper that their model does not account for extinguishing learned associations, and they did not investigate that aspect of classical conditioning, as they "consider that process to involve additional, active, mechanisms." (See Figure 7.)

[8] In Chorley and Seth (2011), they describe percent noise in their discussion as the percent of neurons in the PFC at each timestep that instead of receiving the time-locked input (which is the stimulus-specific time-locked injection of voltage uniformly ranging from -6.5 to +6.5), they receive a random voltage value between -6.5 and +6.5.

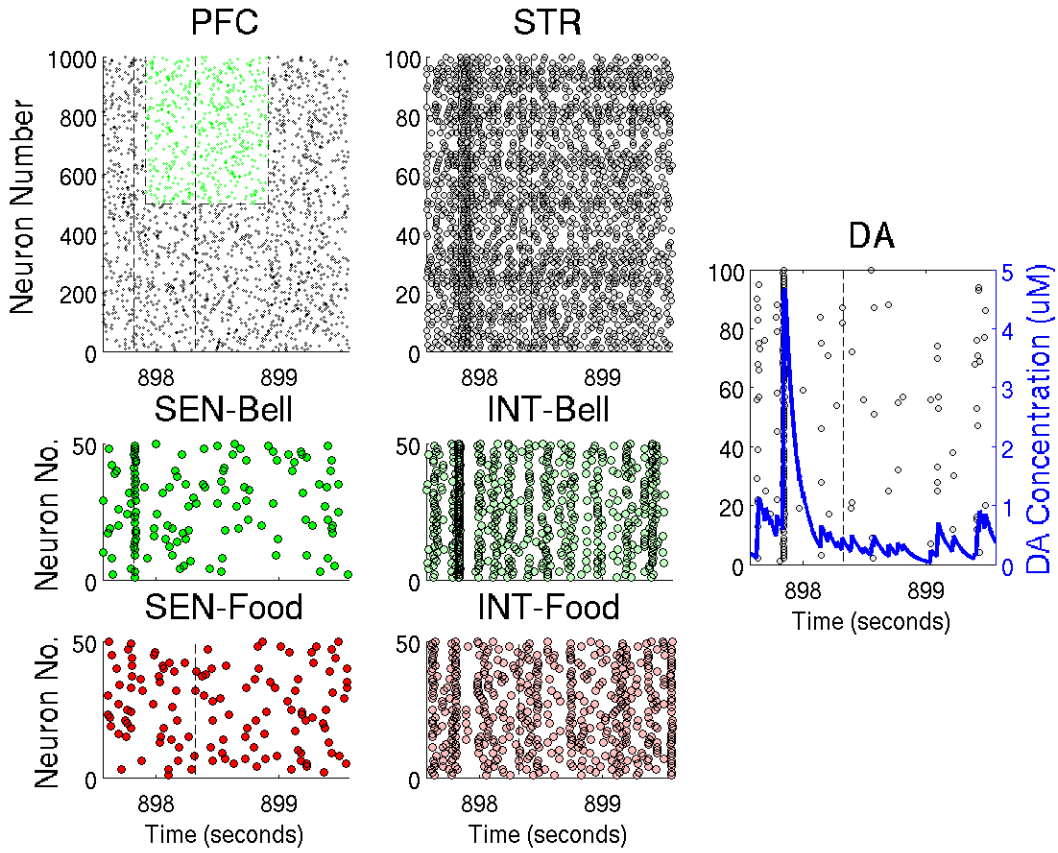


Figure 7: A replication of Chorley and Seth (2011), showing representative data after 200 trials of association learning and 100 trials of extinction. Notice that the dopamine (DA) group of neurons continued to fire for the Bell, even after 100 trials of extinction. In this model, there is nothing that can counteract the learned association of the Bell in the "lower pathway." Because the Bell was unpredictable, the "upper pathway" could not inhibit the dopamine response to the Bell from the lower pathway, as it usually did for the predictable Food.

The Chorley and Seth (2011) model is important because it was the first model of dopamine activity in the SNc that used only spiking model neurons, DA-STDP, and DA-PSF. Most other models required either forced downtime in activity between the conditioned stimulus and the unconditioned stimulus (e.g. Houk et al., 2007), or they used an abstract non-neural formula to save the value of the prediction of reward for use later in time (Barto, 1995; Castro et al., 2009; Florian, 2005, 2007; Potjans et al., 2009; Rao & Sejnowski, 2001; Roberts et al., 2008; Rusu & Florian, 2009). Thus, this model provided progressive work towards a fully biologically-plausible model of reinforcement learning, accomplished solely through a network of spiking neurons, DA-STDP, and DA-PSF.

The Redish Model of Association Learning and Extinction

Work by Redish *et al.* (2007) provided some insights that suggest ways to introduce extinction into the Chorley and Seth (2011) model. The classical understanding of learning and extinction (which doesn't fully account for all learning phenomena) is that the animal learns an association and either directly unlearns it during extinction, or somehow lowers its strength temporarily during extinction (Rescorla &

Wagner, 1972). Redish and his colleagues proposed that, instead, during extinction, learning occurs in relation to a changing internal representation of the state context (Redish et al., 2007). They demonstrated that this model accounts for many phenomena not previously accounted for, such as spontaneous recovery, gradual extinction, and the effect of different schedules of reinforcement on gambling addicts.

The theory behind this model is as follows. The model continually determined what state in state space it was in based on the context information available to it. The contextual cues that determined its state in state space could include the sensory input to the model from its environment, the amount of reward currently being delivered, and the reward-recency, among other things. The model estimated the 'importance' of each cue in determining its current state in state space, weighting the cues that provided more information higher[9]. This was the 'attention' that that cue was given. Uninformative cues were given little or no 'attention' and were not used to determine the current state, while informative cues were given more 'attention' and provided more information about the current state. Each trial the model gathered another data point (another set of cue values) to better categorize and define its state and updated which cues were providing the most information and should, therefore, be valued more. As it gathered more data about a state, it monitored the variance of the most recent set of data points (the cues that made up the most recent set of trials, modified by the 'importance' each cue has), comparing that to the variance of all of the cues in that state (also modified by the 'importance' the cues have). When these measures diverged past a threshold value, the model entered a new discrete state. It is important to note that this comparison was weighted by an 'attention' parameter that made cues more or less important when calculating this difference between all the cues of a state and the most recent cues. This 'attention' parameter was itself modified by the temporal difference (TD) error, or the value-prediction error. Redish and his colleagues theorized that tonicly low levels of TD error caused the model to reassess its current evaluation of the importance of each context cue. Sufficiently frequent low levels of TD error caused the model to change its evaluation of the importance of each cue, so that when it compared the difference between all of the cues that made up the current state and the most recent cues, they would diverge, treating the current situation as a new state. It effectively increased its weighting of previously ignored environmental cues, reducing the importance of previously important environmental cues, and this changed how it represented the current state. When it changed how it encoded the current state, if the collection of weighted cues changed to a sufficient degree, the model treated the current situation as a distinctly (and discretely) separate new state. In this new state, the model had different expectations, and new associations between the state and appropriate actions could be learned. It is important to note that the Redish model ignored positive TD error (unexpected reward) and only used negative TD error (expected reward that did not occur) in its calculations. Doing so, the model would not enter a new state during association trials, when there was positive TD error (the reward was unpredicted and the model was learning to predict it) and would thus build an association from essentially a single state to reward. However, the model would enter new states during extinction,

[9] The entropy (in terms of information theory) of a cue in the cue space for a state was measured and compared with the entropy of all cues observed in that state. If the cue provided information about that state - if a cue increased the entropy of the total cues for a state - then it was weighted as more 'important', and more 'attention' was paid to that cue. Cues that did not increase the entropy of all of the cues observed in that state were ignored, or, at least, given less importance.

because there was negative TD error (the predicted reward did not occur). The Redish model, in all of its detail, is fairly complicated, but it is important to note that these complications are not all necessary to produce similar patterns of behavior. The authors acknowledge this:

> *Any mechanism that produces development of a new state in response to repeatedly low* δ [TD error] *would produce the appropriate extinction with renewal. We have built a model based on increased attention to cues in response to low* δ *for simulation purposes, but it is important to note that any model that produces state changes in response to the undelivered, expected reward would produce similar results. …* (Redish et al., 2007)

Thus, the central insights of this model are quite general. In order to capture the learning effects of interest, it is not important to fully implement this model or use any of the same mechanisms that it uses. It's only important to have a model that changes its internal representation of the current state context in appropriate situations, such as when expected reward repeatedly fails to appear.

CHAPTER 4: REPLICATING THE CHORLEY AND SETH MODEL

In this dissertation, the original accomplishments achieved include replicating the model described in Chorley and Seth (2011). This model, described in detail in Appendix B, was implemented from scratch using MATLAB (The Mathworks Inc., 2016), and the phenomena reported by Chorley and Seth (2011) were reproduced through the execution of simulations. Details concerning the implementation of this replication may be found in Appendix D.

The model in Chorley and Seth (2011) could learn when a conditioned stimulus predicted an unconditioned stimulus, providing the same dopaminergic response pattern as seen in experimental studies (Ljungberg et al., 1991, 1992; Romo & Schultz, 1990; Schultz & Romo, 1990). It could not, however, extinguish those learned associations, and it continued to provide a dopamine response for stimuli that used to predict reward but no longer do. This phenomenon – being able to learn associations but unable to extinguish them – is very similar to what is seen in behavioral studies with animals that have lesions to the hippocampus (Jarrard & Davidson, 1991; Kimble & Kimble, 1970; Quirk & Mueller, 2008; Tracy et al., 2001; Weikart & Berger, 1986).

My replication demonstrated these patterns of responding. Illustrative data, showing the acquisition of an association, are shown in Figures 5 and 6. Notice how the model initially produces a strong DA response only for the US (Figure 5) but, through learning, comes to produce a burst of DA for the CS (Figure 6). Importantly, due to the functioning of the “upper pathway”, learning results in the suppression of a predicted DA response for the US (Figure 6). If the model is subsequently presented with extinction trials, in which the CS is presented but the US is not, the Chorley and Seth (2011) replication fails to lose its association from the CS, continuing to produce a strong DA response when the CS is presented (Figure 7).

In this way, my replication explicitly shows that the Chorley and Seth (2011) model, reflecting aspects of the cortical interactions with the basal ganglia, is able to explain the DA response during the learning of an association but not during extinction trials.

CHAPTER 5: A MODEL THAT INCORPORATES THE HIPPOCAMPUS

In addition to replicating the Chorley and Seth (2011) model, this dissertation presents an augmented model that merges aspects of learning in the basal ganglia with a hippocampally produced internal representation of the state context in order to produce a spiking neural network model that exhibits extinction and related phenomena. This work uses the ideas of Redish and colleagues (2007) to augment the Chorley and Seth (2011) model to be able to replicate animal behavior results. This dissertation combines the model of Chorley and Seth (2011) with the idea of conjunctively coded states, conceptualized as being formed in the hippocampus, from Redish *et al.* (2007). See Figure 8 for information about the connectivity of this augmented model.

In brief, I propose that a hippocampus is needed to unlearn previously learned associations. While the literature contains many models of the hippocampus, the focus, here, is on the generation of sparse conjunctive codes of sensory states (McClelland, McNaughton, & O'Reilly, 1995), as suggested by Redish and his colleagues (2007). As in previous models, the Redish model viewed the dopamine signal as encoding the error in expected reward. This value was high if the model received more reward than expected, and it was low if the model received less reward than expected. This description precisely describes the dopamine signal in the substantia nigra pars compacta (SNc); it is high when an unexpected reward or an unexpected stimulus that predicts reward appears, and is particularly low when an expected reward does not occur.

Methods

The augmented model adds a new hippocampal group of neurons to the Chorley and Seth (2011) model. This group of neurons represents the sensory state and context, mirroring the state representations used in Redish *et al.* (2007). The Redish *et al.* (2007) model used discrete and abstract states without offering a biologically plausible implementation of those states. In the proposed model, these states are seen as encoded across a population of neurons, with the population providing a distributed sparse conjunctive code of the current context. The hippocampus is well suited to learn such a code through an unsupervised learning process (McClelland et al., 1995), but the research proposed here is not focused on representational learning. Thus, as a simplification, the proposed augmented model uses a hardcoded sparse conjunctive code to represent the kinds of contextual states suggested by Redish *et al.* (2007). The idea is that, while other areas of the brain process individual sensory cues separately, the hippocampus processes these cues conjunctively. That is, there are sets of neurons in the hippocampus that activate for a conjunction of multiple cues. To keep things simple, the hard-coded context representation used in the proposed augmented model involves a population of neurons with each neuron responding to the combination of exactly two cues. There are a total of 6 sensory stimuli and contextual cues available to this model: the presence of the US ("Food"), the presence of CS1 ("Light"), the presence of CS2 ("Bell"), the perception of being in Context1 ("Room A"), the perception of being in Context2 ("Room B"), and the reward-recency (an internal physiological cue analogous to *time-since-last-reward,* implemented in Redish et al. (2007); "R"). With 6 contextual cues, there are 15 (6 choose 2 = 15) order-independent conjunctions. For each conjunction of two context cues, we designate a subgroup of 20 neurons in the hippocampus to activate only when both of the two contextual cues are active. There are three types of stimuli into the hippocampus neuron group: sensory stimuli, contextual stimuli, and reward-recency. Sensory stimuli are encoded in the form of short bursts of

