A DUAL PATHWAY APPROACH FOR SOLVING THE SPATIAL CREDIT ASSIGNMENT PROBLEM IN A BIOLOGICAL WAY

by

Patrick Connor

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

at

Dalhousie University
Halifax, Nova Scotia
November 2013

To God, for whom and to whom
I am forever grateful.

Contents

List of Tables

List of Figures

Abstract

To survive, many biological organisms need to accurately infer which features of their environment predict future rewards and punishments. In machine learning terms, this is the problem of spatial credit assignment, for which many supervised learning algorithms have been developed. In this thesis, I mainly propose that a dual-pathway, regression-like strategy and associated biological implementations may be used to solve this problem. Using David Marr's (1982) three-level philosophy of computational neuroscience, the thesis and its contributions are organized as follows:

- Computational Level: Here, the spatial credit assignment problem is formally defined and modeled using probability density functions. The specific challenges of the problem faced by organisms and machine learning algorithms alike are also identified.

- Algorithmic Level: I present and evaluate the novel hypothesis that the general strategy used by animals is to perform a regression over past experiences. I also introduce an extension of a probabilistic model for regression that substantially improves generalization without resorting to regularization. This approach subdues residual associations to irrelevant features, as does regularization.

- Physical Level: Here, the neuroscience of classical conditioning and of the basal ganglia is briefly reviewed. Then, two novel models of the basal ganglia are put forward: 1) an online-learning model that supports the regression hypothesis and 2) a biological implementation of the probabilistic model previously introduced. Finally, we compare these models to others in the literature.

In short, this thesis establishes a theoretical framework for studying the spatial credit assignment problem, offers a simple hypothesis for how biological systems solve it, and implements basal ganglia-based algorithms in support. The thesis brings to light novel approaches for machine learning and several explanations for biological structures and classical conditioning phenomena.

List of Abbreviations Used

APECS - Adaptively Parameterised Error Correcting System

CART - Classification and Regression Tree

CS - Conditioned Stimulus

DPR - Dual Pathway Regression

GPe - Globus Pallidus externa

GPi - Globus Pallidus interna

LASSO - Least Absolute Shrinkage and Selection Operator

LMS - Least Mean Squares

MLP - Multi-layer Perceptron

MSpN - Medium Spiny Neuron

PDF - Probability Density Function

PMC - Premotor Cortex

RDDR - Reinforcement-Driven Dimensionality Reduction

rLMS - rectified Least Mean Squares

RPE - Reward Prediction Error

SFPD - Simultaneous Feature Positive Discrimination

SLIM - Striatal Lateral Inhibition Model

SNc - Substantia Nigra pars compacta

SNr - Substantia Nigra pars reticulata

SOP - Sometimes Opponent Process

STN - Subthalamic Nucleus

SVM - Support Vector Machine

SVR - Support Vector Regression

TD - Temporal Difference Learning

US - Unconditioned Stimulus

VTA - Ventral Tegmental Area

XOR - Exclusive OR

Acknowledgements

I can hardly begin to enumerate the kindnesses which have either brought me to begin, work at, or complete this doctoral degree or those that have contributed to my development as a scholar in the process. But I will try.

I must begin with my loving wife, Sunshine (yes, that's her real name). Without her willingness to pack up our (then) two young children and move yet again, this work would never have commenced. Since then, she has gladly taken on the role of mother to two more children, which can make times when I am out of town on school-related matters a little more hectic. Thank you, my wonderful wife for your encouragement and constant support. You're my biggest fan.

Thomas Trappenberg, my doctoral supervisor, has given me many reasons for thanksgiving. It was partly through Thomas' funding that I was able to embark on this adventure at all. Thomas also provided additional funding to free me to spend more time writing and to send me on a number of conferences and a summer school. Thomas also kindly agreed that I should take up the opportunity to teach an undergraduate course in my third year as well as complete the Dalhousie Certificate in University Teaching. Thomas made himself available to chat about ideas and let me pursue the course of research that interested me most, which I believe has allowed for the breadth and depth (and length, sorry!) of this thesis. Thank you Thomas, for your willingness to challenge ideas and also for balancing criticism with praise and encouragement.

Vincent (Vin) LoLordo has been like a second (unofficial) supervisor to me. I dropped into his office one day and invited him to come to a presentation of some of my research that related to his field. He came. In our early conversations, he offered to look at any manuscript drafts I might prepare regarding animal learning. Thomas and I took him up on this offer and Vin graciously (and sacrificially) spent *many* hours with me over the following two years preparing and revising a manuscript that was ultimately accepted. Vin, I can't express how valuable our time together has been. I have learned a great deal about the field of classical conditioning including the people

behind the research. You have exemplified for me the accurate and careful scholar and have surely helped me develop in the process of scholarly writing. More than that, you have always been an encourager and I have never left your office without having laughed with you numerous times.

Olav Krigolson was a welcome addition to my committee. We invited him to one of our weekly lab meetings and have been connected with him and his lab ever since then. Thank you Olav, for the free lunches at the Grawood and University Club to talk about my work. You did pay for that, right? :)

The Computer Science faculty and staff welcomed me and worked patiently with me over my my time here. Also, Denis Riordan gave me the opportunity to teach an undergraduate course, for which I am deeply grateful. I want to say a special thanks to Alex Brodsky who mentored me through that teaching experience, for which I perhaps gained three times the insight and experience I would have acquired otherwise.

Last but not least, I wish to thank Dr. Chris Watts, from the Dalhousie Faculty of Engineering. Since coming to Halifax he has kindly met with me many times to share his personal wisdom on matters of scholarship and career development from a faith-based perspective. Thank you, Chris, for your mentorship, your care for me and my family, and the godly example you set.

Finally, this work was financially supported by CIHR, NSERC, the Walter C. Sumner Foundation, and Dalhousie University.

Chapter 1

Introduction

Spatial credit assignment is the problem of properly distributing predictive ability among features based on previous experience or data so that one may accurately predict a future outcome for a new combination of inputs. This is a process that is carried out by living things which must learn certain real-world cues that signal the opportunity for sustenance or the impending danger of a predator. In biological systems, learning about such forms of reinforcement has been localized to a few brain areas, namely the amygdala, orbitofrontal cortex, and the basal ganglia (Maia, 2009). A very prominent discovery (Schultz, 1998) supporting this understanding is that the dopaminergic cells of the basal ganglia's substantia nigra pars compacta region and the ventral tegmental area appear to provide a signal that computational models (namely Temporal Difference learning (Montague, Dayan, & Sejnowski, 1996; Schultz, Dayan, & Montague, 1997)) use to learn about reinforcements. However, such studies have mostly been concerned with assigning credit back through time to the earliest predictor rather than assigning credit among many stimuli appearing at the same time. A common interpretation is that the basal ganglia, and particularly the striatum, encodes the reward value of stimuli and actions (Niv, 2009). Whether for representing the expected future reward for being in a specific context or for encouraging the selection of an appropriate action, the basal ganglia appear to have the connectivity and downstream influence necessary for these roles (see Chapter 6). As part of encoding the reward value of a stimulus or potential action, it would seem necessary to carefully assign credit among the many stimuli (or actions) simultaneously present (or taken) prior to the receipt of a reinforcement. How this is done in biological systems has not been thoroughly investigated.

Ultimately, the present work proposes and evaluates a dual pathway, basal ganglia-based approach to spatial credit assignment. In short, the basal ganglia models presented: 1) improve prediction accuracy by making certain assumptions about the

real-world relationships between predictors and outcomes and 2) offer an online way to perform a batch-like regression process over past experiences. To demonstrate this, this thesis covers a variety of topics. This will involve moving freely between the fields of machine learning, animal learning, and neuroscience. This interaction will prove fruitful as the notions from each area will assist and inform the others. For example, the field of machine learning demonstrates the challenging issues associated with learning and offers the theoretical foundations for learning approaches to which biological systems are undoubtedly subject. Also, animal learning and neuroscience provide insight into how biological systems actually solve the problem, which can be used to narrow the search through learning algorithms offered by machine learning.

To show how a *dual pathway approach makes effective assumptions*, several tasks were undertaken. The spatial credit assignment problem is first defined in terms of learning the parameters of a probability density function. A function that captures aspects of the real world was chosen to serve as the "optimal" model from which data were generated for a simple regression task that is used to draw comparisons between learning approaches. Using least mean squares regression simulations of several classical conditioning experiments, it is shown here that mammals appear to accomplish spatial credit assignment by roughly performing a regression over past experience. Taking the regression strategy further, least mean squares with several regularization or Bayesian prior terms is also evaluated. The Dual Noisy OR model, a probabilistic dual pathway model that extends the Noisy OR model of Pearl (1988), is proposed and evaluated. This strategy has no explicit prior (it is implicitly uniform) but naturally accentuates the most relevant or predictive features at the expense of the lesser predictive features, just as occurs in animal learning experiments. The primary mechanisms of this model are then extracted and fitted to the basal ganglia, demonstrating that they are primarily responsible for the model's effectiveness.

A *dual pathway approach ought to learn effectively in an online fashion* for it to be biologically plausible and practical. In biological hardware, performing regression appears implausible at first because it involves batch-learning. However, I propose a novel neural network model of basal ganglia structure with a focus on the striatum that is capable of performing an aspect of our regression task in an online fashion. I also offer the possibility that existing models of classical conditioning are able to

explain other aspects of regression in an online way as well.

Relating our dual pathway models to a number of basal ganglia models on the market, including Frank (2005) and Bar Gad et al. (2001), I suggest a new role for its dual pathway structure. Instead of simply duplicating functionality, the dual pathways support a multiplicative approach to integrating inhibition that reduces prediction errors, which is done in part by separating positive predictors of an outcome and negative predictors (that cancel a prediction) into different pathways. Another possible role for dual pathways, which takes advantage of this separation, is to perform simple non-linear discriminations (e.g., exclusive OR), as will be shown. Preliminary combination of the mechanisms that implement these two roles has proven to be very powerful, substantially reducing the amount of data or experience needed to make such non-linear distinctions.

In sum, the dual pathway approach provides a level of performance in our specfic spatial credit assignment regression task that rivals popular approaches and yet suits the basal ganglia.

1.1 What is Spatial Credit Assignment?

Spatial credit assignment is the task of attributing to features of an environment the ability to predict a future outcome based on previous experience or data. Figure 1.1 depicts this as a black box where each input is the salience of a potential feature and the output is the learner's expectation of a future outcome. The system should learn the relevant input-output associations so that it can make accurate predictions in the future. In more specific terms, spatial credit assignment is the process of learning to predict an outcome value, y, from a vector of attributes or feature values, x, based on a training set where each data point is an x and y pair. In machine learning, this is commonly referred to as supervised learning.

Spatial credit assignment is routinely relied upon by many living things. Although there is much instinctual knowledge for the acquisition of food and escape from predators, much can be learned as well. It becomes important to discern the cues that predict rewarding and punishing outcomes from the spectrum of features in the world. Since people and animals seem to be very effective in this task, machine learning stands to benefit from understanding the biological solution to this

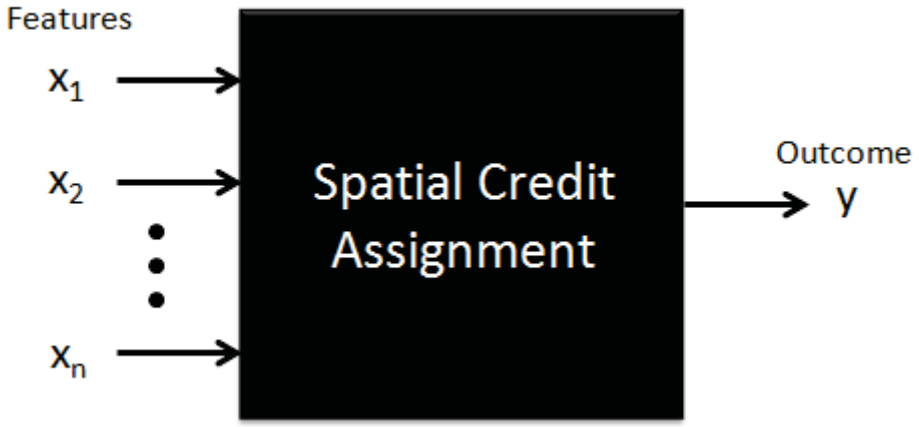

Figure 1.1: An illustration of the spatial credit assignment problem. An algorithm inside the black box relates input features, x, to outcome values, y, based on past experience so as to accurately predict future outcomes from novel inputs.

problem. The purpose of this thesis is to show what biologically-based algorithms and implementations may be inside the black box. In particular I will propose a dual pathway-based structure and show how its associated mechanisms relate to existing algorithms/models addressing the same problem in machine learning and classical conditioning.

Before continuing, it seems appropriate to clarify what spatial credit assignment is *not*. Biological systems appear to use the equivalent of unsupervised deep learning methods to develop representations for stimuli in the world from raw sensory data. Perhaps the most prominent example of this is the visual system. Cortical regions arranged in a hierarchy extract increasingly abstract and transformation-invariant features of the environment (for a review, see DiCarlo, Zoccolan, & Rust, 2012). This ability is crucial for understanding the state of the world around us, encoding the prominence of certain object-level features as the intensity of neural activities. However, this is not spatial credit assignment since no predictions are made (i.e., it is not supervised learning). Yet, spatial credit assignment critically depends on an effective higher-level representation of the predictive features of the world, since these often form the inputs (x) to which we must assign credit. Thus, in this thesis, we rely on the existence of such high-level features but do not examine how such representations are formed.

1.2 Spatial Credit Assignment and Classical Conditioning

The field of classical conditioning studies the biological approach to spatial credit assignment by evaluating animal responses to stimuli after being presented in various arrangements of time and space. In classical conditioning experiments, there are two types of stimuli: the conditioned stimulus (CS) and the unconditioned stimulus (US). In an experiment, a CS such as a light or tone begins as a neutral stimulus that does not elicit any conditioned response (CR) such as salivation or pecking. The US, however, does naturally elicit a response upon presentation since it represents a motivationally salient stimulus such as food pellets or foot shock. In a typical experiment, one or more CSs are presented and followed by a US. For example, in the phenomenon of conditioned excitation, a single stimulus A is followed by reinforcement (A→US, or more commonly expressed as A+). The result of repeated pairings or trials is that the subject will develop a response to the CS, stimulus A, in a subsequent testing phase. In a slightly more involved experiment, two different types of trials are used. In conditioned inhibition, a reinforced CS (A+) is alternated with presentation of a compound or pair of stimuli (A and X) followed by no reinforcement (AX-, where the "–" represents no reinforcement). In this experiment, stimulus A still develops a response, but stimulus X comes to signal no reinforcement and thus inhibits responding that otherwise would occur to A.

Pavlov (1927) established the above nomenclature and identified many of the main phenomena recognized in the field today, which include conditioned excitation and conditioned inhibition. Experiments investigating how simultaneous presentation of multiple CSs affects how each CS is conditioned began in the late 1960's (Kamin, 1968, 1969; Wagner, Logan, Haberlandt, & Price, 1968). Kamin is credited with discovering the forward blocking phenomenon (Kamin, 1969) which involves two phases: 1) repeated presentations of A+ and 2) repeated presentations of AX+. The result was that little conditioning occurred to X in the second phase relative to a control (CX+) group, where the partner stimulus C was not conditioned in the first phase. In essence, conditioning to A in the first phase blocked conditioning to X in the second. Other early work involving multiple stimuli include the phenomena of overshadowing (Pavlov, 1927) and relative validity (Wagner et al., 1968), which we will visit in some detail in the body of this work. In recent years, there have been many human

causal learning studies that study the interaction of multiple CSs, where subjects rate the predictive strength of various stimuli that are sometimes presented together (e.g., Baetu & Baker, 2010; Van Hamme & Wasserman, 1994; McLaren, Forrest, & McLaren, 2012). The important point is that there are many classical conditioning experiments which identify how biological systems distribute predictive ability among multiple features (stimuli) in the environment under a variety of different conditions. It thereby provides a large body of evidence that helps characterize how people and animals solve the spatial credit assignment problem.

The CR expresses an animal's prediction of a future outcome. This prediction can be thought of as an association connecting an environmental cue and a future outcome, such that when the cue appears, the association leads to an internal sense or expectation of the outcome. The field of classical conditioning frequently employs the term "associative strength" to refer to the degree to which a CS predicts another stimulus (usually a US). I will use this term interchangeably with "predictive strength". The strength, rate, or vigor of the animal's CR to a CS presented in the test phase of an animal learning experiment is generally interpreted as being proportional to the degree of associative strength with which the CS predicts the US. Accordingly, this notion is assumed in the later discussion of experiments and simulations.

1.3 Spatial Credit Assignment and Machine Learning

Although little is mentioned in the literature explicitly discussing "spatial credit assignment", the notion is central to machine learning. The black box in Figure 1.1 is representative of supervised learning (i.e., regression and classification), a large segment of the field. Many ways have been invented to develop an effective translation between the inputs, x, and the output, y, based on training data. Some general purpose approaches include multilayer perceptrons, classification and regression trees, and support vector machines.

The multilayer perceptron (MLP) (Rumelhart, Hinton, & Williams, 1986), also referred to as a connectionist network, was one of the first of these algorithms, loosely representing a network of interacting neurons. Usually, there are only two layers of multiple learning units, which are commonly updated in proportion to their prediction accuracy (larger errors lead to larger changes) and the gradient of the prediction

equation (defined by the network) with respect to the learning element or weight to be updated. A sufficiently complex MLP network can represent a non-linear function to an arbitrary accuracy (Cybenko, 1989; Hornik, Stinchcombe, & White, 1989), making this approach very general purpose, in principle. The simpler single-layer perceptron (similar to least mean squares regression) can be substituted when the function to be learned is linear. The classification and regression tree (CART) represents another common prediction model approach (Breiman, Friedman, Olshen, & Stone, 1984). This computationally efficient technique looks to explain data using simple rules (e.g., Is $x > 2.5$?) that branch out. If the first rule is true for a given data point, it is funneled to a secondary rule. If, instead, it is false, it is funneled to another secondary rule. Each data point is evaluated by the secondary rule and then may be sent to a tertiary rule (branch) and so on until a "leaf" node is encountered, where the output value is mapped to the data point. There are several different algorithms that can be used to "learn" effective rules that map the input to the output. Finally, the support vector machine (SVM) (Vapnik, 1995; Cortes & Vapnik, 1995), has gained popularity in recent years. It is quite robust for many problems with default settings and thus does not require careful tuning (unlike the MLP). Like the MLP and CART, it can learn both linear and non-linear relationships. For mapping inputs to outputs, one would use support vector regression (SVR) (Vapnik, 1995), the regression version of SVM that is based on similar theoretical principles. One form of SVR, called the epsilon-SVR involves fitting a hyperplane with margins (i.e., a hyper-rectangular prism) so that it encompasses all data while at the same time minimizing the hyperplane's slope (Smola & Schölkopf, 2004). This is often infeasible, so that outliers are permitted but at a cost, leading to a trade-off between minimization of the slope and the acceptance of outliers.

Much has been learned from machine learning about general problems and solutions in spatial credit assignment. For instance, it is well known that spatial credit assignment usually becomes less accurate as the number of input features becomes large (Bengio & Bengio, 2000; Evangelista, Embrechts, & Szymanski, 2006) relative to the amount of data available, a problem for which many feature reduction algorithms (Fodor, 2002; Guyon, Gunn, Nikravesh, & Zadeh, 2006) have been devised. Naturally, predictions become more difficult as increasing levels of noise are injected

into the data. Machine learning has addressed these general problems in a couple of ways. One way is to increase the number of training data points. This could be done by collecting additional data. In some circumstances, however, acquiring additional data is costly or not feasible. In this case, generating synthetic data may be feasible. Another general solution is to reduce the number of features to only the most important ones, since the number of truly predictive features is often only a fraction of the total. Feature reduction techniques are commonly used to eliminate irrelevant features (e.g., feature subset selection) or express the same data with fewer features (e.g., Principle Components Analysis). We will look at these general problems and machine learning solutions later in greater detail.

While much work has been done in both the fields of classical conditioning and machine learning regarding spatial credit assignment, there has been little formal interaction between the two fields. The important exception to this is a well known model of conditioning by Rescorla and Wagner (1972), which is better known by its temporal cousin, temporal difference (TD) learning (Sutton & Barto, 1990), a key algorithm used in reinforcement learning. In the Rescorla-Wagner model, after each trial, a stimulus or feature's predictive strength is increased whenever the feature is presented and followed by more reinforcement than expected and is decreased when it is presented but followed by less reinforcement than expected. In TD learning, the learning process is similar but is broken down into many time steps within a trial. In an effort to narrow the scope of the present work, we treat predictions and outcomes at a trial-level resolution instead of in"real-time". This allows us to focus on investigating how associations are distributed "spatially", that is, among stimulus features of the environment. In the machine learning subfield of reinforcement learning, this spatial distinction is not frequently made.

In the standard reinforcement learning paradigm, the Markov Decision Process, an agent normally moves across a chess-board style grid, where each square-shaped location is a distinct state that the agent may visit. The agent's goal is to maximize reward and minimize punishment. Beginning in a "start state", the agent moves from state to adjacent state until it arrives at a terminal state, which is typically rewarding or punishing. At this time, it receives and learns from the reinforcement and is then forced to repeat the scenario in a new "episode", with the expectation

that its performance should improve with experience. Reinforcement values are slowly propagated backward through the states traveled leading up to the terminal state(s), which serve as trail through the state grid toward rewarding terminal states (and away from punishing terminal states). For simplicity, each state is normally treated as being distinct, that is, it does not share features or anything in common with other states. In stimulus terminology, it would be like saying that a certain stimulus is only present when the agent is in a specific state and thus learning about that stimulus can only occur at that time. Now, if one needs as many distinct stimuli as states, and there are individual states for every X-Y location on a map, then the number of stimuli required can explode very quickly as a function of the size or resolution of the state map. To avoid this problem, reinforcement learning turns toward function approximation or supervised learning techniques, which employ spatial credit assignment. In this case, each state is now encoded as a vector of features (or stimuli), where similar map locations are represented as similar patterns of input. This allows learning in one "state" to be partially transferred to neighbouring states because they share similar, albeit slightly different, vector representations. So, although spatial credit assignment is not as commonly employed in Markov Decision Processes as, say, in linear regression, it is indispensible as the size of the state-space gets large.

Since we seek to understand the algorithms underlying the biological solution to spatial credit assignment, we focus mostly on machine learning approaches that maintain some level of biological plausibility. Here, for an algorithm to be biologically plausible, it must be plausibly implementable in terms of the neural system it aims to represent. In general, this takes the form of a network of neurons. Individual neurons are known to receive input from and send output to other neurons, but locally compute from their inputs when and how frequently they respond as captured in standard biophysically detailed neuron models (Hodgkin & Huxley, 1952; FitzHugh, 1961; Nagumo, Arimoto, & Yoshizawa, 1962; C. Morris & Lecar, 1981). Updates to neuron parameters also appear to be locally performed since changes in connectivity (synaptic) strengths for a given neuron can be predicted from its specific input and activity (Bi & Poo, 1998; Bliss & Lømo, 1973; Reynolds & Wickens, 2002), which means that one neuron cannot adjust another neuron's parameters apart from its

influence as a potential input. Neurons can be connected in any number of complicated arrangements, potentially involving recurrent, reciprocal, lateral, excitatory, inhibitory, and modulatory connections, etc. However, constraints on the types of connections are imposed when attempting to match an algorithm to a specific brain area, since not all types may be present there.

The biological plausibility factor varies among machine learning approaches, but most fall into the relatively implausible category. For example, even MLPs are criticized (Chinta & Tweed, 2012) because to adjust the parameters of a neuron in one layer (using the gradient), the parameters of downstream neurons must be known. This breaks the local-update criterion. CART is a sequential process of repeatedly making a decision point and branching, all the while maintaining the series of specific rules already defined. It would take a great imagination to realize such a procedure in terms of a neural network, especially a robust one that can scale with a large numbers of features, such as can be accommodated by CART. In this thesis, biological plausibility has been a critical ally in narrowing the evaluation of possible learning algorithms to a few key, standard approaches. In the first half of the thesis, the direct biological plausibility of the algorithms being tested is less critical as abstract approaches that biology may be using are evaluated, although we still prune our investigation to evaluate and relate algorithms that seem somewhat plausible. In the second half of the thesis, the biologically plausible distinction becomes imperative as we seek to understand possible neural implementations of these algorithms. As algorithms are put forward, their levels of biological plausibility will be discussed.

1.4 Chapter by Chapter Thesis Outline

In the present work, I aim to show how biology may solve the spatial credit assignment problem, which involves addressing the general challenges identified by machine learning. I approach this in terms of David Marr's three-level philosophy of computational neuroscience. The first level, called the "computational level" is to identify the problem or task that the brain must solve. This introductory chapter helps accomplish this by defining the spatial credit assignment problem and reviewing initial evidence suggesting that the brain engages in it. In Chapter 2, I formally introduce the spatial credit assignment problem and describe it in terms of probability density

functions. I also define the "world" model used to generate data for later simulations to match the real world in a number of important ways (e.g., sparse activation, relatively few experiences, and rectified outcome) and derive the associated optimal regression model. Finally, I provide simulations to illustrate the key challenges of using few data points, many features, and noise, showing that spatial credit assignment is truly a problem.

The second level, called the "algorithmic/representational level" aims to identify algorithms or general computational approaches that may be used to solve the problem. In Chapter 3, I put forth the novel hypothesis that biological systems treat spatial credit assignment as a regression over past experiences. In support, I compare simulations of classical conditioning phenomena using the Rescorla-Wagner model (Rescorla & Wagner, 1972) and least mean squares (LMS) linear regression, showing how regression can simulate two additional groups of phenomena. Chapter 4 reviews a number of (mostly biologically plausible) ways from the machine learning literature to address the key challenges noted in Chapter 2. The last chapter in this part of the thesis, Chapter 5, offers a novel approach to regression that extends the Noisy OR probabilistic model of Pearl (1988). Here, we show how this principled approach handles many irrelevant features and system noise with little data by forcing parameter values and outputs (probabilities) to be non-negative. The extended, dual pathway, model maintains this principle and allows the integration of inhibitory features by incorporating them in a multiplicative way. In summary, this part of the thesis describes the general regression strategy that biological systems appear to use to solve the problem, shows how certain machine learning strategies overcome the key challenges, and provides a novel dual pathway solution.

The third level in Marr's philosophy, called the "physical level", aims to show how the potential algorithms may be implemented in biological hardware (i.e., networks of neurons). In Chapter 6, I review the anatomy of the basal ganglia and the neuroscience of classical conditioning, which provides the basis for biological implementations of the above algorithms. Chapter 7 presents a novel dual pathway neural network model of the basal ganglia focused largely around the striatum and its lateral inhibitory connectivity, which we call the striatal lateral inhibition model or SLIM. This model is able to explain a number of classical conditioning phenomena

beyond the Rescorla-Wagner model. Specifically, I show that this model is able to explain retrospective revaluation phenomena with only seeing the training data once (i.e., online-learning). This differs from LMS, which must keep and repeatedly cycle through recent and all previous data to explain these same phenomena. Thus, we show how an online approach can be used, making the regression hypothesis more plausible. Chapter 8 introduces Dual Pathway Regression or DPR, a novel way of integrating inhibition into a LMS-like model that is based on our extension of the Noisy OR model in Chapter 5. We find that it is also capable of handling many irrelevant features and additionally explains certain classical conditioning phenomena that could not be explained by LMS alone. Chapter 9 reviews a number of other basal ganglia models and relates them to SLIM and DPR. In particular, I compare SLIM to the models of Frank (2005) and Bar-Gad et al. (2003), which bear a number of similarities. I also show promising preliminary work that combines aspects of SLIM and DPR to perform non-linear discriminations.

1.5 Summary of Contributions to Research

This thesis represents three primary contributions to research:

- The "Regression Hypothesis" of Biological Spatial Credit Assignment - I support with classical conditioning simulations the novel hypothesis that, with regard to spatial credit assignment, animals perform a regression over past experiences. Specifically, but not exclusively, this is expressed in classical conditioning phenomena involving multiple stimuli.

- Generalization by Multiplicative Inhibition - Many models integrate the influence of an inhibitory features (negative parameter values in least mean squares) by summing them with the influence of excitatory features to "cancel" predictions. I show through the Dual Noisy OR model (an extension of Noisy OR) and finally Dual Pathway Regression that prediction errors can be reduced by incorporating inhibition in a shunting, multiplicative fashion. Here the sum of excitatory influences (i.e., the positive prediction) is multiplied by a number between zero and one representing the inhibition level, where higher inhibition is represented by a smaller number.

- The Striatal Lateral Inhibition Model - I developed a novel dual pathway neural network with lateral inhibition that can explain retrospective revaluation phenomena without having to cycle through past trials, thereby supporting the plausibility of the "regression hypothesis".

Other noteworthy contributions include:

- Zero-Peak Regularization/Bayesian Prior - I have defined a new prior distribution form which is the sum of a parameterized constant value and a variable-width Gaussian. It can be also viewed as a regularization technique. Better than a uniform prior, it represents the distribution from which parameters are drawn in our main regression task. Like the LASSO regularization technique, it is also likely effective for tasks that employ some form of feature selection.

- A Realistic Spatial Credit Assignment Task - I defined very specific parameters for the training and test data to reflect features of the real world as it regards the spatial credit assignment task, which included the sparsity of relevant features and reinforcements. For this world, I derived the "optimal" maximum likelihood estimate model as well, which I refer to as rectified least mean squares.

- Comparison of a Variety of Generalization and Regression Methods in Machine Learning - I have evaluated a variety of the machine learning methods from the literature to gage the effectiveness of the novel methods. I also draw attention to the differences in effectiveness between several of the approaches in the literature.

Chapter 2

Modeling the World of Spatial Credit Assignment

2.1 Chapter Summary

This chapter is devoted to describing and defining the real world as it pertains to animal learning and the spatial credit assignment problem. Aspects of the real world are encoded into a mathematical model from which data is generated. This probability density model and data will be used to draw comparisons with other approaches in later chapters. Finally, we examine the spatial credit assignment problem when we have little data but lots of features and noise using a specific regression task that is employed frequently in the thesis. Excerpts are taken from Connor and Trappenberg (2013), ©2013 IEEE, in which I was primarily responsible for developing the theory and simulations as well as drafting the manuscript.

2.2 Modeling the Real World as a Probability Density Function

In the real world, certain environmental features precede certain rewards or punishments. The relationship between a feature and reward or punishment has both deterministic and probabilistic aspects. For example, the stronger the presence of a feature (e.g., redness of the fruit), the larger the reward we expect (deterministic aspect). However there is some uncertainty (probabilistic aspect) in this prediction since the quality of the fruit is not always linked to its redness (e.g., it may be rotten). It is appropriate to model this relationship as a random variable and represent it with a probability density function (PDF). Figure 2.1 shows an imagined PDF for the degree of benefit obtained given the redness of the fruit. The x-axis is the redness of the fruit and the y-axis represents an amount of reward value obtained. Given a certain redness (e.g., x=0.5), there is a one-dimensional PDF (a cross-section of the two-dimensional PDF) indicating the likelihood of each possible amount of reward.

A system that solves the spatial credit assignment problem is able to predict a

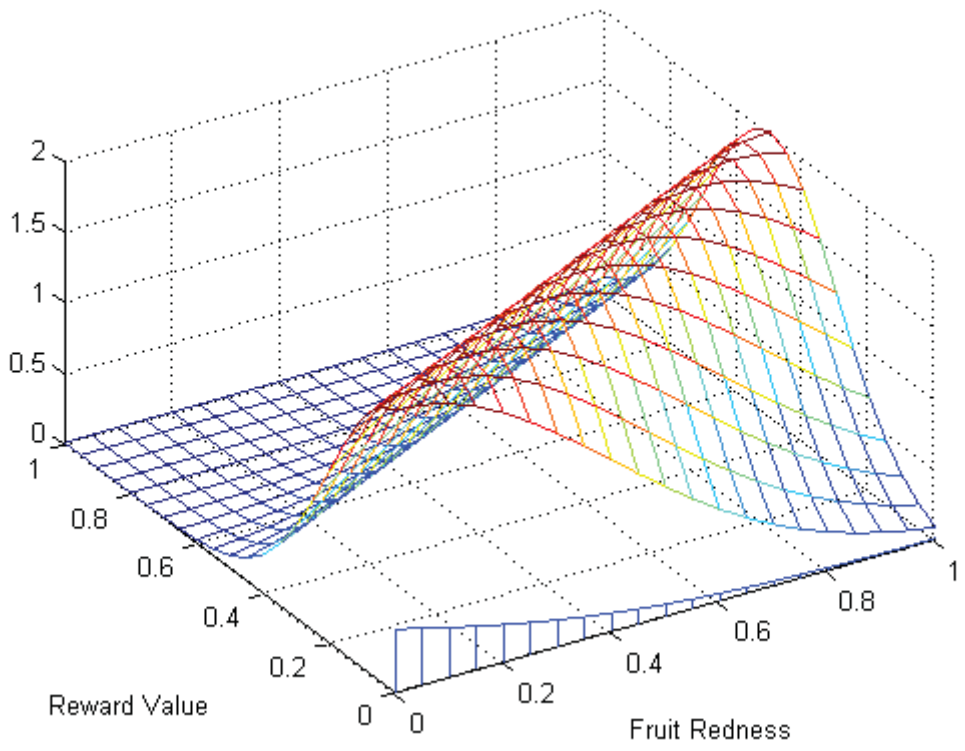

Figure 2.1: A probability (or joint) density function showing the likelihood of certain x and y-value pairs.

future outcome given some input or feature values. Given that we know the true PDF, the best prediction one can make is to compute the expected value of y given x, which is

$$E[Y] = \int_{-\infty}^{+\infty} yP(Y = y, X = x)dy \tag{2.1}$$

Of course, the trouble is that given data, we must somehow approximate the true PDF. *This is the spatial credit assignment problem.* One way to do this is to simply build a probability table, $P(y, x_1, x_2, ..., x_N)$, from the training data, where each data point adds to a probability "bin". However, by having many features or by having only a few real-valued features we run into trouble – there are so many bins to fill that it would take an enormous (infinite) amount of data to give an accurate picture of the true PDF.

If we could assume that the values of the inputs (x) are independent of one another, at least we would not have to consider potential interactions between them. From a probabilistic model point of view, this would allows us to break up the joint distribution, $P(y, x_1, x_2, ...x_N)$, into independent probability density tables, $P(y, x_1)$, $P(y, x_2)$, ... , $P(y, x_N)$. The Naive Bayes model makes use of this assumption to compute $P(y|x)$ as

$$P(y|x) = \frac{P(x_1|y)P(x_2|y)...P(x_N|y)P(y)}{P(x_1)P(x_2)...P(x_N)} \tag{2.2}$$

Although this approach is not as vulnerable, it is still troubled as the number of unique x values increases because it spreads out the available data, making each probability table sampling more inaccurate. Thus, even this approach is only practical when there are either few discrete values of x or when there is a lot of data. To further complicate matters, irrelevant features (ones that do not affect the outcome) will add noise to the computation of $P(y|x)$, since each irrelevant feature's contribution $(\frac{P(x_i|y)}{P(x_i)})$ will deviate from 1 unless there is a lot of data relative to the number of features.

Instead, a common approach to determining the PDF is to assume that the data conforms to a certain PDF *model*, and then the available data is used to configure the model parameters.

2.3 Modeling the World as a Gaussian Random Variable with a Rectified Linear Mean

In the real world, some features are positively predictive of a specific outcome (i.e., a certain reward or punishment). Other features are predictive of no outcome, even in the presence of other features that are positively predictive. Such inhibitory features reduce or cancel otherwise expected outcomes, but importantly do not predict the opposite or a "negative" outcome. For example, consider a classical conditioning experiment, where subjects are accustomed to receiving food in the presence of a tone stimulus. If, in later trials, they receive no food in the presence of a tone-light compound stimulus, they will come to see the light as canceling the food predicted by the tone, *but will not expect a punishment like footshock*, which is the oppositely valenced outcome. Thus, in classical conditioning, an excitatory stimulus is viewed

as being positively predictive of either a reward or punishment whereas an inhibitory stimulus is seen as canceling the prediction of a reinforcement (of either valence). We will later show that making the distinction between such inhibitory stimuli and stimuli predicting oppositely valenced outcomes is beneficial. Therefore, in this work, I have chosen to model the world with a PDF that can only give prediction values ≥ 0 for a specific outcome, that positively predictive stimuli increase and inhibitory stimuli only extinguish. The true PDF, from which the data is generated for simulations, is the rectification of a Gaussian random variable with a linear mean,

$$y = G(\phi^T x + \mathcal{N}(0, \sigma)) \tag{2.3}$$

where x is an input vector, $x_0 = 1$ is the "always-on" input associated with the bias, y is the output, and 0 and σ are the mean and variance of the Gaussian random variable. The function $G()$ is the threshold-linear function (Usher & McClelland, 2001), which returns the argument when it is greater than zero and returns zero otherwise. It essentially eliminates the output of negative values, forcing $y \geq 0$. So, within a certain region ($y > 0$), the rectified linear function is equal to a simple linear function (with noise). An example of this PDF model is pictured in the earlier Figure 2.1. This particular distribution represents the situation where a feature is *linearly* related to the outcome of interest, provided it does not predict an outcome of less than zero. In fact, learning linear relationships is the focus of much of the classical conditioning evidence and an "entry-level requirement" of supervised machine learning methods. Therefore, this work focuses on the linear case.

If we know we are given data generated from Equation 2.3 but do not know the values of ϕ, then a simple solution to spatial credit assignment is to assume this model in our PDF and use the data to infer values of ϕ. We can infer the values using maximum likelihood estimation. We define the probability density function that a data point is generated by a Gaussian random variable with a rectified linear mean (i.e., according to Equation 2.3) as follows

$$p(y, x|\phi) = \begin{cases} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y-\phi^T x)^2}{2\sigma^2}}, & for\ y > 0 \\ \frac{1}{2}(1 - \text{erf}(\frac{\phi^T x}{\sqrt{2}\sigma}), & for\ y = 0 \\ 0, & for\ y < 0 \end{cases} \tag{2.4}$$

The probability or likelihood that a certain training data set is generated from this distribution becomes

$$\begin{aligned} L(\phi) &= p(y^{(1)}, ...y^{(m)}, x^{(1)}, ...x^{(m)}|\phi) \\ &= \prod_{i=1}^{m} p(y^{(i)}, x^{(i)}|\phi) \end{aligned} \tag{2.5}$$

where m is the number of training data points. Our goal is to maximize this likelihood and thereby find the n-dimensional linear function (where n is the number of features in x) that most likely generated the training data. We can maximize this convex function by taking its log and ascending its gradient,

$$\begin{aligned} log\, L(\phi) &= log \prod_{i=1}^{m} p(y^{(i)}, x^{(i)}|\phi) \\ &= \sum_{i=1}^{m} log\, p(y^{(i)}, x^{(i)}|\phi) \\ = -m_{y>0} log(\sqrt{2\pi}\sigma) - \frac{1}{2\sigma^2} &\sum_{i,y^{(i)}>0} (y^{(i)} - \phi^T x^{(i)})^2 \\ -m_{y=0} log(2) + &\sum_{i,y^{(i)}=0} log(1 - erf(\frac{\phi^T x^{(i)}}{\sqrt{2}\sigma})) \end{aligned} \tag{2.6}$$

where $m_{y>0}$ and $m_{y=0}$ are the number of data points when $y > 0$ and $y = 0$, respectively. Taking the gradient of this function with respect to each ϕ_j gives

$$\frac{\partial\, log\, L(\phi)}{\partial \phi_j} = \frac{1}{\sigma^2} \sum_{i,y^{(i)}>0} (y^{(i)} - \phi^T x^{(i)}) x_j^{(i)} - \sqrt{\frac{2}{\pi\sigma^2}} \sum_{i,y^{(i)}=0} \frac{e^{-\frac{(\phi^T x^{(i)})^2}{2\sigma^2}}}{1 - erf(\frac{\phi^T x^{(i)}}{\sqrt{2}\sigma})} x_j^{(i)} \tag{2.7}$$

The gradient can be ascended by iteratively updating the model parameters,

$$\phi_j =: \phi_j + \alpha \frac{\partial\, log\, L(\phi)}{\partial \phi_j} \tag{2.8}$$

where the learning rate, $\alpha = \frac{\sigma^2}{n+2}$ [1]. If instead of using the rectified linear function, we had chosen to use a simple linear function as our PDF model, the derivative of the log likelihood would be the same as the rectified linear function's value when $y > 0$. Maximizing the likelihood of a linear function with Gaussian noise equates to the

[1] We start with this learning rate because it is optimal (Tweed, 2011) for Gaussian random variable with an unrectified linear mean (when the inputs have a zero mean and a variance of one) and it also works well, in practice.

standard practice of minimizing the mean squared error using gradient ascent, which is commonly referred to as least mean squares (LMS). We will refer to the rectified version as rectified least mean squares (rLMS), not to be confused with recursive least mean squares (another well known method). One complicating difference between LMS and rLMS is that rLMS requires knowledge of the additive noise variance (multiplying Equation 4.7 by the learning rate α eliminates some but not all σ terms). Another difference is that rectifying the linear function essentially turns off learning when the model's prediction is negative ($\phi^T x < 0$) and $y = 0$ (with low noise). It is otherwise practically equivalent to ascending the gradient of the linear function itself. Classical conditioning findings suggest that turning off learning under similar circumstances is what animals do, which will be explored further in Section 8.4.

This optimal model is much like the machine learning "perceptron", which shares much in common with the input-computation-output form of a single neuron. The parameters of the model can be loosely seen as the synaptic weights of a neuron and the output be viewed as the neuron's firing rate. In other ways, it is not especially biologically plausible because it allows parameters to change sign and it requires knowing the variance of the noise in the data, which would not be normally known by a biological system. The model, however, will serve as an important approach for comparison purposes, form the basis for certain other approaches, and be a major point in later discussion.

2.4 The Trouble with Small Data, Many Features, and Noise

Given that the training data points are generated from the rectifications of a Gaussian random variable with linear mean, then modeling the true PDF as above and using maximum likelihood is the optimal way to determine its parameters. Although optimal, errors in prediction can still occur. In particular, when either there are few training data points, many features, or the variance of the Gaussian noise is high, rLMS will learn to "predict" the output values of the training data quite well, but will become very inaccurate at predicting outputs for inputs it has never seen, that is, it does not generalize well.

Let us illustrate this in two respects using rLMS in a simple regression task, using training datasets that attempt to mimic aspects of the real world. Each training

data set has 100 data points with each feature being present only 25% of the time. This reflects the reality that rewarding and punishing experiences, and thus their predictors, occur somewhat rarely and that the number of present features is a small fraction of all possible features. Yet it is large enough that we get, on average, 6 data points where both relevant features are present. This is important because in cases where one of the relevant features is inhibitory, we can only learn its identity when it cancels the prediction of the other relevant feature. Each data point is represented by an output, y, and a vector of binary values, x, reflecting the presence or absence of each feature. The rectified linear function that is used to generate the outputs is parameterized by a vector, with each value falling between -0.5 and 0.5. Positive and negative parameter values reflect positive and "inhibitory" predictions, respectively, both of which commonly occur in the world. The bias of the rectified linear function will be set to zero, mimicking the notion that rarely does a motivational outcome consistently occur without some predictive feature attached. Only two of the 99 parameters used to generate the data are chosen to be non-zero. Thus only the inputs associated with two variables actually affect the output value. This represents the usual case that only a small number of features are predictive of reward and that the other features are practically irrelevant. Finally, the outputs of data points that are "negative" (due to the noise) are set to zero (i.e., rectified).

To get a representative sampling of the combinations of the two relevant features in our simulations, we repeat each test for all possible positive-negative parameter combinations (e.g., 0.25, -0.25) and positive-positive parameter combinations (e.g., 0.25, 0.25) from a grid with a resolution of 1/7. To do this we combine every possible positive parameter value (determined by the resolution) with every possible positive and negative parameter value. These combinations provide the mean prediction errors displayed in the figures. Our test data sets have 1000 data points each. Whereas the training data sets use sparsely activated relevant features to correspond to real-world training conditions, we want test data points that are enriched with the presence of relevant features. It is important that an animal makes accurate predictions in these otherwise rare circumstances. Therefore, we use test data sets where the relevant features appear 50% of the time, while the irrelevant features continue to appear only 25% of the time. This can be seen as selectively choosing data points from a larger

dataset where all features appear only 25% of the time to increase the focus on those experiences where relevant features appear. Exactly the same training and test data sets are used for all models. Conveniently, this allows us to compare models using a matched-pairs type of statistical test to determine significance, namely the sign test ($p < 0.01$). This is used in all of the plots of the main regression task now being described. Data points that that are marked with a light gray "x" are significant according to this test with respect to all other curves in the plot. To be marked, the significance must also have the same sign as the difference between the curve means at that point. Curves that are close to one another might be significantly different yet the arrangement of their means may be ordinally opposite to the direction of significance. For clarity's sake, such data points are not marked as significant.

In the top panel of Figure 2.2 we fix the noise at zero and vary the number of inputs between 2 and 1000 while maintaining a training set with only 100 data points. Finding the line of best fit using the rLMS algorithm will solve this linear system of equations exactly, up until the number of inputs crosses some threshold. Experimentally, this occurs at around 50 input features, where the prediction error in the test set begins to rise substantially. In the bottom panel of Figure 2.2 we vary the variance of the additive noise in the training data between 0 and 0.25. We compare the prediction error for the training data with the prediction error for the test data. No Gaussian noise is added to the test sets to more clearly show the differences in model performance. In this scenario, we see that prediction error increases far more rapidly with increasing additive noise for the test data than for the training data. In both panels, we see that rLMS is a better predictive model than the Naive Bayes model described earlier, dramatically showing the value of choosing a specific PDF model and using the available data to configure its parameters. Even when the number of entries in the probability table is few (Top Panel) because the inputs, x, are binary and there is no additive noise, Naive Bayes still struggles tremendously because of the noisy contribution of the irrelevant features. The prediction errors we see for the test set is not an artifact of having chosen a rectified linear PDF model, since even a simple linear PDF model is plagued with this problem (Connor & Trappenberg, 2013) and it is a commonly acknowledged issue in machine learning.

The lower prediction error for the training data seen in both simulations is technically known as overfitting, though this may seem unusual at first for linear regression. Linear regression fits a line to the training data, which is ideal when there is no additive noise and enough data. When either of these requirements are not met, the training data do not perfectly or completely represent the distribution from which the test set is derived. Thus, the line of best fit is crooked to some degree relative to the true linear model which generates the data.

In LMS training, as well as for many other methods evaluated below, if training is terminated before the method "memorizes" the specifics of the data rather than the true linear model, it will reduce overfitting. One approach used to decide when to stop training is to present a "validation" data set to the model after every cycle (epoch) through the data. Validation sets are separate data sets generated from the true model, and often are siphoned from the training sets, making the training sets smaller. In this approach, training will continue as long as the prediction error on the validation set decreases with each epoch (or group of epochs[2]). Once the prediction error begins to grow again, training stops and the test set is evaluated. The biological analogue of a validation set would be having an organism isolate a number of its experiences, not to learn from, but to avoid memorization of other things it learns. Perhaps a more plausible notion would be that learning only runs for a fixed small number of epochs, but in this work the standard practice of employing validation sets will be maintained for sake of drawing more grounded comparisons between machine learning methods. Here, the validation set will contain 10 times more data than the training set since data is freely available and will avoid a potential confound (the use of a small validation set) when explaining results.

In the batch-learning of this work, training data is repeatedly revisited by performing multiple epochs. However, the parameter values of a model are updated after the presentation of each data point rather than waiting until the end of the epoch to perform an average update, as the rLMS update equations call for. This represents another element of biological plausibility since it seems more likely that biological organisms learn all the time rather than saving up their learning for a specific time (e.g., at the end of the day). In general, this has little impact on the final results for

[2]In the case of rLMS and related approaches we run for 25 epochs between evaluations of the validation set to allow for somewhat noisy descents.

small enough learning rates. To be sure, several simulations were run with average updates (not shown) and although these improved rLMS results relative to other approaches that will be shown, it did not affect the conclusions drawn. The exception to this is that one regularization technique (zero-peak) became slightly more effective than another (the LASSO), which are introduced in Chapter 4.

To illustrate the general nature of the problems of small data, many features, and noise in machine learning, Figure 2.3 shows results of the same task using a multi-layer perceptron (MLP), support vector regression (SVR), and a classification and regression tree (CART), which have been manually tuned to get near their best performance. Looking only at test set prediction errors, we see in both the top and bottom panels that the multi-layer perceptron and the support vector regression struggle in much the same way as does rLMS, performing well with low noise and few irrelevant features and breaking down under more extreme conditions. Curiously, the classification and regression tree is quite effective, handling any number of irrelevant features and performing significantly better with high levels of noise.

Recall that in these simulations only two features were relevant (i.e., their generative parameters were non-zero). Thus, only the inputs associated with two variables actually affect the output. CART performs extremely well in the top panel of Figure 2.3 because it ignores all but the most useful features, thereby delivering comparable results regardless of the number of irrelevant features. In fact, not as many data points are necessary to learn the true parameter values for the relevant features under these circumstances because there are more equations than unknowns, if the irrelevant features can be confidently detected (and dropped) somehow. For these reasons, Chapter 4 evaluates existing biologically plausible approaches to "feature selection". Another strategy examined is the artificial augmentation of the training data, since having more data points better informs a regression. Unfortunately for CART and a number of other methods, a neurobiological interpretation is very difficult to conceive, and so they are pruned from further evaluation.

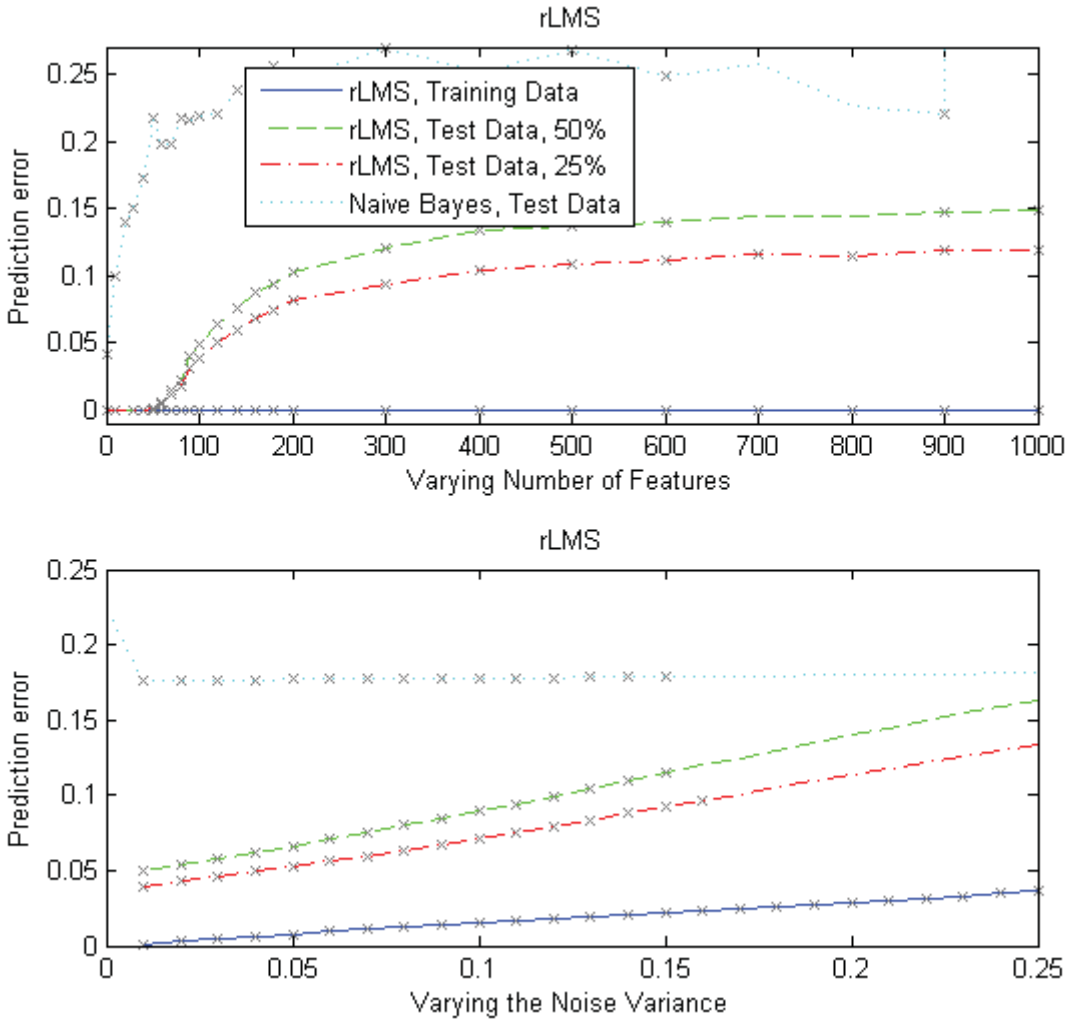

Figure 2.2: (See following page for caption)

Figure 2.2: *Top Panel:* rLMS prediction error as the number of inputs is varied, while keeping the number of data points fixed at 100 and having zero additive noise. The training set's prediction stays low (near zero) regardless of the number of input features. Because there is zero additive noise, the test set's prediction error is small for a low number of irrelevant features. This is the case regardless of whether or not the dataset is enriched with a higher frequency of relevant feature activation (i.e., 50% vs. 25%). As the number of irrelevant features increases, the test set prediction error rises as well. The Naive Bayes model prediction error quickly ascends far above rLMS, showing that a model-based PDF is worthwhile. *Bottom Panel:* As the variance of the additive Gaussian noise is increased, rLMS prediction error in the training and test sets increases. The curves diverge, with the test set increasing in error more rapidly than the training set. Again, the Naive Bayes model prediction error is substantially higher than the rLMS prediction error. In both panels, rLMS test error is larger than the training error, demonstrating the phenomenon of overfitting, where the model fits the training data well but does not generalize as well to unseen data. In both panels, mean values are marked with an X if the associated data is significantly different from the other curves, according to the sign test, $p < 0.01$.

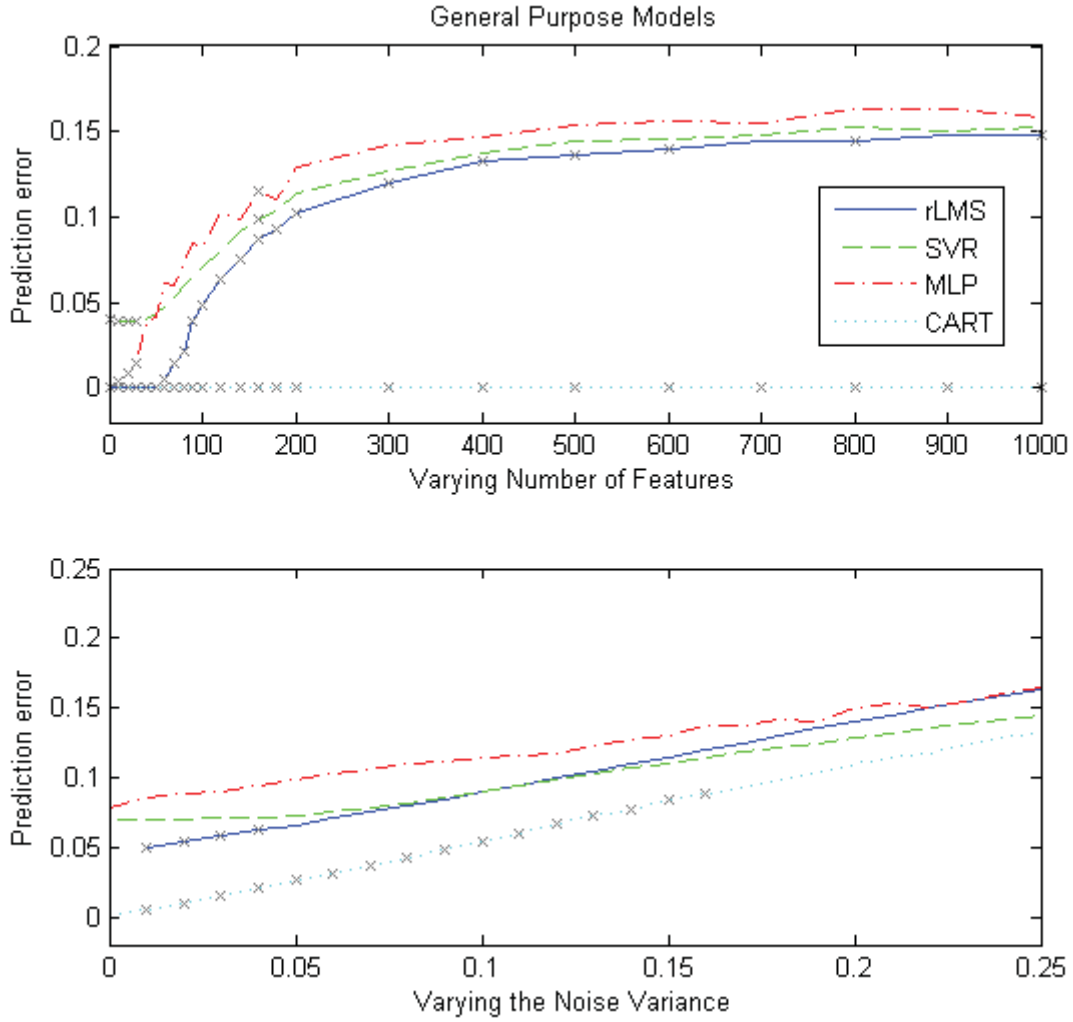

Figure 2.3: *Top Panel:* Test set prediction error as the number of inputs is varied, while keeping the number of data points fixed at 100 and having zero additive noise. SVR and MLP perform comparably to rLMS, whereas CART performs effectively regardless of the number of irrelevant features. *Bottom Panel:* As the variance of the additive Gaussian noise is increased, test set prediction error increases. Again, SVR and MLP perform comparably to rLMS. In this test, CART improves over rLMS but steadily degrades with an increase in noise.

Chapter 3

A Regression Over Experience as a Biological Strategy

3.1 Chapter Summary

In this chapter, we show how a significant number of classical conditioning phenomena appears to suggest that animals learn to predict future outcomes by roughly performing a regression-like process over past experiences, which will be referred to as the "regression hypothesis". This is shown by first describing the close correspondence between LMS and the Rescorla-Wagner model. Then, it is shown how two additional groups of classical conditioning phenomena can be explained by the sole difference between these two models: LMS, a regression model, is allowed to cycle through all prior experiences and train until the US surprisingness (i.e., prediction error) is minimized over the entire set of experiences.

3.2 Comparing LMS and the Rescorla-Wagner Model of Classical Conditioning

Classical conditioning has enjoyed a history of more than a century of experimental work. In the last few decades, it has become more common to develop mathematical models to relate theories and make predictions for future experiments. The Rescorla-Wagner model (Rescorla & Wagner, 1972) was an important early approach to capture the known classical conditioning phenomena and drove a significant number of subsequent experiments. The Rescorla-Wagner model is represented by a single learning rule,

$$\Delta V_i = \alpha_i \beta (\lambda - \sum_j V_j) \tag{3.1}$$

where N is the number of stimuli, V_i represents the associative or predictive strength of a stimulus i, and α_i reflects a stimulus salience which acts like a learning rate. The parameter β is a learning rate associated with the degree of the rewarding/punishing outcome and λ is representative of magnitude of the US. In the equation, the sum of

all V values associated with stimuli present in an experimental trial represents the aggregate outcome prediction for that trial.

Importantly, we can show that the Rescorla-Wagner model is almost identical to LMS (Dawson, 2008), which underlies simple linear regression and machine learning's single layer perceptron. Although we will later perform simulations using rLMS, we will draw comparisons with LMS in this chapter for clarity's sake.

For convenience, the update rule for LMS linear regression can be written as,

$$\Delta\phi_i = x_i\beta(y - \sum_{j=1}^{N} x_j\phi_j) \tag{3.2}$$

where x_i is the strength of an input in a particular trial, ϕ_i represents the weights or parameters of the linear model being learned, β is a simple learning rate, and y is the outcome. Comparing Equations 3.1 and 3.2, the correspondence between parameters is almost one-for-one. The main difference is that the sum in Equation 3.2 is weighted by the current input x whereas in Equation 3.1 the sum is only of the predictive strengths for stimuli present on the trial, V_j, which correspond to the parametric weights, ϕ_j. However, this difference is removed when the x values are binary, which is the case in our simulations and common in classical conditioning experiments (i.e., stimuli are either present or absent but do not vary in salience across trials). There is only one remaining difference: LMS is permitted to repeatedly cycle through the data it has experienced since training began, whereas the Rescorla-Wagner model only sees the data or trial once. Because of the similarities, it comes as no surprise that the Rescorla-Wagner model behaves in much the same way as linear regression.

The Rescorla-Wagner model correctly predicts the findings of many classical conditioning experiments. In excitatory conditioning, a stimulus, A, is followed by a reinforcing outcome (say, $\lambda = 1$), written as A+. Figure 3.1 shows that, given an initial value $V_A = 0$, V_A is increased in the Rescorla-Wagner model with repeated A+ trials until $V_A = \lambda$. Also shown, is the same conditioning simulation using LMS, which gives a similar, although sharper, learning curve. These simulations show the predictive strengths on each trial. In an equivalent classical conditioning experiment, one would measure the animal's responding at the presence of the A stimulus on each trial to get the equivalent curve.

What made the Rescorla-Wagner model novel was how it deals with multiple

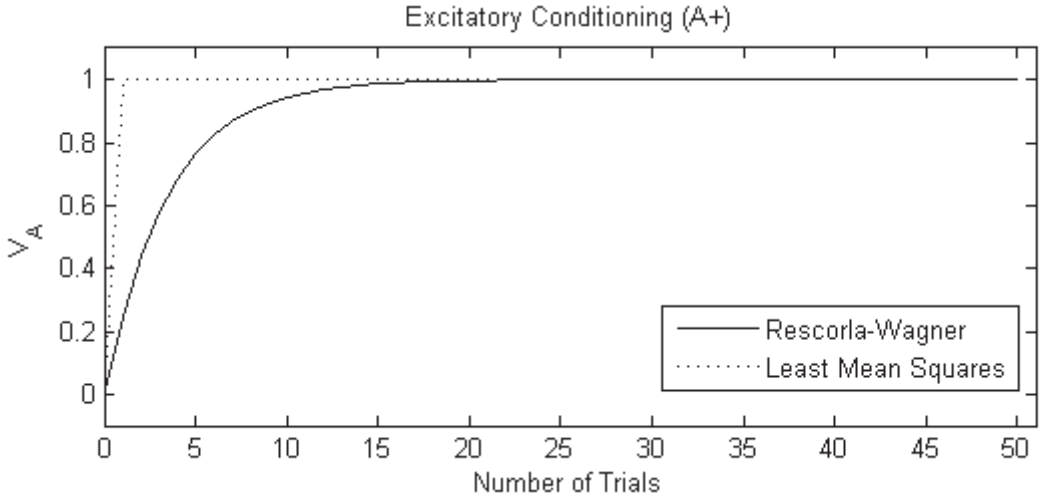

Figure 3.1: Excitatory conditioning simulated using the Rescorla-Wagner model ($\lambda = 1, \alpha_A = 1, \beta = 0.25$) and LMS. Excitatory conditioning develops a linear association between a feature of the environment and a reinforcement, where V represents the strength of the association. The sharp transition of LMS is due to the fact that it repeatedly processes the prior trials until the average trial error cannot be further reduced.

stimuli. It is able to explain a phenomenon called blocking, where one stimulus is first conditioned (i.e., repeated A+ trials), followed by an additional conditioning phase where the first conditioned stimulus is combined with a novel condtioned stimulus (i.e., repeated AB+ trials). Kamin (1968) found that animals conditioned less to the novel stimulus B when it was trained in this Blocking condition than when it was conditioned on its own (i.e., B+). As Figure 3.2 shows, the Rescorla-Wagner model (and LMS) captures this phenomenon. This phenomenon could not be explained by an earlier model of classical conditioning proposed by Bush and Mosteller (1955), which also predicted a negatively accelerated learning curve but the "error term" for a stimulus took only its own associative strength into account. A "blocked" stimulus would therefore appear as a novel stimulus and condition accordingly. The Blocking phenomenon is very similar to, but can be differentiated from, another phenomenon called "overshadowing", which is also simulated in Figure 3.2 using the Rescorla-Wagner model and LMS. In overshadowing, two stimuli are repeatedly presented and reinforced (AB+) in the first (and only) training phase. The result is that they divide the predictive strength between them, where the one with greater salience gains more

of the strength than the other. The phenomenon gets its name, however, from the fact that each stimulus receives less conditioning when conditioned in compound with another stimulus than when conditioned alone.

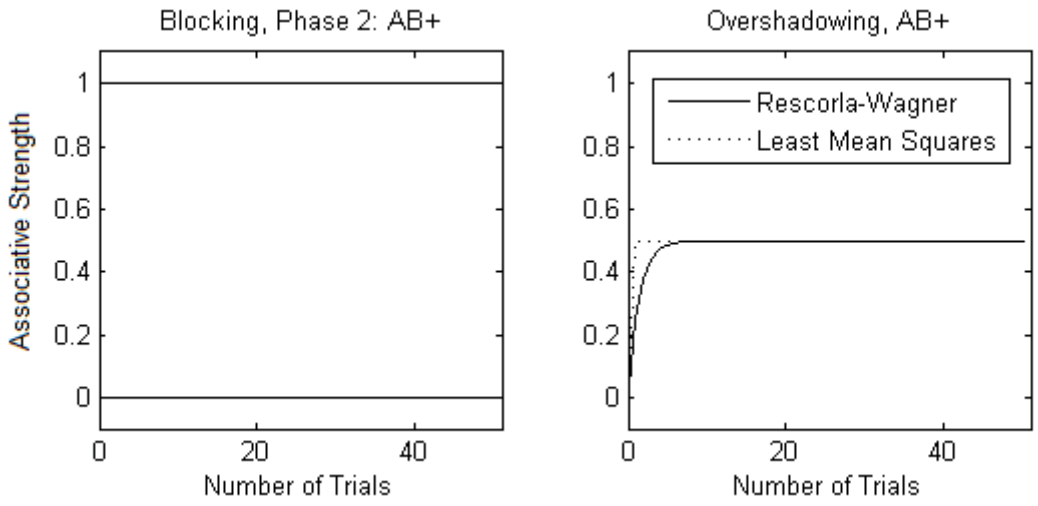

Figure 3.2: Blocking and overshadowing simulated using the Rescorla-Wagner model and LMS (least mean squares linear regression). Both approaches give similar results, having almost identical learning rules. In phase 2 of blocking (the first phase is identical to Figure 3.1), stimulus B gains no associative strength in either model because stimulus A already explains the reinforcement (i.e., there is zero error, so no values will change) in both models. Therefore, the curves for A (horizontal line at y=1) and B (horizontal line at y=0) overlap between models and are indistinguishable. In overshadowing, both models allow both equally salient stimuli to each gain half of the available associative strength supported by the reinforcement, making the individual stimuli indistinguishable in the graph, although the differences between models can be seen.

From the simulations given in Figures 3.1 and 3.2, we see that the Rescorla-Wagner model closely matches the learning ability of LMS. The Rescorla-Wagner model, however, is also known to be limited (Miller, Barnet, & Grahame, 1995) in that it cannot explain all known classical conditioning phenomena that fall within its scope. In the next two sections, we show how two groups of classical conditioning phenomena that cannot be simulated by the Rescorla-Wagner model can be simulated using LMS.

3.3 Retrospective Revaluation Phenomena: Performing Regression

The investigation of retrospective revaluation phenomena arose about a decade after the Rescorla-Wagner model was introduced. An early retrospective revaluation phenomenon is called "recovery from overshadowing" (Kaufman & Bolles, 1981; Matzel, Schachtman, & Miller, 1985). In this experiment, the first phase of conditioning is overshadowing (Phase 1: AB+). In a second conditioning phase, one of the stimuli is presented but not reinforced (Phase 2: A-). Naturally, we would expect that stimulus A would lose predictive strength, and this is seen in the experimental data. The defining feature of this phenomenon is that the data also shows that the predictive strength of the B stimulus increases (above a relevant control group). Stimulus B's predictive strength is revalued (increased) in retrospect after A has been reduced.

A major challenge to the Rescorla-Wagner model is that it cannot explain recovery from overshadowing or other retrospective revaluation phenomena. The Rescorla-Wagner model only changes the predictive strength of its present stimuli, such that in the second phase of recovery from overshadowing, stimulus A loses predictive strength but stimulus B's predictive strength remains unchanged. Figure 3.3 tells a different story for LMS, such that B's predictive strength is increased as A's strength is decreased. Again, the only difference between the LMS approach and the Rescorla-Wagner model in our simulations is that LMS repeatedly cycles through all of its training data. For the Rescorla-Wagner model, this would be the equivalent of cycling through Phase 1 and Phase 2 repeatedly until the model could accurately predict all of the trial outcomes. Table 3.1 includes an inexhaustive list of other retrospective revaluation phenomena (Shanks, 1985; Denniston, Miller, & Matute, 1996; Chapman, 1991; G. Urcelay, Perelmuter, & Miller, 2008; Blaisdell, Gunther, & Miller, 1999; Lysle & Fowler, 1985) that LMS can simulate (see Appendix A for simulation results), but not the Rescorla-Wagner model. Note that each experiment has an appropriate control group but these are not shown in the table for brevity's sake.

Whether or not retrospective revaluation may be accomplished biologically by repeatedly cycling through experiences, we will discuss in Chapter 7. For the moment, it appears that simple animal learning processes such as blocking and overshadowing and more complex phenomena like retrospective revaluation phenomena may be specific expressions of an overall system that performs regression, since LMS is a

Table 3.1: Retrospective revaluation phenomena in classical conditioning. Initial conditions begin with zero responding to the stimuli (i.e., predictive strengths are zero). Known ordinal changes in the response to each stimulus caused by the training in each phase of the experiment are also provided.

Phenomenon	Procedure	A's Change	B's Change
Backward Blocking	P1: AB+	⇑	⇑
	P2: A+	⇑	⇓
Backward	P1: AB-	0	0
Conditioned Inhibition	P2: A+	⇑	⇓
Recovery from	P1: A+	⇑	0
Forward Blocking	P2: AB+	0	0
	P3: A-	⇓	⇑
Recovery from	P1: A+	⇑	0
Conditioned Inhibition	P2: A+, AB-	0	⇓
	P3: A-	⇓	⇑

regression method. To further explore this *regression hypothesis* of biological spatial credit assignment, we consider another category of classical conditioning phenomena.

3.4 Other Revaluation Phenomena: Learning a New Predictive Strength for a Familiar Stimulus

The variety of classical conditioning behaviors that can be brought under the umbrella of a linear regression process includes a second group of phenomena that we will refer to here as other revaluation phenomena. Learning a new predictive strength for a familiar stimulus will require overcoming or overwriting a previous value or predictive strength and thus should be more difficult than learning about a new stimulus. An example phenomenon from this group is latent inhibition, which is also called Conditioned Stimulus (CS) Preexposure. Here, a stimulus is first presented but not reinforced (Phase 1: A-). In the second phase, the same stimulus is presented and reinforced (Phase 2: A+). The stimulus is thus "preexposed" before being associated with reinforcement in the second phase. The outcome is that preexposure slows conditioning relative to a control that has no preexposure (Lubow & Moore, 1959).

In terms of regression, increased predictive strengths due to reinforced trials in phase 2 are partially countered by the preexposure trials in phase 1 as the data from

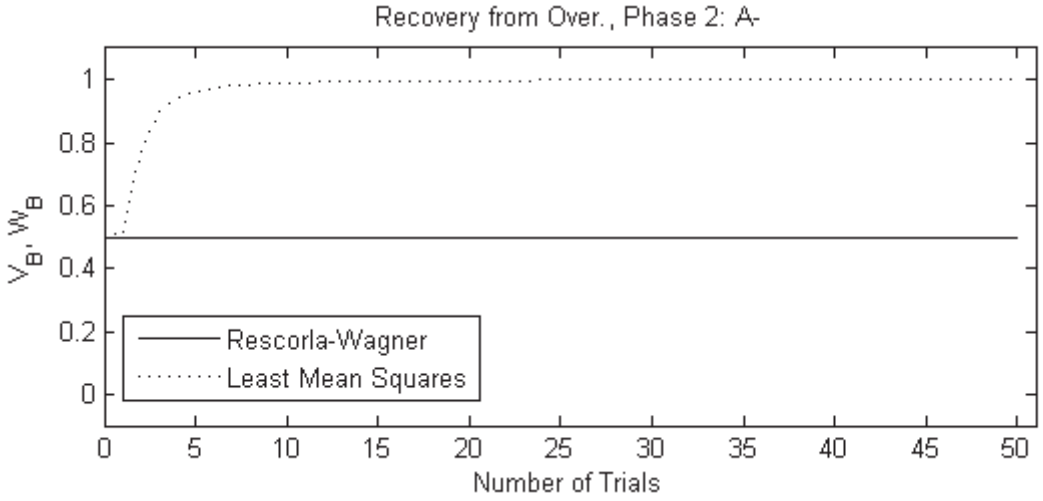

Figure 3.3: Simulation of the second phase of recovery from overshadowing using the Rescorla-Wagner model and LMS, showing stimulus B's predictive strength. The procedure for recovery from overshadowing is Phase 1: AB+, Phase 2: A-. In phase 2, the non-reinforcement of A is expected to lower its predictive strength. Recovery from overshadowing is the finding that this increases the absent stimulus, B's, predictive strength. The Rescorla-Wagner model does not change B's predictive strength, whereas LMS does increase it. The difference is that LMS internally cycles through all trial "data" prior to and including the current trial, training until it can accurately predict or explain its experience to date.

both phases are repeatedly revisited. As shown in Figure 3.4, the preexposure trials "slow down" the learning in LMS requiring a far greater number of conditioning (A+) trials to bring stimulus A's predictive strength near asymptote. For the Rescorla-Wagner model, CS preexposure has no effect (compare with Figure 3.1), counter to the classical conditioning findings. Again, the only formal difference between the two models in our simulations is that LMS repeatedly cycles through all of the trial data up to the current trial.

In some simulations, LMS changes its prediction very suddenly within the first few trials of a training phase. This is the result of the batch-training nature of the algorithm. Such abrupt changes, however, are generally uncharacteristic of conditioning findings (but see Gallistel, Fairhurst, & Balsam, 2004). One reason for this may be simply that the brain uses fewer iterations and thus does not converge as quickly as LMS. Yet, it is also possible to force LMS to "slow down" the learning curve for

a stimulus by seeding the dataset with a few instances where the stimulus is presented but not reinforced. This procedure slowed the learning in the CS Preexposure simulation (Figure 3.4).