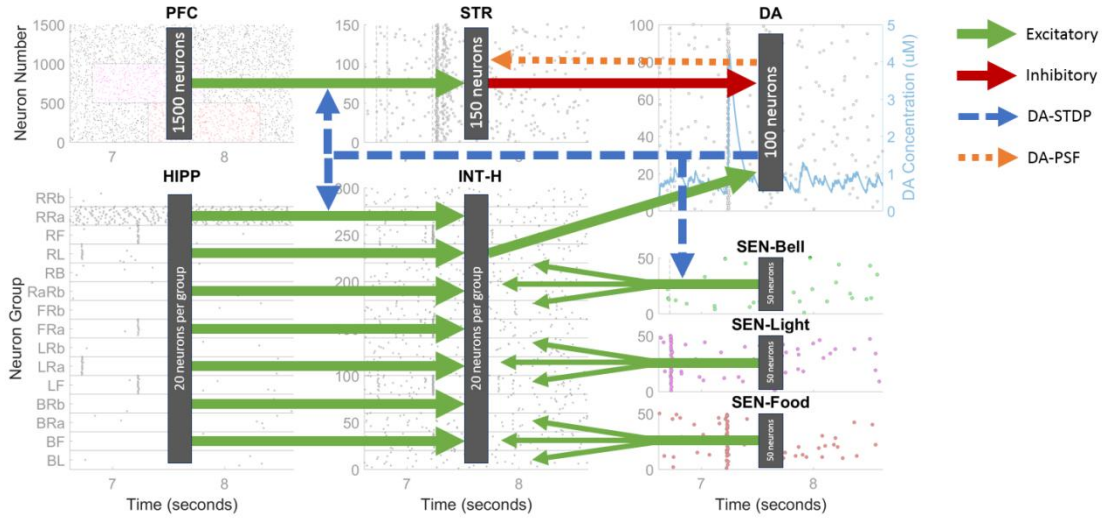


Figure 8: This is a schematic diagram of the augmented model of Chorley and Seth (2011), with a hippocampus component added. The "upper pathway" through the PFC and STR is essentially unchanged. However, in the "lower pathway" a very simple model of sparse conjunctive encoding in the hippocampus is added. In this augmented model, the neurons in the hippocampus are divided into groups, with each group encoding for the conjunction of a pair of cues. The inter-neuron group (INT-H) receives input from the corresponding hippocampal neuron groups, allowing them to represent conjunctions of pairs of cues as well. The cues included information about the sensory cues, as well as other context cues like the Room (A or B), and the reward-recency. Lastly, the "lower pathway" continues to get sensory information from the individual sensory neuron groups. Each sensory input projected to all of the conjunctive groups in INT-H that involved the corresponding sensory cue. Note that the arrows shown here between the sensory groups (SEN-Bell, SEN-Light, and SEN-Food) and the INT-H neuron group is a visual simplification to keep the image concise. In the model, those connections are much less neatly connected to each subgroup in INT-H. Note that, to support simulations involving multiple conditioned stimuli, this model includes both a Bell and a Light as conditioned stimuli, and many of the following simulations involve learning associations from the Light to the Food.

spikes at stimulus onset. Contextual stimuli are encoded by an elevated firing rate for the current context of the learner: Context1 or Context2. The reward-recency is encoded using a variable firing rate, with the rate falling to baseline over time, until a reward (US) is presented, at which time the firing rate is elevated to its maximum value.[10] Each of the neural subgroups in the hippocampus fire conjunctively for two of the six cues mentioned above: when both of those are on/high, then that subgroup shows a burst of firing activity. For neuron subgroups that include one of the other cues (Context1, Context2, reward-recency), the spiking activity is fairly straightforward: the activity of the neuron subgroup is equal to the product of the spiking activity of the two stimuli. For example, the Context1 x US conjunctive hippocampal neurons are only highly active for the 15ms when the model is in Context1 and the US appears (since sensory stimuli are encoded

[10] Notice that Redish et al. (2007) encoded this as time-since-last-reward, and this value slowly increased each millisecond, resetting it to zero when reward was received. In this augmented model, the inverse of time-since-last-reward is used, reward-recency: it is reset to 1 when reward is received and slowly decays to zero. This is a simpler way of modeling it in the context of this model, and bounding the activity is necessary in a spiking neural network.

by bursts of spikes), even if the model is in Context1 for a longer period of time. For conjunctions of two sensory stimuli, there is a brief burst of firing activity on the onset of overlap of their temporally-extended time-locked polychronous group (PNG) in the PFC. That means that if the PNG for the CS starts, and 500ms later the PNG for the US starts, because each PNG is 1000ms in length, both PNGs overlap for 500ms. At the onset of this overlap, the neurons corresponding to the hippocampal conjunction for these two stimuli briefly increase in firing activity. (See Appendix C for details.)

The interneuron groups of the Chorley and Seth (2011) model are changed to be aligned with the hippocampus subgroups, rather than being aligned with each stimulus. Each conjunction subgroup in the hippocampus is connected all-to-all only to the corresponding subgroup in this so-called inter-hippocampus neuron group. This inter-hippocampus group of neurons represents the sub-hippocampal cortex, such as the entorhinal cortex and the bundle of neural fibers that run from the hippocampus up to the fornix, and cingulate cortex, and back down to the entorhinal cortex and hippocampus, called the Papez circuit (Papez, 1937; Wyass & Van Groen, 1992).[11]

The original sensory groups of neurons are still included in the model, and these groups are connected to the inter-hippocampus neuron group in such a way that the subgroups within the inter-hippocampus group that include a contextual cue that is one of the sensory cues have incoming connections from that sensory neuron group.

The inter-hippocampus neuron group then connects to the dopaminergic group of neurons, matching the connectivity of the original interneurons in the Chorley and Seth (2011) model. (See Appendix C for details.) This preserves the ability of the sensory neuron groups to continue to learn the dopaminergic spiking activity behavior shown in Chorley and Seth (2011) and Izhikevich (2007), while maintaining a conjunctive hippocampal state encoding.

Results

Extinction and Reinstatement

Simulations of the hippocampus-augmented model have demonstrated its ability to exhibit the extinguishing of a learned association. The model continues to learn to predict rewards given a predictor stimulus, as also shown by Chorley and Seth (2011) and Izhikevich (2007). However, interestingly, the model now can extinguish previously learned associations.

Simulations of association trials were conducted, showing that the augmented model is capable of learning when a reward (US) is predicted by a conditioned stimulus (CS). In other words, it can be shown that the dopaminergic response for an unconditioned stimulus (US) can move from the US to a US-predicting CS. This is illustrated in Figures 9 & 10, where Figure 9 displays representative firing patterns prior to learning and Figure 10 displays representative firing patterns after 200 association trials in which the Light preceded Food. Notice the shift in the DA response from the time at which Food was presented to the time at which the Light was presented. While this is not novel, it is important to show that the model can still learn associations in this way.

---

[11] There is some research to suggest aspiration or ablative lesioning destroys this pathway and affects behavior (Jarrard & Davidson, 1991) while other inactivation or excitotoxic lesions destroy only hippocampal neurons while leaving axons that pass through the hippocampus and thus, the Papez circuit, intact (Corcoran & Maren, 2001).

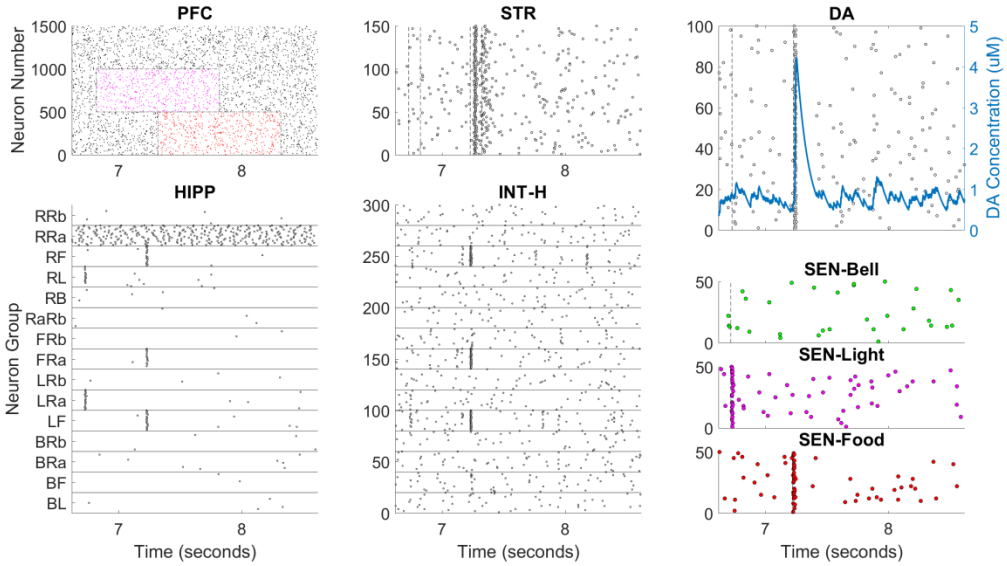


Figure 9: An initial learning trial of the modified model. Notice that the dopaminergic neurons fire strongly for the Food, and not for the Light. Here the upper pathway from PFC to STR to DA remains as it was in Chorley & Seth (2011); however the lower pathway is modified. There are still CS (Bell and Light) and US (Food) neural groups; however they instead are connected to hippocampal interneurons (INT-H). INT-H stands in place for all of the INT neuron groups in Chorley & Seth (2011). INT-H sends excitatory connections to DA. INT-H also receives excitatory input from the hippocampal neuron group. Both INT-H and HIPP are conjunctive neuron groups, where each subsection is a conjunction of 2 of the 6 sensory stimuli (i.e. Bell, Light, Food, Room A, Room B, Reward-recency). The subsections in HIPP are connected only to their corresponding subsection in INT-H, and the three sensory groups for Bell, Light, and Food are connected to each subsection of INT-H that is a conjunction of one of them (e.g. the SEN-Light group is connected to the subsections Reward-recency+Light (RL), Light+Room-A (LRa), Light+Room-B (LRb), Light+Food (LF), and Bell+Light (BL) in INT-H).

Not only does the augmented model learn, but it also can extinguish learned associations (Figure 11). This happens because different hippocampal state representations arise during extinction trials, roughly implementing the ideas of Redish and his colleagues (2007). Redish captured extinction by having the state representation change over extinction trials in a way that it didn't during association trials. The Redish model accomplished this through a specific mechanism that noted when reward was predicted but did not arrive, using this as a trigger for the generation of new discrete states. In the augmented Chorley and Seth model, a simpler approach is taken. The hippocampal state in this model is sensitive to the reward-recency (perhaps related to satiation), and this internal cue remains fairly constant over association trials, since Food arrives regularly. In contrast, the reward-recency declines linearly over extinction trials, causing the hippocampal state to substantially shift over those trials. In this way, the augmented Chorley and Seth model succeeds at producing a changing state representation during extinction, as in the Redish model, but it does so without introducing a special mechanism for detecting the failure of reward predictions.

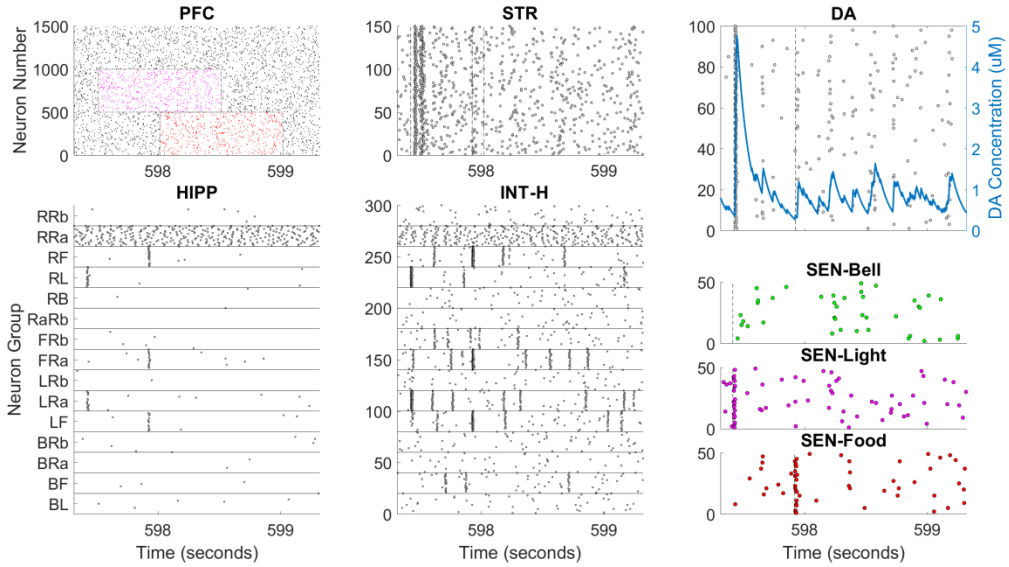


Figure 10: The augmented model after learning has occurred. Displayed is data from a representative trial in which the system had been presented with 200 association trials between the Light and Food. Notice that the dopaminergic neurons now fire strongly for the Light, and no longer for the Food.

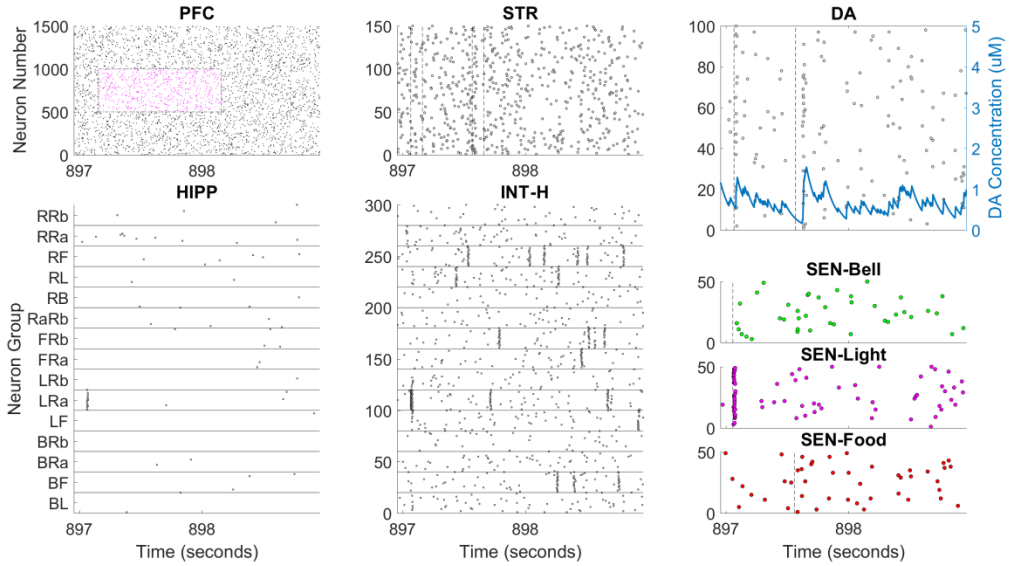


Figure 11: The model can extinguish learned associations when the state changes. This diagram shows an example trial after 200 association trials between the Light and Food, followed by 100 extinction trials, in which the Light is presented but no Food is delivered. Notice that the hippocampal groups that are firing for the Light have changed from Figure 10. Also notice that the dopaminergic neurons do not fire very strongly at the time of the presentation of the Light.