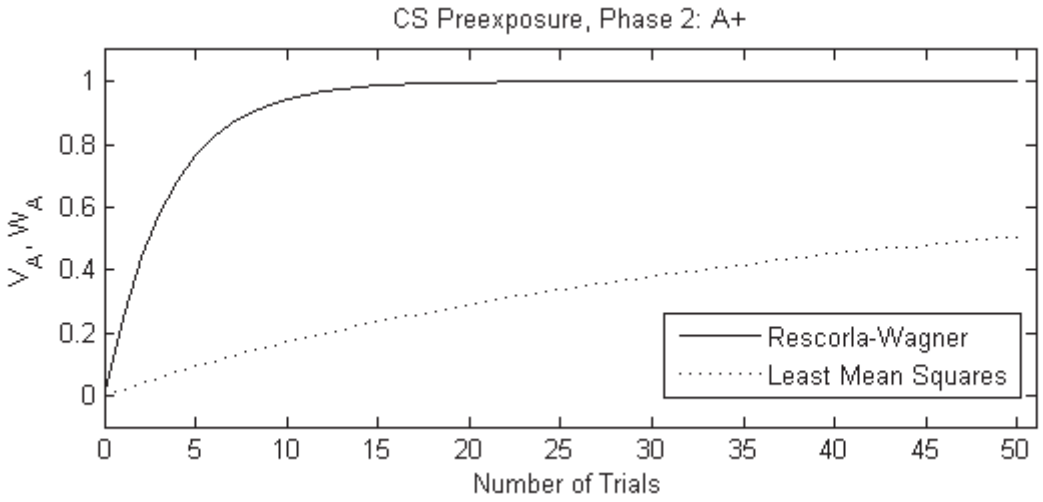

Figure 3.4: CS Preexposure simulated using the Rescorla-Wagner model and LMS. Experimentally, prior exposure to a conditioned stimulus (CS) without reinforcement slows subsequent excitatory conditioning. Again, we see that the Rescorla-Wagner model cannot explain this phenomenon whereas LMS does so by repeatedly cycling through all training data prior to and including the current trial.

Table 3.2 describes a few additional phenomena (Hall & Pearce, 1979; Rescorla, 1969) which involve stimuli whose predictive nature is changed midstream. These phenomena can also be modeled by an LMS approach (see Appendix A). The thin upward arrows reflect a smaller rate of increase in predictive strength than for a control group.

3.5 Latent Cause Normative Models

Another normative approach to explaining a variety of classical conditioning phenomena is based on determining the latent causes of CS-CS and CS-US combinations (Courville, Daw, & Touretzky, 2006; S. J. Gershman & Niv, 2012). Figure 3.5 pictorally describes the approach. For each trial, it is presumed that an unobserved or latent cause(s) induces the CS-US arrangement. This is inferred using Bayes' Rule based on experience over trials. In subsequent trials, the presentation of a CS triggers

Table 3.2: Reversal phenomena. Initial conditions begin with zero conditioning to the stimuli (i.e., predictive strengths are zero). Known ordinal changes in the response to each stimulus caused by the training in each phase of the experiment are also provided.

Phenomenon	Procedure	A's Change	B's Change
Hall-Pearce	P1: A+ (G1), B+ (G2)	⇑ (G1) 0 (G2)	0 (G1) ⇑ (G2)
Negative Transfer	P2: A++	↑ (G1) ⇑ (G2)	0
Retardation test for	P1: A+	⇑	0
Conditioned Inhibition	P2: A+, AB-,	0	⇓
	P3: B+ (G1), C+ (G2)	0	↑ (G1) ⇑ (G2)

the relevant or probable latent cause, which can then be used to predict the future outcome.

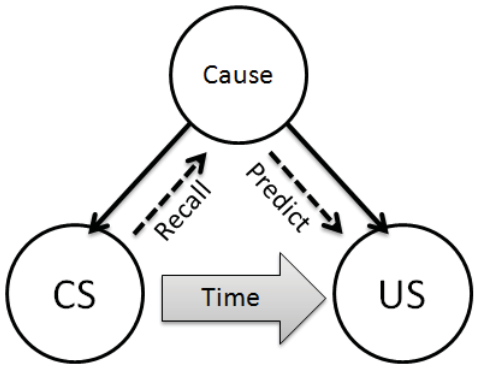

Figure 3.5: The latent cause theory of conditioning proposes that animals learn the unobservable causes of stimulus-outcome relationships. Then, when a stimulus appears, it recalls the related cause from which it can predict future stimuli or outcomes.

This normative approach is capable of explaining many diverse phenomena, from many of those captured by the Rescorla-Wagner, some configural phenomena, the revaluation phenomena described in this chapter, and more. It is an elegant theory suggesting that rather than learning to predict future outcomes based directly on the presence of stimuli, that a higher cause must first be identified from which the future outcomes can be predicted. It seems slightly less parsimonious than the regression hypothesis, however, which learns direct predictive strengths from stimulus to outcome.

The more novel a stimulus and the more volatile the world appears to be, the

more uncertain one should be about the predictive nature of a stimulus. The more uncertain the predictive nature of a stimulus, it would seem the more one should adjust its prediction based on additional, recent data. From Courville et al. (2006), the latent cause approach can deal with predictive uncertainty and change over time by influencing how quickly model parameters are able to change (faster for uncertain stimuli). In slight contrast, the regression hypothesis as defined here only more quickly learns about novel stimuli and assumes that the world does not change over time. However, it would be possible to augment the present hypothesis to weight more recent data to better reflect a changing world or include stimulus-specific learning rates that depend on the uncertainty. Doing so would substantially increase the complexity of this approach, however.

3.6 Blaisdell, Bristol, Gunther, and Miller (1998)

Although many classical conditioning phenomena appear to reflect a regression process, not all relevant phenomena in classical conditioning support this view. Blaisdell et al. (1998) found that CS preexposure and overshadowing counteract one another (see Blaisdell, Savastano, & Miller, 1999 and G. P. Urcelay & Miller, 2006 for related findings). In CS preexposure, learning is slowed because of the prior exposure. In overshadowing, the asymptotic level of learning is reduced because the association is being split among two or more stimuli. So, both processes, on their own, reduce or slow the acquisition of the predictive strength of a CS. The natural intuition is that preexposing and overshadowing the same stimulus (e.g., stimulus A in an experiment with Phase 1: A-, Phase 2: AB+) should lower its associative strength below that of either treatment alone. However, Blaisdell et al. found that preexposing a stimulus that is later overshadowed causes it to gain a larger association than if only one of these treatments is given.

This finding is significant because it seems related to a common situation encountered by biological systems. Often, we may be in an environment that on its own has no significance (i.e., CS preexposure, A-). Later on in that environment, a novel stimulus appears and is followed by reinforcement (i.e., Overshadowing, AB+). A natural intuition would be that the preexposure of the environment should hinder an increase in its associative strength, leaving the novel stimulus B to gain the majority

of the associative strength. The results of Blaisdell et al., however, would suggest the opposite: that the preexposure of the context A would lead to an enhancement in its associative strength when the overshadowing phase/trial (AB+) occurs. Yet, there is an important difference between this scenario and the procedure used in Blaisdell et al. In their work, stimulus A was a punctate stimulus rather than an environmental context. It is therefore uncertain what might occur if the context were preexposed and overshadowed stimulus in their experiment and would need to be evaluated empirically. This would help to both understand the specific results of trading a distinct stimulus for a context and also to confirm the findings of Blaisdell et al., which have been questioned by one group that found that CS Preexposure and Overshadowing do sum to further reduce conditioning in conditioned taste aversion in rats (Nagaishi & Nakajima, 2008; Nakajima, Ka, & Imada, 1999; Nakajima & Nagaishi, 2005).

Chapter 4

Machine Learning Strategies

4.1 Chapter Summary

In this chapter, we review and evaluate several machine learning techniques using the same regression task defined in Chapter 2 to show how machine learning theory addresses the key challenges to spatial credit assignment as identified earlier. In short, machine learning both allows us to determine the optimal results for our regression task and shows us that feature selection is critical to reducing prediction error when there are many irrelevant features or noise in the system.

4.2 Feature Reduction and Feature Selection

Because of its usefulness, feature reduction is a large area of research (Guyon & Elisseeff, 2003; Fodor, 2002; Saeys, Inza, & Larrañaga, 2007). The general notion of feature reduction can be broken down into two main categories: feature extraction and feature selection. Generally speaking, feature extraction takes a set of features and remaps them into a lower dimensional space, that is, as fewer features. This can be done in principled ways such as in principle components analysis (Jolliffe, 1986), where the data is reexpressed as a linear combination of a few n-dimensional axes about which there is the most variance in the data distribution. Feature extraction can also be accomplished in adhoc ways. For example, one might fit a geometric shape to an object in an image and use the parameters of best fit as features instead of the raw pixel intensities. Feature selection, on the other hand, is not about transforming the data. Instead, it chooses a subset of features to use verbatim. This is the form of feature reduction employed in spatial credit assignment since the goal is to eliminate irrelevant real-world features from consideration.

Feature selection methods can be further categorized as either filter, wrapper, or embedded methods (Saeys et al., 2007). A filter method gives a score for each feature

or subset of features, often using a statistical measure such as mutual information. Features with the highest scores are then selected according to some criterion. The Bayesian model selection approach used in the next section can be seen as a filter method, where we compute a score (the likelihood) for each combination of 2 relevant features and then select the combination with the highest likelihood. To use a wrapper method, an intermediate subset of features is selected and a score is computed from the prediction error of a chosen supervised learning method. Then, the wrapper method iteratively adds or remove features from this subset to improve the score. The specific combination of supervised learning technique and wrapper method (e.g., add or remove) influences results and the number of potential combinations is large. Finally, embedded feature selection methods are supervised learning approaches that have built-in feature selection techniques. For example, weight decay is a technique that is added to LMS to reduce the ϕ values of features that contribute little to the prediction. This feature is embedded in the LMS update by slightly changing the gradient ascent learning rule. CART is another example of an embedded feature selection technique since, for each branch, it creates a rule from a single feature.

We do not further consider wrapper methods, which serially evaluate a plethora of feature combinations, since they appear very biologically implausible. Instead, we discuss two filter methods and a family of embedded methods that can be used in conjunction with rLMS. These methods strongly relate to an embedded method we will propose and evaluate in the following chapter.

4.3 Bayesian Modeling and Optimal Feature Selection

How effective can feature selection possibly be? The optimal performance benchmark for a specific learning task comes from a Bayesian model, whenever it is analytically tractable. Unfortunately, Bayesian models are rarely biologically plausible because they can become very complicated very easily. Yet because of their benchmarking ability and theoretical importance, it is worthwhile to investigate them here.

As we saw above, simply computing the maximum likelihood estimate for rLMS' log likelihood function $log(P(y|x))$ did not take into account irrelevant features. One way of doing this is to find a way to determine which of the features are irrelevant and ignore these. For this, Bayesian modeling offers a solution.

In the regression task used earlier to evaluate rLMS, we used only 2 relevant features and a number of irrelevant features. Here, we use a Bayesian model selection approach to discern which 2 features are the relevant ones. We treat each combination of 2 features as a separate hypothesis, giving us $\binom{n}{2}$ hypotheses total. The probability (posterior) that a certain hypothesis (combination of 2 features) is responsible for generating the training data can be described using Bayes' rule,

$$P(h_j|D) = \frac{P(D|h_j)P(h_j)}{P(D)} \tag{4.1}$$

where h_j is the hypothesis under test and D represents the training data (instances of x- and y-values). Here, the model will be endowed with as much prior knowledge or assumptions as possible about how the data is generated in the task, to perform optimal feature selection and give us a prediction error benchmark for our regression task. We know that the 2 features are chosen and that this is done at random with a uniform distribution. Thus, the prior, $P(h_j) = \frac{1}{\binom{n}{2}}$, is a constant for all hypotheses. In the model selection task, only knowing which of the hypotheses is most likely is necessary. As such, it is not necessary to include the uniform $P(h_j)$. Nor is it necessary to compute P(D), since it will also be the same for all hypotheses. Thus, the posterior will be proportional to the the likelihood term, $P(D|h_j)$. This likelihood is equal to the rLMS $L(\phi)$ from Equation 2.5 except that here there are only two parameters, those associated with hypothesis h_j. So, if we want to find the best hypothesis (for rLMS and our specific simulations), we simply find the one with the highest likelihood. But this requires knowing the optimal parameter values for each hypothesis. Bayes' rule speaks to this as well.

Given the two features from a hypothesis, we create a new hypothesis space where each combination of the two ϕ values is a separate hypothesis. Conveniently, we do not need to enumerate all of these. The probability that a particular pair of ϕ values are the ones that generated the data is given by

$$P(\phi|D) = \frac{P(D|\phi)P(\phi)}{P(D)} \tag{4.2}$$

Again, we can simplify this problem by the fact that we are merely interested in picking the most likely parameters (i.e., performing a MAP estimate). This allows us to not compute $P(D)$. It is also unnecessary to include $P(\phi)$ because this probability

is uniformly distributed. This leaves us with

$$P(\phi|D) \propto P(D|\phi) \tag{4.3}$$

So, in this case, the MAP estimate is the same as the maximum likelihood estimate. This ultimately means that we can use the rLMS descent algorithm on the selected parameters to find the optimal parameter values for a given selection of two features.

So, to select the best features overall, we could first evaluate every feature-pair hypothesis by running rLMS to find the optimal parameter values for each feature pair and then select the hypothesis with the highest likelihood. However, this would be extremely computationally expensive. One way around this is to compute the "expected" likelihood by marginalizing over the possible feature values to evaluate our hypotheses. This turns out not to be straight forward, however. Marginalizing our likelihood function over the appropriate range for each of the two unknown parameters does not appear to have an analytical solution for rLMS, although one does exist for LMS. Yet, Bayes can still help us. Bayes tells us that the most likely hypothesis for the rLMS model in our simulations is the one with the highest likelihood function value. This will generally be the correct hypothesis (i.e., the hypothesis containing the truly relevant features), although a relevant parameter that contributes very little may be passed-over for an irrelevant feature that appears more relevant, and especially so when there is little data. Thus, we would say that the Bayesian feature selection technique would choose the correct hypothesis most of the time. So, if we want to define the "optimal" results for our regression task according to Bayesian theory, we can simply manually choose the correct hypothesis, find the best ϕ parameter values by gradient ascent, and record the prediction error.

Now, we can repeat the earlier rLMS simulations and determine the optimal results for our regression task. In the upper panel of Figure 4.1, we vary the number of features but maintain zero noise. The prediction errors are consistently near zero, since the fact that we manually choose the correct hypothesis does not change, regardless of the number of irrelevant features. In the lower panel, we see that prediction errors do change, slowly ramping upward with increasing noise. Both results are a substantial improvement over rLMS with many irrelevant features. Why is this? In the process of deriving Bayesian optimal parameters for a given feature-pair, we

indicated that the prior probability, $P(\phi)$, was uniform since the pair of true parameter values were drawn from a uniform distribution. This led to the realization that we could use the rLMS algorithm to find such parameters. So, although it was not said earlier, *the rLMS model assumes that each parameter is selected from a uniform distribution.* However, this is not actually the case, since the parameter values of irrelevant features, comprising most of inputs, are actually set to zero. This begs the question of whether or not using a more accurate prior distribution in rLMS will improve results. In the next section, this will be evaluated. For now, however, notice that this Bayesian feature selection process can be used to get around applying an accurate prior distribution. With feature selection, we can use the uniform prior distribution to get optimal results, when the actual prior distribution without feature selection has a spike around 0 (since the many irrelevant features are given parameter values of zero).

4.4 Bayesian Priors and Regularization

Weight decay is the idea of reducing the strength of a weight or parameter over time or with each learning update in proportion to the current weight. Intuitively, it seems it should shrink the effects of features which do not highly correlate with the outcome, providing an embedded type of feature selection method. It is also biologically plausible since synaptic weights representing model parameters could conceivably decay according to some natural process and would require only local information.

Perhaps a better way to think of weight decay is as the application of a non-uniform prior for each feature to the PDF model. A simple such prior is the Gaussian distribution. We could multiply Equation 2.4 by $exp^{-\lambda\phi_j^2}$ for each feature, j, dropping the normalizing term of the Gaussian and seeing λ as equivalent to the usual $\frac{1}{2\sigma^2}$. If we do this, and derive the learning rule using maximum likelihood estimation as before, we get

$$\frac{\partial\, log\, L(\phi)}{\partial\phi_j} = \frac{1}{\sigma^2}\sum_{i,y^{(i)}>0}(y^{(i)}-\phi^T x^{(i)})x_j^{(i)} - \sqrt{\frac{2}{\pi\sigma^2}}\sum_{i,y^{(i)}=0}\frac{e^{-\frac{(\phi^T x^{(i)})^2}{2\sigma^2}}}{1-erf(\frac{\phi^T x^{(i)}}{\sqrt{2}\sigma})}x_j^{(i)} - \lambda\phi_j \quad (4.4)$$

where the prior introduces right-most term, $\lambda\phi_j$. Here, not only are weights influenced by the data, but weights are also reduced in proportion to their size, since the Gaussian

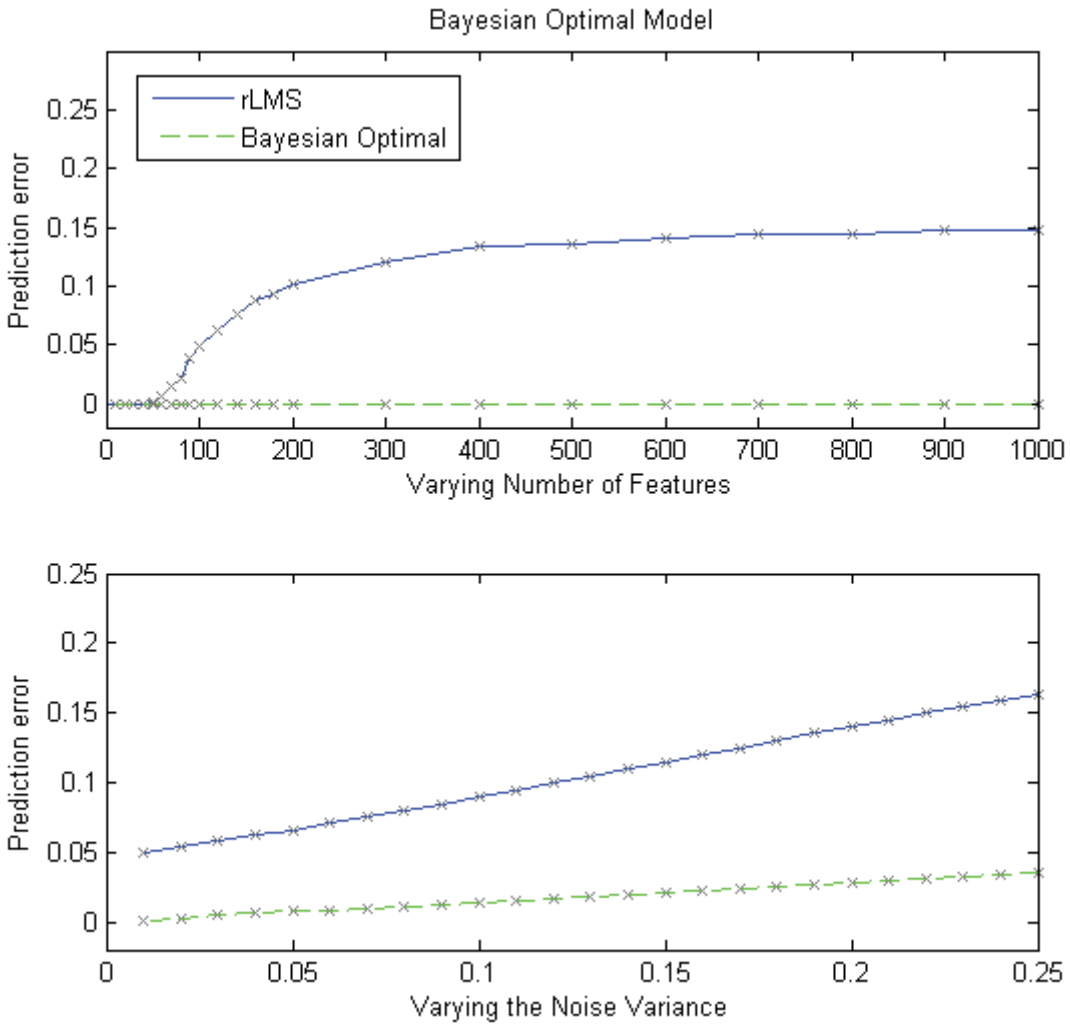

Figure 4.1: *Top Panel:* As the number of input features is increased, the Bayesian optimal model continues to achieve near zero error, since it always selects the correct 2 features and is therefore unaffected by the total number of features. *Bottom Panel:* As the variance of the additive Gaussian noise is increased, the Bayesian optimal model gives far less prediction error than rLMS.

shaped prior suggests that it is more likely that a parameter value be small than large. This prior is equivalent to "ridge regression" (Hoerl & Kennard, 1970), a special case of Tikhonov regularization (Tikhonov, 1963), which employs an L_2 norm penalty on the parameters. Krogh and Hertz (1992) showed that the optimal value of λ for LMS is the variance of the Gaussian random variable divided by the average squared generative parameter, $\lambda = \frac{\sigma^2}{\frac{1}{M}\sum_i \psi_i^2}$.

Instead of using a prior based on the Gaussian, another option is to base it on the parameters' absolute values, giving a prior of $exp^{-\lambda|\phi_j|}$. This is equivalent to the "least absolute shrinkage and selection operator" (LASSO) approach (Tibshirani, 1996), which employs an L_1 norm penalty on the parameters. This formulation is non-differentiable at parameter values of $\phi_i = 0$, such that standard gradient ascent cannot be accurately used to reach the global maximum. A number of solutions involving quadratic programming have been devised (Tibshirani, 2011). However, it is still a convex optimization problem, such that we might approximate the gradient and find local minima that are good approximations of the global minimum. In the following simulations, this direction is taken in an effort to remain more biologically plausible. Given that the initial parameter values are set to some negligibly small random value, there will be no cases of $\phi_i = 0$ with the following learning rule,

$$\frac{\partial \, log \, L(\phi)}{\partial \phi_j} = \frac{1}{\sigma^2} \sum_{i, y^{(i)} > 0} (y^{(i)} - \phi^T x^{(i)}) x_j^{(i)} - \sqrt{\frac{2}{\pi \sigma^2}} \sum_{i, y^{(i)} = 0} \frac{e^{-\frac{(\phi^T x^{(i)})^2}{2\sigma^2}}}{1 - erf(\frac{\phi^T x^{(i)}}{\sqrt{2}\sigma})} x_j^{(i)} - \lambda sign(\phi_j) \tag{4.5}$$

In ridge regression, a weight decays in proportion to its current value with each iteration, whereas in the LASSO, every weight is reduced at the same constant rate. One difference between this approximation and true LASSO solutions is that our approximation forces weights to always maintain at least some small non-zero value proportional to λ whereas true solutions can force some parameters to zero. In our approximation, parameters whose absolute values of ϕ are less than λ jump back-and-forth across zero with each iteration.

When considering the shapes of the priors implied by ridge regression and the LASSO, we see that neither of these match well the distribution from which the parameter values in the regression task are actually drawn. Figure 4.2 compares these two with an additional prior distribution, where the likelihood function is multiplied

by

$$\prod_j (z + \frac{(1-z)}{\sqrt{2\pi}\sigma_\phi} e^{-\frac{\phi_j^2}{2\sigma_\phi^2}})^\lambda \tag{4.6}$$

where z is the probability that a parameter is chosen randomly with a uniform distribution (say, between -0.5 and 0.5) and σ_ϕ is a small standard deviation of a Gaussian centered about zero. In short, this distribution represents a uniform distribution plus a sharp Gaussian distribution, suggesting that parameter values are drawn most frequently around 0 but some are drawn with a uniform distribution. This "zero-peak" prior represents the true distribution from which the parameters were drawn, and leads to the following learning rule,

$$\frac{\partial \, log \, L(\phi)}{\partial \phi_j} = \frac{1}{\sigma^2} \sum_{i, y^{(i)}>0} (y^{(i)} - \phi^T x^{(i)}) x_j^{(i)} - \sqrt{\frac{2}{\pi\sigma^2}} \sum_{i, y^{(i)}=0} \frac{e^{-\frac{(\phi^T x^{(i)})^2}{2\sigma^2}}}{1 - erf(\frac{\phi^T x^{(i)}}{\sqrt{2}\sigma})} x_j^{(i)} + \frac{\lambda}{\sigma_\phi^2} \frac{\phi_j (z-1)}{z(\sqrt{2\pi}\sigma_\phi e^{\frac{\phi_j^2}{2\sigma_\phi^2}} - 1) + 1} \tag{4.7}$$

In this scenario, weights decay rapidly around zero and almost not at all for larger values of ϕ. When $z = 1$ this approach reverts back to a uniform distribution and when $z = 0$, it is equivalent to ridge regression (if $\sigma_\phi = 1$). This is reminiscent of Elastic Nets (Zou & Hastie, 2005), which combine the LASSO and ridge regression in a similar weighted way. However, Elastic Nets do not employ sharp Gaussians since the weighted value, which takes the place of λ, never gets above 1.

For ridge regression, the LASSO, and zero peak, we repeat the earlier simulations as shown in Figures 4.3, 4.4, and 4.5, respectively. The upper panels of these figures show the results of varying the number of features when there is zero additive noise. Here, the three methods give very different results. The optimal λ for ridge regression is zero because there is zero additive noise. Therefore, in the top panel in Figure 4.3, we see that increasing λ only increases prediction errors. In contrast, increasing λ for the LASSO completely eliminates the rapid climb. Nevertheless, a λ value must be carefully chosen since larger λ values increase the base level of prediction error. Finally, we see that the zero-peak prediction error climbs in proportion to its value of σ_ϕ, providing prediction errors between ridge regression and the LASSO. The lower panels show the results of varying the variance of the additive noise while

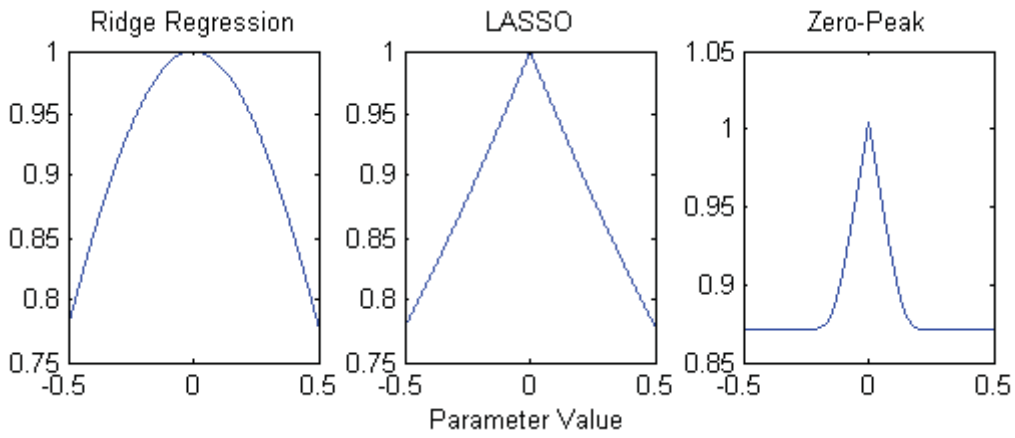

Figure 4.2: Prior distributions for a single parameter. Ridge regression and LASSO are common approaches used to reduce prediction errors. These can be viewed as priors on the parameters of a PDF model. Ridge regression and LASSO, however, do not match the true parameter distribution very well. The "zero-peak" distribution proposed here better matches the true distribution of our regression task, where a few parameters are drawn from a uniform distribution but most are set to zero. Using this prior should lower prediction errors because it matches the true distribution better than the others.

maintaining the same number of features. In all figures, by increasing the value of λ or σ_ϕ, the prediction errors are reduced for high noise levels whereas the prediction errors are increased for low noise levels. As noted earlier, in ridge regression, it has been shown that the optimal λ value depends on the true variance of the error (Krogh & Hertz, 1992). In the end, the best LASSO and zero-peak curves give smaller overall prediction errors in the lower panels than ridge regression.

In summary, ridge regression helps to improve generalization (lower test set prediction error) in the presence of noise, but does not help as the number of features gets large. In contrast, the LASSO improves generalization substantially in both cases. In particular, it generalizes extremely well when there are many features relative to the amount of data. Finally, the zero-peak prediction errors are a little lower than the LASSO in the varying noise case, but far worst when there are large numbers of features. Although technically the best fitting prior, zero-peak is inferior to the LASSO if it must consider the possibility of noise, which requires a widening of the Gaussian for effective performance. Without noise, the Gaussian can be thinner and will perform better in the varying numbers of features case.

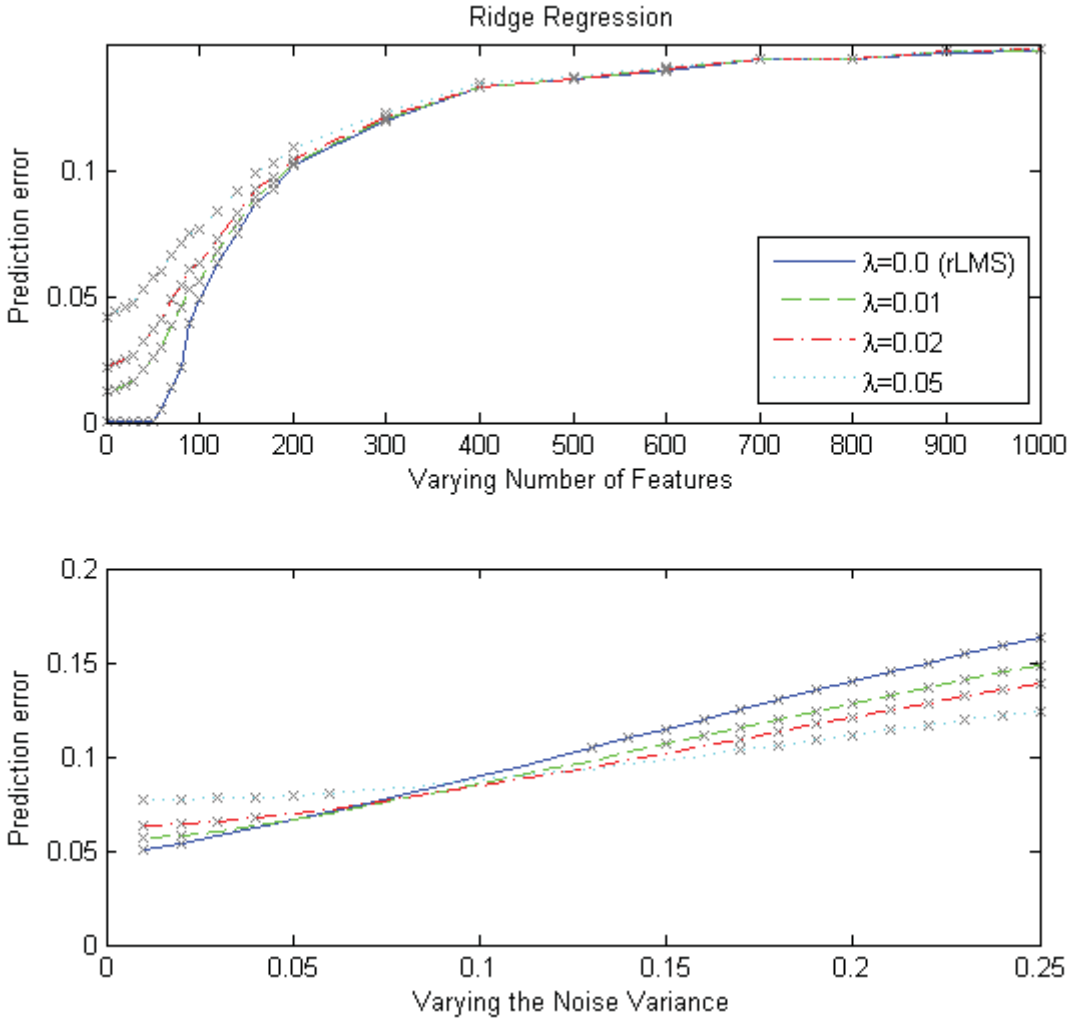

Figure 4.3: *Top Panel:* Ridge regression prediction error as the number of features is varied. Here, increasing λ only increases prediction errors. See text for explanation. *Bottom Panel:* Ridge regression prediction error as the variance of the additive noise is varied. Larger values of λ lead to lower prediction errors when noise is large but larger prediction errors when noise is small.

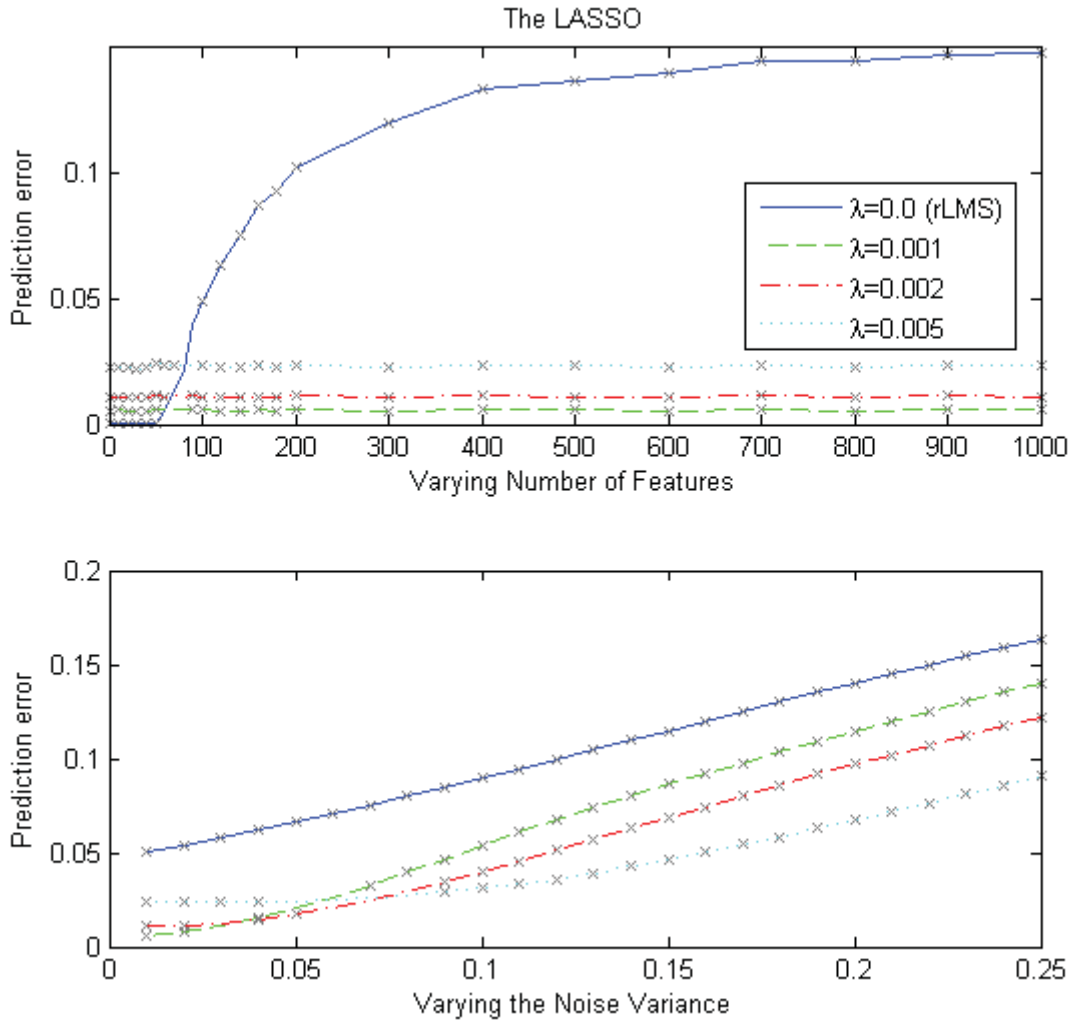

Figure 4.4: *Top Panel:* The LASSO prediction error as the number of features is varied. Here, we see that by increasing the λ value we both eliminate the dramatic climb of the prediction error and raise the baseline level of prediction error. *Bottom Panel:* The LASSO prediction error as the variance of the additive noise is varied. Larger values of λ lead to lower prediction errors when noise is large but slightly larger prediction errors when noise is small.