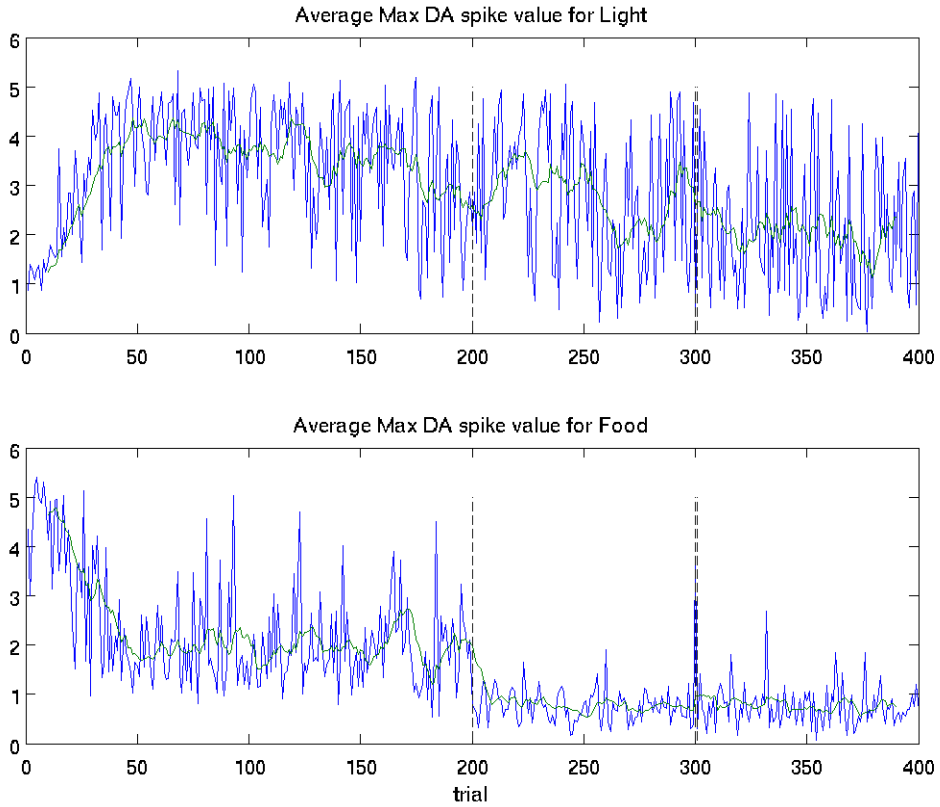


Figure 12: The average maximum DA value when the Light and Food appear *for the replication of Chorley and Seth (2011)*. This figure graphs the maximum DA value within a 50ms window after either the Light (upper) or Food (lower) was presented on each trial throughout the experiment. The Y axis is the amount of DA in the model and the X axis is the trial number. The blue line is the DA value, while the green line is the averaged DA value, averaging over the last 10 data points. The experiment is first associative: the Light precedes and predicts the Food. Then extinction trials begin at trial 201 (the first vertical dotted line), during which only the Light is presented. The second and third (very close) vertical dotted lines at trials 301 and 302 are a single trial of just the Food presented, followed by more extinction trials of only the Light. Notice that, initially, the DA levels rise for the Light and drop for the Food, but, during extinction and even after the Food is presented on trial 301, the DA levels for the Light (upper graph) remain relatively high.

The overall learning and extinction profile of the model can be more succinctly summarized. (See Figures 12 and 13.) By recording the maximum dopamine response in the 50ms after the Food appears, we can measure the model's response to Food over many trials. We can do the same for the 50ms after the Light appears. By examining these measures over trials, we can assess how well the model learns an association, how well it extinguishes it, and how well it can reinstate a learned association. Figure 12 shows this information for my replication of Chorley and Seth (2011). This is the model that could learn associations but could not extinguish them. Notice that the dopamine (DA) level for the Food (lower) drops after 50 trials of learning, showing that the model correctly simulates dopamine levels in animals: there is no dopamine response for food that is predicted by a light. Notice the further drop in DA below baseline during

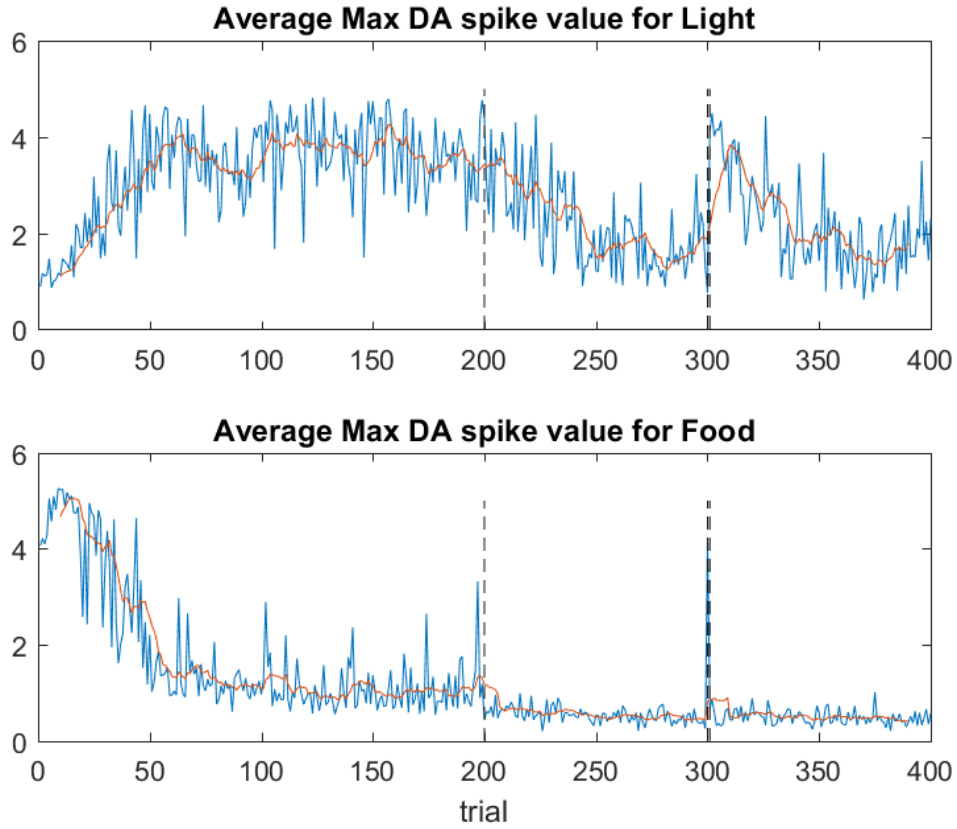


Figure 13: The average maximum dopamine (DA) value when the Light and Food appear *for the hippocampally augmented Chorley and Seth (2011) model*. Notice that the DA levels rise for the Light and drop for the Food. However, during extinction trials, the dopamine activity for the Light drops precariously, exhibiting extinction. Also, after the Food is presented on trial 301, higher DA levels for the Light are reinstated. Finally, after more trials of extinction (302-400), the DA levels drop down again, nearing baseline levels.

extinction. This is also what we see in animals. When a predicted US does not occur, there is a dip in dopamine at the time it was expected.

It is also important to notice the upper part of Figure 12. This displays the DA response at the time that the Light is presented. Notice that the dopamine level for the Light rises during the first 50 trials of learning and then remains noisy and fairly high throughout the remainder of the experiment, over the subsequent trials of learning (trials 50-200), extinction trials (trials 201-300), a single trial of Food (trial 301), and further extinction trials (trials 302-400). Notice the lack of strong effects on the DA response to the Light that these changes in reinforcement schedules have, once the association is learned.

Figure 13 shows representative results for the same sequence of trials, this time presented to the augmented model. We can compare these results with those from the replicated Chorley and Seth model (Figure 12): the lower portions are similar, if a little less noisy in the augmented model. What is important to note, however, is that the dopamine level for the Light (upper) drops off during extinction in the augmented model that has a hippocampus. This shows that this augmented model can unlearn associations.

Extinction arises in this model due to changes in the state representation encoded in HIPP that arise over the course of extinction trials. As the extinction period advances, the reward-recency cue declines, causing fewer and fewer spikes in the neuron clusters that encode cue conjunctions involving reward-recency. The DA-STDP strengthened synapses from neurons in these HIPP clusters, which came to help drive a conditioned DA response through INT-H, stop contributing to the generation of corresponding INT-H activity. In this way, as the state representation changes, the DA response to the Light declines, demonstrating extinction.

These simulations also demonstrate the presence of another established conditioning effect in the augmented model which is absent in the Chorley & Seth (2011) replication. This is the effect of reinstatement, in which an extinguished association returns simply by providing the US, alone. Reinstatement can be seen at trial 301, when the model is given a single trial of Food. After that trial, the dopamine level for the Light jumps back up, just as you would expect for a healthy subject. Further extinction trials (302-400) extinguish the association once again, as we also see in animals. Note that this effect does not appear in the model lacking a hippocampus (Figure 12), since extinction does not occur. It arises in the augmented model, however, due to a large change in the HIPP state representation upon the presentation of Food. The single trial of Food presentation resets the reward-recency cue, returning the HIPP state representation approximately to what it was before the onset of extinction trials. The sudden increase in spiking in HIPP neurons associated with reward-recency allows the synapses that were strengthened during the association trials to, once again, play a role in producing a strong DA response to the Light.

Thus, simulations of the augmented model show that the hippocampus can explain extinction and can still learn associations. It also exhibits the phenomenon of reinstatement. There is good reason to believe that this model will also account for a number of other phenomena related to classical conditioning, which are explored next.

Spontaneous Recovery

Recall that spontaneous recovery is the phenomenon of classical conditioning where the animal spontaneously recovers a previously extinguished association when brought back to the original learning context after a period of absence. After learning an association and subsequently extinguishing it, the animal is brought to its home cage, and the next day when it is returned to the original experimental setup the animal continues to predict that the unconditioned stimulus (US) will occur after the conditioned stimulus (CS) occurs. This is strange because the previous day, by the end of extinction, the animal no longer predicted the US after the CS.

This phenomenon was replicated in the augmented model by making the following modifications to the experimental setup. As before, the first 100 trials were learning trials in Room A where we presented the Light (CS) before the Food (US). The next 100 trials were extinction trials, still in Room A, where we presented just the Light (CS) and no Food (US). The difference in this experimental setup was that these 200 trials were followed by 100 trials in Room B with no stimuli, no CS or US. This represented bringing the animal back to its cage that night. After that, there were 50 trials of just the Light, but back in Room A, to test the dopaminergic activity of the model in response to the Light. The results are displayed in Figure 14. Notice that the dopaminergic activity in response to the Light was initially strong in trials 301-350, when the simulated animal was returned to the training environment. The dopaminergic activity

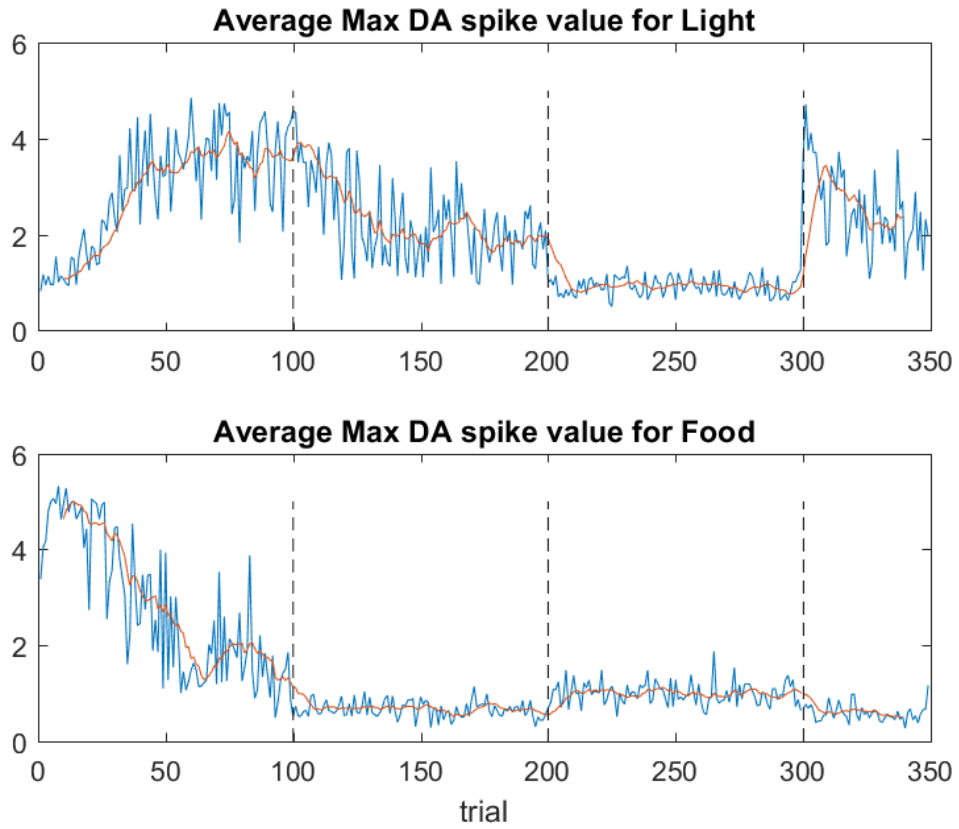


Figure 14: The dopaminergic activity of the augmented model during a spontaneous recovery experimental setup used to test the model's ability to spontaneously recover previously extinguished associations. The first 100 trials are learning trials in Room A. Trials 101-200 were extinction trials in Room A. Trials 201-300 were trials in Room B without any stimulus: Light, Bell, nor Food were presented. Trials 301-350 were trials back in Room A with the presentation of just the Light (i.e., further extinction trials). Notice that trials 300-350 show similar dopaminergic activity for the Light as trials 101-150, the trials when extinction originally occurred. This means that when the model was brought back to the experimental setup, it spontaneously recovered its previously extinguished association.

for the Light was about as high as it was during the first 50 trials of extinction. This shows that the model spontaneously recovered the previously extinguished association.