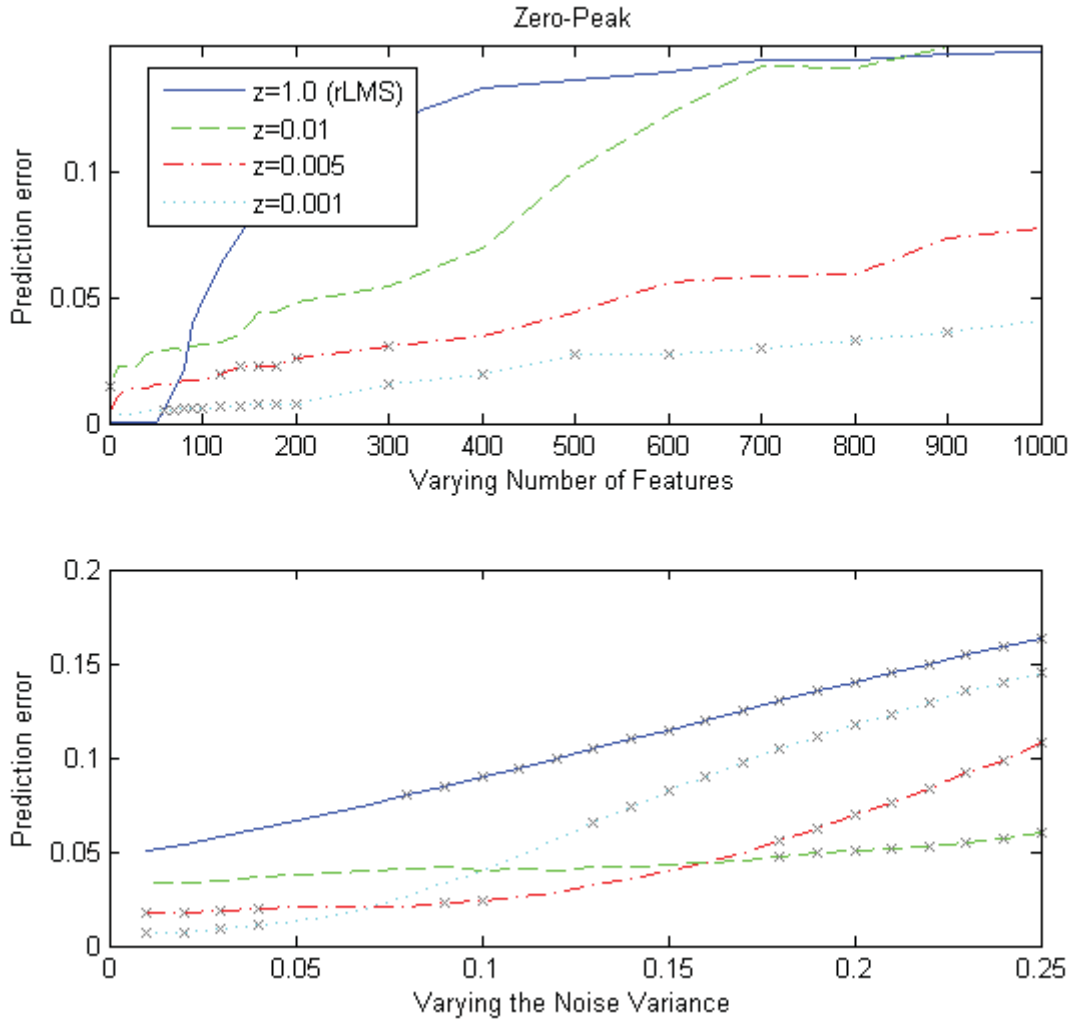

Figure 4.5: *Top Panel:* The zero-peak prediction error as the number of features is varied. Here, we see that decreasing the σ_ϕ-value reduces the dramatic climb of the prediction error. Smaller values of σ_ϕ better approximate the true distribution from which the underlying parameter values are drawn. *Bottom Panel:* The zero-peak prediction error as the variance of the additive noise is varied. It is now larger values of σ_ϕ that lead to lower prediction errors than small values. As noise increases, irrelevant parameters are seen as being more relevant and thus a wider Gaussian curve is needed (for the same learning rate) to envelope and shrink irrelevant parameter values, which leads to lower prediction errors.

4.5 Feature Selection via Feature Correlation

The correlation between an input feature and the outcome can be used to identify relevant features, falling under the filter-type feature selection method category. Correlation could be seen as being accomplished by Hebbian learning ("neurons that fire together, wire together"), which can be seen as a generalized form of spike-time dependent plasticity (Bi & Poo, 1998, 2001), a biological synaptic plasticity protocol.

The sample correlation coefficient can be used to compute the correlation directly from data,

$$r_{x_j y} = \frac{\sum_{i=1}^{m}(x_j^{(i)} - \mu_{x_j})(y^{(i)} - \mu_y)}{\sqrt{\sum_{i=1}^{m}(x_j^{(i)} - \mu_{x_j})^2 \sum_{i=1}^{m}(y^{(i)} - \mu_y)^2}} \tag{4.8}$$

where μ_{x_j} and μ_y are the means of an individual feature and the output, respectively. When $r_{x_j y} = +1$, the feature x_j perfectly correlates with the output, y. A value of -1 represents a perfectly anti-correlated feature and a 0 value represents an uncorrelated feature. This approach identifies only linear features, although such model-independent feature selection can be generalized to higher-order feature combinations (i.e., mutual information (Back & Trappenberg, 2001)).

To illustrate this feature selection technique, we again simulate the same two scenarios: one that varies the number of input features and one that varies the variance of the additive noise. Prior to training, we select the 2 features with the highest correlations/anti-correlations, eliminating all others, and then use rLMS to find the associated parameters for these 2 features. The results show that this works very well. In the both panels of Figure 4.6 the prediction error closely tracks the Bayesian optimal model's prediction error. In both models, the exact number of relevant features were specified, which gives these approaches an advantage over the others.

4.6 Data Augmentation

Given more data, we would expect rLMS to perform better. However, getting this data can be costly to sample in the real world. One possibility, however, is to synthesize some new data from existing data by adding noise. This seems biologically plausible, since there are many stages in the neuron signal transmission process by which we might expect noise to enter. If we add Gaussian noise to the input, this

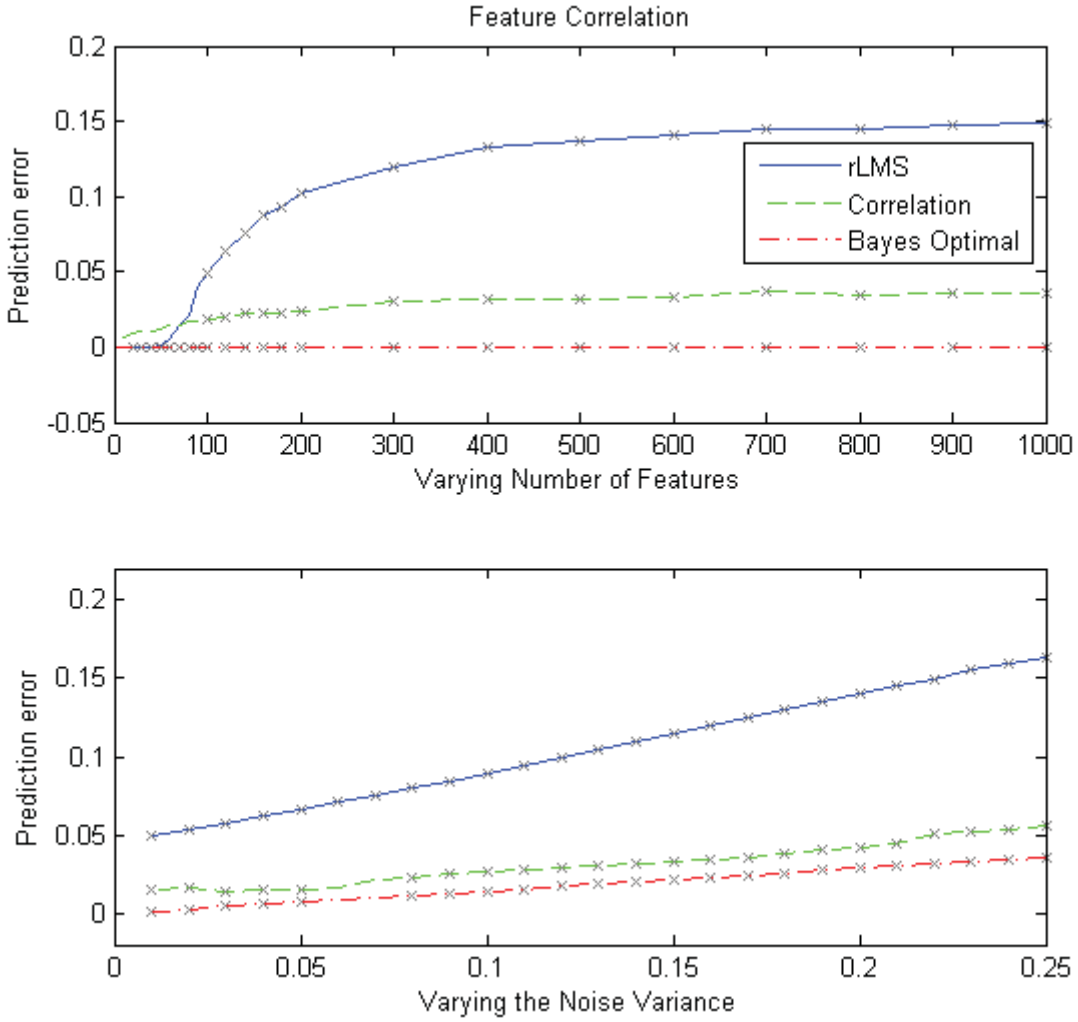

Figure 4.6: *Top Panel:* Prediction error as the number of features is varied. Feature correlation as a selection technique gives low prediction error relative to rLMS and similar, albeit larger than, the Bayesian optimal model. *Bottom Panel:* Prediction error as the variance of the additive noise is varied. Again, feature selection via feature correlation leads to a low prediction error relative to rLMS and is similar to the Bayesian model.

will multiply by the ϕ values and give us a noisy prediction. The prediction error will thereby be noisy as well, leading to a "noisy" update. Equivalently, one could simply add noise to the output to get a noisy update. Since the sum of two Gaussian random variables is another Gaussian random variable, temporary injection of artificial noise into the data is equivalent to producing a new dataset with more additive noise. We can use this process to synthesize as much data as we like.

Applying this to the simulations of rLMS, Figure 4.7 shows the results for when the number of parameters (upper panel) and the variance of the noise (lower panel) are varied. Like ridge regression, the prediction error in the upper panel either stays about the same or increases with the variance of the injected noise. Like ridge regression, it can be shown (Bishop, 1995) that data augmentation is also equivalent to a form of Tikhonov regularization. The lower panel of Figure 4.7 shows that augmenting the data set is at best insignificant (sign test, $p < 0.01$) with varying amounts of additive noise in the data.

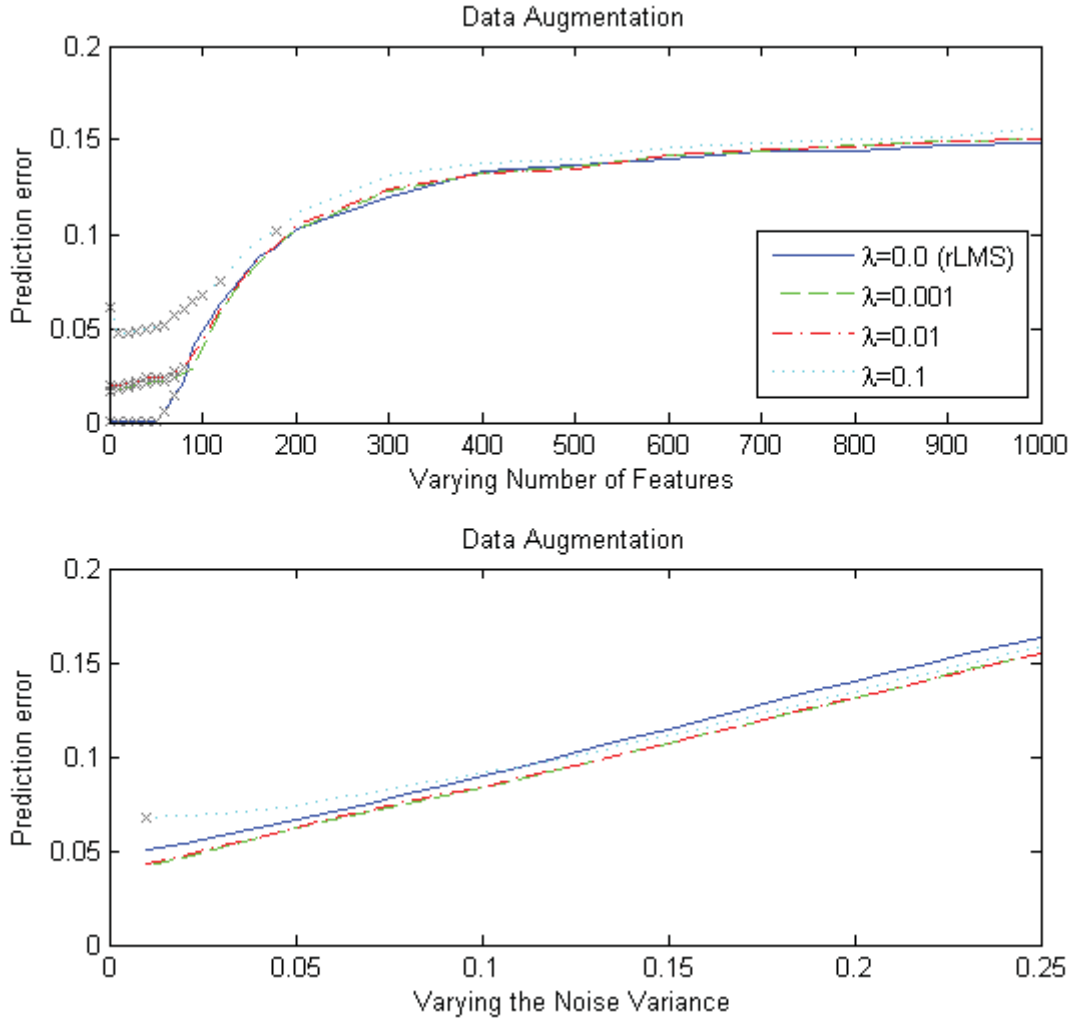

Figure 4.7: *Top Panel:* Prediction error as the number of features is varied when the training data is augmented by an unlimited supply of synthesized data, created by injecting small amounts of noise into the y-values. Like ridge regression, increasing λ does little but increase prediction errors. *Bottom Panel:* Prediction error as the variance of the additive noise is varied when the training data is augmented by an unlimited supply of synthesized data. A small variance on the noise is able to only slightly decrease the prediction error here, but even this is not significant.

Chapter 5

A Dual Pathway Approach

5.1 Chapter Summary

Here, we review and extend the Noisy OR model (Pearl, 1988) as a constrained approach to regression in the specific regression task. We describe how this approach improves results by reducing the residual parameter values for irrelevant features, something that can also be accomplished by regularization techniques. It is also shown that employing one of these solutions explains the conditioning phenomenon of relative validity, which suggests that mammals are likewise doing something to reduce their residuals as well. Some of the text in this chapter is taken from Connor and Trappenberg (2013), in which I was primarily responsible for developing the theory and simulations as well as drafting the manuscript.

5.2 The Noisy OR Model (Pearl, 1988)

For the moment, let us consider the case where we are given binary inputs representing features of the world (1 for present, 0 for absent) and binary outputs representing whether or not an outcome occurs. This is the input and output arrangement used by the Noisy OR model (Pearl, 1988) shown in Figure 5.1. The Noisy OR model, like LMS, makes predictions based on parameters inferred from data. The subtle difference is that its prediction is the probability that an outcome will occur rather than the expected value of the outcome.

The Noisy OR model is intended to capture the probability of an outcome occurring ($y = 1$), where an individual probability indicates its feature's contribution to this likelihood. To get the probability of the outcome occurring, we simply union the individual probabilities of the present features. This model is not obviously biologically plausible because it deals in probabilities which are sharply constrained (between 0 and 1) and are normalized during learning, etc.. Extensions of this model

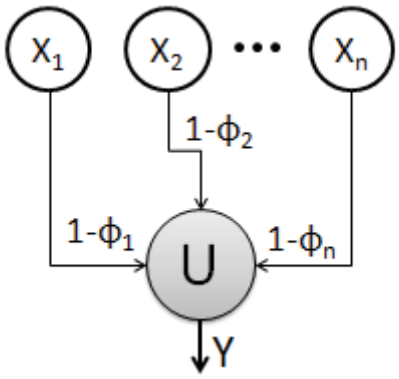

Figure 5.1: The Noisy OR model. For inputs, x_j, that are present (=1), we take the union of the associated individual probabilities $1 - \phi_j$ to get the probability of the outcome, y. Used by permission, ©2013 IEEE.

described later, however, will be shown to be more plausible.

Using an equivalent but slightly different formulation of the Noisy OR model than Pearl (1988), we write the probability or likelihood of an outcome occurring as

$$P(y = 1|x; \phi) = 1 - \prod_j \phi_j^{x_j} \tag{5.1}$$

where x is a binary input vector and ϕ is here a vector of probabilities (each represents 1 minus the probability that the associated input will be followed by an outcome). The probability of getting no outcome ($y = 0$) becomes $1 - P(y = 1|x, \phi)$. Thus the probability that a certain output (y) occurs given an input (x) is written as

$$P(y|x; \phi) = (1 - \prod_j \phi_j^{x_j})^y (\prod_j \phi_j^{x_j})^{(1-y)} \tag{5.2}$$

The Noisy OR model is probabilistic in nature, but can be seen as a sort of discrete version of the rLMS model. To visualize this, Figure 5.2 overlays the PDFs of each model for the case of a single relevant feature (no irrelevant features). The mass of each discrete point of the Noisy OR PDF (circled spikes in the figure corners) represents the mass of a representative zone in the rLMS PDF space. Formally, this means interpreting output values, y, as probabilities. For example, an output value of $y = 0.79$ is now seen as a probability of outcome of $P(y = 1) = 0.79$. So, although, the Noisy OR model was originally designed for representing the probabilities of outcomes, this reinterpretation allows it to be used for linear regression to a limited

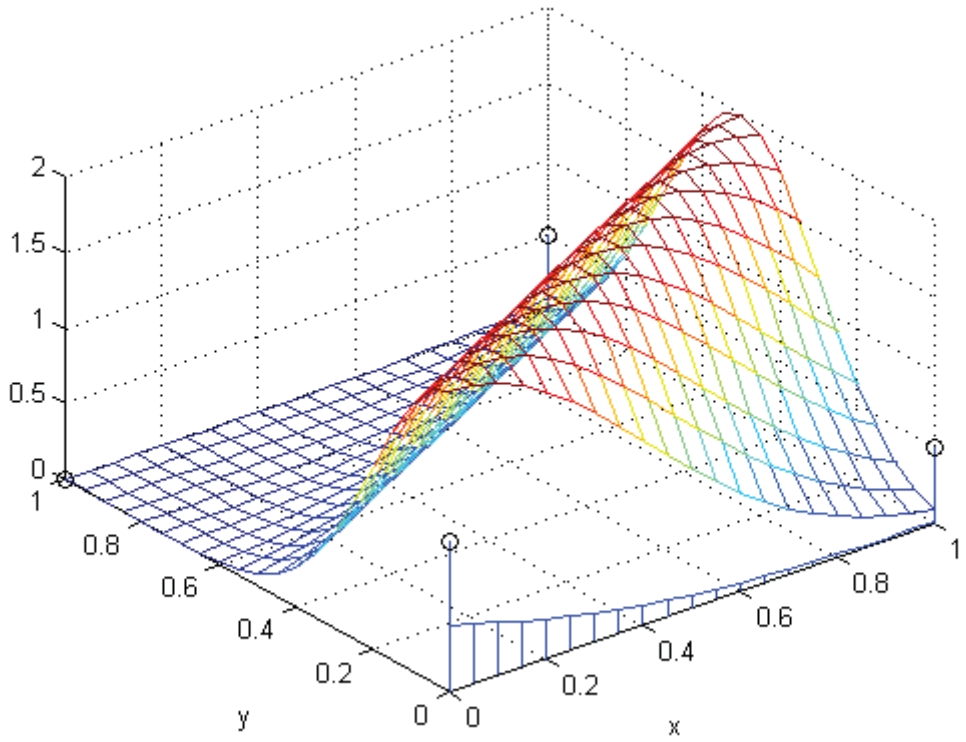

Figure 5.2: The probability density functions (PDFs) of the Noisy OR (circled spikes at the figure corners) and rLMS (wave-shape with discrete values at $y = 0$) models overlaid on one another for the case of a single relevant feature (no irrelevant features) whose underlying parameter value is 0.5. The Noisy OR model can be seen as an approximate discretization of the rLMS model, where the Noisy OR's discrete probabilities roughly express the rLMS's probability mass over the region they represent.

degree. The model's binary output is simply relaxed to include positive real values between 0 and 1 without any reformulation and an "always-on" input is added to represent the linear function bias term. Although the Noisy OR model seems to nicely fit our simulation paradigm, there is reason to expect it will perform linear regression worse than rLMS under ideal conditions because the Noisy OR model is *not linear*. Instead of a weighted sum, the Noisy OR model provides a union of probabilities, which subtracts out the "common area" covered by more than one present input. The larger the individual probabilities, the worse this non-linearity becomes.

One fundamental problem remains. The Noisy OR model has no way of representing negative linear parameters since values of $1-\phi$ principally represent probabilities.

In the following section, a solution is proposed.

5.3 The Dual Noisy OR Model

As shown in Figure 5.3, the proposed Dual Noisy OR model extends the Noisy OR model to represent the negative parameters of a linear function (i.e., global inhibitory influences) by having both a "positive" Noisy OR model that represents the probability that an outcome will occur (P_+) and a "negative" Noisy OR model that represents the probability that the outcome will be inhibited or canceled (P_-). The probability of the outcome occurring then becomes the probability of the outcome multiplied by the probability of it not being canceled or,

$$P(y|x;\phi) = P_+(1 - P_-) \tag{5.3}$$

where P_+ and P_- are expressed by Equation 5.1, where each model has its own distinct set of ϕ values. The associated likelihood function can be drawn from this, just as was done for Equation 5.1. The likelihood becomes

$$\begin{aligned} L(\phi) = \prod_{i=1}^{m}((1 - \prod_j \phi_{+,j}^{x_j^{(i)}}) \prod_j \phi_{-,j}^{x_j^{(i)}})^{y^{(i)}} \\ (1 - (1 - \prod_j \phi_{+,j}^{x_j^{(i)}}) \prod_j \phi_{-,j}^{x_j^{(i)}})^{(1-y^{(i)})} \end{aligned} \tag{5.4}$$

The gradient for the positive model parameters becomes

$$\begin{aligned} \frac{\partial \, log \, L(\phi)}{\partial \phi_{+,k}} &= -\sum_{i=1}^{m} x_k^{(i)} \frac{\prod_{j \neq k} \phi_{+,j}^{x_j^{(i)}} (y - (1 - \prod_j \phi_{+,j}^{x_j^{(i)}}) \prod_j \phi_{-,j}^{x_j^{(i)}})}{(1 - \prod_j \phi_{+,j}^{x_j^{(i)}})(1 - (1 - \prod_j \phi_{+,j}^{x_j^{(i)}}) \prod_j \phi_{-,j}^{x_j^{(i)}})} \\ &= -\sum_{i=1}^{m} x_k^{(i)} \frac{\prod_{j \neq k} \phi_{+,j}^{x_j^{(i)}} (y - P(y|x;\phi))}{(1 - \prod_j \phi_{+,j}^{x_j^{(i)}})(1 - P(y|x;\phi))} \end{aligned} \tag{5.5}$$

and the gradient for the negative model parameters becomes

$$\begin{aligned} \frac{\partial \, log \, L(\phi)}{\partial \phi_{-,k}} &= \sum_{i=1}^{m} x_k^{(i)} \frac{(y - (1 - \prod_j \phi_{+,j}^{x_j^{(i)}}) \prod_j \phi_{-,j}^{x_j^{(i)}})}{\phi_{-,k}(1 - (1 - \prod_j \phi_{+,j}^{x_j^{(i)}}) \prod_j \phi_{-,j}^{x_j^{(i)}})} \\ &= \sum_{i=1}^{m} x_k^{(i)} \frac{y - P(y|x;\phi)}{\phi_{-,k}(1 - P(y|x;\phi))} \end{aligned} \tag{5.6}$$

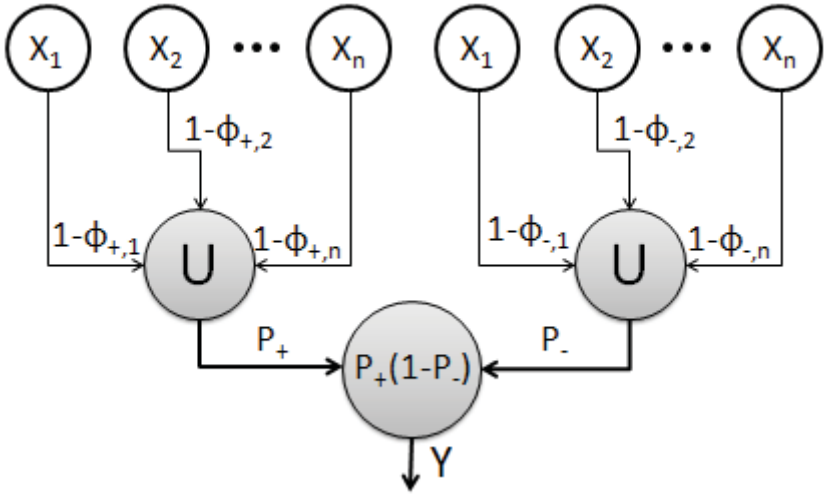

Figure 5.3: The Dual Noisy OR model is an extension of the Noisy OR model that is able to incorporate global inhibitory influences and thereby represent the negative parameters of a linear function. It is essentially two Noisy OR models, where one represents the probability that an outcome will occur and the other represents the probability that the outcome will be canceled. These models join in predicting the outcome as the probability of occurrence multiplied by the probability that the outcome is not canceled. Used by permission, ©2013 IEEE.

which are used to update model parameters, as in Equation 2.8.

Figure 5.4 shows results for the regression task comparing the Dual Noisy OR model with rLMS, the LASSO, and the optimal benchmark. The Dual Noisy OR model performs well relative to rLMS in both simulations, whether with or without regularization (the LASSO), even though the Dual Noisy OR model is not the optimal model for this task. The rLMS world was chosen as the testbed for the present work as a useful compromise between incorporating aspects of the real-world and being able to relate sufficiently with standard machine learning algorithms and principles. However, it is possible that the real-world is less like the continuous, rLMS model and more like the discrete, Dual Noisy OR model. In such a case, we would expect the Dual Noisy OR model to be even more effective than shown here. Another potential advantage of this model is that it does not require knowledge of the additive noise variance, unlike how rLMS does (to be optimal).

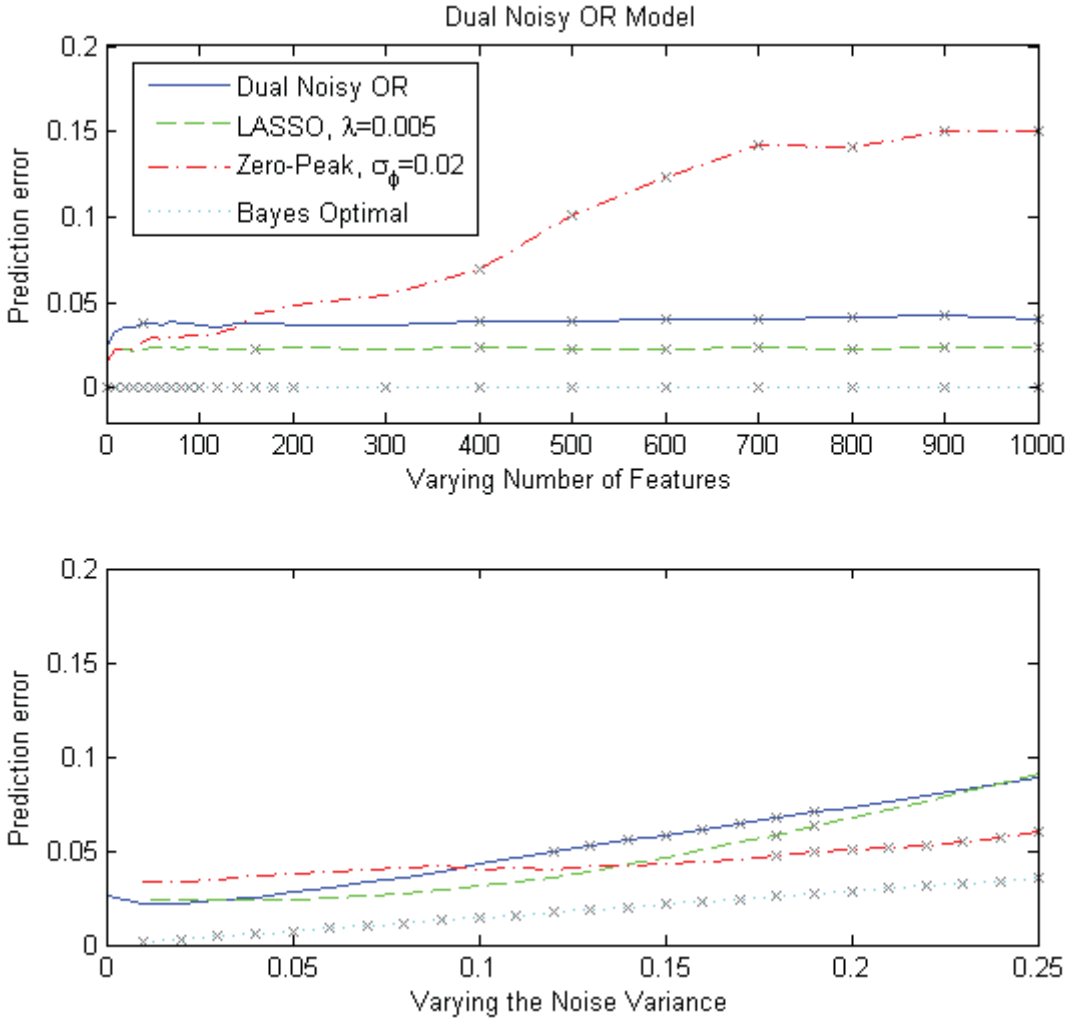

Figure 5.4: *Top Panel:* Test set prediction error of the Dual Noisy OR model, the LASSO, and zero-peak as the number of features is varied. The Dual Noisy OR has a slightly larger average prediction error than the LASSO, but far less than zero-peak at large numbers of irrelevant features. *Bottom Panel:* Test set prediction error of the Dual Noisy OR model, the LASSO, and zero-peak as the variance of the additive output noise is varied. The Dual Noisy OR model is comparable to the other methods except that zero-peak shows significantly less error at high levels of noise.

5.4 Eliminating a Primary Source of Generalization Error

When examining the math of the Noisy OR model, we see that $P(y|x;\phi)$ increases by increasing values of $1-\phi_j$ for $x_j = 1$ when $y = 1$, that is, by increasing the influence of features that correlate with the output. $P(y|x;\phi)$ also increases by decreasing values $1-\phi_j$ for $x_j = 1$ when $y = 0$, that is, by decreasing the influence of the lesser correlated and uncorrelated features. Something very similar happens in rLMS. Its $P(y|x;\phi)$ increases and decreases as ϕ values are increased and decreased for $x_j = 1$ when $y = 1$ and $y = 0$, respectively. Thus, maximizing $L(\phi)$ in the Noisy OR model and rLMS maximizes the likelihood of the input generating the output by increasing the influence of the features that correlate most with the output and by reducing (and in rLMS, even making negative) the influence of features that least correlate with the output. This process will be referred to as relative correlation.

There is evidence of relative correlation in classical conditioning. Simultaneous Feature Positive Discrimination (SFPD) (Ross & Holland, 1981) is a conditioning phenomenon where there are two stimuli which are correlated with reinforcement. In this single phase experiment, reinforced presentations of a compound stimulus (AX+) are alternated with non-reinforced presentations of one of the elements (X-). Stimulus A is followed by reinforcement every time it is presented, whereas stimulus X is followed by reinforcement only half of the time. After training, the animal learns that the most highly correlated stimulus predicts the reinforcement and that the other stimulus does not, even though it is correlated with reinforcement half of the time. We know that this is not because a partially reinforced stimulus cannot be conditioned, since it can be (Jenkins & Stanley Jr, 1950).

The simulation of SFPD using the Rescorla-Wagner model, shown in Figure 5.5, gives a useful explanation of why this may occur in animals. In the AX+ trials, both stimuli increase in associative strength. In the X- trials, stimulus X decreases in associative strength. If the associative strength of the AX compound reaches the asymptotic level, a subsequent X- trial will reduce this. The next AX+ trial will have new room for associative strength to grow once again and A and X will split the gains. This cycle of stimulus X losing in X- trials, while gains are split between A and X in AX+ trials leads to A slowly draining all of X's associative strength. This is relative correlation in action, where the most highly correlated stimulus steals away

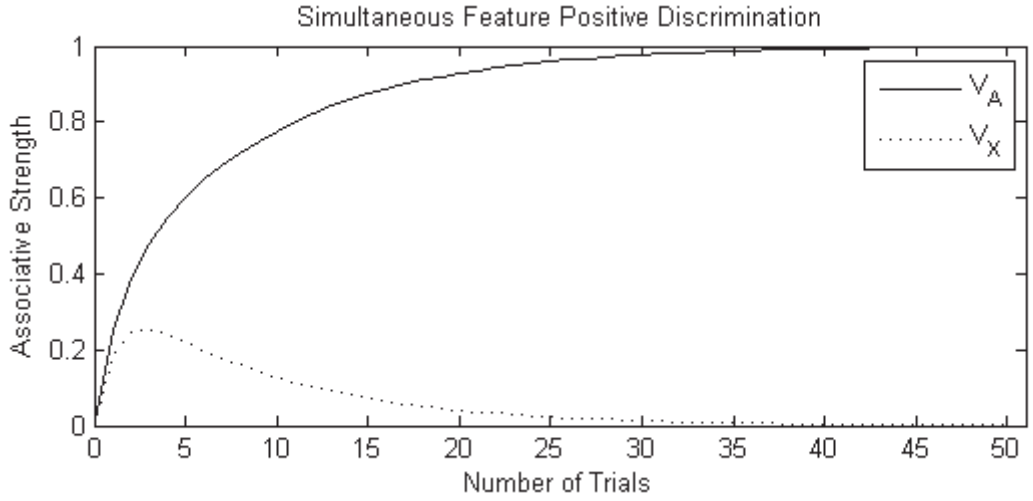

Figure 5.5: Simulation of simultaneous feature positive discrimination (SFPD) using the Rescorla-Wagner model. The procedure is described by two trial types in a single phase experiment, AX+, X-. Although in early trials, the associative strength of X is increased, it is eventually extinguished in favour of A even though X is reinforced half of the time. In this simulation, the associative strength to X is reduced in X- trials and split between A and X in AX+ trials such that A slowly steals all of X's early associative strength.

the associative strength of the lesser correlated stimulus. Indeed, it could be said that X is an irrelevant feature and that its partial correlation with reinforcement was circumstantial, due only to its simultaneous presentation with A.

Problems arise in the Rescorla-Wagner model in regard to relative correlation as we begin to add more irrelevant features. Simulating the phenomenon of relative validity (Wagner et al., 1968) captures this. In relative validity, there are two groups, each with three stimuli: A, B, and X. In group "correlated" we have the trials AX+, BX-, AX+, BX-, such that A and B are perfectly correlated with reinforcement and non-reinforcement, respectively. Stimulus X, however, is only partially correlated. In group uncorrelated, we have the trials AX+, BX-, AX-, BX+, where all of the stimuli are equally partially reinforced (50%), except that X is presented twice as often. Wagner et al. (1968) discovered that responding to X in group correlated was less than in group uncorrelated. This is just what relative correlation would predict since X is less correlated with the reinforcement than A. Figure 5.6 shows the result of simulating this phenomenon with the Rescorla-Wagner model. In neither

simulated group is the associative strength of X mostly extinguished[1] as occurred in the case of SFPD and as occurs in relative validity experiments as well (Wagner et al., 1968; E. A. Wasserman, 1974; Pearce, Dopson, Haselgrove, & Esber, 2012). After an AX+ trial increases the predictive strength for A and X, a BX- trial reduces the associative strength of B and X. Since B is never reinforced, B only decreases and becomes an inhibitor. At asymptote, the associative strength of A and X sum to predict the full reinforcement and B becomes inhibitory enough to cancel the residual association in X so that BX predicts zero reinforcement. Such a residual association in an apparently irrelevant or non-predictive feature becomes a bigger problem as we increase the number of irrelevant features.

This can be illustrated by examining the parameters of rLMS after being trained in the regression task. Figure 5.7 displays one such case (100 data points, 99 features, 0.0 noise variance). Although there is no additive noise, irrelevant parameters retain a small residual value. When there are many parameters, a new random binary activation of the inputs can add up and significantly throw off a prediction, leading to substantial prediction error.

How can we solve this problem? One simple solution is providing more data. In principle, the optimal rLMS model will give the best result with sufficient data. Since biological systems usually do not have this luxury, we must consider alternatives. The use of priors or regularization methods we examined in the previous chapter battled this issue by reducing all parameter values to varying degrees. In the case of the LASSO, every update involved a constant reduction in the absolute value of all parameter values. In this way, such residual parameter values were extinguished, allowing the relevant features to acquire from the loss and improving generalization. The reason for the Dual Noisy OR model's success is twofold. First, the model never has negative parameters, since they are probabilities. This avoids the problem of positive and negative parameter values canceling one another to support residual values. Now, a small residual negative pathway probability can only slightly decrease the total prediction probability because it is applied multiplicatively instead of additively (e.g., 0.1(1 - 0.1) instead of 0.1 - 0.1). Second, Equation 5.6 indicates that the change

[1]The Rescorla-Wagner model can explain relative validity if it uses a different learning rate for trials on which there is reinforcement than for trials on which there is none (Rescorla & Wagner, 1972), but in keeping with LMS theory, this detail was not incorporated here.

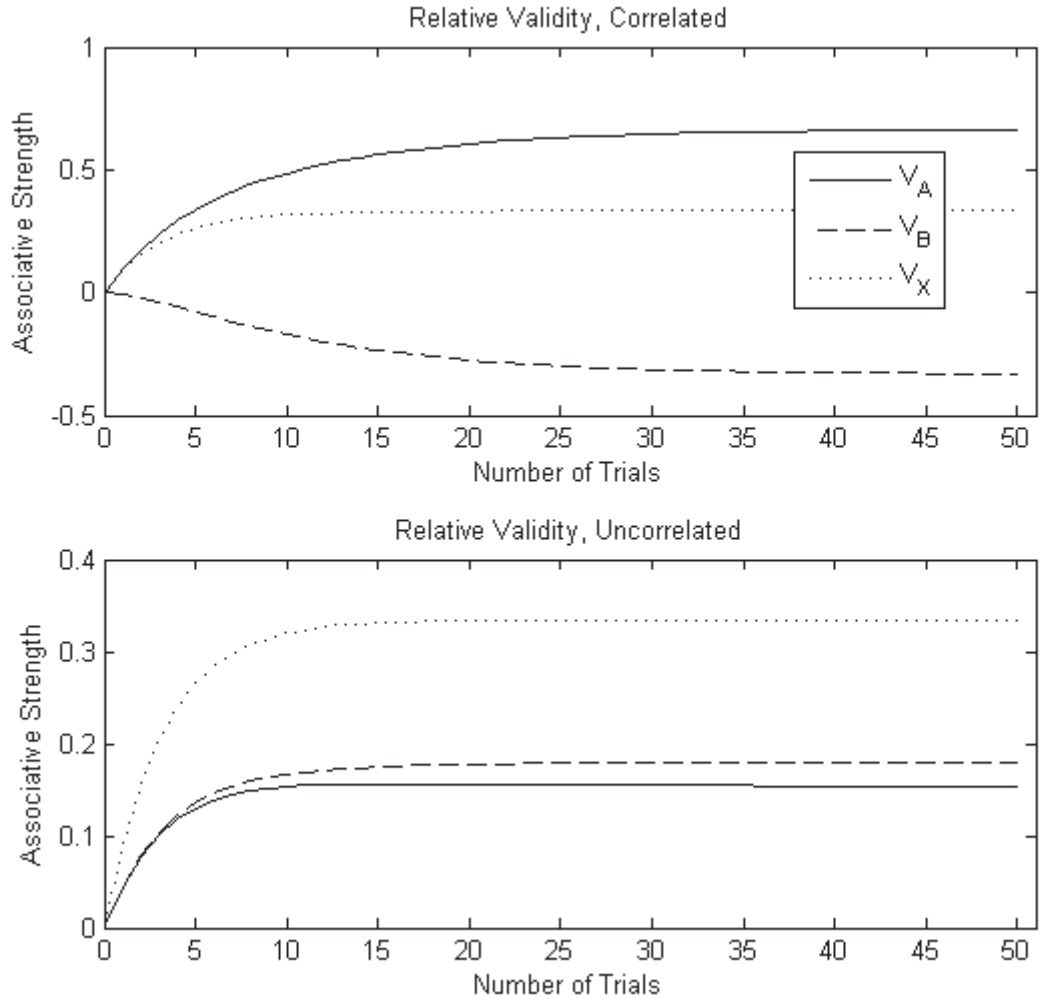

Figure 5.6: Simulation of relative validity using the Rescorla-Wagner model. *Top Panel:* The associative strengths of the stimuli in group "correlated", which receive AX+, BX-, AX+, and BX- trials in a single phase of conditioning. *Bottom Panel:* The associative strengths of the stimuli in group "uncorrelated", which receive AX+,BX-,AX-, and BX+ trials in a single phase. Notice that the asymptotes for V_X are the same in both groups, despite that X could be perceived as an irrelevant predictor in group correlated.

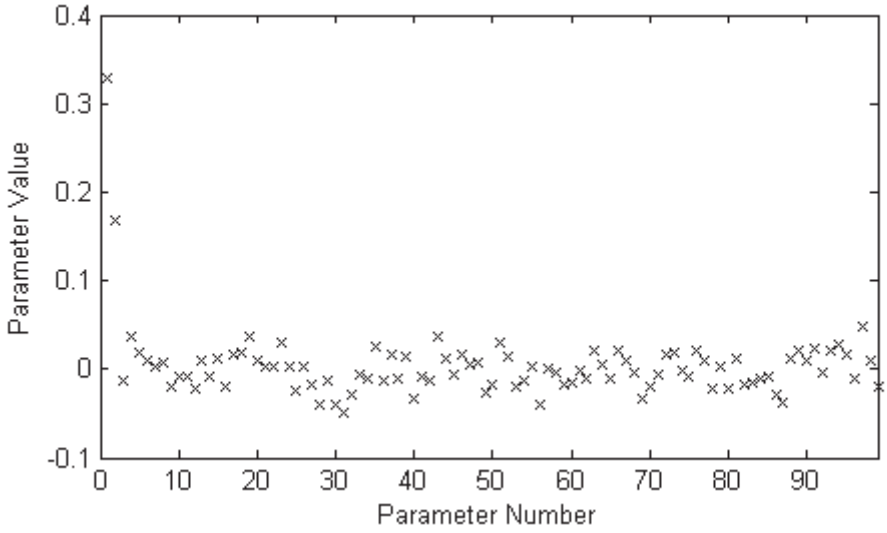

Figure 5.7: Residual parameter values after training rLMS with 99 features, 100 data points, and zero noise. The two relevant features are on the far left and have generative values of 0.4 and 0.2. We see that these parameter values do not quite reach their true generative values and that the irrelevant features retain a residual parameter value, even though they have generative values of zero. These residual values cancel-out one another so that the prediction error is zero on training set input vectors where relevant features are absent. With test data, however, they add noise to the predictions and increase prediction errors.

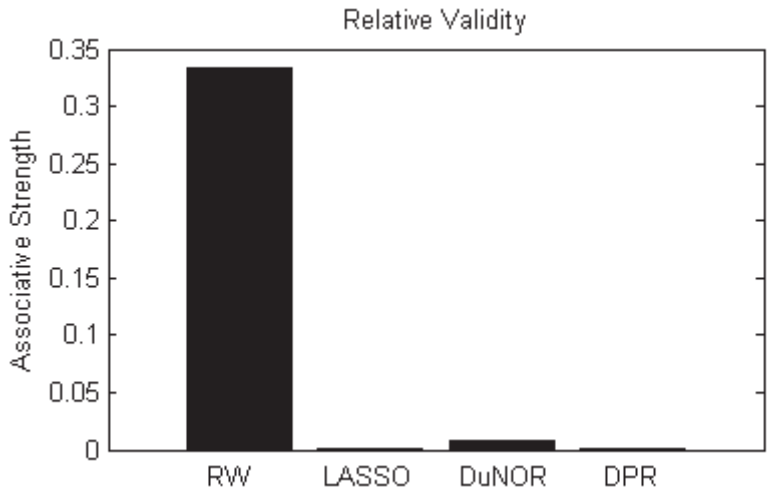

Figure 5.8: The associative strength or parameter value of stimulus X following correlated relative validity conditioning (AX+, BX-) for the Rescorla-Wagner (RW) model/LMS, the LASSO, Dual Noisy OR (DuNOR), and Dual Pathway Regression (DPR) (introduced in Chapter 8). All methods except the Rescorla-Wagner model effectively extinguish this irrelevant stimulus.

in negative parameter values is proportional to the total model prediction[2]. This means that the small positive predictions made by irrelevant positive residuals will have a small learning rate, further discouraging negative residuals from forming.

So, residual associations to irrelevant parameters are reduced when either regularization is used or negative parameters are not permitted and there is a multiplicative way of representing inhibition. Consequently, this leads to a simulation of relative validity that matches what we see in classical conditioning experiments. Figure 5.8 shows the asymptotic associative strength of X in group correlated in simulations of relative validity using the LASSO regularization, Dual Noisy OR, and Dual Pathway Regression (a derivative of Dual Noisy OR that is the focus of Chapter 8). In all cases, the association to X is extinguished, unlike in the Rescorla-Wagner/LMS case.

That relative validity animal learning experiments (Wagner et al., 1968; E. A. Wasserman, 1974; Pearce et al., 2012) mostly extinguish the irrelevant stimulus, and approaches that implicitly assume a uniform prior distribution on stimulus relevancy (e.g., the Rescorla-Wagner model, rLMS) do not, suggests that biological systems

[2]Changes in the positive pathway are also proportional to the full model prediction. However, they are also inversely proportional to the positive pathway prediction (which is large whenever the total model prediction is large), such that these two terms counteract one another.

must be using some strategy (e.g., regularization or non-negative parameterization with multiplicative inhibition) to reduce residual associations for irrelevant parameters.

Chapter 6

The Neuroscience of Stimulus Learning

6.1 Chapter Summary

In this chapter, we seek to describe the neurobiology after which the subsequently presented models of spatial credit assignment will be fashioned. In particular, the salient features of the basal ganglia are described with a focus on the striatum. Requiring models to fit this specific neurobiology narrows the possible spatial credit assignment solutions. We will later rely on this exposition to compare and contrast these models in terms of their biological faithfulness.

6.2 Basal Ganglia Anatomy

6.2.1 Organization of Nuclei

The basal ganglia are a group of interconnected nuclei in the midbrain, as shown in Figure 6.1. This collection of nuclei has input and output stages. Specifically, the striatum and subthalamic nucleus serve as the input regions and the substantia nigra pars reticulata (SNr) and the globus pallidus interna (GPi) serve as the output regions. The globus pallidus externa (GPe) receives inputs from and projects to other basal ganglia nuclei. The substantia nigra pars compacta (SNc) and ventral tegmental area (VTA) receive projections as though they were output nuclei but provide special feedback connections to the striatum from their dopaminergic neurons. The basal ganglia receive inputs from all over the neocortex, providing them with the opportunity to integrate CSs from widely varying sources. Other inputs include the amygdala, the hippocampus, and the thalamus. In the following subsections, we describe some of the major features of the basal ganglia. For additional details, the reader is referred to Wilson (2004) and Parent and Hazrati (1995a, 1995b).

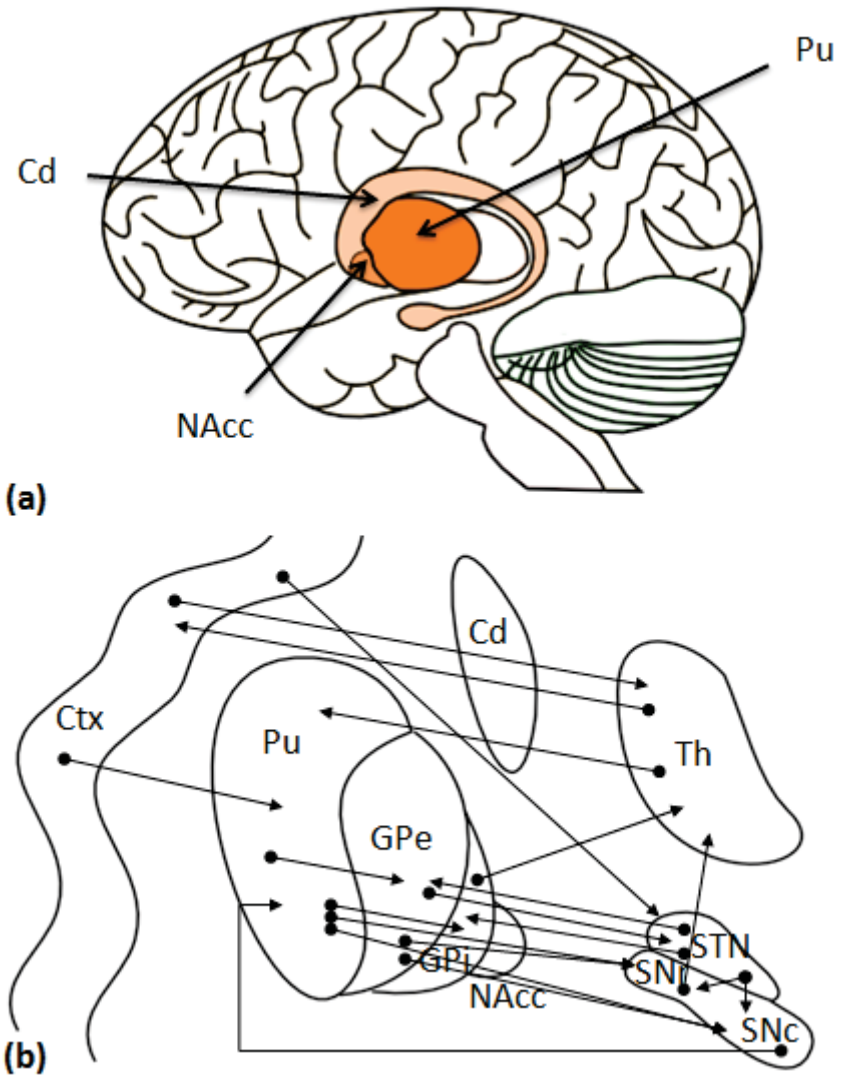

Figure 6.1: (a) A side-view of the striatal nuclei situated among other brain regions: Cd - caudate, Pu - putamen, and NAcc - nucleus accumbens. (b) A rear-view (a composite coronal slice) schematic of basal ganglia nuclei, the thalamus, and the cortex and their interconnections. Of the striatal regions, only the putamen is connected for clarity's sake. Note that the connections are highly-schematic and do not represent the breadth of the connectivity between areas. Abbreviations: Ctx - neocortex, Th - thalamus, GPe - globus pallidus externa, GPi - globus pallidus interna, STN - subthalamic nucleus, SNr - substantia nigra pars reticulata, SNc - substantia nigra pars compacta

6.2.2 Basal Ganglia Pathways

The striatum is mostly composed of medium spiny neurons (MSpNs), which receive excitatory input primarily from the cortex and thalamus and project to downstream basal ganglia nuclei. The striatum also contains several types of interneurons that are believed to help sculpt MSpN activity (Tepper, Bolam, et al., 2004; Tepper, Wilson, & Koos, 2008; Pakhotin & Bracci, 2007), though we do not discuss them at length here. As shown in Figure 6.2, each MSpN projects along one of two major routes commonly known as the direct and indirect pathways. Direct pathway MSpNs connect directly to the basal ganglia output nuclei (GPi and SNr) and SNc/VTA while the indirect pathway MSpNs project first to the GPe before arriving at these same destinations. An unusual feature of these projections is that they are all inhibitory, including the influence of the output nuclei on the main target, the thalamus. Because the indirect pathway involves three sequential inhibitory projections (Striatum→GPe→GPi/SNr→Thalamus), its net effect on the thalamus is inhibitory, while the direct pathway projection has only two sequential inhibitory projections (Striatum→GPi/SNr→Thalamus), causing a net excitatory effect. Thus, these two pathways tend to have opposing effects. The direct and indirect pathways differ in a few ways besides their different routes through the basal ganglia. MSpNs primarily express either D1 (direct pathway) or D2 (indirect pathway) dopamine receptors (Surmeier, Song, & Yan, 1996; Wilson, 2004). They receive cortical inputs from neurons in either the more superficial (direct pathway, layer 3 and upper layer 5) or deep layers (indirect pathway, layer 5) whose axons have either broad (direct pathway) or narrow (indirect pathway) branching patterns in the striatum, as found in rats (Lei, Jiao, Del Mar, & Reiner, 2004; Cowan & Wilson, 1994). There is a further dimension of organization within each of the direct and indirect pathways according to each final destination nucleus. In the striatum there are so called "matrix" and "striosomal" zones which project ultimately to either the SNr/GPi or SNc/VTA, respectively (Joel & Weiner, 2000). One challenge to the direct and indirect pathway dichotomy is that MSpNs of the direct pathway also send collateral projections to the GPe (Kawaguchi, Wilson, & Emson, 1990). While this seems to cloud the distinction,

we will later provide a potential algorithmic interpretation of this feature that may help in maintaining the dual pathway distinction.

A "hyper-direct" pathway has also been proposed (Nambu, Tokuno, & Takada, 2002), where the cortex provides input to the subthalamic nucleus, which excites many basal ganglia nuclei. It has been suggested that the hyper-direct pathway is useful in canceling actions and suppressing competing motor actions besides the one voluntarily selected. It provides a direct and diffuse route for exciting activity in the GPi, which inhibits the thalamus and other subcortical nuclei.

6.2.3 Lateral Inhibition in the Striatum

The projection neurons of the striatum send out lateral inhibitory (Plenz, 2003; Tunstall, Oorschot, Kean, & Wickens, 2002) axon collaterals which freely contact other MSpNs, within and between the direct and indirect pathways (Yung, Smith, Levey, & Bolam, 1996). Some MSpN collaterals cover a local area, whereas others branch out relatively broadly (Kawaguchi et al., 1990). There appears to be, at most, about a $\frac{1}{3}$ probability of one neuron sending an inhibitory connection to another (Taverna, Ilijic, & Surmeier, 2008). As a result, these lateral inhibitory connections are usually one-way (at best only 1 in 9 connections are reciprocated in a randomly connected network). Although profuse, the lateral inhibitory connections are believed to be weak compared with the inhibition of fast spiking inhibitory interneurons (Tepper et al., 2008; Gruber, Powell, & O'Donnell, 2009).

6.2.4 Division of Labour in the Striatum

Another feature of the basal ganglia is that there are parallel, functionally segregated channels passing through it (Alexander, DeLong, & Strick, 1986). Segregation is fine enough that somatotopic organization exists for both somatosensory (Flaherty & Graybiel, 1991) and motor cortical inputs. These channels exist for both the direct and indirect pathways and allow for focused computations.

A number of basal ganglia models (Mink, 1996; Gurney, Prescott, & Redgrave, 2001; Humphries, Stewart, & Gurney, 2006; Frank, 2005; Houk, Adams, & Barto, 1995) support the notion that the basal ganglia takes part in the selection of appropriate actions by lowering its tonic inhibition in channels associated with the action(s)

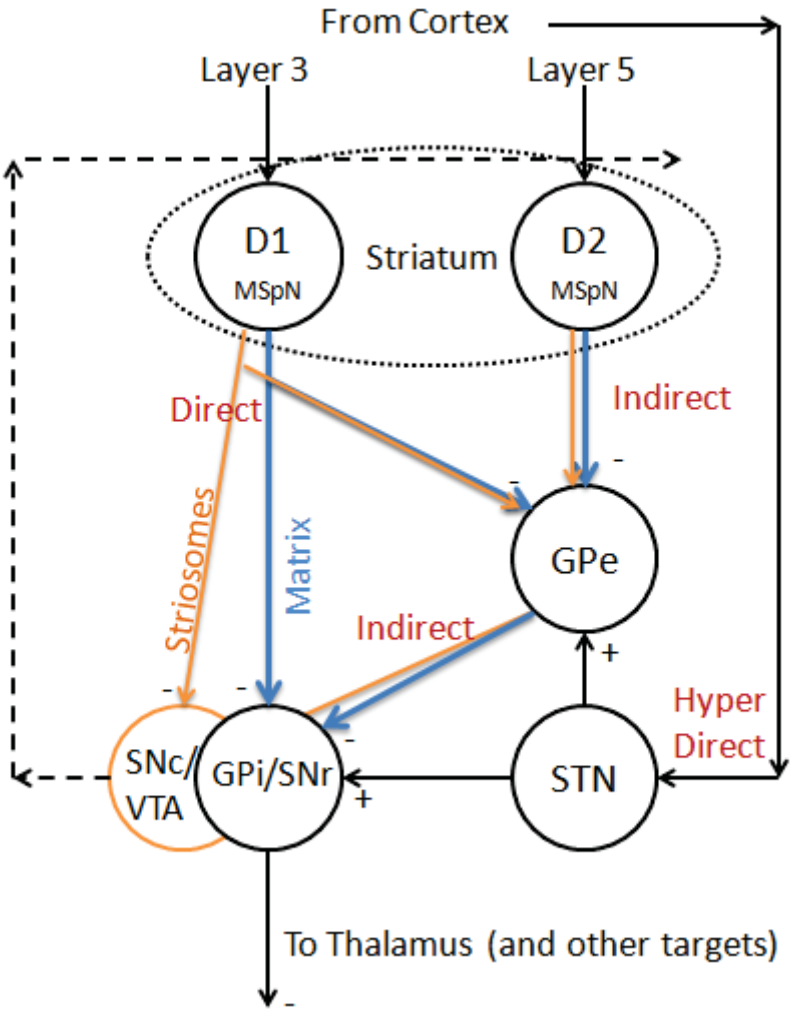

Figure 6.2: Diagram of salient basal ganglia features. There are three pathways through the basal ganglia: the direct, indirect, and hyperdirect pathways. The direct and indirect pathways have distinctive features including dopamine receptor subtypes (D1 or D2) and input from the cortex. Separate channels (or striatal compartments), known as the striosome and matrix, innervate the SNc/VTA and SNr/GPi, respectively. Abbreviations: MSpN - medium spiny neuron, GPe - globus pallidus externa, GPi - globus pallidus interna, STN - subthalamic nucleus, SNr - substantia nigra pars reticulata, SNc - substantia nigra pars compacta, VTA - ventral tegmental area

and that learning occurs in the striatum in the course of rewarding and punishing experiences. While the basal ganglia is often thought in terms of its role in action selection, parallel channels flow through the basal ganglia from non-motor cortical and subcortical areas as well. One reinforcement learning model proposes an actor-critic dichotomy (see Joel, Niv, & Ruppin, 2002 for a review), where the actor (mapped onto the dorsolateral (putamen) and dorsomedial (caudate) striatum) is responsible for action selection, and the critic (mapped onto the ventral (nucleus accumbens) striatum) is used to track the expected future reward in each situation. This coordinates with the difference between instrumental and classical conditioning, which learn about stimuli with and without the subject's action, respectively. The striatal areas associated with the actor can be further subdivided between the dorsolateral striatum and dorsomedial striatum according to whether it supports habit-driven action or goal-directed action, respectively (Balleine & O'Doherty, 2009). We focus here on learning the predictive values of environmental stimuli rather than the appropriate actions that maximize future rewards, though the similarity in the striatal structures commonly associated with each purpose supports the notion that there are common underlying mechanisms for both (Pennartz, Ito, Verschure, Battaglia, & Robbins, 2011; Matamales et al., 2009). In addition, cortico-striatal synaptic plasticity and the dual pathway nature of the basal ganglia are consistent with the notion that the ventral striatum is learning and expressing associative strength, as described in more detail below.

6.3 Classical Conditioning and the Basal Ganglia

Two areas believed to play a strong role in classical conditioning are the cerebellum and the amygdala (for a review, see Fanselow & Poulos, 2005). The cerebellum is the locus of Pavlovian eyeblink conditioning. The amygdala, however, has featured prominently in the conditioning literature as a key region for the acquisition and expression of fear conditioning (LeDoux, 2007; Maren, 2001), which is more relevant to the present discussion. Specific evidence of this includes that long-term potentiation (LTP) (a sustained increase in synaptic strength) appears to occur in the frontotemporal amygdala (Clugnet & LeDoux, 1990) and that this happens during fear conditioning to a tone (McKernan & Shinnick-Gallagher, 1997).

Yet, the amygdala sends a projection to the nucleus accumbens, a region of the ventral striatum (McDonald, 1991; Russchen, Bakst, Amaral, & Price, 1985), which has also been found to facilitate Pavlovian conditioning in a number of scenarios (Ito, Robbins, Pennartz, & Everitt, 2008; Haralambous & Westbrook, 1999; Bradfield & McNally, 2010; Phillips, Setzu, & Hitchcott, 2003). Several studies (Stuber et al., 2011; Ambroggi, Ishikawa, Fields, & Nicola, 2008; Fadok, Darvas, Dickerson, & Palmiter, 2010; Popescu, Popa, & Paré, 2009; Setlow, Holland, & Gallagher, 2002, but see Shiflett & Balleine, 2010) found that inactivation of this amygdaloid-striatal projection decreased responding to predictive cues. In regard to complex conditioning phenomena, a number of studies have implicated the ventral striatum. Two imaging studies (Corlett et al., 2004; San-Galli, Marchand, Decorte, & Di Scala, 2011) have suggested that the ventral striatum takes part in retrospective revaluation. Also, several studies suggest that the ventral striatum may contribute in complex experimental designs where there are competing predictors (Bradfield & McNally, 2010; Iordanova, McNally, & Westbrook, 2006; Iordanova, Westbrook, & Killcross, 2006; McNally & Westbrook, 2006).

The ventral striatum and the amygdala both receive dopaminergic projections from the mid-brain. Schultz and others (1997; 1998) found that phasic bursts of dopamine occur with the receipt of unpredicted reward and that dopamine dips occur when reward is expected but omitted, which resembles a US-surprisingness or prediction error signal[1]. Many experiments have now demonstrated this (see Niv, 2009 and Maia, 2009 for reviews), a view referred to as the reward prediction error (RPE) hypothesis of dopamine (Montague et al., 1996; Schultz, 1998). Most notably, in conditioned acquisition, the phasic dopaminergic response elicited by a conditioned stimulus is proportional to the magnitude and probability of the reward it predicts (Fiorillo, 2003; G. Morris, Arkadir, Nevet, Vaadia, & Bergman, 2004; Tobler, Fiorillo, & Schultz, 2005). In blocking and appetitive conditioned inhibition, the phasic dopamine response also appears to largely fulfill predictions made by the RPE hypothesis (Tobler, Dickinson, & Schultz, 2003; Waelti, Dickinson, & Schultz,

[1]Temporal Difference (TD) learning, a temporally extended version of the Rescorla-Wagner model, bears a signal called the reward prediction error which is equivalent to US surprisingness in the single time-step case. It is the reward prediction error signal that has been correlated with phasic dopamine signals of the midbrain.

2001). Although some reports suggest dopamine is involved in conditioned acquisition (Flagel et al., 2010; Lex & Hauber, 2010; Tsai et al., 2009) the evidence is not entirely conclusive since others suggest that conditioned acquisition is possible without dopamine-based learning (Young, Moran, & Joseph, 2005; Berridge, 2007).

Although the RPE hypothesis is strongly supported in terms of quantity and multi-directional quality of the evidence (Niv, 2009), there are competing views of what the phasic dopamine signal represents (see Berridge, 2007 for a review). One is the hypothesis that the short latency phasic dopamine signal that occurs with the appearance of novel/salient sensory stimuli is used to promote actions which lead to further experiences of such novel sensory stimuli (Redgrave, 1999; Redgrave & Gurney, 2006). Since the onset of the phasic dopamine burst is apparently too early (70 ms after stimulus presentation) for a stimulus signal to ascend the inferior temporal cortex and be interpreted as rewarding, the RPE hypothesis seems insufficient to entirely account for the phasic dopamine activity. Yet the duration of the phasic burst appears to depend on the initiator, namely that short phasic bursts (100 ms duration) are indicative of novel/salient stimuli, whereas a more sustained phasic activation (200 ms duration) is reward driven (Horvitz, Choi, Morvan, Eyny, & Balsam, 2007), which is time enough to interpret the reward value of a stimulus. This suggests that both reward and novel/salient stimuli contribute to the dopamine signal. Another challenge to the RPE hypothesis is that phasic dopamine neuron activity and increased dopamine concentrations have been detected in response to aversive stimuli (Ungless, 2004; Matsumoto & Hikosaka, 2009; Young, 2004), whereas the RPE hypothesis predicts a reduction in dopamine activity under these circumstances.

One unifying possibility is that the dopamine signal encompasses more than the RPE hypothesis. Both appetitive and aversive stimuli appear to induce phasic dopamine signals and prediction error-like responses in the striatum (see Delgado, Li, Schiller, & Phelps, 2008 for a review). This suggests that the same US surprisingness-like dopamine signal observed in appetitive conditioning may be useful in learning about aversive stimuli. Such a scenario would conveniently suit the standard classical conditioning notion that excitatory conditioning can be accomplished in the presence of either an appetitive or aversive US, although the way in which conditioned responding is measured depends on this detail. Furthermore, extracellular

dopamine concentrations increase (in the striatum) during sensory preconditioning, where one neutral stimulus is followed by another, but concentrations do not increase during presentation of a single neutral stimulus alone (Young, Ahier, Upton, Joseph, & Gray, 1998; Young, 2004). Dopamine may thus even encode a form of CS surprisingness, accommodating the perspective (common in classical conditioning) that, in sensory preconditioning, conditioning alters CS-CS associative strength rather than a prediction of the amount of expected future reward or punishment per se.

6.3.1 Cortico-Striatal (MSpN) Synaptic Plasticity

The discovery of the correlation between prediction error and dopamine activity has spawned much research aimed at understanding what is learned by neurons in the striatum. Again, this nucleus receives a strong projection from the dopamine neurons of the SNc/VTA, which provides the prediction error needed to learn predictions (i.e., values of ϕ). Below, we will review at least four major factors that influence learning in MSpNs. These are the input (presynaptic) activity, MSpN (postsynaptic) activity, dopamine levels, and the temporal relationship between pre- and postsynaptic spikes.

Reynolds and Wickens (2002) reviewed a number of studies that evaluate some or all of the first three of these factors. They summarized their findings with a "three-factor rule" which says that long-term depression (LTD) (a sustained decrease in synaptic strength) occurs when there is presynaptic activity and postsynaptic depolarization with a normal level of dopamine whereas LTP occurs when there is presynaptic activity and postsynaptic depolarization with a high phasic increase of dopamine.

Shen et al. (2008) provides a detailed account of how dopamine and the temporal relationship between pre- and postsynaptic neural impulses/spikes (also called spike-time dependent plasticity) influences learning in striatal neurons. Since we assume that a conditioned stimulus will send presynaptic spikes to neurons and induce postsynaptic spikes, we limit our discussion to this pre-post (positive) timing. Shen et al. show that when D1 dopamine receptor agonists are applied in the presence of positive spike timings, LTP is induced in direct pathway neurons, where D1 dopamine receptors are most common. However, when D1 receptor antagonists are applied, LTD occurs. The opposite is true of indirect pathway neurons, which dominantly express

D2 dopamine receptors. Positive timings in the presence of D2 dopamine receptor agonists lead to LTD whereas D2 receptor antagonists elicited LTP.

Dopamine enhances activity in direct pathway neurons when they are above a certain membrane potential, and suppresses otherwise (Hernandez-Lopez, Bargas, Surmeier, Reyes, & Galarraga, 1997). It also appears that dopamine suppresses synaptic currents of indirect pathway neurons and thus excitability (Hernandez-Lopez et al., 2000). Essentially, with phasic bursts of dopamine, the direct pathway neurons become more active and take more of the responsibility for learning whereas with dopamine dips, the indirect pathway neurons would appear to bear more of the weight. It may be, then, that when a phasic dopamine burst occurs, the direct pathway is primarily activated, leading to the LTP. This would help to reconcile the conclusions of Reynolds and Wickens (2002) with the findings of Shen et al. (2008).

Chapter 7

The Striatal Lateral Inhibition Model

7.1 Chapter Summary

In this chapter, we begin by highlighting the seemingly implausible batch-learning nature of LMS, that is, that LMS cycles through all of its past experiences (data) until it reaches convergence. Then, I present a biologically plausible model of the basal ganglia that can explain retrospective revaluation (a snapshot of regression in action) without resorting to batch-learning. We also relate this to a number of other existing classical conditioning models that explain this phenomena without batch-learning. Finally, several experimental predictions are made. The majority of the text in this chapter is an taken from Connor, LoLordo, and Trappenberg (2013), in which I was primarily responsible for developing the theory and simulations and drafting the manuscript.

7.2 Batch-learning versus Online-learning

LMS relies on batch-learning, the repeated reprocessing of all the data (or "trials" in the classical conditioning simulations) until some low prediction error threshold is reached. At face value, this appears to be biologically implausible, since there would be the need to store all of this data and repeatedly process it while processing new experiences. In animal learning, the theory of "rehearsal" resembles batch-wise training in machine learning to a certain degree. It may involve the recall and processing ("rehearsing") of surprising trials (Wagner, Rudy, & Whitlow, 1973), a recent or special grouping of trials (Ratcliff, 1990), or trials related to the present trial (Chapman, 1991). At odds with this notion is that rehearsal may interfere with the processing of the constant flow of incoming data. One way to resolve this would be to perform rehearsal in restful periods. In memory consolidation, hippocampal memories are transferred over time to cortical areas for longer term storage (McGaugh, 2000).

In one study, recordings of hippocampal and visual cortex neurons showed that rats replayed their memory of traveling on a track as a number of spiking cells that fired in the same sequence both during task execution and afterward during sleep (Ji & Wilson, 2006). It has also been shown that replay occurs in the ventral striatum during sleep (Lansink, Goltstein, Lankelma, McNaughton, & Pennartz, 2009), where the model presented in this chapter finds its focus. A simple proposal of batch-like learning in the brain, then, is that animals learn immediately from their experiences and then, during sleep or restful periods, subconsciously replay and reprocess past experience to further hone predictions.

If batch-like learning were the sole facility through which regression-like animal learning phenomena were achieved, we should expect that phenomena such as retrospective revaluation would not occur over short, mentally active periods when the mental hardware is presumably needed for other processing (e.g., incoming signals) and not performing replay. Yet, there are a number of examples where retrospective revaluation phenomena have appeared under such conditions (e.g., Luque, Flores, & Vadillo, 2013; McLaren et al., 2012; S. Gershman, Markman, & Otto, 2012). Specialized online-learning models serve as an alternative explanation of phenomena that can be explained by batch-learning. For example, most models from the classical conditioning literature that explain retrospective revaluation phenomena do so in an online-learning way. Even if the brain were to do batch-learning, say, during sleep, an online-learning method could still benefit from such repetition, as does the Rescorla-Wagner model. Indeed, employing a method that can be batch-like during online operation and then benefit from offline batch-processing would be more flexible than a strictly batch-learning approach. We now turn to review a number of the online models of classical conditioning that explain retrospective revaluation and present an online-learning dual pathway model of the basal ganglia (with a focus on the striatum) that does the same.

7.3 Online Models of Classical Conditioning that Explain Retrospective Revaluation

The associative classical conditioning models in the early days of retrospective revaluation findings were unable to account for these phenomena. In response, the Rescorla-Wagner model and the "Sometimes Opponent Process" (SOP) model of conditioning (Wagner, 1981) were retrofitted to explain them (Dickinson & Burke, 1996; Aitken & Dickinson, 2005; Van Hamme & Wasserman, 1994). Both models make use of within-compound associations (associations between stimuli presented simultaneously) presumably developed in the first phase of the phenomena to later associatively retrieve the absent stimulus of a subsequent phase. Then each model uses a mechanism to revalue the absent stimulus. In the case of the Van Hamme and Wasserman extension of the Rescorla-Wagner model, the absent stimulus is retrieved and given a negative alpha value. This leads to revaluation of the absent stimulus in the opposite direction as the presented stimulus. The Van Hamme and Wasserman model was recently elaborated upon by Witnauer and Miller (2011). Their model further develops the use of within-compound associations to enable it to account for second-order retrospective revaluation phenomena, which we will discuss later. Within-compound associations are also featured in other associative models of retrospective revaluation (e.g., Kasprow, Schachtman, & Miller, 1987; Kutlu & Schmajuk, 2012; Jamieson, Crump, & Hannah, 2012). For example, in the comparator hypothesis (Kasprow et al., 1987; Miller & Matzel, 1988; Stout & Miller, 2007), the within-compound associations become important during the test phase. The presence of a stimulus at test evokes previously paired stimuli through within-compound associations. The presented stimulus' response-evoking power becomes the associative strength of the presented stimulus B minus a fraction of the product of A's associative strength and the strength of the A→B within-compound association. In recovery from overshadowing, subsequent extinction of cue A makes the product of the B→A association and A→US association smaller than in a control group, thereby increasing the response-evoking power of cue B at test. All of these within-compound models are online learners.

A few online-learning models have taken a different approach, explaining retrospective revaluation phenomena apart from within-compound associations. The

"Adaptively Parameterised Error Correcting System" (APECS) model (McLaren, 1993; Le Pelley & McLaren, 2001; McLaren, 2011) explains these phenomena using configural units that represent memories of compound trials. A few elemental associative models (Ghirlanda, 2005; Dawson, 2008) also explain the phenomena apart from within-compound associations. Yet, the models of Ghirlanda and Dawson bear a key flaw. As will be shown, substantial revaluation in these models can occur without an associative history between the elements, an apparent failure to match the experimental data (e.g., Matzel et al., 1985) and the present general understanding of these phenomena. In this chapter, we present a model which overcomes this particular issue. The result is an online-learning neural network that can explain retrospective revaluation phenomena without relying on within-compound associations. Importantly, it also appears to suit basal ganglia structure and function.

7.4 Online Elemental Models of Retrospective Revaluation

The models of Dawson (2008) and Ghirlanda (2005) represent simple elemental models of retrospective revaluation phenomena. Dawson (2008) offers a model similar to the Van Hamme and Wasserman (1994) extension but gives negative α values to all absent stimuli. Ghirlanda (2005) took a different approach, which we will now discuss in some depth. This model represents stimuli in a distributed format instead of the usual one-to-one stimulus-input arrangement. Each stimulus in this model is described as a compound of many stimulus elements (mini-stimuli). For a feature such as colour, we could model 100 stimulus elements spanning the visible colour spectrum, each element representing a different wavelength. Here, Ghirlanda represents each punctate stimulus (e.g., stimulus A) as a Gaussian pattern of stimulus elements, as shown in Figure 7.1. Formally, the input provided to Ghirlanda's model is

$$S_i = K + \sum_j \alpha_j e^{-\frac{(\frac{i}{N} - \mu_j)^2}{\sigma^2}} \tag{7.1}$$

where S_i represents the i^{th} stimulus element's input salience, α_j is the salience of the j^{th} stimulus (analogous to the Rescorla-Wagner model's α term), and N represents the number of stimulus elements. Each stimulus' Gaussian pattern of stimulus elements is centered about a specific feature value (e.g., wavelength) between 0 and 1 (μ_j), and

has a certain width (σ). In simulations of Ghirlanda's model and our proposed model, we use 100 stimulus elements spanning between feature values of 0 and 1, and use $\sigma = \frac{1}{10\sqrt{2}}$ to fit several Gaussian-shaped stimuli into this range. The environmental context's representation does not have a Gaussian shape. Instead, it is represented as a flat function, such that all 100 stimulus elements have the same value, $K = 0.2$. In a simulated trial, the input provided to Ghirlanda's model (S_i) is the sum of the Gaussian patterns for each presented stimulus and the flat background context. The Gaussian shaped stimuli L, T, and C that we use in our simulations are shown in Figure 7.1, as well as an example input LTC compound, which incorporates the context (X).