The way in which the model accomplishes spontaneous recovery involves the way that it models moving from one environmental context to another. In Redish et al. (2007), when the model was brought into a new context, the time-since-last-reward context cue (here, reward-recency) was reset. This can be justified by one of two reasons: The animal has been brought back to its home cage where there is food, or the animal possesses an "optimistic critic" that initially predicts the presence of food in new environments. Because this reward-recency cue is reset, when the simulated animal is brought back to Room A, the hippocampal state is similar to the state it was in during the acquisition of the association. During initial association learning, the reward-recency cue is high, and during the initial few trials of returning to Room A the reward-recency cue is also high. Because conjunctions of other cues and reward-recency drop when reward - recency drops, the hippocampal activity relies heavily on reward-recency. Thus, as in reinstatement, the augmented model explains spontaneous recovery as arising from a

return to a hippocampal state representation similar to that present during association learning trial and unlike that present during extinction trials. The return to the previous state then drives DA-STDP strengthened synapses to produce INT-H activity which, in turns, results in a recovered DA response.

Blocking

Blocking is the phenomenon where previous knowledge of an association from one conditioned stimulus limits the learning of an association from a second conditioned stimulus when both are presented together prior to reward. To simulate blocking with the augmented model, we made use of both the Light and the Bell sensory inputs. The model was trained for blocks of 50 trials, instead of the previously used 100 trials, to avoid over-learning effects. The model was presented with 50 learning trials involving the Light (CS1) and Food (US), followed by 50 learning trials in which the Light (CS1) and the Bell (CS2) were presented at the same time, shortly followed by the US. After this sequence of learning trials, the model was tested by additional association trials, focusing on different CS stimulus items during different simulation runs. During some runs, the testing trials involved presenting CS1 followed by US, and others involved presenting CS2 followed by US. We focused specifically on the 50 testing trials (trials 100-150), looking for indications of associations between each individual CS and reward. Including the US in the testing trials allowed us to see if, in the context of each CS, the US generated a DA response. The lack of a DA response at the time of the US would suggest that the US was successfully predicted by the CS. If the DA response was high when the US was presented, then the model failed to inhibit the dopamine response to the US, which means it did not predict the US, and, thus, did not learn that that particular CS predicts the US. (The model context consistently indicated that the simulated animal was in Room A during the entire experiment.)

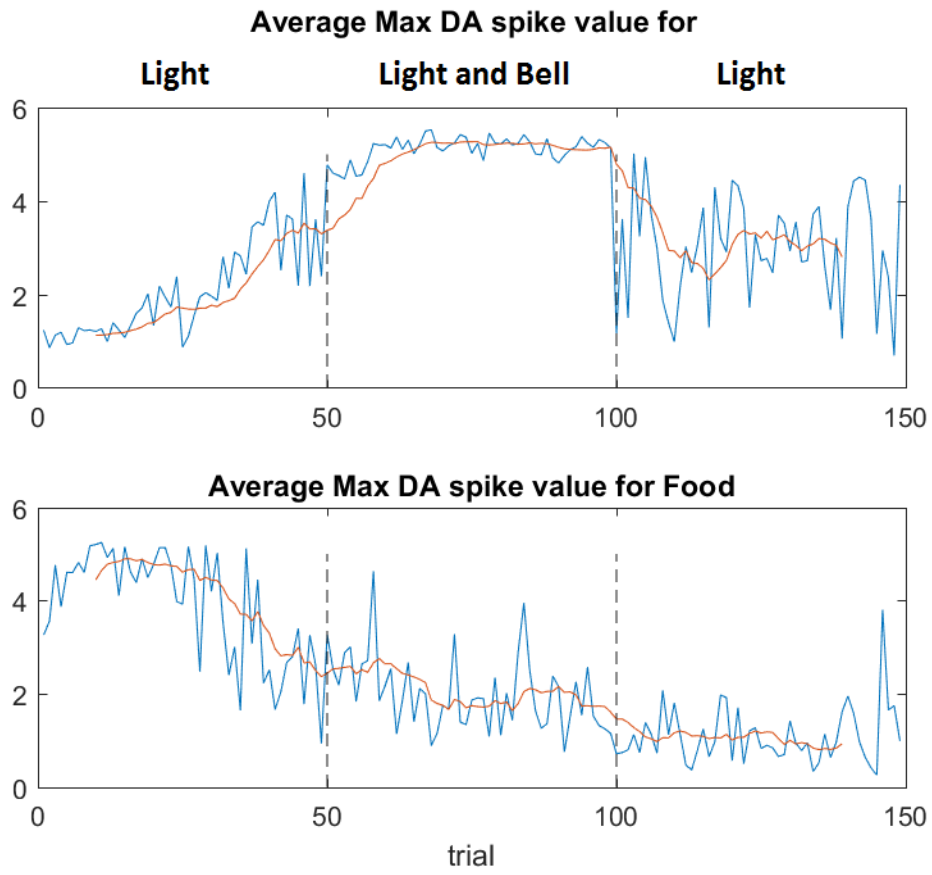


Figure 15: *Testing Blocking for CS1 (Light)*. The dopaminergic activity of the model after learning CS1-US in the first 50 trials. During trials 51-100, both CS1 (Light) and CS2 (Bell) were presented simultaneously directly before the US. Finally, there were 50 test trials incorporating only CS1 and US. The dopamine activity in response to the Food is particularly salient, as it indicates whether or not the model predicted the US given CS1 (Light). Because there is little dopamine activity during the last third of the experiment, the model predicted that the US would arrive after CS1 and, thus, inhibited the dopamine response to the US. This indicates that the model retained the association between CS1 (Light) and the US.

Figures 15 and 16 show the results of the blocking experiment. The lower graph in each figure displays the DA response at the time of US presentation. The first 50 trials were CS1-US learning trials. Trials 51-100 involved the simultaneous presentation of CS1 and CS2 followed by the US. The last 50 trials in Figure 15 are CS1-US trials, and the last 50 trials of Figure 16 are CS2-US trials.

Figure 15 shows that the model retained its knowledge that CS1 predicts the US. When presented with CS1-US during the last 50 trials, the dopamine activity remained low. Remember that, when Food occurred, it intrinsically sent strong activity to the dopamine neuron group through its very strong synaptic weights. This means that the observed low dopamine activity when the US was being presented during the last 50 trials must have been the result of inhibition from the upper pathway, produced as a result of the CS1 representation in PFC. In this way, Figure 15 shows that CS1 predicted the US via the upper pathway.

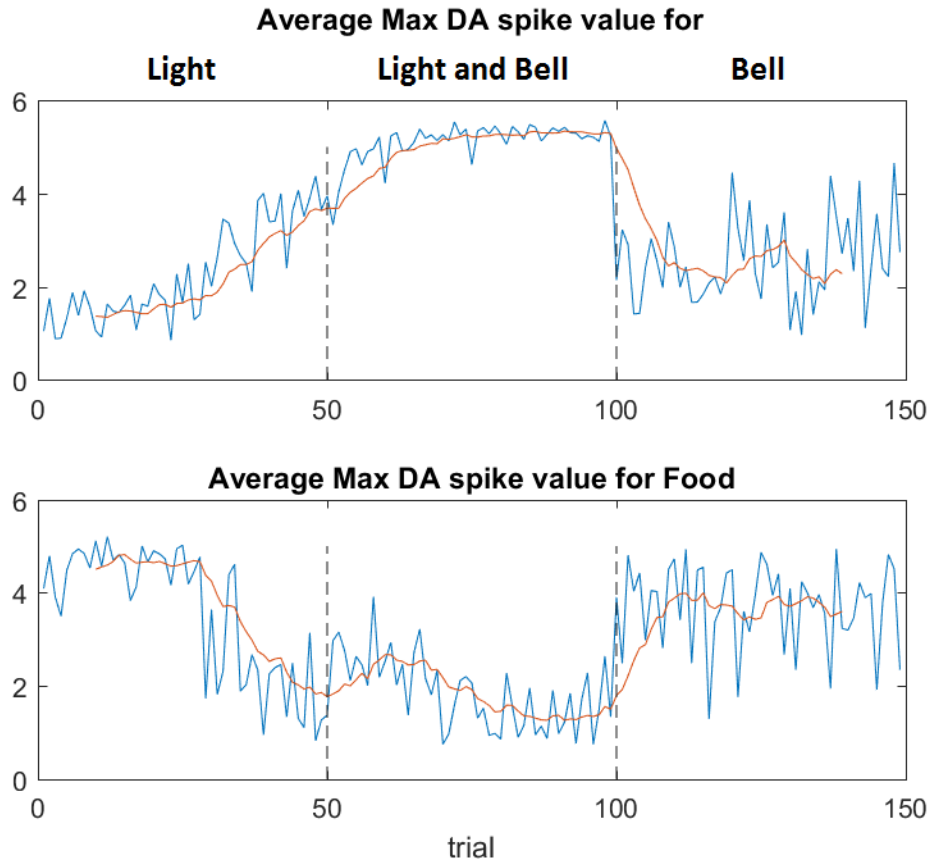


Figure 16: *Testing Blocking for CS2 (Bell)*. The dopaminergic activity of the model after learning CS1-US in the first 50 trials. During trials 51-100, both CS1 (Light)and CS2 (Bell) were presented simultaneously directly before the US. Finally, there were 50 test trials incorporating only CS2 (Bell) and US. The dopamine activity to the Food is particularly salient, as it indicates whether or not the model predicted the US given CS2 (Bell), which was first introduced together with CS1 at trial 50. The dopamine activity in response to the Food is particularly salient, as it indicates whether or not the model predicted the US given CS2. The high level of dopamine activity during the last third of the experiment indicates that the model did *not* inhibit the dopamine response for the US after the presentation of CS2. Thus, the model did *not* predict that the US would arrive after *CS2* (Bell) was presented. No strong association between CS2 (Bell) and the US was acquired during trials 51-100.

In contrast, Figure 16 displays high dopamine activity during the last 50 trials. In this case, CS2-US pairings were presented during these trials. The high dopamine activity during the last 50 trials is similar to that present during the initial trials of learning. This suggests that the US generated dopamine response was not inhibited by the upper pathway, driven by the CS2 representation in PFC. In short, the model did not learn that CS2 predicts the US. This shows that the learned CS1-US association blocks learning of a CS2-US association during trials 51-100, when CS1 and CS2 were presented simultaneously (Kamin, 1969).

The augmented model exhibits blocking by suppressing the dopamine response during the conjoined CS1-CS2 trials, leaving little DA to drive DA-STDP in synapses from the CS2 related neurons. After 50 trials of learning that CS1 predicts the US, the upper pathway in the model comes to inhibit the dopamine burst at the onset of the US.

When CS2 is subsequently added to the trials, there is no DA signal to strengthen the weights from neurons driven by CS2. In this way, further learning does not occur. CS2 is blocked by CS1.

It is interesting to note that CS2 produced a much stronger DA response during the last 50 trials than would be expected if no association from CS2 had been learned. The dopamine levels during these trials are much higher than they are for CS1 at the very beginning of training, for example, suggesting that CS2 produced a stronger response than a neutral baseline. This shows that the model did learn a little bit about the CS2-US pairing during the 100 trials of simultaneous CS1-CS2 presentation. However, the upper pathway still did not learn to inhibit the dopamine burst upon the onset of the Food when presented with CS2 alone. This potential weakness of the model is left as a topic for future investigation, but there is one possible explanation that is worth mentioning. It is possible that the dopamine response generated by CS1, after initial learning, provided support for DA-STDP strengthening of synapses from CS2 related neurons. The representation in HIPP of the simultaneous presentation of CS1 and CS2 involved a sudden burst of firing of CS2 associated neurons, potentially producing large eligibility traces for many CS2 related synapses. The DA response driven by CS1 might have coincided with these large eligibility traces, strengthening the association from CS2. In contrast, the representation of CS2 in PFC is a delayed and temporally distributed pattern of spikes, and the CA1-driven DA response is not appropriately timed to strengthen weights from these spikes to STR so as to inhibit DA at the time of the US. The DA response from CA1 is too early to allow the upper pathway to learn to inhibit the US DA response, but it might contribute to the strengthening of CA2 related weights in the lower pathway. This is one hypothesis that should be investigated in the future.

Lesioning the Hippocampus

One of the primary motivations for augmenting the original Chorley and Seth (2011) model with a hippocampus was the existence of empirical observations that hippocampal lesions in animals impair extinction, a capability lacking in the original model, but do not impair the learning of associations, a capability that is robust in the original model. The way in which a hippocampus was implemented in the augmented model makes fundamental changes to how the state of the animal is represented, however, introducing reasons to wonder if the hippocampus implementation left the independent association learning capabilities of the Chorley and Seth (2011) model intact. Thus, an important test of the augmented model involves its ability to continue to learn associations when the hippocampus component is “lesioned” in simulation. If the behavior of the augmented model with a hippocampal “lesion” matches the behavior of the model described in Chorley and Seth (2011), then there is evidence that the added hippocampus does not remove the learning capabilities of the original model.

In order to model a lesion of the hippocampus in the augmented model, the injected current noise driving the sensitivity and base firing rate of the hippocampus neurons (see Appendix A) was reduced to zero. Also, the activity of the hippocampus neuron group was not affected by the presence of any stimulus. This effectively removed all spiking activity in the HIPP neurons. The lesioned model was simulated, using the previously employed experimental design of association learning trials, followed by extinction trials, followed by a test for reinstatement.