Learning proceeds after each trial as in the Rescorla-Wagner model except that each stimulus element has an associative strength that is updated instead of associative strengths for punctate CSs,

$$\Delta W_i = S_i \beta (\lambda - r_s) \tag{7.2}$$

$$r_s = \sum_i W_i S_i \tag{7.3}$$

where W_i represents the i^{th} distributed stimulus element's associative strength, and r_s computes the total associative strength for all stimuli including the context on a given trial. For simple Pavlovian conditioning (i.e., AX+, X-, where X- represents the extinction of the context during the intertrial interval), the AX+ trials pull the W_i values upward toward $r_{AX} = 1$ while the X- trials pull W_i values downward toward $r_X = 0$. At the end of this tug-of-war, W_i values are found that satisfy both pulls and thus asymptotes are reached. The resulting associative strengths can be pictured as a Gaussian curve shifted downward (negatively) by an amount similar to the context value, K.

One of the earliest investigations of recovery from overshadowing was reported by Matzel et al. (1985). In Experiment 3, they paired a light and tone in the first phase, followed by reinforcement (TL+). They also reinforced separate presentations of a click stimulus (C+). In the second phase, they separated subjects into three groups: Group ET received non-reinforced presentations of the tone, Group EC received non-reinforced presentations of the click stimulus, and Group O was placed in conditioning chambers (as was done for the other groups) but no additional stimulus presentations

were made. In the third phase, testing was performed. Results from their experiment and results from a simulation of this procedure using Ghirlanda's model are shown in Figure 7.2. The first phase of simulation (TLX+, CX+, X-) leads to a set of W_i values that could be depicted as three negatively shifted Gaussians ($r_T = 0.50$, $r_L = 0.50$, $r_C = 1.0$). In the second phase, the extinction of the tone ($r_T = 0.0$) in Group ET inflates responding to the light ($r_L = 0.61$), which corresponds to the ordinal findings in Matzel el al. (1985). However, when we examine Group EC, where the separately conditioned click stimulus is extinguished, we find that responding to the light stimulus has also been inflated ($r_L = 0.71$), which is at variance with the experimental data in Figure 7.2. The extinction of the tone stimulus in Group ET also inflated responding to the click stimulus above Group O, the control (Group ET: $r_C = 1.11$, Group O: $r_C = 1.0$) and extinction of the click stimulus in Group EC also inflated the tone above Group O (Group EC: $r_T = 0.71$, Group O: $r_T = 0.50$). These two revaluations also disagree with the experimental findings. In summary, these simulations show that retrospective revaluation in Ghirlanda's model does *not* require a history of compound conditioning. Instead, it predicts that the revaluing of a conditioned stimulus will substantially affect the associative strength of even separately conditioned stimuli. Dawson's (2008) model makes the same prediction, apparently employing a negative α for *all* absent stimuli. These simple associative models not only disagree with the findings of Matzel et al. (1985) and others (Miller, Barnet, & Grahame, 1992; Cole, Barnet, & Miller, 1995) but also the current general understanding of these phenomena (but see Escobar, Pineño, & Matute, 2002; Amundson, Escobar, & Miller, 2003). In practical terms, if conditioning a stimulus could substantially alter the responses to unrelated stimuli, such interference could accumulate and confuse an organism about what each stimulus actually predicts.

As we will show, the present model overcomes the problems described above and yet, like the simple associative models, does not rely upon within-compound associations. In what follows, we will describe our model and then look at the contributions of each of its mechanisms by enabling them one at a time while simulating several classical conditioning phenomena. Ultimately, we will arrive at an explanation for retrospective revaluation phenomena.

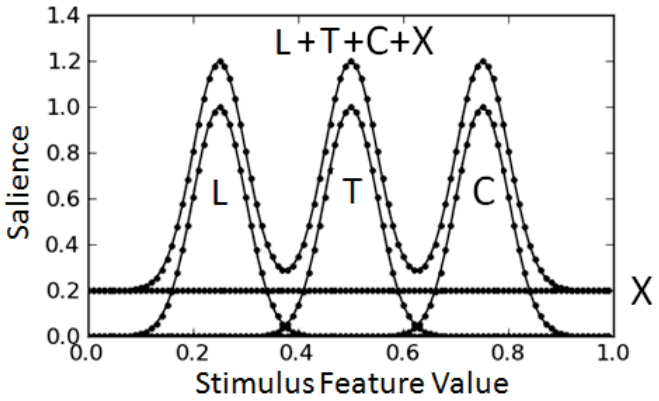

Figure 7.1: Distributed stimuli used in a simulation of the Ghirlanda (2005) model. Gaussian-shaped stimuli L, T, and C represent conditioned stimuli and the flat function X represents the context. The input to Ghirlanda's model is the sum of the present stimuli and the context, of which the example LTCX is given. There are 100 stimulus elements.

7.5 The Striatal Lateral Inhibition Model (SLIM)

Here we show how associative strength is computed and updated in our neural network model, which we refer to as the striatal lateral inhibition model (SLIM). It is illustrated in Figure 7.3. Where possible, we attempt to maintain biological plausibility and will briefly mention where this motivation influences certain design decisions. Possible neurobiological correspondences reflecting specific structural and learning rule details of the basal ganglia are discussed in the next section.

We represent a stimulus in the same Gaussian form as in Ghirlanda's model. Unlike the Ghirlanda model, however, we do not use a flat function to represent the context. Instead, the context is expressed by its own Gaussian-shaped curve, to represent that an environmental context is really a collection of stimuli itself. This approach also allows the possibility of having different contexts, if so desired. Given a certain distributed CS as input, the model responds with activity in its neurons, which equals the excitatory input minus the lateral competition. Upon stimulus presentation, each neuron is allowed to settle into an internal activity (u_j) according to

$$\Delta u_j = \frac{1}{\tau}(-u_j + (\frac{1}{N}\sum_{i=1}^{N} S_i w_{ij}^I - \frac{1}{M}\sum_{k=1}^{M} r(u_k) w_{kj}^L)) \tag{7.4}$$

where $\tau = 10$, $N = 100$ is the number of stimulus elements, and $M = 2500$ is

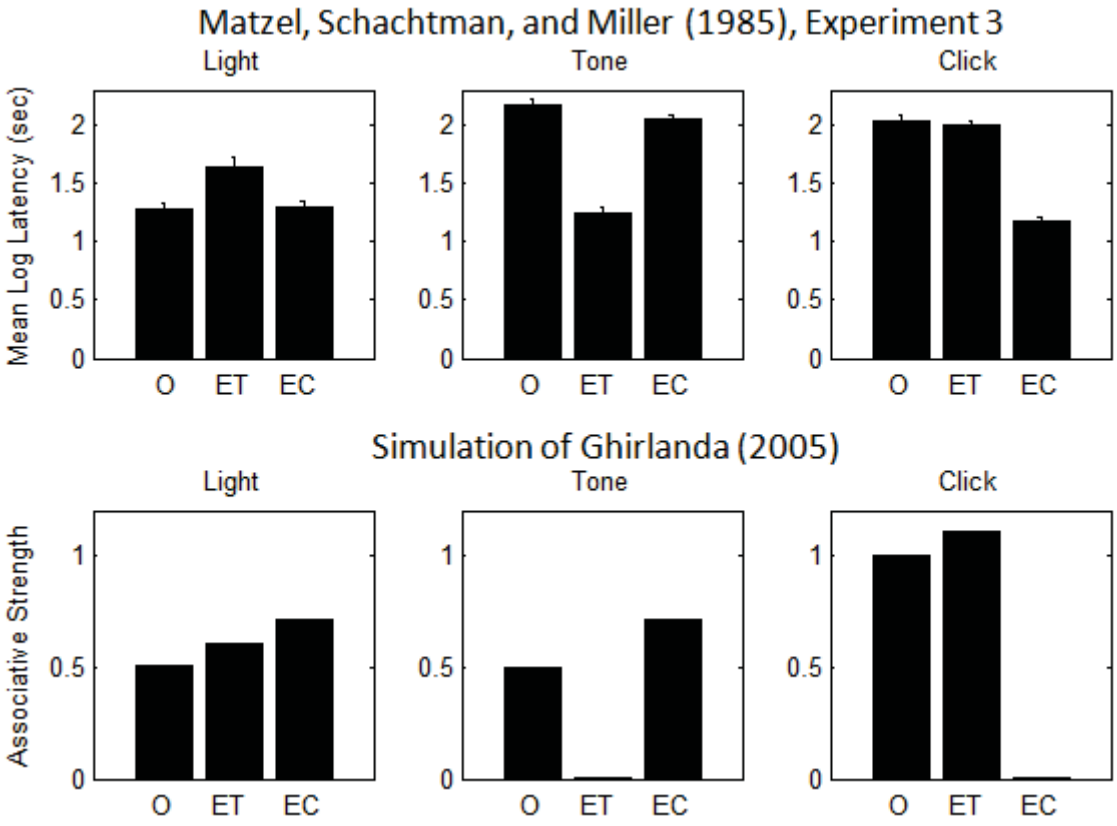

Figure 7.2: Results of a lick suppression experiment (3) in Matzel et al. (1985) and its simulation using the model of Ghirlanda (2005). Responding shown in the upper panels is in terms of mean log latency (in seconds) to make 25 licks in the presence of the light stimulus. Longer latencies indicate greater suppression and greater associative strength. Corresponding simulations of associative strengths from Ghirlanda's model are provided in the lower panels. In the simulations a procedure similar to the experiment was used ('X' is the context): Phase 1: TLX+, X-, CX+, X-, Phase 2: Group O: X-, X-, Group ET: TX-, X-, Group EC: CX-, X-, Phase 3: LX-, TX-, CX- (all groups). Sufficient trials were used in each phase of simulation to ensure that responses to a stimulus reached asymptotic levels. In Ghirlanda's model, extinction of the tone in phase 2 (Group ET) inflated the light above the overshadowing control group (Group O), which corresponds to the findings of Matzel et al.. The extinction of the click (Group EC) in simulation, however, also strongly inflated the light, which is a failure to predict the associated experimental data. The extinction of the tone in the model also inflated the click and vice versa, but this also fails to occur in the data. Experimental data from Matzel el al. (1985), Experiment 3, used by permission.

the number of neurons in the model. We use a large number of model neurons because it improves the consistency of the simulation results. The distributed stimulus elements, S_i, are connected to each neuron with a certain connection probability ($P^I = 0.25$). Subsequent equations appear to invoke full connectivity. However an absent connection is represented as an immutable connection weight of zero, which helps to simplify both the formal description and implementation of the model. This partial connectivity allows certain neurons to prefer activating in response to one stimulus or another and agrees with the reality that neural projections are not fully connected. The synaptic weights receiving stimulus input, w^I_{ij}, are initialized with a value of 20 for each connection made. Note that the indices i and j represent specific stimulus elements and specific model neurons respectively. So, instead of having a single weight per stimulus element, as in the Ghirlanda model, there is a single weight per stimulus element *for each neuron* in the model. Each neuron also has lateral synaptic weights, which receive inhibitory inputs from competing neurons. The lateral weight w^L_{kj} is located on neuron j and receives input from competing neuron k. These weights are also initialized to 20 for connections made and recurrent connections are permitted and the connection probability is $P^L = 0.25$. The term $r(u_k)$ is an activation function transforming the internal activation into a mean neuron firing rate,

$$r(u_k) = G(u_k)^2 \tag{7.5}$$

where $G(u_k) = 0$ when $u_k < 0$ and $G(u_k) = u_k$ otherwise is the threshold-linear function (Usher & McClelland, 2001), which means that neurons become silent when their internal activation goes below zero (analogous to real neurons). The outputs of these model neurons converge as a sum of the neuron firing rates, $r(u_j)$, with half of the neurons increasing the output and the other half decreasing it,

$$\Sigma V = \frac{\lambda_S}{M} \sum_{j=1}^{M} T(j) r(u_j), \tag{7.6}$$

$$T(j) = sgn(\frac{j}{\frac{M}{2}} - 1) \tag{7.7}$$

where the $sgn()$ function returns the sign of its argument and λ_S is a factor that translates the output from a sum of activities into units of associative strength ($\lambda_S =$

2500). The function $T(j)$ is $+1$ for half of the neurons (hereafter called "positive" neurons) and -1 for the other half ("negative" neurons) based on their index, j. The final sum represents the combined associative strength of the input stimuli, or the expected associative strength (ΣV), the analogue to r_s in Equation 7.3 from Ghirlanda's model. This approach permits both positive and negative associative strengths by using a population of positively contributing and negatively contributing neurons. For a positive associative strength, the positive neurons are (on average) more active than the negative neurons. Negative associative strengths are expressed by the opposite difference of activity. The segregation of neurons into these two groups allows all input connection weights to be consistently positive, which is more plausible biologically speaking. There are few instances where a real neural connection can switch from having a positive to a negative influence. Also, as we will see, having two segregated groups of neurons contributes to generating a novel configural mechanism from elemental inputs.

Importantly, the ensemble of neurons that wins the competition and remains active is determined by the stimulus input provided. Because it is the activities of model neurons that ultimately combine to express associative strength, the active neural ensembles come to represent the associative strengths of the stimuli that evoke them. The learning rules cooperate by only updating the active ensemble neurons and only for non-zero stimulus elements (i.e., present stimuli). The learning rule used to update the input weights of each neuron is,

$$\Delta w_{ij}^{I} = T(j) S_i \beta (\lambda - \Sigma V) G(u_j) \tag{7.8}$$

Notice that updates to a neuron are proportional to its internal activation when it is above zero only (i.e., $G(u_j)$) and thus is part of the ensemble of active neurons. Also notice that weights associated with distributed stimulus elements that have zero salience will also not change. This ensures that *input* weights are only updated for presented stimuli. The learning rule for the lateral weights, which receive inputs from other neurons, is

$$\Delta w_{kj}^{L} = T(j) H(u_k) \rho (\lambda - \Sigma V) G(u_j) \tag{7.9}$$

where $H()$ is the Heaviside or unit step function (1 when the argument is greater than zero and 0 otherwise), which means that learning will only occur if the sending neuron,

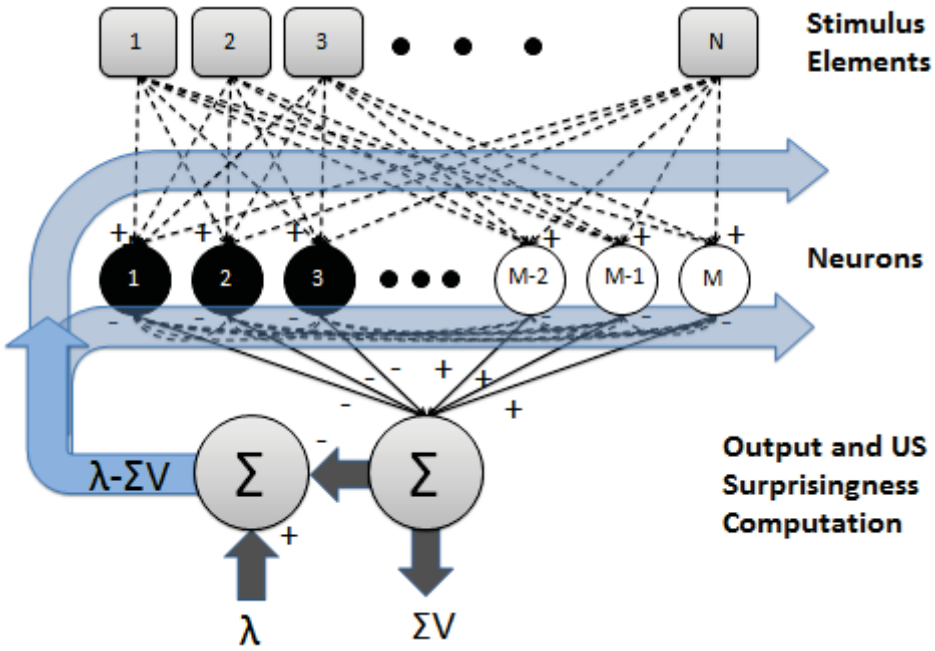

Figure 7.3: The striatal lateral inhibition model (SLIM). The stimulus element inputs represented by the rounded boxes take exactly the same distributed input as used in Ghirlanda's model, except that the context here is also modeled as a Gaussian pattern. Each dashed line in the model represents a connection that will (or will not) be established upon model initialization with some fixed probability. Neurons in the model, represented by circles, receive input and become excited. The connections between the neurons are inhibitory. These connections induce competition between the neurons, which reduces neuron activities and leads to a subset of neurons that dominates and suppresses all other neurons. The activities of the neurons are accumulated (bottom-center circle), where one half of these neurons add and the other half subtract from the sum. The total is appropriately scaled and represents the sum of associative strengths (ΣV) for the input stimuli. Conditioning is accomplished by changing the connection weights of model neurons. This is a function of the several factors including the US surprisingness (computed in the bottom-left circle), which is represented by the broad arrow leading back to the input and lateral connections. Importantly, the stimuli presented on a trial determine the ensemble of active neurons that develops through competition. Since it is the sum of activities of model neurons that gives the associative strength, the active neural ensembles come to represent the associative strengths of the stimuli that evoke them.

indexed as k, is active. The parameter ρ in Equation 7.9 is the learning rate parameter for lateral weights ($\rho = 0.5\beta$). In this model, an individual neuron's weights, w_{ij}^I and w_{jk}^L, are always positive. This keeps the stimulus input influence excitatory and the lateral influence inhibitory in Equation 7.4. Weight changes must have opposite signs for the positive and negative neurons, so that these opposing pathways learn cooperatively. The function $T(j)$ defined in Equation 7.7 achieves this. Equation 7.8 is essentially a Hebbian learning rule (from its pre- and postsynaptic activity terms) modulated by the US surprisingness error term.

7.6 SLIM's Relationship to the Neurobiology of the Basal Ganglia

The present model, although described in an abstract way, can be readily related to features of the neurobiology of the ventral striatum and basal ganglia. The positive and negative neurons map to the striatal projection neurons belonging to the direct and indirect pathways of the basal ganglia, respectively. The input weight learning rule (Equation 7.8) corresponds with experimental findings regarding the effects of dopamine ($\lambda - \Sigma V$) and pre- (S_i) and postsynaptic activity ($G(u_j)$) on cortico-striatal synapses (Frank & Fossella, 2011; Reynolds & Wickens, 2002; Schultz et al., 1997; Shen et al., 2008). In short, SLIM represents a small patch of striatum that receives a small portion of topologically distributed cortical inputs.

The key feature of the present model that sets it apart from related neurobiological models of the basal ganglia (Houk et al., 1995; Frank, 2005) is its lateral inhibition and lateral learning. SLIM's lateral connectivity mirrors the rarely reciprocated lateral connections of the more common local laterally projecting MSpNs in the striatum (Kawaguchi et al., 1990). As noted earlier, although the lateral inhibitory connections are believed to be weak, fast spiking inhibitory interneurons provide additional inhibition (Tepper et al., 2008; Gruber et al., 2009). Therefore, the lateral inhibitory component of this model could roughly be viewed as representing the contributions of both types of inhibition. Neurobiology concerning the lateral learning rule, however, is less clear than evidence concerning its input weight counterpart. Long-term learning has been found to occur in the lateral synaptic connections of striatal projection neurons (Rueda-Orozco et al., 2009), although how this relates to a dopamine-based error term is not clear. In the model, the learning rule for these lateral connections is

similar to that of the input connections, except that the CS intensity, I, is replaced by a term related to the activity of one of its laterally inhibiting neurons.

7.7 Classical Conditioning Simulations and SLIM

SLIM is readily integrated into trial-based simulations of classical conditioning experiments. A single trial consists of presenting stimuli, presenting the outcome (US or no US), computing the surprisingness, and adjusting synaptic weights according to the learning rules. Although we simulate phenomena that develop CS-US associations, the model does not explicitly exclude the notion of developing CS-CS associations, though these do not occur in the present simulations. Our model, like many others, does not define any CS-US timing, thereby excluding certain temporal phenomena (e.g., serial feature-positive discrimination) from its scope. Experimental findings and model predictions are ordinal in nature, so the usual assumption that associative strength is monotonically translated into conditioned responding is made here.

Salience levels play a role in our simulations. The US ($\beta = 0.1$) has a value of $\lambda = 100$ when the US is present and $\lambda = 0$ when it is absent. Conditioned stimuli have a salience of $\alpha = 1.0$, while the context (X) has a salience of $\alpha = 0.2$. Each intertrial interval is simulated like a single conditioning trial, where the context is presented but not reinforced (X-), just as in simulations of Ghirlanda's model. Associative strength accrued to the context is partly extinguished during these intertrial intervals. Parameters of the model used in simulations have been specified in the preceding section. A similar version of the present model was described in Connor and Trappenberg (2011) and shows how performance of the model varies within appropriate ranges for selected parameters.

An example of how conditioning proceeds in this network is shown in Figure 7.4. In excitatory conditioning (CS→US) trials, the synaptic weights of active positive neurons are increased while the synaptic weights of active negative neurons are decreased. This results in increased activity in the positive neurons and decreased activity in the negative neurons for subsequent trials. The difference between the activity in these two pathways gives a final positive associative strength. As conditioning trials continue, the associative strength will grow until it matches that supportable by the US. More and more neurons are also silenced through lateral inhibition as

an ensemble of neurons increasingly dominates. During extinction, the opposite process happens, cutting down positive neuron activity and restoring negative neuron activity. Also, the active ensemble will grow to include more neurons once again. In short, increases and decreases in associative strength track increases and decreases of synaptic weights in the positive neurons, and the opposite relationship exists between associative strength and the synaptic weights of negative neurons.

In the following sections we use additional classical conditioning simulations to show how certain model mechanisms affect model behaviour. Beyond the acquisition example in Figure 7.4, we do not revisit demonstrations of the mechanisms borrowed from the Rescorla-Wagner model but rather focus on the unique mechanisms of the present model. Building upward, we first show how the combination of the activity dependent learning term $G(u_j)$ and having dual pathways (i.e., positive and negative neurons) develops configural representations from individual stimuli. Then we demonstrate how adding lateral inhibition sculpts ensembles of active neurons, and finally how learning in these lateral connections enables retrospective revaluation effects.

7.7.1 Activity Proportional Learning and Dual Pathways Perform Configuration

Recall that there are positive and negative neurons in the model whose influences sum to provide the overall associative strength. Changes to the weights of these neurons are made in proportion to their internal activation, $G(u_j)$ (Equations 7.8 and 7.9). The combination of these two mechanisms leads to the development of configural cues. To demonstrate this, we simulate the negative patterning procedure, which in a single phase interleaves trials of AB- with A+ and B+ trials. The ordinal finding is that responding to the compound AB during a subsequent test is less than responding to either A or B alone (Woodbury, 1943; Delamater, Sosa, & Katz, 1999; Harris, Gharaei, & Moore, 2009; Redhead & Pearce, 1995). To demonstrate the combined efforts of the two mechanisms, Figure 7.5 shows simulated negative patterning results for our model with and without each of them. In addition, results when lateral inhibition and lateral learning are enabled are also given to show that these additional mechanisms do not interfere. Note that to disable the dual pathway nature of the model, we eliminate excitatory input to the negative neurons to silence

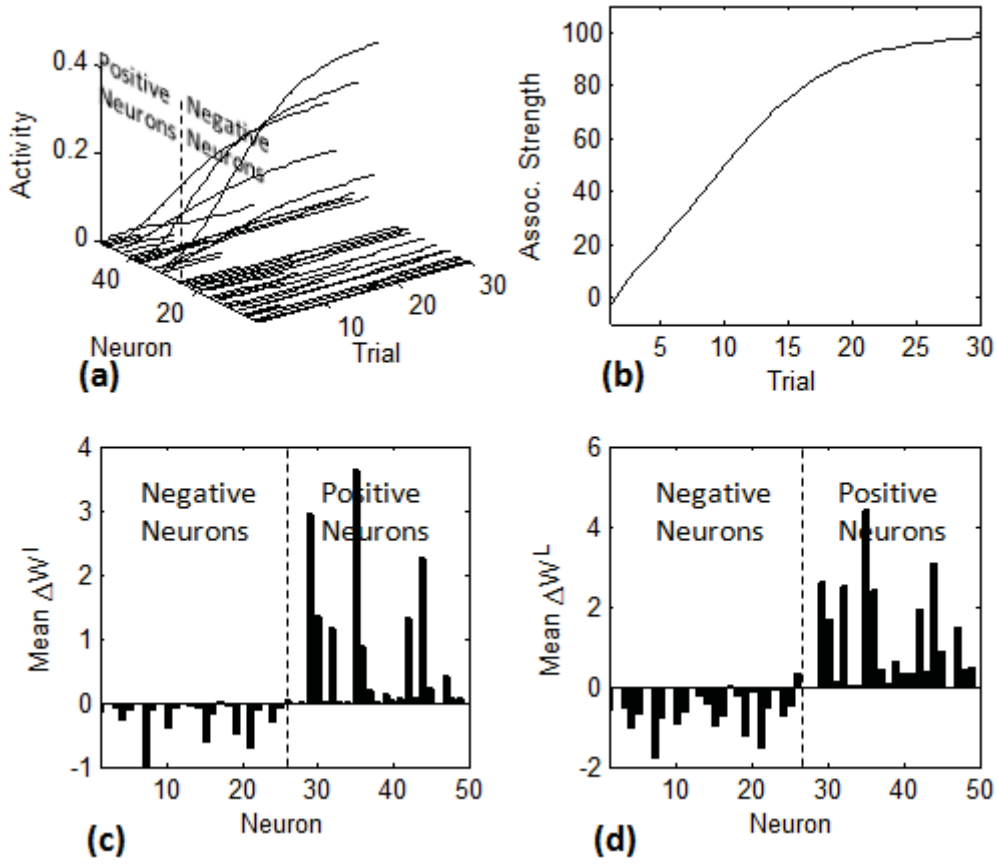

Figure 7.4: SLIM during excitatory conditioning, simulated using only 50 neurons for demonstration purposes. a) Activity in some positive neurons (neurons 26-50) increases with the number of trials. Other neurons lose the competition and are silenced. Negative neurons (1-25) are either suppressed or very weakly active. b) Overall associative strength increases, approaching asymptote within 30 trials. c) The average change in input synaptic weights for each neuron between the first and last trials shows a substantial increase for positive neurons and a slight decrease for negative neurons. d) Lateral synaptic weights also increase for positive neurons and decrease for negative neurons.

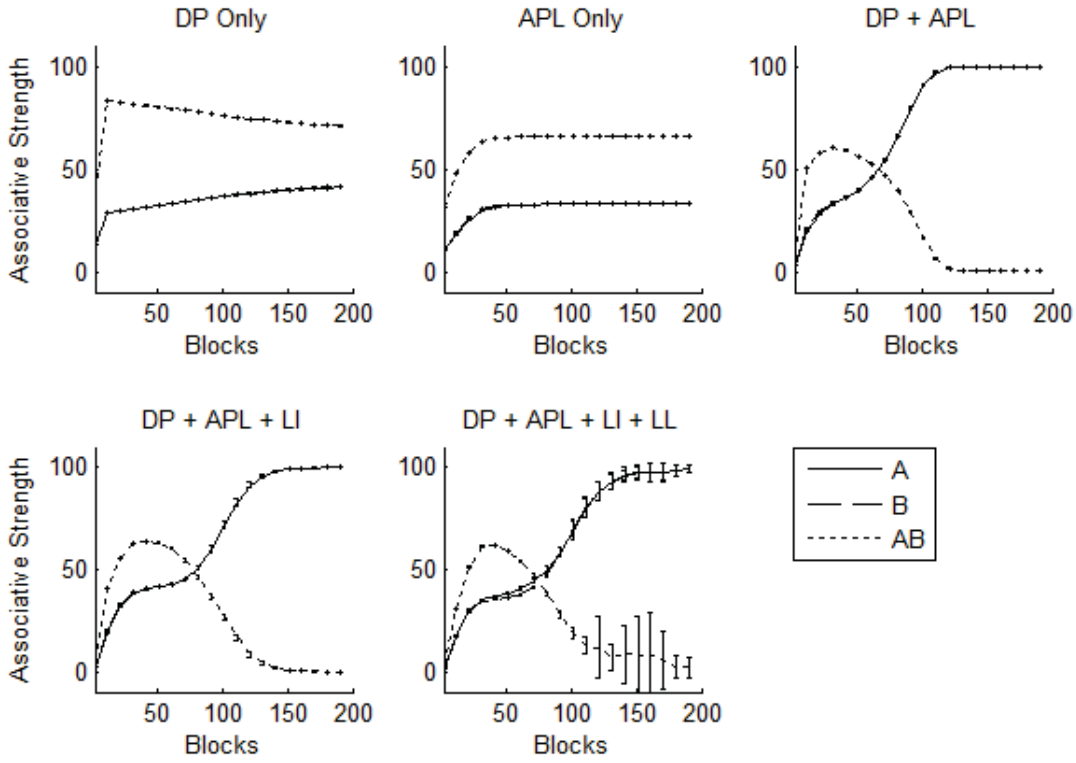

Figure 7.5: Simulation of negative patterning using various configurations of the present model for 15 differently initialized models (stat rats). Each block consists of 3 trials (A+, B+, AB-). Negative patterning requires that both the positive and negative neurons exist and that there is activity proportional learning. The lateral inhibition and lateral learning mechanisms do not assist but also do not substantially interfere. Acronyms: DP - Dual Pathway, APL - Activity Proportional Learning, LI - Lateral Inhibition, LL - Lateral Learning

them. To disable activity proportional learning, we simulate without the last term in Equations 7.8 and 7.9. Disabling lateral inhibition is accomplished by fixing all lateral weights to zero and disabling lateral learning is done by setting $\rho = 0$.

As Figure 7.5 shows, when either the dual pathway or activity proportional learning mechanisms work alone, negative patterning fails. When both mechanisms are engaged, however, the phenomenon emerges. The way in which the model accomplishes this can be seen from the input weights. Figure 7.6 shows that when activity proportional learning and both pathways are enabled, positive neuron weights specialize for either stimulus A or B, while negative neuron weights grow similarly for each stimulus. When only a single stimulus (A or B) is present, the specializing positive neurons activate strongly, whereas the unspecialized negative neurons activate little.

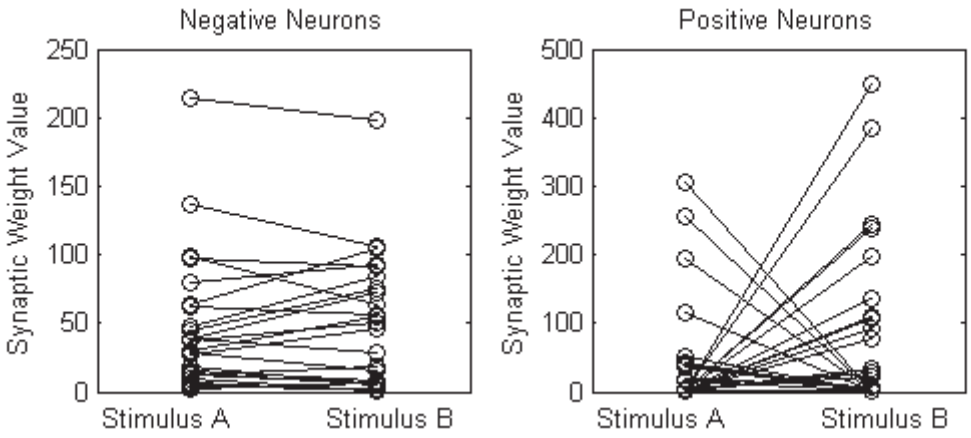

Figure 7.6: Correlation between the weights in a random selection of model neurons and stimuli A ($\Sigma S_i^A w_{ij}^I$) and B ($\Sigma S_i^B w_{ij}^I$) when both pathways and activity proportional learning are enabled (i.e., lateral inhibition and lateral learning are disabled). Negative neurons grow relatively evenly for both stimuli A and B, making them respond substantially more to the compound AB than to A or B alone. In contrast, positive neurons' weights tend to specialize (increase) for either stimulus A or B and decrease for the other stimulus.

The sum of large positive neuron activations minus small negative neuron activations results in asymptotic (λ) conditioned responding. Now, recall that the activation function is $r(u_j) = G(u_j)^2$, which means that doubling the stimulus input (which doubles u_j) will quadruple a neuron's output. So, when both stimuli are present (AB) the negative neurons' activations are increased exponentially, which enables them to balance out the positive neuron activations, resulting in zero conditioned responding. The weights develop as follows: initial stages of training show that for positive neurons which are activated more strongly for a certain stimulus (e.g., A), the input weights increase more on A trials than they decrease on AB- trials. Conversely weights receiving connections from the stimulus for which a positive neuron's activation is weaker (e.g., B) will have a net decrease because their reinforced trial increases the weights less than the decrease occurring from AB- trials. Negative neuron weights decrease in early stages because the reinforced trials have a larger positive error term (which decreases negative neuron weights) than the negative error term on AB- trials. As associative strength increases to the individual stimuli, however, this situation reverses and negative neuron weights begin to grow more on the AB- trials and do so roughly evenly for A and B.

Although not a focus of this thesis, this approach to developing configural cues is quite novel. Its elemental basis puts it in the same realm as Harris' elemental model (2006) and the replaced elements theory (Brandon & Wagner, 1998; Wagner & Brandon, 2001; Wagner, 2003). In such models, each stimulus is represented by a set of elements. Elements that are active (or within an attentional buffer) during conditioning receive larger changes in associative strength than the others. Certain elements for each stimulus are allowed to become activated depending on whether the stimulus is presented alone or in compound. Therefore, in the negative patterning procedure, some stimulus elements are primarily conditioned in the single stimulus trials but not the compound trials and vice versa. This allows some elements to encode the single-stimulus associative strength and others to help represent an opposite compound associative strength. The present model departs radically from this idea, not needing to deactivate stimulus elements, but instead deriving its configural ability from its dual pathways and activity proportional learning in the context of a squared activation function. Additional work is needed to evaluate this approach by simulating other experimental paradigms and drawing thorough comparisons with the other configural models on the market. For the present work, however, we have focused on the model's ability to explain retrospective revaluation phenomena, which is supported by lateral inhibitory connections and the learning therein, to which we now turn.

7.7.2 Adding Lateral Inhibition

Figure 7.7 shows the neuron response following negative patterning for a model with only 200 neurons for illustration purposes, where both panels show models that have activity proportional learning and both pathways enabled. In the top panel lateral inhibition is disabled and in the bottom panel it is enabled (without learning, $\rho = 0$). Although the difference in associative strength between the two conditions will be small (e.g., see comparison in Figure 7.5), the neural activity takes a new form. Without lateral inhibition, all neurons are active for every input. When lateral inhibition is enabled, a unique ensemble of active neurons takes shape in response to the presentation of each stimulus or compound. On the surface, it may seem that this mechanism is very similar to the replaced elements mechanism of models noted above

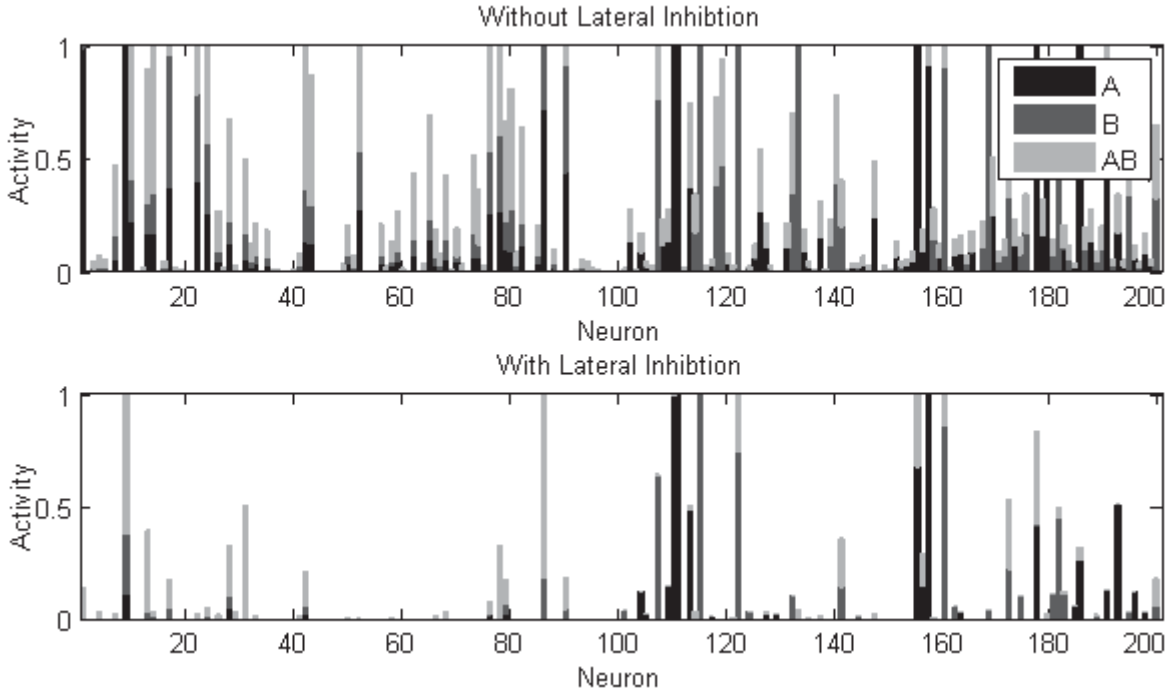

Figure 7.7: Model neuron activity after negative patterning. Using only 200 neurons for demonstration purposes, the simulated activity for each stimulus or compound is computed and drawn as a stacked column in the bar graph, where each column represents one neuron. The length of each colored bar in the stack is the amount of activity observed for the condition it represents. The left half of the neurons (1-100) are negative neurons and the right half (101-200) are positive neurons. When lateral inhibition is disabled, all neurons respond to some degree for every stimulus, and thus take part in representing every stimulus' associative strength. When lateral inhibition is enabled, however, only a fraction of the neurons are active for any given stimulus. This means that each neuron takes part in representing only certain stimuli's associative strengths.

which perform configuration. Although lateral inhibition may technically be able to behave in this way, it is ancillary in our simulations. In fact, lateral inhibition does not appear to explain any additional phenomena in this context, besides taking part in helping lateral learning explain retrospective revaluation phenomena, which will be discussed later. A potential benefit, however, is that because it uses fewer neurons to represent the same information, the overall capacity of the system to learn further stimulus-outcome relationships should increase.

With lateral inhibition enabled, unique active ensembles emerge for specific input stimuli, such that when the input stimuli change, so will the active ensemble to some

degree. Roughly speaking, the weights of the neurons in the ensemble match the distributed stimulus input profile more closely than do the weights of neurons that are silenced. Thus, the similarity between two ensembles depends on the similarity of the two stimuli, and is reduced as stimuli become dissimilar. This is shown in Figure 7.8, where the ensemble similarity between a previously conditioned stimulus having a specific feature value (e.g., green light) and all feature values (i.e., green, red, blue, yellow, etc.) is computed, both for when lateral inhibition is disabled and enabled (but no lateral learning). Although there is no difference in associative strength with and without lateral inhibition (upper panel), the similarity between the neural activations of vastly different features (e.g., 0.5 vs. 0.2, or 0.5 vs. 0.9) is smaller with lateral inhibition (lower panel). Without lateral inhibition, there will be a greater similarity between the neural activations of different features because every neuron is active for every feature.

Until now, we have discussed similarity between two distinct stimuli. Consider a related case in which one stimulus is joined by a second stimulus to make a compound. Because there is substantial similarity between the compound and its constituents, the activation of neurons by the compound will be more similar to the activation evoked by one of its constituents than to the activation evoked by an unrelated stimulus.

7.7.3 Adding Lateral Learning

Simulations of recovery from overshadowing using the present model are shown in Figure 7.9 following Matzel et al. (1985) and the simulations of Ghirlanda's model described earlier. In particular, these simulations show: (1) that recovery only occurs when lateral learning is enabled and (2) that revaluing a conditioned stimulus only significantly affects the associative strengths of stimuli with which it was previously paired, and not unrelated stimuli.

To understand how lateral learning accomplishes all of this, we will focus on two positive neurons and explain how recovery from overshadowing can occur, as shown in Figure 7.10. Negative neurons do not play a major role in this phenomenon, but may take a more significant role in other retrospective revaluation phenomena (e.g., recovery from conditioned inhibition by extinction of the excitor). Excitatory conditioning of a compound (Phase 1) increases the input weights of its active neurons.

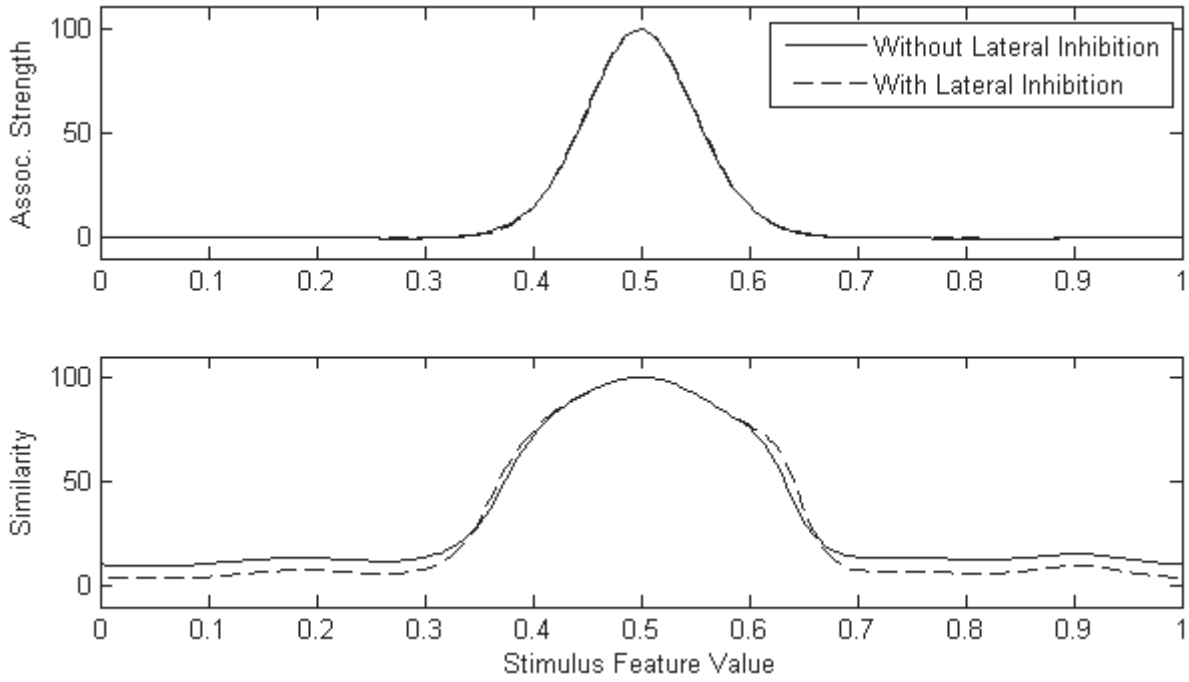

Figure 7.8: Measures of associative strength and active ensemble similarity between a previously conditioned stimulus (feature value = 0.5) and all other feature values (0 to 1) with and without lateral inhibition. In both cases, we see that CSs with similar feature values evoke substantially similar ensembles and thus associative strengths. Adding lateral inhibition tends to lower the similarity between the ensembles activated by unrelated stimuli. Similarity is computed as the cosine of the angle (i.e., the normalized dot product) between the neural ensembles activated for the previously conditioned stimulus and the test stimulus.

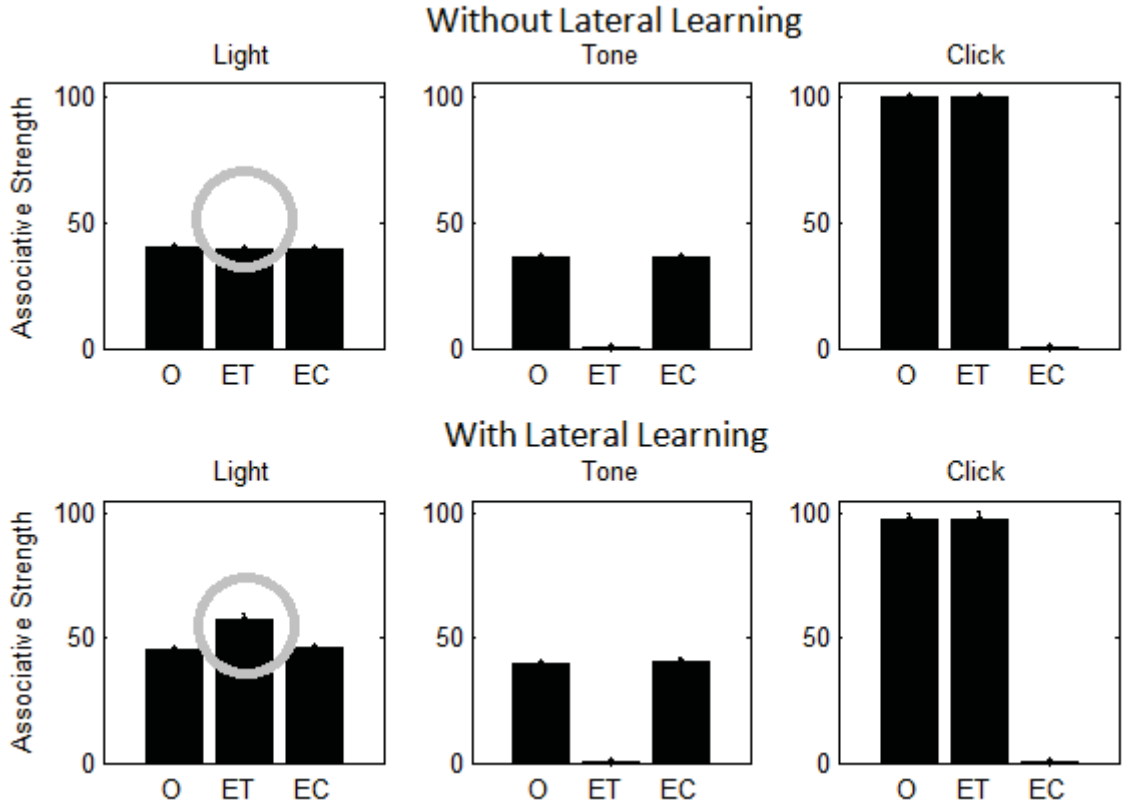

Figure 7.9: Simulations of recovery from overshadowing (Matzel et al. (1985), Experiment 3) using the present model when lateral learning is disabled ($\rho = 0$) and enabled ($\rho > 0$). Error bars represent the small deviation in results for 15 differently initialized models (stat rats). The simulation procedure matches that used for earlier simulations of the Ghirlanda (2005) model: Phase 1 (50 trials): TLX+, X-, CX+, X-, Phase 2 (200 trials): Group O: X-, X-, Group ET: TX-, X-, Group EC: CX-, X-, Phase 3 (1 trial): LX-, TX-, CX- (all groups). Circled in the results, we see that extinction of the tone in phase 2 of the simulation (Group ET) revalued (inflated) the light above the control group (Group O) when lateral learning is enabled, but not when it is disabled. Also in agreement with the experimental data, the simulations did not substantially revalue any other stimuli (regardless of whether or not lateral learning was enabled), in contrast to the simulations of Ghirlanda's model.

Because $\rho > 0$, the lateral inhibitory connections between the active neurons grow as well. In Phase 2, only one of the constituents (A) is presented. Because of its history of activating the ensemble associated with the compound (AB), there is a substantial degree of similarity between the ensembles activated by AB and A and thus these same two positive neurons are activated again. When the presentation of A is followed by no reinforcement, these active neurons' A-specific input weights and their lateral weights are decreased. In Phase 3, when the absent stimulus (B) is tested, we detect a change. Although the B-specific input weights did not change (because there was no input from stimulus B in Phase 2), its active ensemble's lateral inhibitory weights are smaller. As a result, there is less inhibition, which increases these positive neurons' overall activities and thereby increases the associative strength of B. Intuitively speaking, excitatory conditioning of a compound ties its constituents together in terms of causing them to activate similar ensembles of neurons in future trials. Then, interactions occur between these stimuli through lateral learning in their shared connections. The result is that extinguishing one increases the associative strength of the other (i.e., recovery from overshadowing), and increasing one's associative strength will decrease the other's (i.e., backward blocking). In this way, the shared lateral connections play a similar role as the within-compound associations found in other models, but do not retrieve explicit stimuli per se. Without the compound conditioning step, as is the case for unrelated stimuli, there would be fewer shared neurons (and thus lateral connections) in the ensembles of the individual stimuli. As a result, there would be far less change in the lateral inhibition for an absent stimulus were an unrelated stimulus presented and revalued.

7.8 Second-order Retrospective Revaluation and Relation to Other Models

In recent years, theorists have focused on the phenomena of second order retrospective revaluation. The second-order retrospective revaluation procedure involves conditioning, in successive phases, two compounds that share a common element (i.e., Phase 1: AX+, Phase 2: XB+) and in a third phase revaluing one of the non-shared stimuli (A).

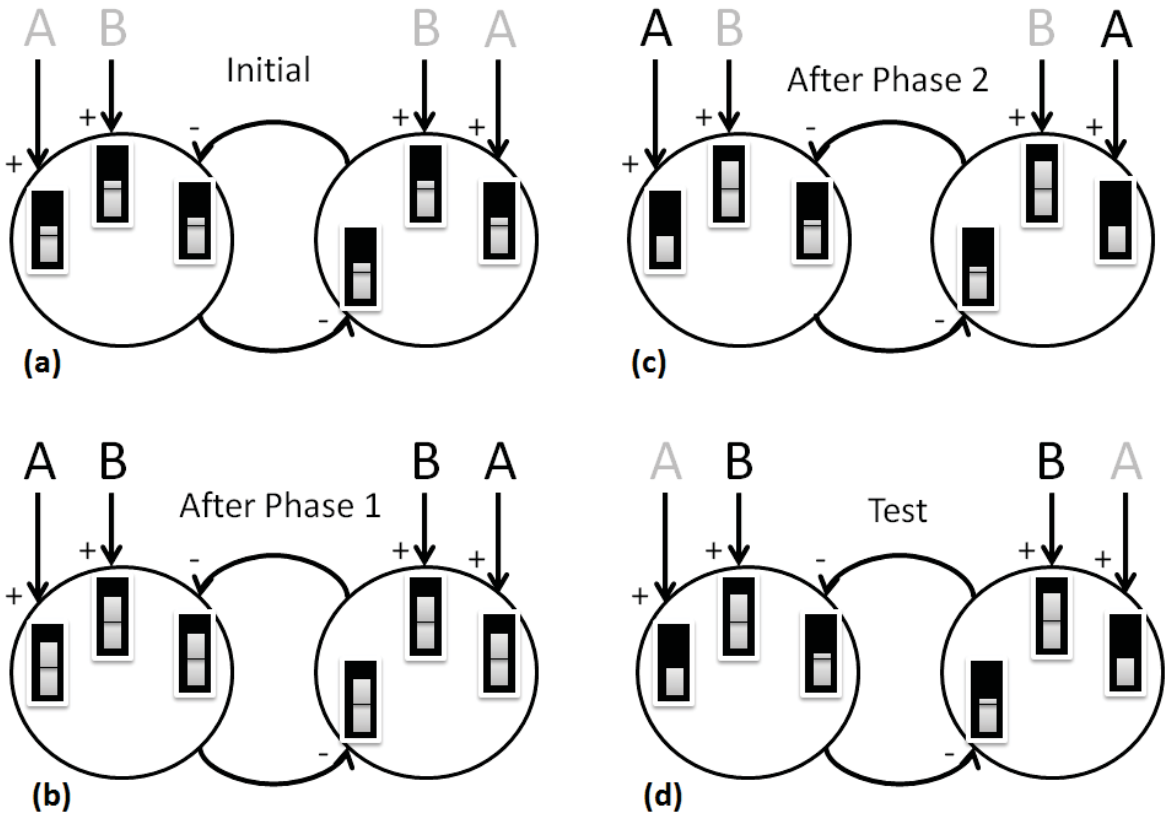

Figure 7.10: Recovery from overshadowing as demonstrated in the present model. This diagram focuses on two positive neurons represented by circles that are active whenever A, B, or AB are presented. Each neuron receives excitatory inputs from stimuli A and B and an inhibitory connection from the other neuron. (a) The neurons' synaptic weights, which are represented thermometer style in the rectangles associated with each connection, are initialized to about half value. (b) After conditioning to compound AB (Phase 1), input weights connecting A and B to the neurons are increased. Also increased are the lateral weights between these active neurons. (c) In the second phase, A is presented but not reinforced, which decreases its input weights and lateral weights. (d) Subsequent testing of B shows an increase of associative strength. Although B's input weights are unchanged, its lateral weights have decreased. Less inhibition means greater activity in these positive neurons, which translates into more associative strength (Equation 7.6).

DeHouwer and Beckers (2002) ran three experiments, using a weapons/tanks procedure. In the first experiment they did Phase 1: CT1+, Phase 2: T1T2+, and Phase 3: either C+ or C- between groups. Group C+ had a much higher rating of C, a lower rating of T1, and a higher rating of T2. The next experiment looked at third-order retrospective revaluation: Phase 1: CT1+, Phase 2: T1T2+, Phase 3: T2T3+, and Phase 4: C+ or C- between groups. Group C+ had higher ratings for C and T2 than Group C- did, but lower ratings for T1 and T3. All the effects were substantial. The next experiment looked at second-order retrospective revaluation in a within-subjects design, essentially Phase 1: C1T1+, C2T3+, Phase 2: T1T2+, T3T4+, and Phase 3: C1+, C2-. Ratings of C1 were higher than C2, and T3 was rated higher than T1 (first-order retrospective revaluation). There was also a big second-order retrospective revaluation effect: T2 was higher than T4. Melchers, Lachnit and Shanks (2004) obtained results similar to DeHouwer and Beckers in a within-subjects experiment within the foods/allergies setting. Their experiment 3 looked at second-order retrospective revaluation (e.g., Phase 1: AB+, BC+, Phase 2: C+ vs. Phase 1: DE+, EF+, Phase 2: F-) and its direct analogue (e.g., where the element trials came before the compound trials). First-order retrospective revaluation occurred (e.g., B<E), and second-order retrospective revaluation was in the opposite direction (e.g., A>D). Denniston et al. (2001) using rats, employed a between groups paradigm: Phase 1: CA+, Phase 2: BA+, then either C- or nothing. The conditioned response to B was lower in the C- group than in the controls. This finding is consistent with DeHouwer and Beckers (2002) and Melchers, Lachnit and Shanks (2004).

McLaren, Forrest and McLaren (2012) reported an experiment on retrospective revaluation using the foods/allergies setting. First- and second-order retrospective revaluation were assessed in a within-groups design, so Phase 1: BC+, DE+, Phase 2: AB+, EF+, and Phase 3: A+, F-. Ratings of B and C both declined, relative to D and E, and the first-order effect was about as big as the second-order effect. Their second order result is opposite of the findings described above. McLaren et al. reported that if they instead provided all the data at once on handouts, which they interpreted as entailing a low memory load, then they got a different result; ratings of B and C moved in opposite directions after A+, and ratings of D and E moved in opposite directions after F-. They further suggested that the findings of Melchers

et al., which were opposite to their own, were the result of a relatively low memory load. This hypothesis needs to be tested in an experiment that varies memory load.

So, we have a data conflict, but what do the models predict? McLaren et al. said that their participants who received all the data on handouts reported using rational inference to derive their conclusions, so that after Phase 1: BC+, Phase 2: AB+, and Phase 3: A+, they reasoned that if food A was responsible for the allergy, then food B must not have been, and if food B was not responsible, then food C was responsible.

Witnauer and Miller (2011) compared the second-order retrospective revaluation predictions that are made by the Van Hamme and Wasserman extension (1994) of the Rescorla-Wagner model (1972) with their own extension that involved more development of the role of within-compound associations. The Van Hamme and Wasserman extension modeled retrospective revaluation by updating an absent stimulus' associative strength with a negative learning rate, whenever a stimulus was presented with which it had a within-compound association. Witnauer and Miller's extension additionally multiplied this by the sum of within-compound associations between each of the present stimuli on a trial and the absent stimulus. Witnauer and Miller show that while both models demonstrate first-order retrospective revaluation effects, only their extension demonstrates the most commonly observed second-order retrospective revaluation effects, in which the first- and second-order associates move in opposite directions. It appears that the critical difference is that Witnauer and Miller's enhanced within-compound model encodes the sign of the within-compound associations (i.e., the inhibitory association between the non-shared elements of the two compounds), whereas the Van Hamme and Wasserman extension does not. Witnauer and Miller note that Stout and Miller's Sometimes-Competing Retrieval model (Stout & Miller, 2007) also predicts the second-order (and higher-order) effects, and that Dickinson and Burke's modification of SOP does not. They conclude that all models that can explain the most commonly observed higher-order retrospective revaluation effects use within-compound associations.

In a second-order retrospective revaluation experiment with Phase 1: AB+, Phase 2: BC+, Phase 3: A-, the present model predicts a different result than would be made by within-compound models. Because of lateral learning, recovery from overshadowing will occur to the shared element (i.e., B's associative strength will increase), but

the model also predicts that the other, non-shared element (C) will also elicit more responding, when tested. The reason for this is that when BC is conditioned, it will gravitate toward using a relatively similar ensemble of neurons as the previously conditioned AB. As a result, as A is extinguished, C's somewhat similar ensemble will have its lateral inhibition lowered as well. This leads to greater positive neuron activity upon presentation of C and thus greater associative strength.

If retrieval by within-compound association is the mechanism by which retrospective revaluation occurs, then we would expect that large within-compound associations should lead to greater retrospective revaluation than weak within-compound associations. Consider the following procedure: Phase 1: AX+, Phase 2: AX-, Phase 3: AY+, BX+, Phase 4: A-, Phase 5: X-, Y- (Test). After the first two phases, the within-compound associations between A and X should be relatively large, despite the fact that responding to AX after the second phase should have returned to near initial conditions (i.e., low responding). In the third phase, A and X are separated but conditioned in compound with Y and B, respectively. Given that the AX within compound association is stronger than the AY within-compound association after phase 3, then within-compound models predict that stimulus X should be revalued more than stimulus Y. The present model makes the opposite prediction, that Y will be revalued more than X. The first phase develops a neural ensemble for AX but the second phase extinguishes this, essentially restoring the network to initial conditions. The third and fourth phases are then seen as a simple recovery from overshadowing paradigm, where Y is revalued more than X. Rational inference makes the same prediction as the present model because at the end of phase 2, the inference would be that neither A nor X predicts the US. In this way the third and fourth phases become a simple recovery from overshadowing paradigm.

SLIM differs from other models of retrospective revaluation that do not employ within-compound associations. Although it revalues an absent stimulus according to associative mechanisms, it does so only when the stimulus presented in the second phase was previously paired with the absent stimulus to be revalued (i.e., unlike Ghirlanda, 2005 and Dawson, 2008). SLIM also does not make use of memory retrieval, though this is another route apart from within-compound associations to explain the phenomena. For example, the APECS model (McLaren, 1993; Le Pelley

& McLaren, 2001; McLaren, 2011) takes this approach. It is a neural network approach that recruits a new hidden layer neuron for each unique trial it experiences (e.g., separate nodes for A+, A-, and AX+). Although a detailed description of the model is not feasible here, the bias of a node representing a compound behaves in much the same way as our lateral inhibition mechanism. In recovery from overshadowing, the first phase establishes a compound node ("AB+") and associates it with the US. During the intertrial intervals of this phase, the 'bias' weight for this node is made negative, to offset the increased prediction made by the node when the inputs are absent. In the second phase, a new node is established ("A-") and during the intertrial intervals of this phase, the "AB+" node's bias is increased. This increases the "retrievability" of node "AB+", which then leads to an increased response upon presentation of stimulus B (i.e., recovery from overshadowing). The bias of the APECS model functions like the lateral inhibition in SLIM except that it has the opposite sign: in our model during extinction of stimulus A in the second phase, lateral inhibition is decreased, making the positive neuron response to B larger. In both models, the second phase does not change the input weights associated specifically with the absent (B) stimulus, but rather the lateral weights for SLIM and the bias for APECS.

Having two opposing pathways to compute associative strength is also a feature of the comparator hypothesis (Kasprow et al., 1987; Miller & Matzel, 1988). However, the comparator hypothesis uses the second pathway to evoke CS-CS associations and compare the associative strengths of different stimuli, while the present model simply uses the second pathway to help represent negative associative strengths. The dual pathway structure also bears resemblance to the division of CS-US and CS-no-US associations discussed in Le Pelley (2004). A model of spontaneous recovery from extinction by Pan et al. (2008) uses positive and negative weights, which are changed in opposite directions and are summed to produce a measure of responding.

7.9 Application to other retrospective revaluation findings

In Figure 7.11, we show that the present model can also explain the backward blocking effect (Shanks, 1985; Denniston et al., 1996; E. Wasserman & Berglan, 1998). Using the backward blocking procedure in Shanks (1985) along with an additional control group (see Figure 7.11 for details), we correctly simulate the effect ($p < 0.001$,

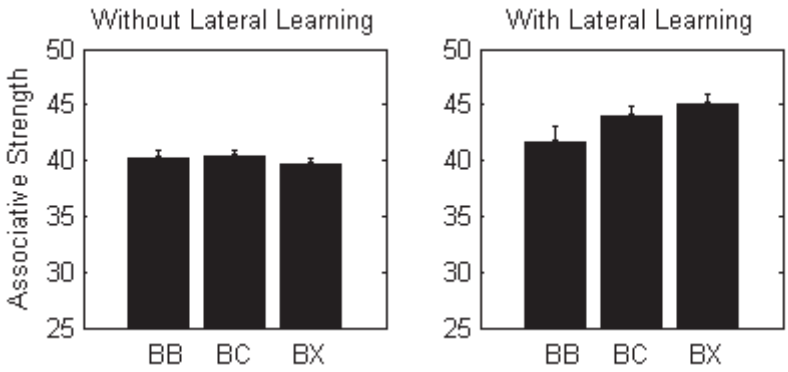

Figure 7.11: Responding to stimulus B in the test phase of backward blocking simulations when lateral learning is disabled (i.e., $\rho = 0$) and enabled (i.e., $\rho > 0$) using the paradigm of Shanks (1985) and an additional control group. With lateral learning enabled, the backward blocking group (Group BB: Phase 1 (50 trials): ABX+, X-, Phase 2 (200 trials): AX+,X-, Phase 3 (1 trial): BX- (Test)) expressed lower responding ($p < 0.001$, Wilcoxon signed-rank test, 15 differently initialized simulations or stat rats) than both control groups, Group BC and Group BX. In Group BC, phase 2 trials reinforced a novel stimulus (Phase 2: CX+ X-) while in Group BX, phase 2 trials did not involve any stimulus presentations (Phase 2: X- X-). In the other phases, these groups received the same treatment and test as Group BB. Note that in Phase 2, conditioning of A and the novel stimulus C reached asymptotic levels of responding in their respective groups. This simulation shows that lateral learning leads to a weak but significant backward blocking effect.

Wilcoxon signed-rank test, for 15 different model initializations or stat rats) with SLIM. In the figure, it appears that without lateral learning the procedure increases rather than decreases responding to the blocked stimulus relative to a control group (BX). However, this is simply due to greater extinction of the context in the control group, which is overwhelmed when lateral learning is enabled.

Backward conditioned inhibition (Chapman, 1991; G. Urcelay et al., 2008) refers to the paradigm in which a non-reinforced compound is presented in the first phase (AX-) followed by phase where one element is reinforced (A+). The result is that the other element becomes inhibitory relative to a control group. An experiment by Espinet et al. (Espinet, Iraola, Bennett, & Mackintosh, 1995) preexposed compounds AX and BX (AX-, BX-). In the second phase, conditioning to one of the non-shared constituents was conducted (A+). The result of these manipulations was that stimulus B's association with illness was either weakened or became inhibitory.

This is called the Espinet effect. Formally, the Espinet paradigm is the second-order analogue of the backward conditioned inhibition paradigm. The present model does not explain these effects. For the same reasons as the Rescorla-Wagner model, preexposure has no effect on subsequent conditioning phases. As a result, no sharing of neural representations occurs, unlike compound conditioning when a US is presented. If SLIM was extended to develop similar neural representations during a preexposure phase (as occurs in a conditioning phase), then the model might also come to explain these two effects.

Reminder-induced recovery from overshadowing (Kraemer, Lariviere, & Spear, 1988) is the finding that presentation of the overshadowed stimulus somewhere between the conditioning sessions and test sessions (the day following conditioning and two days prior to test in Kraemer et al., 1988) enhances responding to an overshadowed stimulus. Corresponding reminder-induced recovery has also been discovered in the blocking (Schachtman, Gee, Kasprow, & Miller, 1983), relative validity (Cole, Denniston, & Miller, 1996), and latent inhibition (Kasprow, Catterson, Schachtman, & Miller, 1984) paradigms. One prominent interpretation of the reminder-induced recovery from overshadowing findings is that the overshadowed stimulus' associative strength is not reduced by being conditioned in compound, as the notion of cue competition suggests, since later we find that responding has "recovered". More formally, the interpretation says that overshadowing is due to a deficit in performance (e.g., memory retrieval failure in the test phase) rather than a deficit in acquisition through cue competition in the conditioning phase. The question remains, however, as to what mental processes the reminder treatment might invoke. One remaining potential acquisition-deficit explanation (but see Schachtman et al., 1983) is that the reminder treatment strengthens a within-compound association between the overshadowed stimulus and the overshadowing stimulus. Then, when the overshadowed stimulus is later tested, the overshadowing stimulus is thereby retrieved and submitted as internal input to the associative learner, thereby generating a greater (or "recovered") level of responding. A similar mechanism might also explain spontaneous recovery from overshadowing (Kasprow, Cacheiro, Balaz, & Miller, 1982), which has also been thought to indicate a performance-deficit rather than an acquisition-deficit in learning. In this phenomenon, responding to the overshadowed stimulus is greater

after a retention interval. The acquisition-deficit explanation would say that time, instead of a reminder treatment, may lead to stronger within-compound associations. There is some evidence within a sensory preconditioning paradigm, however, that within-compound associations degrade rather than strengthen when there is a delay between conditioning and test (Pineo, Urushihara, & Miller, 2005). Additional experimental work testing the strength of within-compound associations after reminder treatments and post-acquisition delays may better discriminate between the performance-deficit and acquisition-deficit explanations.

Retrospective revaluation effects are not always observed (Shevill & Hall, 2004; Dopson, Pearce, & Haselgrove, 2009). Within-compound association-based approaches can often explain this as a failure in within-compound association-based retrieval either during conditioning or test. In SLIM, retrospective revaluation phenomena are reduced when the conditioned stimuli in the initial pairing are similar, that is, when there is significant overlap between their distributed input representations. Consider when stimuli are nearly identical. This will generate strongly similar input representations such that subsequent conditioning or extinction of one will similarly affect the other, because the model treats them as essentially the same stimulus. This is the opposite of retrospective revaluation behaviour and is referred to as mediated conditioning, which has been found to occur when the paired stimuli are strongly similar (Liljeholm & Balleine, 2009). Then to observe neither mediated conditioning nor retrospective revaluation in the present model, one explanation is that the stimuli making up a compound stimulus have some middle-ground degree of similarity. From our model, another possible explanation of why retrospective revaluation is sometimes not observed is that the lateral learning rate (ρ) may change dynamically. In the present model, we set $\rho > 0$, which supports retrospective revaluation phenomena. As noted in Figure 7.9, when $\rho = 0$, no retrospective revaluation occurs. Furthermore, if we set $\rho < 0$, the opposite of retrospective revaluation would occur (i.e., mediated conditioning) because lateral learning would change weights in the opposite direction. For example, recall the process in Figure 7.10. Instead of reducing lateral weights, which increased the absent stimulus B's associative strength, the lateral weights would be increased, which would reduce activity and associative strength when B is presented. Thus as stimulus A is extinguished, so would be its

previously partnered stimulus B.

In a recovery from overshadowing experiment, Liljeholm and Balleine (2006) found that the extinction of the more salient element of the compound revalued the less salient element more than the other way around. In the present model, the more salient stimulus takes a larger share of the associative strength due to cue competition. This means that the more salient stimulus will have more associative strength to extinguish and its lateral weights will be reduced proportionally. Thus extinction of a salient stimulus will lead to more revaluation of the absent stimulus than will extinction of a weakly salient stimulus. In more general terms, the larger the change in a present stimulus' associative strength, the more the absent stimulus is revalued.

7.10 SLIM Predictions

In Section 7.8, I described a procedure that would pit rational logic against within-compound conditioning models but that lines up with SLIM's prediction. In short, SLIM predicts that the extinction of a conditioned compound largely unties the constituent stimuli from future revaluation of one another in subsequent procedures, whereas within-compound models predict that extinction instead increases the tie between the stimuli.

In general, attenuating lateral learning (pharmacologically or otherwise) should attenuate retrospective revaluation phenomena. The model also assumes that, during an update, the lateral inhibitory weights of a neuron change in the same direction (positively or negatively) as its input weights. In contrast, if lateral weights were changed in the opposite direction, the model would not generate retrospective revaluation effects. It will be interesting to see whether or not neuroscience confirms this key feature of the model, which might be done by evaluating the changes of synaptic efficiency of the lateral connections using a paradigm similar to Shen et al. (2008).

Predictions may be made from the configural capability of the model as well. In positive patterning, a configural cue must be invoked so that A, B, and AB can achieve their target responses. So, in the present model, we have a configural cue pulling up (AB+, an increase in direct pathway activations), which generalizes somewhat to A and B, while the constituents themselves are being pulled downward (A-, B-, an increase in indirect pathway activations). This tug of war results in significant activity

in the indirect and direct pathways upon presentation of the constituents (A,B) so that a balance between these pulls will result in no conditioned responding. To contrast, in simultaneous feature-positive discrimination (AB+,B-), constituent B will be pulled in two directions at first, but in the end conditioning will accrue to A alone and B will activate model neurons with about the intensity of a novel stimulus. Because of the configural-cue-generating nature of the model, it predicts that the neural activity that B induces should be greater following positive patterning than following simultaneous feature-positive discrimination, even though the amount of conditioned responding to B is the same at the end of the procedures (zero). If physiological recordings are able to confirm this, it would suggest that the dual pathways are indeed used to represent configural cues and further support the notion that the basal ganglia is involved in classical conditioning.

Chapter 8

Dual Pathway Regression

8.1 Chapter Summary

Here, we show a way of implementing features of the Dual Noisy OR model in an LMS-like format that suits most of the high-level structural details of the basal ganglia. This model is named Dual Pathway Regression and performs comparably to the Dual Noisy OR model in the regression task. Dual Pathway Regression is then translated into a classical conditioning format and is used to explain additional phenomena beyond those explained by LMS. Finally, several experimental predictions are made. Some of the text in this chapter is taken from Connor and Trappenberg (2013), ©2013 IEEE, in which I was primarily responsible for developing the theory and simulations as well as drafting the manuscript.

8.2 Dual Pathway Regression (DPR): An LMS-like Dual Noisy OR Model

The Dual Noisy OR model of Chapter 5 is an abstract model, not clearly related to neural structure or function. Here we seek to remedy this issue and at the same time highlight the mechanisms of the Dual Noisy OR model that are primarily responsible for its ability to lower prediction errors.

Earlier, we noted that the success of the Dual Noisy OR model was that it prevents negative residuals on irrelevant parameters from developing. This prevents the canceling of the effects of positive parameters by negative parameters that allows residual parameters to persist. As noted in Chapter 5, one of the mechanisms behind overcoming overfitting in the Dual Noisy OR model is to change the way in which inhibition is integrated. In principle, there are at least two ways in which inhibition can be integrated. One is to sum negatively weighted inputs with positively weighted

inputs (i.e., LMS). The other, put forward by the Dual Noisy OR model, is to multiply the positive prediction of the outcome by some number between 0 and 1, where more inhibitory influence reduces this multiplier. To integrate this second approach into LMS, we will create a positive and negative model as in the Dual Noisy OR model. This dual pathway LMS-like model is formulated as

$$y = 2\phi_+^T x(1 - \frac{1}{1 + e^{-\phi_-^T x}}) \tag{8.1}$$

where ϕ_+ and ϕ_- are the positive model and negative model parameters respectively (though both are forced to have positive values only), which both receive the same set of inputs, x. The sigmoid function ensures that the range of values for the inhibitory influence is contained between 0 and 1, where the slope is 1 when all parameters are 0. The parameter values can be learned using gradient ascent, as in the other methods evaluated here. This is the natural outcome of performing maximum likelihood estimation on a Gaussian random variable with Equation 8.1 at its mean. Now, the gradient of a parameter depends on its associated pathway. For the positive pathway parameters, the gradient becomes

$$\frac{\partial L(\phi)}{\partial \phi_{+,j}} = \sum_{i=1}^{m} (y^{(i)} - 2\phi_+^T x^{(i)}(1 - \frac{1}{1 + e^{-\phi_-^T x^{(i)}}}))x_j^{(i)} \frac{1}{1 + e^{\phi_-^T x^{(i)}}} \tag{8.2}$$

and for the negative pathway parameters, it becomes

$$\frac{\partial L(\phi)}{\partial \phi_{-,j}} = -\sum_{i=1}^{m} (y^{(i)} - 2\phi_+^T x^{(i)}(1 - \frac{1}{1 + e^{-\phi_-^T x^{(i)}}}))x_j^{(i)} \frac{e^{\phi_-^T x^{(i)}} \phi_+^T x^{(i)}}{(1 + e^{\phi_-^T x^{(i)}})^2} \tag{8.3}$$

where a positive or negative parameter is updated by

$$\phi_j =: \phi_j - \alpha \frac{\partial L(\phi)}{\partial \phi_j} \tag{8.4}$$

A few interesting properties can be seen in these learning rules. As in the Dual Noisy OR model, the positive pathway's parameters are changed in the opposite direction as the negative pathway parameters during an update. Also, because initial parameters are set very near zero, initial parameter changes are almost identical to LMS for the positive pathway. Like the Dual Noisy OR model, negative pathway parameters are updated in proportion to the existing model prediction. This is how the approach distinguishes between inhibitory and irrelevant features: inhibitory features tend to reduce existing model predictions whereas irrelevant features do not.