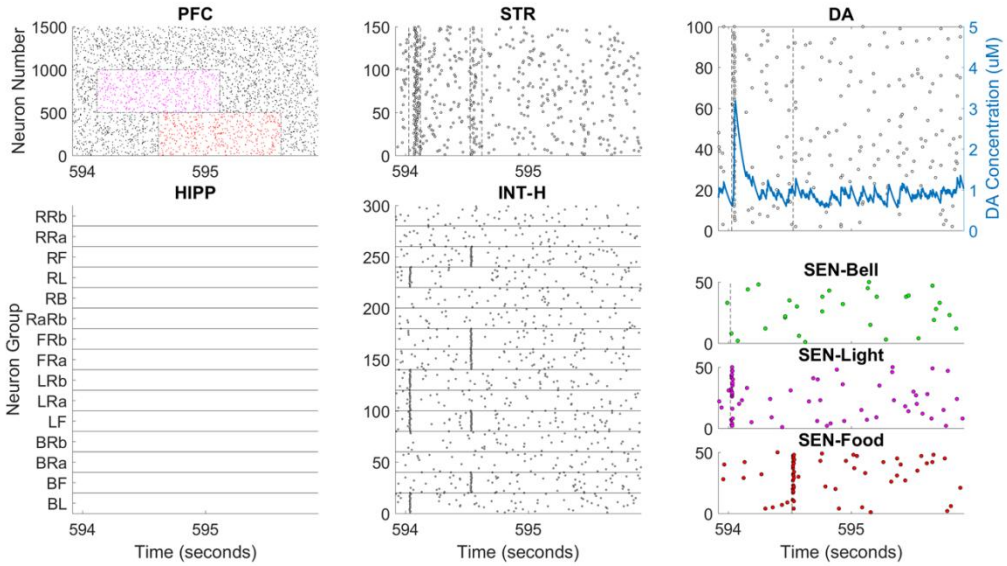


Figure 17: One trial of the model after almost 200 trials of CS-US pairings. The first thing to notice is that the hippocampal neuron group displayed no activity. It had no background activity, and it was not active when either the Light (CS) or the Food (US) arrived. Also notice that the model learned to produce a burst of dopamine activity for the Light, and it inhibited the dopamine burst for the Food. This shows that the augmented model can still learn associations with the hippocampus lesioned.

Figure 19 shows a snapshot of a single trial after the learning of the association between the Light (CS) and the Food (US) had occurred. After learning, the model characteristically produced a dopamine burst for the CS and inhibited a dopamine burst for the Food. This means that the augmented model can still learn associations when the hippocampus is lesioned.

Figure 20 shows the dopamine activity throughout the experiment, showing that the dopamine activity for the Light increased over time during learning trials (trials 1-200) and did *not* drop during extinction trials (trials 201-300). Also, the dopamine activity for the Food dropped to baseline during learning trials, and it didn't change after that point.

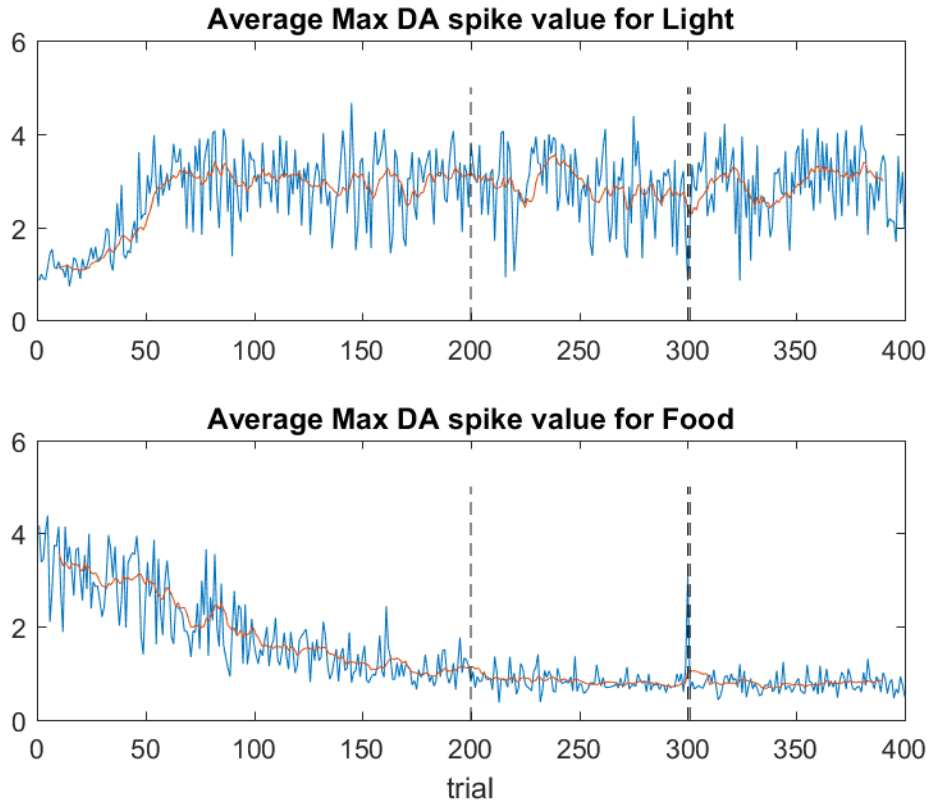


Figure 18: These graphs show the dopamine activity throughout the entire experiment, for both the Light and the Food. Trials 1-200 were CS-US pairing association trials, trials 201-300 were extinction trials presenting only the CS, trial 301 was a reinstatement trial presenting only the Food (US), and trials 302-400 were further extinction trials. This is the same experimental structure as the extinction and reinstatement experiment reported early in this chapter. Notice that the dopamine activity for the Light rose during early trials, and the dopamine activity for the Food dropped off. This demonstrates that the model learned that the Light predicted Food. Also notice that during extinction trials 201-300 the dopamine levels for the Light remained high, showing that the augmented model with the hippocampus lesioned did not extinguish a learned association.

With the hippocampus lesioned, it cannot provide contextual information, particularly about the recency of reward, in order to encode a state change during extinction trials. Because the lower pathway of the model does not represent the increasing time since last reward during extinction trials, its state remains relatively constant during those trials, and the lower pathway continues to excite dopamine when the Light appears. In contrast, in the augmented model with a working hippocampus, the lower pathway of the model uses the recency of reward to change the state representation over extinction trials in the HIPP and INT-H neuron groups, allowing extinction of the association to occur.

Analysis of Hippocampal Activity

The augmented model was partially motivated by the Redish et al. (2007) account of extinction, in which a change in state representation plays a central role. The sparse conjunctive code used in the hippocampus component of the augmented model was intended to produce a similar pattern of state change, with the state representation

changing little during association trials but changing substantially during extinction trials. While this was the intention, it is worthwhile to examine the patterns of activity in the modeled hippocampus more carefully in order to demonstrate the role of changing representations in extinction. Here, an analysis of modeled hippocampus activity patterns is provided with the goal of illuminating how the changes in activity in the hippocampus provide the necessary tools for the model to learn and extinguish associations.

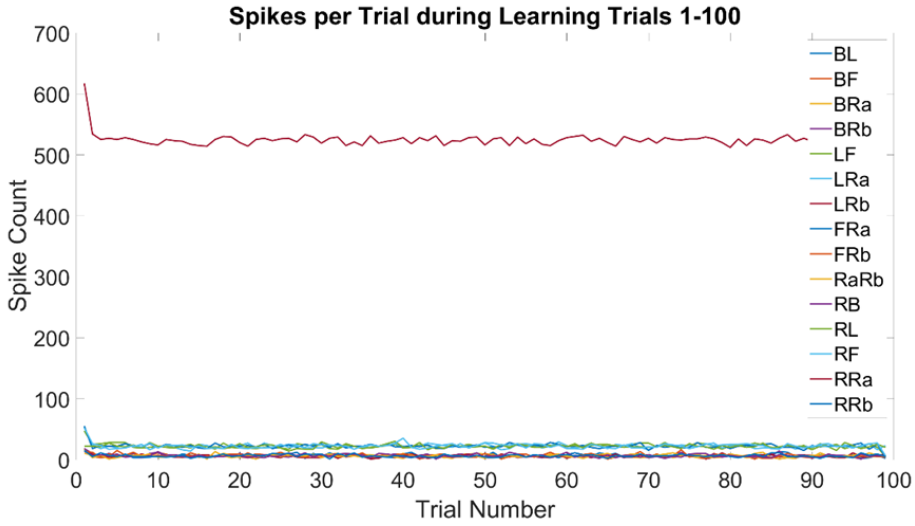
The main idea is that the hippocampal activity doesn't change very much during learning an association, largely because the reward-recency cue remains high. The food reward is provided during every trial. This allows the model to associate the active neurons in this relatively constant state representation with reward, using DA-STDP. During extinction, however, the reward-recency cue drops, since food is not delivered during any of these trials. The dropping reward-recency cue affects the activity of each subsection of the hippocampus neuron group that encodes a conjunction between reward-recency and some other cue. Because a number of the hippocampus subsections substantially decrease in activity during extinction trials, the aggregate pattern of hippocampal activity becomes different. Thus, the internal state representation becomes different.

These state representation changes can be examined by recording the number of spikes generated in each conjunctive bin in HIPP over the course of learning and extinction trials. Such an analysis would show which neurons are generating many spikes, providing opportunities to strengthen synapses to INT-H, as well as which neurons that might already drive INT-H strongly fade in firing rate over extinction trials.

Data of this kind is displayed in Figures 21 and 22. These two figures are worth comparing. During learning trials, the hippocampus provided one context representation. During extinction trials, the activity in the hippocampus changed to provide a different (shifting) context. The subsections in the hippocampus for conjunctions LF, FRa, RL, and RF have higher than baseline activity levels during learning, but during extinction, they all drop down to baseline activity levels. This difference in hippocampal activity is what allows the model to extinguish learned associations. The DA response is driven by hippocampus neurons with synapses strengthened during learning trials, but these same neurons become silent over the course of extinction trials, removing support for a DA burst.

It is important to note that this process is also what allows the model to spontaneously recover and reinstate previously extinguished associations. In this model, extinction arises simply from a change the internal representation of the current context. If the model is brought back to a similar internal state representation as when it was learning the association (such as what happens in both reinstatement and spontaneous recovery), the model will, once again, activate neurons that have had their synapses strengthened during learning, and thereby show the previously extinguished association, once again.

a)



b)

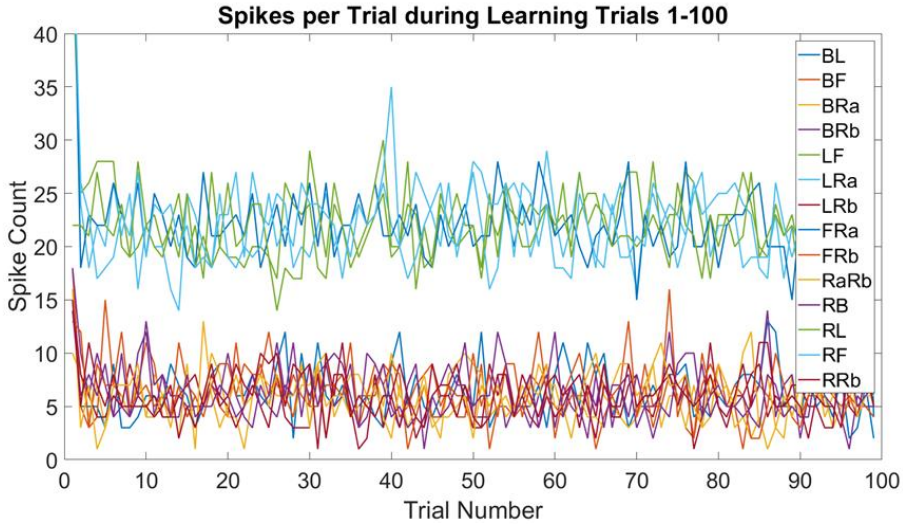
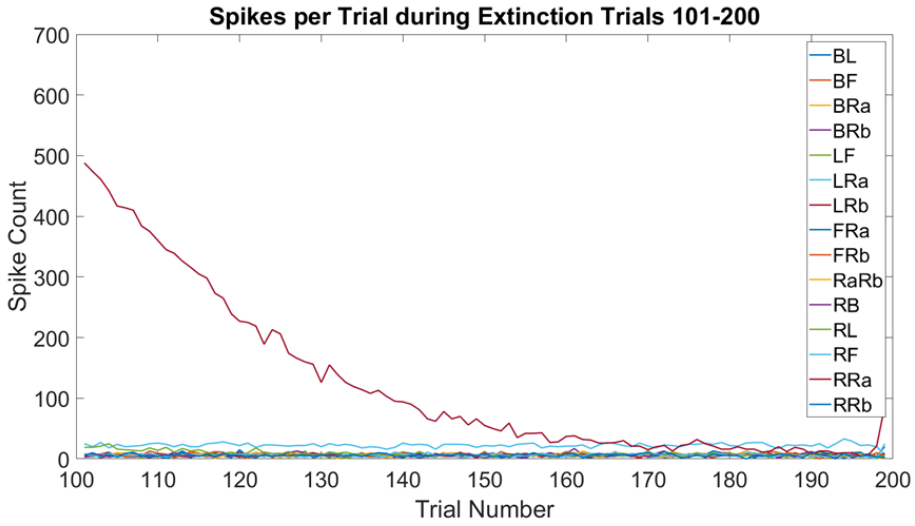


Figure 19: Spiking activity of the 15 subsections in the hippocampus neuron group during association trials. **a)** is all 15 subsections of the hippocampus, while **b)** removes outlying subsection RRa to more easily see the activity in the other subsections. This is the hippocampal activity during learning the association. Notice that we can group the activity of the subsections into three different activity levels – 1) very high activity for RRa, because this hippocampus neuron subgroup is a conjunction of two cues, and the cue Ra (Room A) is constantly present, and the cue R (reward-recency) is high for the entire learning period. 2) medium activity for subsections LF, LRa, FRa, RL, and RF. These are the subsections that receive small bursts of activity at the onset of the conjunctions of these two cues. During learning, the onset of Food while the Light is on (LF), the onset of the Light while in Room A (LRa), the onset of Food while in Room A (FRa), the onset of the Light and the current level of Reward-Recency (RL), and the onset of the Food and the current level of Reward-Recency (RF). 3) very low background noise activity levels because the conjunction of the two context cues never appears during learning. What is important to take away from this graph is that the activities of these hippocampal subsections do not change very much over time, and that the subsections that have very high and medium activity levels during learning are different during learning trials (Figure 21) than during extinction trials (Figure 22).

a)



b)

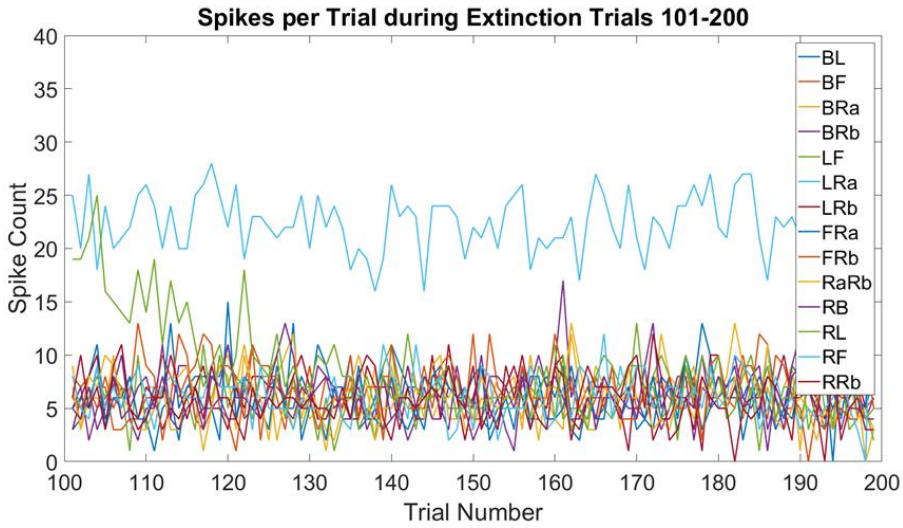


Figure 20: Spiking activity of the 15 subsections in the hippocampus neuron group during extinction trials. **a)** is all 15 subsections of the hippocampus, while **b)** removes outlying subsection RRa to more easily see the activity in the other subsections. In comparing Figure 21 with Figure 22, we can see two changes, 1) conjunctions RRa and RL drop over time during extinction, and 2) the other subsections that had medium activity during learning now have low activity during extinction (LF, RL, FRa, and RF). Because the hippocampal activity during learning is different than during extinction, the model essentially moves into a new state during extinction trials, and, thus, produces different expectations of reward. Contributions from the differential neuron subgroups contribute to a DA response during learning trials, but they fall away during extinction, extinguishing DA activity.

CHAPTER 6: DISCUSSION

This dissertation discusses a variety of topics, including the theories of learning, the phenomena seen in classical conditioning studies, and the effects of lesioning the basal ganglia and hippocampus on learning. The document also offers a short history of related computational modeling efforts, from temporal difference learning to the modeling of it biologically. The first computational model of spiking neurons that replicates dopamine activity during association is reviewed, along with some of the computational models that led up to it. Some published research on the importance of the hippocampus for extinction and spontaneous recovery (Redish et al., 2007) is explained, and I show how combining this understanding of the role of the hippocampus with the spiking neural network model of Chorley and Seth (2011) produces a new augmented spiking model that can both learn associations and extinguish them.

By augmenting the Chorley and Seth (2011) model with a simple hippocampus, I have produced a spiking model without abstract variables that captures learning and extinction. This augmented model exhibits extinction and reinstatement, phenomena the original model of Chorley and Seth failed to show, and there are other learning phenomena that this model is able to capture. Specifically, this dissertation project offers the following original contributions:

- a replication of the model of Chorley & Seth
- the presentation of a new model, based on Chorley & Seth but incorporating a simple spiking neuron model of the hippocampus
- a demonstration that the hippocampal model continues to exhibit association learning
- a demonstration that the hippocampal model exhibits extinction
- a demonstration that the hippocampal model exhibits reinstatement
- a demonstration that the hippocampal model exhibits spontaneous recovery
- a demonstration that the hippocampal model exhibits blocking
- a demonstration that the hippocampal model mirrors the results of lesion studies when the hippocampus component of the model is removed
- an analysis of how the hippocampus model represents context

There are a variety of other classical conditioning phenomena that could be explored in the future using the augmented model, given time.

For example, in *overshadowing*, the salience of one stimulus is stronger than that of another. Both stimuli are presented at the same time right before the US, in order to build associations, but the association is only strongly established for the more salient stimulus. This could be simulated by increasing the amount of injected current in the sensory neuron groups for the more salient stimulus, and similarly reducing the injected current in the less salient stimulus. It would also need respective increases and decreases in the number of neurons in the PFC allocated to each stimuli, or changes in the patterns of spikes produced by the two stimuli in PFC, as differences in the inhibition of the predicted stimuli is also necessary to capture this effect.

Reversal is an interesting phenomenon that could be studied using the augmented model. In reversal, initially two stimuli are used, with one being presented on each learning trial, but only one of the two predicts reward. When the *reversal* occurs, the two stimuli reverse in their ability to predict reward: the predictive stimulus now no longer predicts reward and the other begins to predict reward. Animals undergoing reversal can learn this change in predictability when it occurs. Reversal may be harder to model than extinction, because the reward-recency cue will remain approximately

constant both before and after the reversal of contingencies. Thus, even the hippocampally augmented model may not be able to tell the difference between contexts, and both stimuli may come to have a strong association with the reward.

Second-order conditioning has a similar problem, worthy of future exploration. In this case, after a CS1-US pairing is learned, a second CS2-CS1 pairing can be learned, and, as a result, a CS2-US prediction occurs. Attempting to model this phenomenon might introduce problems when attempting to learn the CS2-CS1 association. While, due to the learned CS1-US association, dopamine may be present during CS2-CS1 trials, supporting synaptic change, the reward-recency cue may introduce state representation problems, since it is high in the CS1-US situation but not in the CS2-CS1 situation. The differences in state representation might interfere with learning. Also, the CS1-US association might become extinguished during CS2-CS1 trials, since CS1 is being present without a following predicted reward. If CS1 stops producing a DA response, learning CS2-CS1 associations through DA-STDP will become impossible, in this model.

Renewal is a phenomenon that is similar to spontaneous recovery. Both phenomena involve learning an association and extinguishing it in one environmental context and then, later, testing for a response to the conditioned stimulus. In spontaneous recovery, the animal is returned to *the original context after a long delay*, typically spent in its home cage. In renewal, the animal is tested in *a new context after a short delay*. While these are similar phenomena, the model may have trouble generalizing a learned association from one context to a different context. Renewal tests for the generalization of a learned response to other new contexts, and thus may be harder to implement in this model.

Finally, gradual extinction is also a phenomenon worthy of exploration within the framework of this model. Gradual extinction has only been recently discovered (Gershman et al., 2013). This phenomenon arises when, after learning an association between a CS and a US, trials of association learning of CS followed by US are interspersed within increasing numbers of extinction trials containing just the CS. This type of extinction contrasts with traditional extinction, where the US is suddenly withheld, and the CS is presented alone many times until the animal no longer shows a response to the CS. Animals experiencing gradual extinction show much less spontaneous recovery and reinstatement than those given traditional extinction. This may be a harder phenomenon to implement because our hippocampus model implements the ideas of Redish et al. (2007) in a hard-coded way with a set number of conjunctive states, instead of a dynamically malleable INT-H and HIPP neuron group that can create new conjunctions on the fly. Such new state representations might be needed in order to capture the interplay between state change and association learning.

These last two phenomena suggest a more abstract characterization of some of the weaknesses of this model. The model exhibits difficulty in *generalizing what is learned*. The inability to *flexibly modify hippocampal state representations* is another weakness. Focusing on the second of these weaknesses first, in the hippocampally augmented model, learning is context-driven from the lower pathway, using the representation produced by the hippocampus. For example, the model would learn that when food present, more food will appear, but when food is absent, more food will not appear. This sort of association occurs because the major tool that the model uses to tell the difference between states is the reward-recency cue. When reward-recency drops, the hippocampus changes states, and the learning that occurred in the state when the reward-recency was high is made dormant, unable to produce a dopamine response. Instead, different hippocampal conjunctions become available. This poses a problem for

the hard-coded conjunctive codes that were used in the reported implementation of the model. Ideally, the hippocampus would offer a dynamic encoding of cue conjunctions that could change over time and be modified depending on the reward prediction error, as described in Redish et al. (2007). This dynamicity is required in order to capture most of the phenomena described above that were not explored in this dissertation.

The other major weakness of this model is its difficulty accounting for generalization. This problem might be addressed by a dynamic state-encoding in the hippocampus, as just described, but the weakness most likely stems from some additional properties of learning and extinction. One theory is that learning is a general purpose mechanism that learns something for all contexts, while extinction is a discriminatory mechanism that identifies the contexts in which the generalized learning is not applicable (Baum, 2012). This idea is an interesting one, but it is somewhat counter to the learning mechanisms of the augmented model. In the augmented model, the lower pathway learns associations specific to the context encoded by the hippocampus, while the upper pathway simply inhibits the predictable stimuli (irrespective of whether extinction has occurred or not). One approach to addressing this weakness might involve having the hippocampus representation affect the upper pathway of this model, allowing the lower pathway to learn generalizable associations, while the upper pathway inhibits the learned associations that have been extinguished in certain contexts.

Using a conjunctive code in the PFC was explored, but it has yet to bear fruit. The problem is that the hippocampus would have to learn to inhibit the dopamine spike from unpredictable stimuli like Light or Bell. If we implemented a hippocampus in the upper-pathway of the model instead of the lower-pathway and kept the lower-pathway the same from Chorley and Seth (2011), the lower-pathway would come to learn that conditioned stimuli that predict reward are just as rewarding as unconditioned stimuli (unconditioned stimuli are modeled as stimuli that have strong connections to the interneurons). After many trials of learning, conditioned stimuli come to have the same strong connections to the interneurons as the unconditioned stimuli, and thus are no different. The upper pathway inhibits the dopamine bursts from the strong connections from the unconditioned stimuli by used stimuli that occur directly before it. This means that a hippocampus affecting the upper pathway would have to inhibit the conditioned stimuli that occur randomly, with nothing to predict them, and the current way the upper pathway inhibits dopamine spikes is by using predictable stereotyped spikes in the PFC directly before the dopamine burst. With no predictable stererotyped spikes before a conditioned stimulus, the upper pathway cannot inhibit the dopamine bursts from the conditioned stimuli, and thus this modification of the upper pathway seems to lead to a dead end.

The research in this dissertation is a novel contribution above the work of Redish *et al.* (2007) in the sense that it offers a wholly spiking model and shows how neurons could be capable of producing these phenomena. Ultimately, this model challenges the dominant view that the method of temporal differences can explain the full range of learning phenomena by showing the computational importance of the hippocampus.

APPENDIX A: Izhikevich Neuron Model of Spiking and Synaptic Change

In 2003 Izhikevich developed a model of neural firing activity that accurately captured the myriad of neural bursting types that occur in neurons in the brain. They were derived from the complex but complete functions that fully modeled the different activities of squid neurons developed by Hodgkin and Huxley (1952). The complex multi-equation Hodgkin and Huxley model was distilled into two much simpler differential equations:

$$v' = 0.04v^2 + 5v + 140 - u + I$$

$$u' = a(bv - u)$$

As well as the spike resetting mechanism: if $v \geq 30$mV, then set v to c and set u to $u + d$. These two functions and spike-resetting mechanism require four parameters, and are able to model almost all of the different spiking behaviors observed in neuroscientific studies by modifying the four parameter values accordingly. Here, v is the membrane potential, and u represents an abstract membrane recovery ability of the neuron to recover from large amounts of input, while I is input current from other neurons' action potentials as well as any simulation-specific current injected into the neuron.

Each of the four parameters represents a different aspect of a neuron's response to input. The parameter a represents the speed of the recovery of the membrane potential back to baseline. The parameter b represents how sensitive the recovery variable is to changes in the membrane potential. The parameter c is the reset value of the membrane potential after a spike. And the last parameter, d, represents how the membrane recovery variable u is affected after a spike. Together these four parameters are all that's needed to model the spiking behaviors of many different types of neurons. A few types of spiking behavior showcased in the Izhikevich (2003) publication include: regular spiking, intrinsically bursting, chattering, fast spiking, low-threshold spiking, thalamo-cortical, and resonator. For more about how well this model can simulate the behavior of neurons, see Izhikevich (2004). For the neuron models in the papers cited in this dissertation, all used the parameters for modeling regular spiking neurons (a = 0.02, b = 0.2, c = -65, and d = 8).

In 2007 Izhikevich produced a model for synaptic change sensitive to the presence of dopamine. Izhikevich developed eligibility traces to model how dopamine affects synapse strength. This dopamine-modulated spike-timing-dependent plasticity (DA-STDP) that was used in this research, and which is heavily based on the work by Izhikevich (2007) is described in more detail here.

$$STDP(t_{post} - t_{pre}) = \begin{cases} if\ (t_{post} - t_{pre}) > 0: & 0.1 * e^{\frac{-(t_{post} - t_{pre})}{20}} \\ if\ (t_{post} - t_{pre}) \leq 0: & -0.15 * e^{\frac{(t_{post} - t_{pre})}{20}} \end{cases}$$

$$c = \left(c + STDP(t_{post} - t_{pre})\right) * \left(1 - (1/\tau_c)\right)$$

$$D = (D + 0.05k) * \left(1 - (1/\tau_d)\right)$$

$$s = s + 0.05 * \left(\frac{(cD^2)}{5}\right)$$

Here, t_{post} is the time of the post-synaptic action potential, and t_{pre} is the time of the pre-synaptic action potential. k is the number of DA spikes, c is eligibility trace, s is synaptic weight, and D is dopamine level. τ_c is the time constant for eligibility traces that determines the rate of decay, while τ_d is the time constant for that determines the rate of decay of the dopamine level. τ had multiple values. τ_d was set to 100, however τ_c had different values for the synapses from the striatum (STR) neuron group to the prefrontal cortex (PFC) neuron group than the other synapses (the synapses SEN to INT and HIPP to INT). τ_c for PFC to STR synapses was 200, while τ_c for all other synapses was 1000.

APPENDIX B: The Chorley and Seth model

In 2011 Chorley and Seth developed a model of dopamine activity that occurs during learning that was built upon the model of synaptic change from Izhikevich (2007). For an illustration of the model see Figure 3. The following equations, modified from Chorley and Seth (2011), were used to replicate the Chorley and Seth (2011) model:

$$I_j = \sum_i^N w_{ij} + \xi$$

$$\xi = U(-6.5,6.5)mA$$

$$L = U(1,10)ms$$

Input I for unit j is the sum over all incoming synapses from i to N of the weights w of those synapses, plus the noise parameter ξ. The noise ξ injected into each neuron at each millisecond is a random uniform value between -6.5mA and +6.5mA. This was enough to produce low level (1-5Hz) background noise in the firing activity of all neurons. The synapse delays L are all randomly selected from a uniform distribution between 1ms and 10ms.

$$\xi = \xi + 0.5mA \qquad for \qquad t_{stim} < t < (t_{stim} + 10ms)$$

When sensory units are stimulated for 10ms, the noise is increased by 0.5mA every ms, linearly increasing the noise from being selected from $U(-6.5,6.5)mA$ to $U(-1.5,11.5)mA$. This created a brief increase of their firing rate, without specifying a spike ordering.

$$b = 0.01D^2 + 0.19$$

And the calculation for dopamine-modulate post-synaptic facilitation (DA-PSF) effect in the STR neuron group was implemented by increasing the b parameter of the Izhikevich neuron model according to the squared dopamine level D^2 at each timestep. The b parameter is the sensitivity of the membrane potential recovery parameter (u) to fluctuations in the membrane potential (v).

For the PFC units, there were 1000 excitatory neurons that had about 1-5Hz of random spiking activity over time. These neurons represented a portion of the prefrontal cortex (PFC), and they were used to maintain information about presented stimuli over time (see Figure 6). When a stimulus was presented to the model, both the lower

pathway and the upper pathway were affected, but they were affected in different ways. The lower pathway briefly increases its firing rate in the neural group associated with the stimulus presented for 10ms, as previously mentioned. However, in the PFC portion of the upper pathway, the firing activity was not simply increased, but it was changed to be a stereotyped (i.e., identical whenever triggered) 1000ms time-locked pattern of neural spikes, starting when the stimulus was presented. Of the 1000 PFC neurons, there were 500 neurons that fired in a specific pattern whenever the conditioned stimulus appeared, and there were 500 other neurons that fired in a (different) specific pattern when the unconditioned stimulus appeared. Thus, every time a particular stimulus was presented, the same spatio-temporal pattern of neural spikes was generated across a population of neurons in PFC dedicated to that stimulus. The way this happens is that the input currents to these PFC neurons are pre-calculated as a matrix of injected current at each timestep for each neuron, specific to each stimulus. The currents are pulled from the same distribution of input currents of the background noise so that the overall firing rate of the PFC does not change over time. This results in the PFC presenting the same spatio-temporal pattern for the neurons that represent a stimulus every time that stimulus occurs.

For the connectivity, the connections from the PFC neuron group to the STR neuron group are feed-forward, and have a 10% connectivity to the STR neuron group; each PFC neuron is connected to 10% of the STR neurons. This balances the input activity into STR because PFC is 10x larger than STR. These connections are plastic, initially set to small but non-zero strengths. All other connections are feed-forward and have 100% connectivity (all-to-all). The neurons in the STR group are feed-forward and all-to-all connected to the DA neuron group, however their connections are *not* plastic. The synaptic strengths are not learnable and are fixed at 1mA. Each sensory neuron group is feed-forward and all-to-all connected to their corresponding interneuron group with plastic weights, initially set to the minimum weight strengths for the unconditioned stimulus, and the maximum weight strengths for the conditioned stimulus. Both interneuron group is feed-forward and all-to-all connected to the dopamine neuron group, but these connections are also *not* plastic. They are fixed at 0.6mA.

All-to-all means each neuron in the first group (for a stimulus) is connected to each neuron in the second group (the interneurons for the same stimulus). In this context, feed-forward means that these connections are only made from the sensory neurons to the interneurons, and not back from the interneurons to the sensory neurons. It is important to note that these connections are only from the sensory neurons of a stimulus to the interneurons for the same stimulus. Sensory neurons for one stimulus are not connected to the interneurons of the other stimulus.

Modifications to Chorley and Seth for the Replication

The following simplifications to the equations from Chorley and Seth (2011) were used to replicate the Chorley and Seth (2011) model:

$$\xi = U(-4.0,8.0)mA \qquad for \qquad t_{stim} < t < (t_{stim} + 15ms)$$

When sensory units are stimulated for 15ms, the noise is increased by 2.5mA so that the injected noise for each unit is taken from a uniform random distribution between -4mA and 8mA, $U(-4.0,8.0)mA$. In my simulations, I found that a uniform increase in injected noise of 2.5mA over 15ms was simpler to code and produced a similar result to linearly

increasing the input activation by 0.2mA each ms over 10ms, as described in Chorley and Seth (2011).

In this replication of the Chorley and Seth model, for the PFC units that are stimulated to represent different stimuli, the random seed used to calculate the injected random noise ξ is changed depending on the stimulus in order to provide stereotyped PFC activation specific to each stimulus. When the CS is presented, the unique random seed for the CS is used to calculate the injected random noise, while another unique random seed is used when the US occurs. After the 1000ms presentation of the stereotyped activation in the PFC of either the CS or US, the unique random seed for that stimulus is reset, so that the next time the CS or US occurs, the random noise injected into those neurons will be the same stereotyped noise. The random seed used to calculate background noise when the CS or US isn't presented is simply replaced with the unique random seed for the specific stimulus. This effectively means that the injected noise into PFC remains fairly steady, but specific neurons will get specific injected input when either a CS or US occur.

APPENDIX C: A Hippocampal Model of Dopamine Activity

In 2011 Chorley and Seth modeled the dopamine activity during learning an association. (See Appendix B) This appendix details the modifications done to introduce to it a hippocampal model of context. The fully modified model is illustrated in Figure 8.

The Hippocampus

The hippocampus added to the model has 15 subgroups of neurons, each subgroup corresponding to a conjunction of the 6 different context cues: Reward-Recency, Light, Bell, Food, Room A, and Room B. 6 choose 2 is 15. Each subsection has 20 neurons in it, and the injected noise ξ into all subsections increases by 2.5mA for 15ms on the onset of the conjunction of the two cues. For some subsections, we encoded this differently, so here we will review the spiking activity for each hippocampal subsection.

FRa, FRb, BRa, BRb LRa, LRb: When Food (or Light or Bell) is presented in Room A (or Room B), the injected noise into the conjunction of Food and Room A subsection of the hippocampus increases by 2.5mA for 15ms on the onset of the Food. This also goes for conjunctions between the Food, Light, Bell, and Room A, Room B.

LF, BF, BL: For conjunctions between the Light, Bell, and Food stimuli, the injected noise into the conjunction of the two in the hippocampus increases by 2.5mA for 15ms when their temporally-extended time-locked polychronous group (PNG) in the PFC overlap. This timing was chosen so that the hippocampus provided a short burst of activity, otherwise being active for the entire overlap of the two PNGs would provide too much activation to INT-H and thus to DA for 500ms. This creates both a small burst of activity, and allows for stimuli that aren't presented at the exact same time to provide information about when they overlap.

RF, RL, RB: For these conjunctions, the value reward-recency (or time-since-last-reward, see Redish et al. (2007)) was treated as a variable that ranged from 1 to 0: it was 1 when the reward-recency was reset, and that value decayed down to 0 over time. The injected activity of the subsections RF, RL, and RB increased by the product of the reward-recency variable and 2.5mA for 15ms on the onset of the related (Food, Light, Bell) context cue. This means that activity bursts in RF, RL, and RB could range from as high as an increase in 2.5mA in activity to no increase in activity.

RRa, RRb: These conjunctions were the conjunctions of the reward-recency and whether the model was in Room A or Room B. Instead of short bursts of activity in these subsections, they showed prolonged increases of activity over time. The injected noise into these subsections were also equal to the product of reward-recency and 2.5mA, however this lasted the entire time the model was in Room A. When the model entered Room B, injected noise into RRa returned to baseline (because Ra, if treated like a Boolean, was now 0), and the injected noise into RRb became the product of reward-recency and 2.5mA.

And finally, the RaRb conjunction was only provided for completeness, because in these experiments it is not possible to be in both Room A and Room B, and so this conjunction never changed from baseline firing rates.

INT-H

The neuron group INT-H simply took in the input from the three sensory neuron groups and from the hippocampus subsections and wasn't otherwise modified. What was controlled was the specific connectivity from those neuron subgroups to INT-H.

For the three sensory neuron groups, they were fully feed-forward connected to each hippocampal subsection whose conjunction contained that sensory cue. For the hippocampus neuron group, each conjunction was fully connected in a feed-forward connection to each corresponding subsection in the INT-H neuron group. E.g. the neurons in subsection RRa of the hippocampus were fully connected only to the neurons in the RRa subsection of INT-H.

Weights

The connections from the hippocampus neuron group to INT-H were all initially set to small non-zero values of 0.15 (weight strengths ranged from 0 to 10), except for all subsections that conjuncted with Food, in which case they were set to the maximum weight value of 10.

The connections from the sensory neuron groups to INT-H were all set to zero, except for connections from the Food sensory neuron group, in which case it was set to the weight maximum of 10.

Maximum Dopamine Burst Analysis

In order to create graphs of the maximum dopamine bursts over the entire experiment, we took the maximum dopamine level between 20ms and 70ms after the onset of the stimulus. The delay of 20ms was used because the start of dopaminergic bursting consistently began 20ms after the stimuli were presented. Note: with a delay range of all connections are 1-10ms, one might think with two connections (from HIPP/SEN to INT-H, and from INT-H to DA) it would on average only take 11ms for activity to reach DA (5.5ms + 5.5ms, the average of 1-10). It's important to remember that neurons need multiple spikes to fire; a single spike is not enough to cause it to fire. So, given the weight values and the spikes that occur, it takes around 20ms for spiking activity to travel from the HIPP or SEN neuron groups to the DA neuron group.

APPENDIX D: Implementation

The replication of Chorley and Seth in this dissertation as well as the implementation of the hippocampal model was programmed in MATLAB version 7.7.0.471. This code made use of an implementation of the Izhikevich neuron produced by William Benjamin St.Clair and made available as part of his dissertation work. Code

for these simulations is available by contacting the author at jjrodny@msn.com or on GitHub at https://github.com/JJrodny/Spiking-Networks.

Bibliography

Albus, J. S. (1975). A new approach to manipulator control: The cerebellar model articulation controller (CMAC). *Journal of Dynamic Systems, Measurement, and Control*, (SEPTEMBER), 220–227. https://doi.org/10.1115/1.3426922

Barto, A. G. (1995). Adaptive Critics and the Basal Ganglia. *Models of Information Processing in the Basal Ganglia*, (1994), 215–232. Retrieved from http://books.google.com/books?hl=en&lr=&id=q6RThpQR_aIC&oi=fnd&pg=PA215&dq=Adaptive+Critics+and+the+Basal+Ganglia&ots=zQVwZfEm1n&sig=BaJIPv0J-_MFCytgU6tPux1ljgs

Barto, A. G., Sutton, R. S., & Anderson, C. W. (1983). Neuronlike adaptive elements that can solve difficult learning control problems. *IEEE Transactions on Systems, Man, and Cybernetics*, *SMC-13*(5), 834–846. https://doi.org/10.1109/TSMC.1983.6313077

Baum, W. M. (2012). Extinction as discrimination: The molar view. *Behavioural Processes*, *90*(1), 101–110. https://doi.org/10.1016/j.beproc.2012.02.011

Boyan, J. A. (1992). *Modular neural networks for learning context-dependent game strategies*. Retrieved from http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Modular+Neural+Networks+for+Learning+Context-Dependent+Game+Strategies+by+by#0

Boyan, J. A., & Moore, A. W. (1995). Generalization in reinforcement learning: Safely approximating the value function. *Advances in Neural Information Processing …*. Retrieved from http://books.google.com/books?hl=en&lr=&id=M9WuI6tiqRcC&oi=fnd&pg=PA369&dq=Generalization+in+Reinforcement+Learning:+Safely+Approximating+the+Value+Function&ots=Gu4tfY5zWz&sig=7gqqhMotBya1W0iK8WOt8Od5n6o

Bozarth, M. A., & Wise, R. A. (1981). Intracranial self-administration of morphine into the ventral tegmental area in rats. *Life Sciences*, *28*(5), 551–555. https://doi.org/10.1016/0024-3205(81)90148-X

Castro, D., Volkinshtein, D., & Meir, R. (2009). Temporal difference based actor critic learning-convergence and neural implementation. *Advances in Neural Information …*, 2–9. Retrieved from http://papers.nips.cc/paper/3517-temporal-difference-based-actor-critic-learning-convergence-and-neural-implementation

Chorley, P., & Seth, A. K. (2011). Dopamine-signaled reward predictions generated by competitive excitation and inhibition in a spiking neural network model. *Frontiers in Computational Neuroscience*, *5*(May), 21. https://doi.org/10.3389/fncom.2011.00021

Collingridge, G. L., & Bliss, T. V. P. (1987). NMDA receptors - their role in long-term potentiation. *Trends in Neurosciences*, *10*(7), 288–293. https://doi.org/10.1016/0166-2236(87)90175-5

Corbett, D., & Wise, R. A. (1980). Intracranial Self-Stimulation in Relation to the Ascending dopaminergic Systems of the Midbrain: A Moveable Electrode Mapping Study. *Brain Research*, *8*(185), 1–15.

Corcoran, K. a, & Maren, S. (2001). Hippocampal inactivation disrupts contextual retrieval of fear memory after extinction. *The Journal of Neuroscience*, *21*(5), 1720–1726. https://doi.org/21/5/1720 [pii]

Dayan, P. (1992). The convergence of TD(lambda) for general lambda. *Machine Learning*, *8*, 341–362. https://doi.org/10.1007/BF00992701

Florian, R. (2005). A reinforcement learning algorithm for spiking neural networks.

Symbolic and Numeric Algorithms for Scientific …. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1595864

Florian, R. (2007). Reinforcement learning through modulation of spike-timing-dependent synaptic plasticity. *Neural Computation*, *19*(6), 1468–502. https://doi.org/10.1162/neco.2007.19.6.1468

Frank, M. J. (2005). Dynamic Dopamine Modulation in the Basal Ganglia: A Neurocomputational Account of Cognitive Deficits in Medicated and Non-medicated Parkinsonism. *Journal of Cognitive Neuroscience*, *17*, 51–72.

Froemke, R. C., & Dan, Y. (2002). Spike-timing-dependent synaptic modification induced by natural spike trains. *Nature*, *416*(6879), 433–438. https://doi.org/10.1038/416433a

Gerrig, R., & Zimbardo, P. (2010). *Psychology and Life* (19th ed.). Boston: Allyn & Bacon. Retrieved from http://www.apa.org/research/action/glossary.aspx

Gershman, S. J., Jones, C. E., Norman, K. a, Monfils, M.-H., & Niv, Y. (2013). Gradual extinction prevents the return of fear: implications for the discovery of state. *Frontiers in Behavioral Neuroscience*, *7*(November), 164. https://doi.org/10.3389/fnbeh.2013.00164

Hebb, D. O. (1949). *The Organization of Behavior*. *New York: Wiley* (Vol. 911).

Hodgkin, A., & Huxley, A. (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. *The Journal of Physiology*, (117), 500–544. Retrieved from http://www.fenomec.unam.mx/pablo/parciales/HH.pdf%5Cnpapers3://publication/uuid/34C15183-A87D-4329-9954-089C693F8BFD

Holland, P. C., & Rescorla, R. a. (1975). Second-order conditioning with food unconditioned stimulus. *J Comp Physiol Psychol*, *88*(1), 459–467. Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=1120816

Houk, J. C., Bastianen, C., Fansler, D., Fishbach, a, Fraser, D., Reber, P. J., … Simo, L. S. (2007). Action selection and refinement in subcortical loops through basal ganglia and cerebellum. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *362*(April), 1573–1583. https://doi.org/10.1098/rstb.2007.2063

Hull, C. L. (1943). Principles of Behavior: An Introduction to Behavior Theory. In *The Journal of Abnormal and Social Psychology* (Vol. 39, pp. 377–380). https://doi.org/10.1037/h0051597

Izhikevich, E. M. (2003). Simple model of spiking neurons. *IEEE Transactions on Neural Networks / a Publication of the IEEE Neural Networks Council*, *14*(6), 1569–72. https://doi.org/10.1109/TNN.2003.820440

Izhikevich, E. M. (2004). Which model to use for cortical spiking neurons? *IEEE Transactions on Neural Networks / a Publication of the IEEE Neural Networks Council*, *15*(5), 1063–70. https://doi.org/10.1109/TNN.2004.832719

Izhikevich, E. M. (2006). Polychronization: computation with spikes. *Neural Computation*, *18*(2), 245–82. https://doi.org/10.1162/089976606775093882

Izhikevich, E. M. (2007). Solving the distal reward problem through linkage of STDP and dopamine signaling. *Cerebral Cortex (New York, N.Y. : 1991)*, *17*(10), 2443–52. https://doi.org/10.1093/cercor/bhl152

Izhikevich, E. M., & Hoppensteadt, F. (2009). Polychronous wavefront computations. *International Journal of …*, *19*(5), 1733–1739. Retrieved from http://www.worldscientific.com/doi/abs/10.1142/S0218127409023809

Jarrard, L. E., & Davidson, T. L. (1991). On the Hippocampus and Learned Conditional Responding : Effects of Aspiration versus Ibotenate Lesions, *1*(1), 107–117.

Kamin, L. J. (1969). Predictability, surprise, attention, and conditioning. *Punishment and Aversive Behavior*, *279296*, 279–298.

Kimble, D. P., & Kimble, R. J. (1970). The effect of hippocampal lesions on extinction and “hypothesis” behavior in rats. *Physiology & Behavior*, *5*(7), 735–8. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/5522489

Kravitz, A. V, Freeze, B. S., Parker, P. R. L., Kay, K., Thwin, M. T., Deisseroth, K., & Kreitzer, A. C. (2010). Regulation of parkinsonian motor behaviours by optogenetic control of basal ganglia circuitry. *Nature*, *466*(7306), 622–6. https://doi.org/10.1038/nature09159

Liu, X., Ramirez, S., Pang, P. T., Puryear, C. B., Govindarajan, A., Deisseroth, K., & Tonegawa, S. (2012). Optogenetic stimulation of a hippocampal engram activates fear memory recall. *Nature*, *484*(7394), 381–5. https://doi.org/10.1038/nature11028

Ljungberg, T., Apicella, P., & Schultz, W. (1991). Responses of monkey midbrain dopamine neurons during delayed alternation performance. *Brain Research*, *567*(2), 337–341. https://doi.org/10.1016/0006-8993(91)90816-E

Ljungberg, T., Apicella, P., & Schultz, W. (1992). Responses of monkey dopamine neurons during learning of behavioral reactions. *Journal of Neurophysiology*, *67*(1), 145–163. https://doi.org/10.1016/j.tins.2007.03.003

Maguire, E. A., Gadian, D. G., Johnsrude, I. S., Good, C. D., Ashburner, J., Frackowiak, R. S., & Frith, C. D. (2000). Navigation-related structural change in the hippocampi of taxi drivers. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(8), 4398–4403. https://doi.org/10.1073/pnas.070039597

Martinez, R., & Paugam-Moisy, H. (2009). Algorithms for structural and dynamical polychronous groups detection To cite this version : Algorithms for structural and dynamical polychronous groups detection. In *ICANNN’2009, International Conference on Artifical Neural Networks* (Vol. 5769, pp. 75–84).

McClelland, J. L., McNaughton, B. L., & O’Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, *102*(3), 419–457. https://doi.org/10.1037/0033-295X.102.3.419

Minsky, M. (1961). Steps toward Artificial Intelligence. *Proceedings of the IRE*, *49*(1), 8–30. https://doi.org/10.1109/JRPROC.1961.287775

Montague, P. R., Dayan, P., & Sejnowski, T. J. (1996). A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *16*(5), 1936–1947. https://doi.org/10.1.1.156.635

Morris, R. G. M. (1984). Developments of a water-maze procedure for studying spatial learning in the rat. *Journal of Neuroscience Methods*, *11*(1), 47–60. https://doi.org/10.1016/0165-0270(84)90007-4

Moser, E., Moser, M. B., & Andersen, P. (1993). Spatial learning impairment parallels the magnitude of dorsal hippocampal lesions, but is hardly present following ventral lesions. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *13*(9), 3916–25. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8366351

O’Reilly, R. C., & Frank, M. J. (2006). Making Working Memory Work: A Computational Model of Learning in the Prefrontal Cortex and Basal Ganglia. *Neural Computation*,

18(2), 283–328. https://doi.org/10.1162/089976606775093909
Packard, M., & Knowlton, B. (2002). Learning and Memory Functions of the Basal Ganglia. *Annual Review of Neuroscience*, *25*(1), 563–593. https://doi.org/10.1146/annurev.neuro.25.112701.142937
Packard, M., & McGaugh, J. (1996). Inactivation of hippocampus or caudate nucleus with lidocaine differentially affects expression of place and response learning. *Neurobiology of Learning and Memory*, *65*(6), 65–72. https://doi.org/10.1006/nlme.1996.0007
Packard, M., & Teather, L. (1998). Amygdala modulation of multiple memory systems: hippocampus and caudate-putamen. *Neurobiology of Learning and Memory*, *69*(2), 163–203. https://doi.org/10.1006/nlme.1997.3815
Papez, J. W. (1937). A proposed mechanism of emotion. *Archives of Neurology & Psychiatry*, *38*(4), 725–743.
Pavlov, I. P. (1928). Lectures on conditioned reflexes: Twenty-five years of objective study of the higher nervous activity (behaviour) of animals. In *Experimental psychology and psycho-pathology in animals* (pp. 47–60). https://doi.org/10.1037/11081-001
Pfister, J., & Gerstner, W. (2006). Triplets of Spikes in a Model of Spike Timing-Dependent Plasticity, *26*(38), 9673–9682. https://doi.org/10.1523/JNEUROSCI.1425-06.2006
Potjans, W., Morrison, A., & Diesmann, M. (2009). A spiking neural network model of an actor-critic learning agent. *Neural Computation*, *339*, 301–339. Retrieved from http://www.mitpressjournals.org/doi/abs/10.1162/neco.2008.08-07-593
Quirk, G. J., & Mueller, D. (2008). Neural mechanisms of extinction learning and retrieval. *Neuropsychopharmacology : Official Publication of the American College of Neuropsychopharmacology*, *33*(1), 56–72. https://doi.org/10.1038/sj.npp.1301555
Rao, R. P., & Sejnowski, T. J. (2001). Spike-timing-dependent Hebbian plasticity as temporal difference learning. *Neural Computation*, *13*(10), 2221–37. https://doi.org/10.1162/089976601750541787
Redgrave, P., Rodriguez, M., Smith, Y., Rodriguez-Oroz, M. C., Lehericy, S., Bergman, H., … Obeso, J. A. (2010). Goal-directed and habitual control in the basal ganglia: implications for Parkinson's disease. *Nature Reviews. Neuroscience*, *11*(11), 760–72. https://doi.org/10.1038/nrn2915
Redish, A. D., Jensen, S., Johnson, A., & Kurth-Nelson, Z. (2007). Reconciling reinforcement learning models with behavioral extinction and renewal: implications for addiction, relapse, and problem gambling. *Psychological Review*, *114*(3), 784–805. https://doi.org/10.1037/0033-295X.114.3.784
Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. *Classical Conditioning II Current Research and Theory*, *21*(6), 64–99. https://doi.org/10.1101/gr.110528.110
Roberts, P. D., Santiago, R. A., & Lafferriere, G. (2008). An implementation of reinforcement learning based on spike timing dependent plasticity. *Biological Cybernetics*, *99*(6), 517–23. https://doi.org/10.1007/s00422-008-0265-6
Romo, R., & Schultz, W. (1990). Dopamine Neurons of the Monkey Midbrain: Contingencies of Responses to Active Touch During Self-Initiated Arm Movements. *Journal of Neurophysiology*, *63*(3), 592–606. Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Ci

tation&list_uids=2329364
Rusu, C., & Florian, R. (2009). Exploring the link between temporal difference learning and spike-timing-dependent plasticity. *BMC Neuroscience*, *10*(Suppl 1), P201. https://doi.org/10.1186/1471-2202-10-S1-P201
Schultz, W. (1998). Predictive Reward Signal of Dopamine Neurons. *Journal of Neurophysiology*, *80*(1), 1–27. https://doi.org/10.1111/j.1471-4159.1989.tb09224.x
Schultz, W., & Romo, R. (1990). Dopamine neurons of the monkey midbrain: contingencies of responses to stimuli eliciting immediate behavioral reactions. *Journal of Neurophysiology*, *63*(3), 607–624. Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=2329364
Scoville, W. B., & Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. *The Journal of Neuropsychiatry and Clinical Neurosciences*, *12*, 103–113. https://doi.org/10.1136/jnnp.20.1.11
Shohamy, D., Myers, C. E., Hopkins, R. O., Sage, J., & Gluck, M. A. (2009). Distinct hippocampal and basal ganglia contributions to probabilistic learning and reversal. *Journal of Cognitive Neuroscience*, *21*(9), 1821–33. https://doi.org/10.1162/jocn.2009.21138
Skinner, B. F. (1938). The Behavior of Organisms: An experimental analysis. *The Psychological Record*, 486. https://doi.org/10.1037/h0052216
St. Clair, W. B., & Noelle, D. C. (2013). Implications of Polychronous Neuronal Groups for the Nature of Mental Representations. In *Cognitive Science Conference Proceedings* (pp. 1372–1377).
St. Clair, W. B., & Noelle, D. C. (2015). Implications of polychronous neuronal groups for the continuity of mind. *Cognitive Processing*, *16*(4), 319–323.
Strausfeld, N. J., & Hirth, F. (2013). Deep homology of arthropod central complex and vertebrate basal ganglia. *Science (New York, N.Y.)*, *340*(6129), 157–61. https://doi.org/10.1126/science.1231828
Sutton, R. S. (1988). Learning to predict by the methods of temporal differences. *Machine Learning*, *3*(1), 9–44. https://doi.org/10.1007/BF00115009
Sutton, R. S. (1996). Generalization in reinforcement learning: Successful examples using sparse coarse coding. *Advances in Neural Information Processing Systems*, 1038–1044. Retrieved from http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Generalization+in+Reinforcement+Learning:+Successful+Examples+Using+Sparse+Coarse+Coding#0
Tesauro, G. (1992). Practical issues in temporal difference learning. *Machine Learning*, *8*(3–4), 257–277. https://doi.org/10.1007/BF00992697
The Mathworks Inc. (2016). MATLAB - MathWorks. https://doi.org/2016-11-26
Thorndike, E. L. (1898). Animal intelligence: An experimental study of the associative processes in animals. *Psychological Review*, *2*(4), 1–107. https://doi.org/10.1097/00005053-190001000-00013
Thorndike, E. L., & Jelliffe. (1912). Animal Intelligence. Experimental Studies. *The Journal of Nervous and Mental Disease*. https://doi.org/10.1097/00005053-191205000-00016
Tracy, A. L., Jarrard, L. E., & Davidson, T. L. (2001). the Hippocampus and Motivation Revisited - Appetite and Activity, *127*, 13–23.
Weikart, C. L., & Berger, T. W. (1986). Hippocampal lesions disrupt classical conditioning of cross-modality reversal learning of the rabbit nictitating membrane

response. *Behavioural Brain Research*, *22*, 85–89. Retrieved from http://www.sciencedirect.com/science/article/pii/0166432886900835

Wyass, J. M., & Van Groen, T. (1992). Connections between the retrosplenial cortex and the hippocampal formation in the rat: A review. *Hippocampus*, *2*(1), 1–11. https://doi.org/10.1002/hipo.450020102