Figure 8.1 illustrates the effectiveness of this approach, which will be referred to as Dual Pathway Regression (DPR). The top and bottom panels compare this approach in the familiar regression task, in which the number of features and the noise variance are varied. In both cases, we see that DPR performs comparably to the Dual Noisy OR model.

8.3 DPR's Relationship to Neurobiology of the Basal Ganglia

Figure 8.2 illustrates basal ganglia anatomy and DPR. For clarity, the positive and negative pathways will be described as P_+ and P_-, respectively. The dual pathway nature of DPR can be fitted to the dual pathway structure in the basal ganglia, with the positive and negative pathways mapping onto the direct and indirect pathways, respectively. Appropriately, the inhibitory projections of the indirect pathway support the inhibitory $1-P_-$ term, where the "1" is contributed by excitatory input from the STN. The pathways converge in the output nuclei and SNc/VTA, providing an inhibitory product ($-P_+(1-P_-)$) to add to the tonic activity supported by the STN (in SNr/GPi) or the reinforcement signal, λ (in SNc/VTA), giving $1-P_+(1-P_-)$ for the SNr/GPi (assuming again that STN contributes the "1") and $\lambda-P_+(1-P_-)$ for the SNc/VTA. This inhibitory product could be computed in two different locations, since the goal is that the output nuclei incorporate an inhibitory prediction computation of $P_+(1-P_-)$, which expands to $P_+-P_+P_-$. One possibility is that the product is performed in the output nuclei and the SNc/VTA. Mathematically, the product of two negative numbers is positive, but in terms of neural effects, the two inhibitory projections could conceivably amplify one another's inhibitory influence. Perhaps a more parsimonious location, however, for this computation is in the GPe. We know that there is an axon collateral projection from the direct pathway to the GPe (Kawaguchi et al., 1990). If, again, the two inhibitory projections amplify one another's inhibitory influence, the computation at the level of the GPe would become $1-P_+P_-$. Then, only a simple summation of the inputs to the output nuclei and SNc/VTA (P_+ from the direct pathway plus $1-P_+P_-$ from the indirect pathway) would yield the necessary computation there. Thus, only one multiplicative integration (in the GPe) would be needed instead of three or four (SNr, GPi, and SNc/VTA). Yet, it appears that the striato-nigral projection collaterals to the GPe

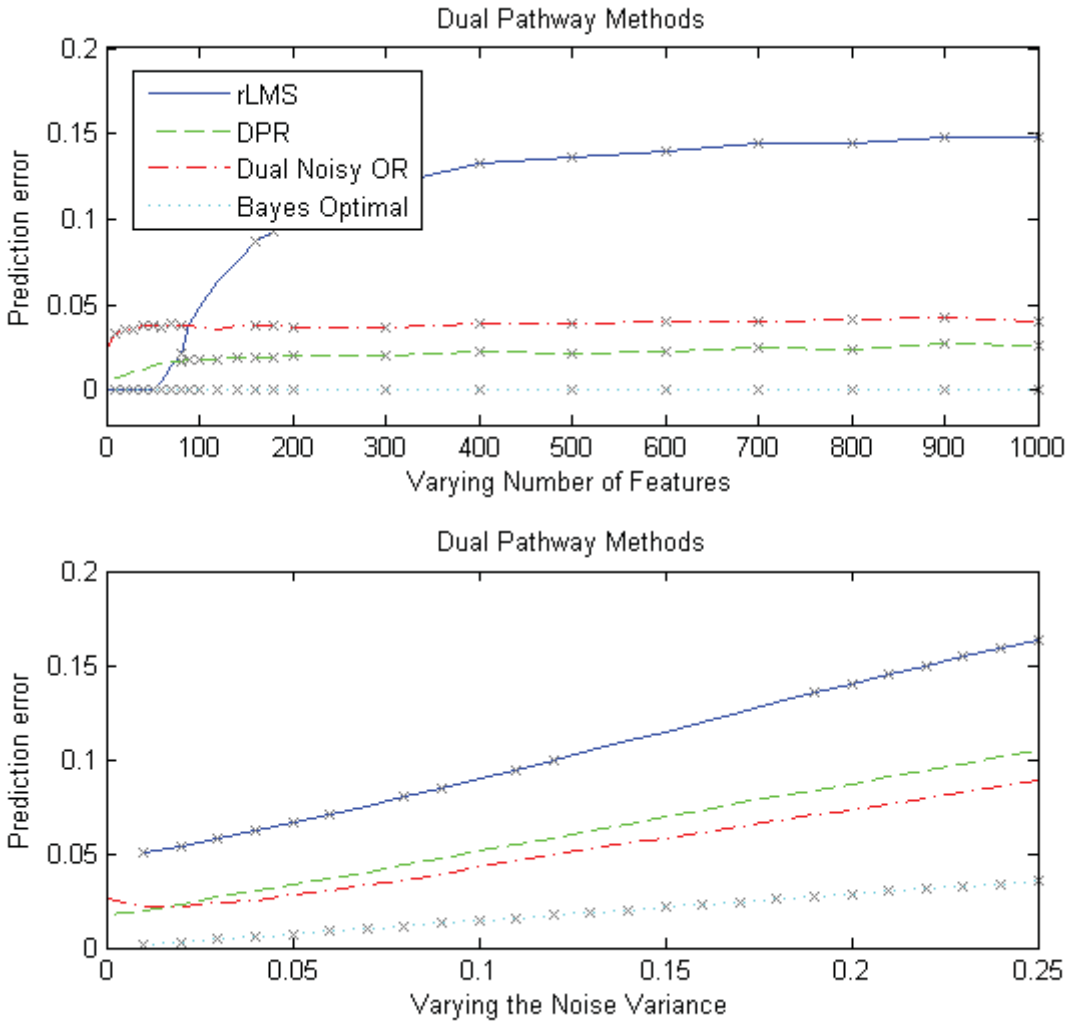

Figure 8.1: *Top Panel:* Prediction error as the number of features is varied. DPR performs a little better overall than the Dual Noisy OR model, yet still substantially worst than the optimal. *Bottom Panel:* Prediction error as the variance of the noise is varied. Here, DPR's performance is very similar to the Dual Noisy OR model.

are more weakly connected than are the GPe-GPi connections (Wu, Richard, & Parent, 2000; Parent & Hazrati, 1995a). In fact, it seems that the GPe to GPi projections are strong, since they frequently contact the soma of GPi neurons, seemingly giving this projection a modulatory role (Nambu, 2007). The same connectivity is also seen between the GPe and SNr (Smith & Bolam, 1989) and between the GPe and SNc (Smith & Bolam, 1990) (although more sparsely so). The form of the GPe to VTA connectivity appears to be unknown. So, although structurally it seems that a multiplication in the GPe would be simpler, multiplication in the output and dopaminergic nuclei seems better supported at present. It also appears that such a shunting inhibition from the GPe projection makes a summation approach to integrating the two pathways less tenable than a multiplicative approach.

As discussed above, learning rules that invoke opposite signs for positive and negative pathways in DPR nicely fit with what is known about cortico-striatal synaptic plasticity (Shen et al., 2008; Frank & Fossella, 2011), including the effects of the dopamine prediction error signal on the direct and indirect pathways, respectively. According to DPR's mathematical formulation, however, additional learning signals are needed. For the positive pathway, the negative pathway's prediction is needed and, for the negative pathway, the positive pathway's prediction is needed. One simple possibility is that this information is transferred through the lateral inhibitory connections in the striatum, although this may conflict with SLIM's proposal for such connections. There is at least one other possibility. The way we arrived at these learning rules was to find the maximum likelihood estimate, which results in taking the derivative of the overall prediction with respect to positive and negative pathway parameters. This allows us to ascend a gradient to reach the minimum prediction error. However, it is not absolutely necessary to take the exact gradient of this function to get to near the minimum. Figure 8.3 shows that we can modify the learning rules and still perform comparably. In short, all that must be done is to remove some negative pathway prediction terms from each of the learning rules giving

$$\frac{\partial L(\phi)}{\partial \phi_{+,j}} = \sum_{i=1}^{m} (y^{(i)} - 2\phi_+^T x^{(i)} (1 - \frac{1}{1 + e^{-\phi_-^T x^{(i)}}})) x_j^{(i)} \tag{8.5}$$

and

$$\frac{\partial L(\phi)}{\partial \phi_{-,j}} = -\sum_{i=1}^{m} (y^{(i)} - 2\phi_+^T x^{(i)} (1 - \frac{1}{1 + e^{-\phi_-^T x^{(i)}}})) x_j^{(i)} \frac{\phi_+^T x^{(i)}}{(1 + e^{\phi_-^T x^{(i)}})} \tag{8.6}$$

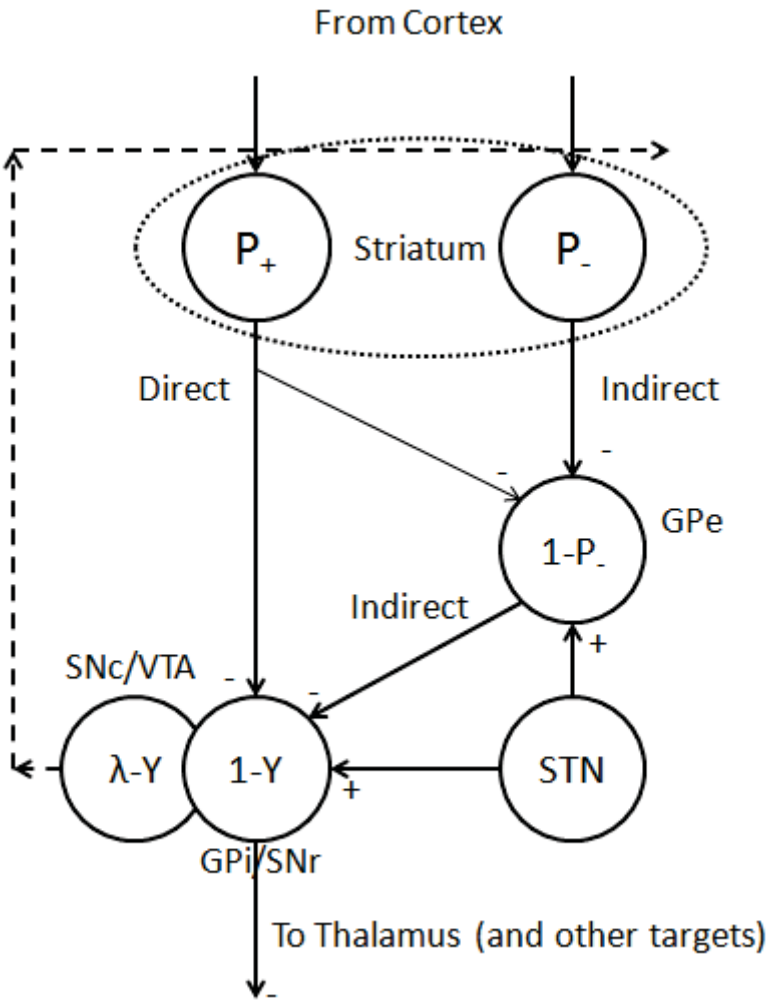

Figure 8.2: Mapping DPR onto the basal ganglia. The positive and negative pathway contributions (P_+ and P_-) follow the direct and indirect pathways, out of the striatum and into their targets, the GPi/SNr/SNc/VTA and GPe, respectively. Here, they subtract from tonic activity (supported by the STN). Multiplicative inhibition may occur in either the GPe or the output targets and SNc/VTA (see text for details). If this occurs, the output nuclei and SNc/VTA will receive a prediction $Y = P_+(1-P_-)$, which they will use to compute their output signals. Cortico-striatal learning agrees with the notion that the two DPR pathways learn with opposite signs in tandem with a dopamine signal from the SNc/VTA that encodes prediction error. Additional signals appear necessary for proper DPR learning as well, though perhaps only one provided by the thalamus is necessary, if DPR is slightly simplified (see text for details).

These two equations strip the model down to the essential terms or mechanisms also at play in the Dual Noisy OR model. The similar effectiveness of this reduced model suggests that it represents the set of mechanisms responsible for reducing prediction error in both DPR and the Dual Noisy OR model. The first simplified equation allows the positive pathway to learn based on the prediction error and input activity only, just like LMS. The second equation requires the negative pathway to additionally have access to the model's prediction. One possibility is that this is conveyed by the output nuclei. The output nuclei express the model's prediction and the thalamus, a target of the output nuclei, is known to provide inputs directly to the striatum (see Parent & Hazrati, 1995a for a brief review), which are believed to be excitatory. However, the thalamic input does not make preferential contact with one pathway or the other (Doig, Moss, & Bolam, 2010; Huerta-Ocampo, Mena-Segovia, & Bolam, 2013) whereas this model supposes that thalamic connections make contact mostly with the indirect pathway. Recalling the gradient terms in the Dual Noisy OR update rule (Equations 5.5 and 5.6), we see that parameters from both pathways are changed in proportion to the model's full prediction. However, positive pathway parameters are also changed inversely proportional to the positive pathway prediction. A simple substitute for this may be that the parameters of a direct pathway neuron are updated inversely proportional to the neuron's activity. Although the DPR model does not use multiple neurons and there is no postsynaptic activity proportional term in the learning rule, such is not incompatible with DPR, as will be shown in Section 9.5.

8.4 Classical Conditioning Simulations and DPR

DPR is capable of explaining some classical conditioning phenomena beyond the ability of LMS. To demonstrate this, we will first express DPR in standard classical conditioning modeling terms. This will also help to describe novel classical conditioning predictions made by DPR in the next section. Given the following constraints,

- must guarantee that $P_- \leq 1$, as DPR does by employing the sigmoid function
- inputs have a binary value (i.e., stimuli are either present or not present)
- positive and negative model associative strengths are not permitted to go below zero

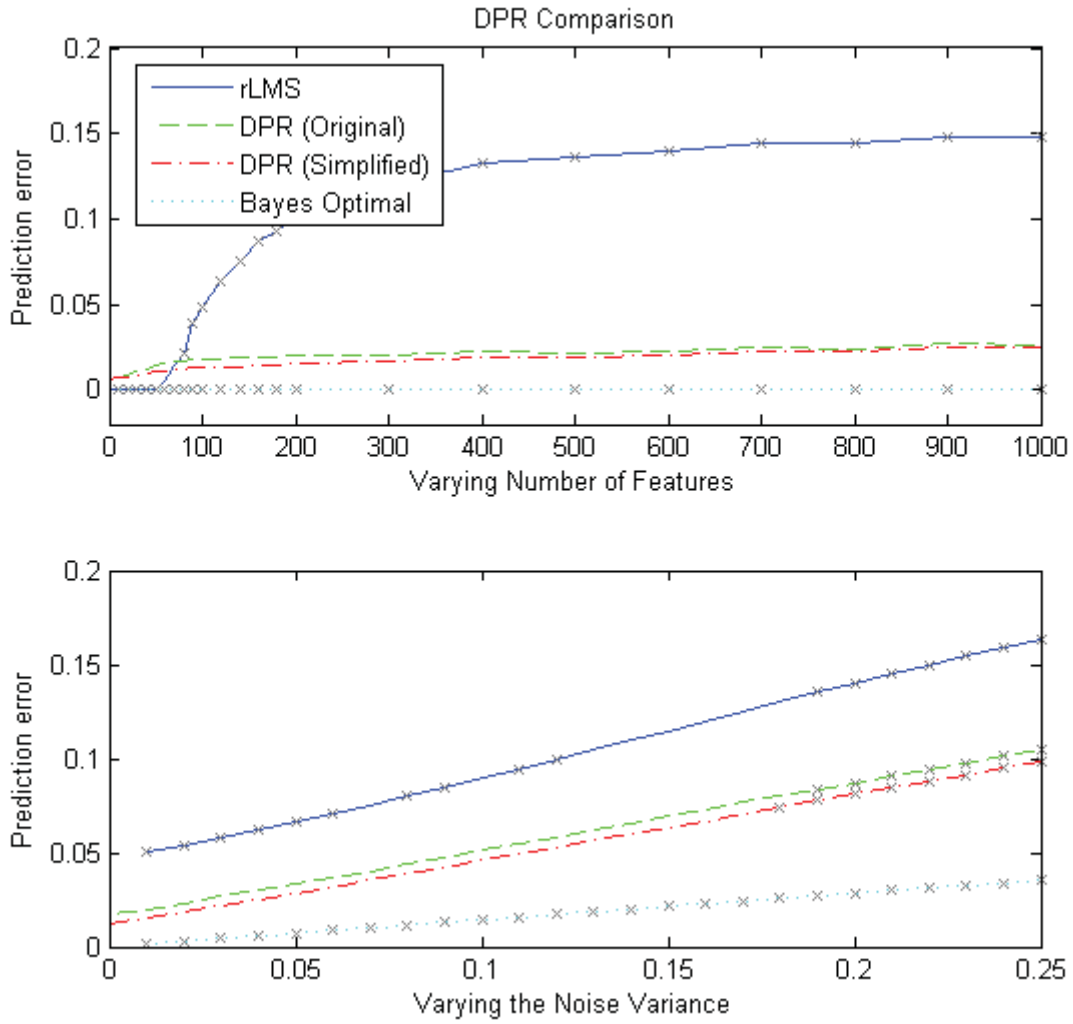

Figure 8.3: Comparison of two DPR models, where one employs the original update equations (Equations 8.2 and 8.3) and the other employs the simplified update equations (Equations 8.5 and 8.6). The performance of the models is very similar in both panels except that the prediction error for the simplified version is slightly less than the original in the bottom panel for high noise, bringing it a little closer to the results of the Dual Noisy OR model shown in Figure 5.4.

we can write the associative strength of a compound stimulus AB as

$$V_{AB} = (V_A^+ + V_B^+)(1 - V_A^- - V_B^-) \tag{8.7}$$

where V_A^+ and V_A^- are the positive pathway and negative pathway contributions for stimulus A, respectively. The compound stimulus associative strength is here defined to convey how multiple stimuli interact. If instead of V_{AB}, we wanted to compute V_A, we would drop the V_B^+ and V_B^- terms from Equation 8.7. Following the simplified update rules for DPR, the update for the positive and negative associative strengths for a single stimulus (A), after a trial in which compound AB was presented, is

$$\Delta V_A^+ = \alpha_A \beta (\lambda - V_{AB}) \tag{8.8}$$

and

$$\Delta V_A^- = \alpha_A \beta (\lambda - V_{AB}) V_{AB}, \tag{8.9}$$

respectively. Note that the update rule for V_A^+ is the Rescorla-Wagner model update rule and the update rule for V_A^- is additionally proportional to the associative strength of the compound. The first example of this model used in a conditioning experiment was already shown in Figure 5.8, demonstrating that it is capable of explaining relative validity, just as can rLMS with the LASSO and the Dual Noisy OR model.

8.4.1 Recovery from Conditioned Inhibition: Inhibitory Residuals are Expected

Recall that rLMS is the optimal model for the data in our simulated world (assuming a uniform prior distribution). The derivation of this optimal model indicated that for this simulated world, it is appropriate to turn off learning when the prediction for the current input is negative. It was also noted that animals appear to behave in the same way. In the classical conditioning phenomenon of conditioned inhibition (Rescorla, 1969), subjects receive trials of A+, AB-. The result is that subjects view stimulus A as predictive of reinforcement and that B cancels reinforcement or predicts no reinforcement. Classical conditioning experiments confirm B's inhibitory quality (Rescorla, 1969), such that the subject's response is smaller when B is combined with a separately conditioned stimulus than when this stimulus is presented alone (i.e., $C + B < C$). Now, if stimulus A were presented but not reinforced in a second phase,

its prior conditioning would be extinguished and the animal would no longer respond to it. Likewise, it seems reasonable to expect that if the inhibitor stimulus B were presented but not reinforced that this would extinguish its inhibitory quality as well. In fact, this is a direct prediction of the Rescorla-Wagner model and is referred to as recovery from conditioned inhibition by extinction of the inhibitory stimulus.

This prediction and slight variations thereof have been evaluated experimentally a number of times (Baetu & Baker, 2010; DeVito & Fowler, 1986; DeVito & Fowler, 1987; Hallam, Grahame, Harris, & Miller, 1992; Lotz & Lachnit, 2009; Williams, 1986; Williams, Travis, & Overmier, 1986; Witcher & Ayres, 1984; Zimmer-Hart & Rescorla, 1974). In the first of these experiments, Zimmer-Hart and Rescorla (1974) conditioned rats using the following procedure (Experiment 1) for each of 3 daily sessions: A+ (tone, 6 trials), B+ (clicker, 6 trials), and AX- (tone+light, 12 trials). They divided the rats into two groups. Group E received 24 trials of non-reinforced presentations of X-, while Group C spent an equal amount of time in the conditioning chamber without stimulus presentations. Testing of AX relative to A demonstrated X had become a conditioned inhibitor. The attempted extinction of X, however, failed since there was no difference in responding to compounds AX or BX between Groups E and C. In their Experiment 2, they used a within-subjects design but found the same result. They therefore failed to extinguish the conditioned inhibition. Williams (1986) arrived at similar results despite using a different procedure to establish the conditioned inhibition, which was accomplished by interspersing presentations of the to-be inhibitor stimulus and US-alone in an explicitly unpaired way. Devito and Fowler (1987) also arrived at similar results, except that a moderate inhibitors' strength was enhanced, rather than extinguished, in non-reinforced presentations (see Williams et al., 1986, Experiment 4 for a similar result). Witcher and Ayers (1984) used a similar paradigm to Zimmer-Hart and Rescorla, but instead of only non-reinforcing the inhibitor X in the attempted extinction phase, they also made non-reinforced presentations of the excitor A and the compound AX. Yet, they similarly found that inhibition to X remained (see Hallam et al., 1992 for a similar procedure and result). In their second experiment, they found that by presenting X and the US randomly and independently of one another they could extinguish the inhibition, presumably by making X appear irrelevant or non-predictive (see DeVito

& Fowler, 1986 for a similar procedure and result).

More recent work has disagreed with the bulk of these results. Lotz and Lachnit (2009), used a human causal learning paradigm where, on each trial, the subject was presented with a food and asked to indicate how it affected a hormone level, and was then told the outcome. In Group Unidirectional, subjects could indicate that the hormone level would be either elevated or unchanged whereas the Group Bidirectional could additionally indicate that the hormone level would be decreased. Conditioned inhibition was established using the standard paradigm (A+,AX-) in the first learning phase. This was followed by a phase attempting to extinguish the inhibitor (A+, X-). The result was that extinction occurred in Group Bidirectional but not in Group Unidirectional. So, by providing the option to rate foods as decreasing the hormone level, they found extinction. Baetu and Baker (2010) followed up with a very similar experimental paradigm. The key difference between their experiment and that of Lotz and Lachnit (2009) is that they dropped an intermediate test phase and modified the test phase to ask for a rating value rather than selection of specific (increase, neutral, decrease) outcomes. As a result, Baetu and Baker found that extinction of inhibition occurred in both the Unidirectional and Bidirectional groups.

In summary, although the results are mixed, the burden of the evidence, involving the majority and most straightforward experiments suggest that inhibition is not extinguished by mere non-reinforced presentations of the conditioned inhibitor.

In Figure 8.4, we show the second phase of a simulation of recovery from conditioned inhibition by extinction of the inhibitory stimulus (i.e., Phase 1: A+, AB-, Phase 2: B-). In this simulation, we conditioned the models in an online-learning mode (data is not repeatedly reprocessed) to better show the changes that occur to the parameters. The results confirm that the Rescorla-Wagner model extinguishes the inhibitory strength of stimulus B. For the Rescorla-Wagner model, a second phase B- trial results in a positive prediction error, encouraging the inhibitory B stimulus to become neutral instead of remaining an inhibitor. The same is true of LMS, which would give results equivalent to the Rescorla-Wagner model except for their being compacted along the x-axis. In contrast, the attempted extinction of the inhibitory stimulus (B) is ineffective in DPR and rLMS models. In rLMS, the presence of an inhibitor alone gives a negative prediction and thus shuts off its learning according

to Equation 4.7, which prevents the inhibitor's extinction. Apparently, this is the optimal thing to do in a world where there are no "negative" or anti-reinforcement outcomes (e.g., less than zero food pellets), since rLMS is the optimal model for performing spatial credit assignment in such a world. In a way, this makes intuitive sense. Since the inhibitory stimulus is non-reinforced in both phases, there is little reason to change one's belief about the value of that stimulus in the second phase. This contrasts with excitatory conditioning and then subsequent extinction (A+, then A-), where the outcome differs between phases. DPR also does not extinguish inhibitory stimuli when presented alone. According to Equation 8.7, the value for V_{AB} will *always be positive* because it is the product of two positive numbers. Since, $\lambda = 0$ in a non-reinforced trial, the prediction error $(\lambda - V_{AB})$ in the second phase (B-) will always be negative (or zero). The other terms of Equation 8.9 will always be positive, so V_B^- can only be reduced in a non-reinforced trial. Thus, an inhibitor stimulus presented alone will never extinguish in DPR.

8.4.2 A Stimulus can be an Excitor and an Inhibitor at the Same Time

Miller, Barnet and Grahame (1995) grouped the failed extinction of conditioned inhibition among several other failed predictions of the Rescorla-Wagner model. Among these is the prediction that a single stimulus cannot have both excitatory and inhibitory properties at the same time. One experiment in opposition to this notion is Matzel, Gladstein, and Miller (1988). They partially reinforced a stimulus and it had excitatory properties when presented alone but also passed the standard summation and retardation tests for conditioned inhibition. DPR can accommodate this finding, because it has separate positive and negative parameters associated with each input (stimulus) and employs a multiplicative learning rule. In fact, partially reinforcing a stimulus will lead to non-zero positive and negative pathway parameters in DPR (not shown). So, for example, suppose a partially reinforced stimulus' DPR parameters were 0.5 for the positive parameter and 0.5 for the negative parameter. The perceived associative strength of the stimulus, when presented alone would be excitatory (i.e., $0.5(1 - 0.5) = 0.25$). Also, when paired with a continuously reinforced stimulus with a small positive parameter (say, 0.25) and a zero negative parameter, it will still appear to be excitatory

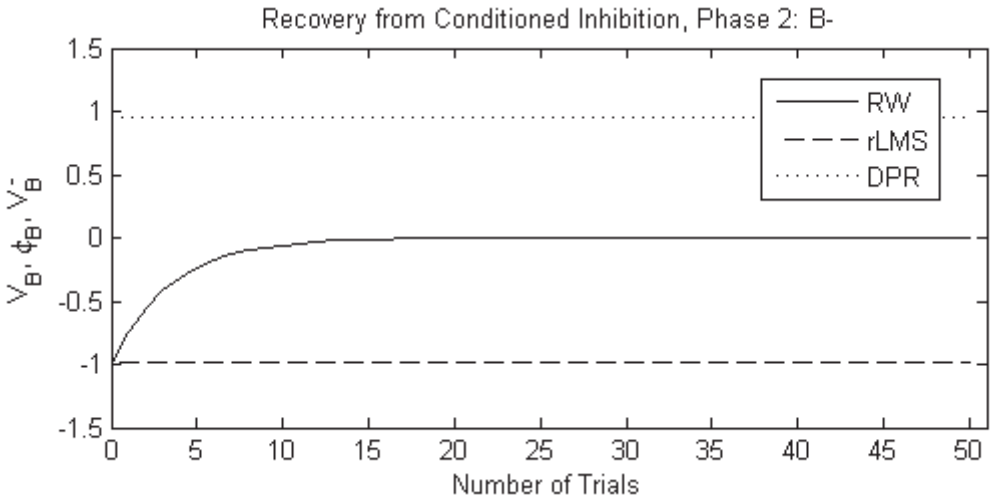

Figure 8.4: Recovery from conditioned inhibition by extinction of the inhibitory stimulus (Phase 1: A+, AB-, C+, Phase 2: B-). Shown, are the stimulus B parameter values for the Rescorla-Wagner (V_B) and rLMS (ϕ_B) models and B's negative pathway prediction strength for DPR (V_B^-). Not shown are the curves for a control group that received no Phase 2 presentations, which would then test at the same levels as the first trials in the figure for these models. Generally, this procedure does not seem to extinguish the inhibitory stimulus in either animal learning experiments nor in the rLMS or DPR models. In contrast, the Rescorla-Wagner model and LMS predict that the inhibitory strength will be extinguished with non-reinforced presentations of the inhibitory stimulus. See text for explanation.

(i.e., $(0.25 + 0.5)(1 - 0.0 - 0.5) = 0.375$) relative to a novel control whose parameter values are zero (i.e., $(0.25 + 0.0)(1 - 0.0 - 0.0) = 0.25$). However, if paired with a stimulus whose positive and negative parameters were 1.0 and 0.0 respectively, the perceived associative strength would be inhibitory ($(1 + 0.5)(1 - 0.0 - 0.5) = 0.75$) relative to a novel control (i.e., $(1.0 + 0.0)(1 - 0.0 - 0.0) = 1$). DPR can therefore represent a stimulus so that it appears either excitatory or inhibitory, dependent on the circumstances of its presentation.

8.5 DPR Predictions

In addition to the explanations of classical conditioning phenomena provided by DPR, this model can make novel classical conditioning experimental predictions.

Again, in several models the aggregate prediction or associative strength is simply the sum of the parameters associated with the input stimuli. DPR would agree that for excitatory stimuli, a sum-like operation occurs. However, it also says that the integration of inhibition is multiplicative. One important test of DPR, then, is to evaluate this feature of the model, as expressed in Equation 8.7. The following experiment attempts to contrast the summation vs. multiplication hypotheses for inhibition: Phase 1: A+, B+, BX-, Phase 2: ABX-, ACD- (test). In the first phase, A and B become excitors and X becomes an inhibitor, fully canceling responding to B in BX. In a summation scheme like the Rescorla-Wagner model, B and X would have equal associative strengths but opposite signs, to give zero responding for presentations of BX. So, when A is added to the compound, responding should be equivalent to the combination of stimulus A and two novel stimuli (to account for generalization decrement due to external inhibition). In a multiplication scheme, however, the inhibitor induces a shunting effect, substantially reducing responding to all present stimuli. Thus, DPR would predict that responding to ABX will be significantly less than to ACD. One potential confound, however, is generalization decrement due to the 3-stimulus compound. Although perhaps unlikely, it may be that the BX compound is more potent than a novel stimulus at inducing external inhibition and explains low responding rather than shunting or multiplicative inhibition.

Learning in DPR is also distinctive because it has two different variables per stimulus and different ways in which to update them. Above, we showed that recovery

from conditioned inhibition does not occur by extinction of the inhibitory stimulus (X-). It does occur, however, by the conditioning of the inhibitory stimulus (X+) (Rescorla, 1969). However, the subsequent conditioning is known to be slower than for conditioning a novel stimulus. Notice again that the negative pathway learning rule (Equation 8.9) increases negative pathway parameters in proportion to the prediction of the model for the present stimuli. This is important because it says that stimuli can only become inhibitory inasmuch as they reduce existing predictions. Turning this around, however, we can say that if a strong inhibitor (X) were subsequently conditioned in compound (e.g., AX+), it would condition at a rate proportional to the model prediction as well. The proposed experiment becomes: Phase 1: A+, AX-, B+, D+, Phase 2 (few trials): G1: BX+, G2: CX+, Phase 3: DX- (both groups). In the first phase, the excitors and an inhibitor are established. In phase 2 a few trials are used to extinguish the conditioned inhibitor by conditioning it in compound with a previously conditioned stimulus (group 1) and a novel stimulus (group 2). DPR predicts that there will be more extinction of inhibition in group 1 than group 2 during phase 2 and thus more responding during the test phase (phase 3). In contrast, the Rescorla-Wagner model predicts that more extinction will occur in group 2 because it will have a larger US surprisingness term than will group 1.

Chapter 9

Relationship to Other Basal Ganglia Models

9.1 Chapter Summary

Here, we review certain basal ganglia models in the literature. The purposes of such models range from time encoding to sequence learning. The evaluation is restricted to the most related models, and in particular, we will examine two models that bear the most resemblance to SLIM. In so doing, the uniqueness of the proposed models and their relationship other models will be shown.

9.2 Models with a Focus on the Dual Pathway

Today, models of the basal ganglia explain how the structure may be involved in a variety of computations and roles from sequence learning to principle components analysis (Houk, 2007). Their development really began in earnest around 1990. The Albin-DeLong model (Albin, Young, & Penney, 1989; DeLong, 1990), shown in Figure 9.1, was the first dual-pathway model of the basal ganglia. It offered an explanation for the hyper and hypo-kinetic disorders as a difference in the effectiveness of pathway-specific striatal neurons on their targets. One feature of this model is that the indirect pathway here flows through the STN whereas some more recent models (including SLIM and DPR) instead make use of the projection from the GPe directly to the output nuclei. This trend follows Parent and Hazrati (1995a), which suggest that the STN-GPe reciprocal connection is likely segregated from the STN-output nuclei connections. Nevertheless, several models have followed and extended the Albin-Delong's interpretation of the direct pathway. Mink (1996) notes that because the connections from the STN to the output nuclei are diffuse and the direct pathway's connections are focused, that one can expect a center-surround effect. Such a mechanism could act to disinhibit the central action and suppress the neighbouring actions and thus resolve competitions between them. Gurney et. al (2001) reformed

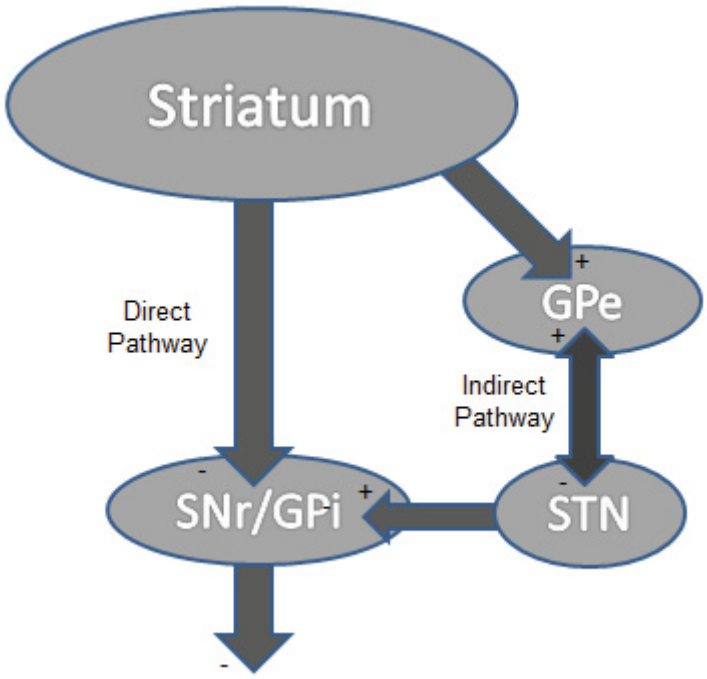

Figure 9.1: The Albin-Delong model of the basal ganglia, consisting of striatial input and GPi/SNr output with a dual pathway structure. This early model offered explanations for hypo- and hyper-kinetic disorders. Its indirect pathway is routed through the STN. Terms: GPe - globus pallidus externa, STN - subthalamic nucleus, SNr - substantia nigra pars reticulata, GPi - globus pallidus interna

the direct and indirect pathway paradigm to instead consist of selection and control pathways. They suggest that the striatal neural circuitry of the selection (direct) pathway expresses a salience for each action and that the control (indirect) pathway scales the output activity via the diffuse connectivity of the STN, providing a inter-nuclei center-surround effect.

One common feature of these models is that the direct and indirect pathways are considered the means of facilitating or suppressing certain actions. In contrast, Houk et al. (1995) used the same structure to explain the process by which the reinforcement values of stimuli could be learned. They superimposed this model on the striosomal compartments that link the striatum and the midbrain dopamine system. In their model, the pathways convey the expected future reward, where the indirect pathway coveys this expectation at the present time step and the direct pathway delivers the prediction of the previous time step. These two signals plus the primary

reward signal are necessary to compute the temporal difference error in TD learning. Houk et al. also go on to explain how this system can be combined with an action-learning module to make an "actor-critic" system, a concept from the machine learning subfield of reinforcement learning (Sutton & Barto, 1998). Figure 9.2 illustrates this concept in terms of basal ganglia anatomy. The critic learns to predict the expected future reward for a given state. The actor uses an analogous learning structure for each action and learns to value certain actions more than others. Importantly, the learning signal for both modules is computed by the critic. So, when the critic gets more reward than expected, it increases the "value" of actions that were selected just prior to the time of the reward. A number of actor-critic basal ganglia models have been described (e.g., Barto, 1995; Suri & Schultz, 1999; Baldassarre, 2002; Houk, 2007), although some of their biological interpretations appear problematic (Joel et al., 2002). The important point is that in the models described so far, the direct and indirect pathways have been modeled in two distinct ways: 1) as facilitating and suppressing actions (respectively) and 2) as the expectation of future reward for two slightly separated moments in time.

9.2.1 Why have Dual Pathways?

In the models of action selection, the direct pathway facilitates and the indirect pathway suppresses. In principle, however, it is not necessary to have two pathways to complete this dual function. A single pathway that has an initial bias of positive activity can be reduced to express suppression or increased to show facilitation. In Frank's (2005) model of the basal ganglia, which we examine in more detail below, the dual pathway structure appears to only increase the rate at which associations are learned. It appears that instead of the indirect pathway having a distinct role, it simply duplicates that of the direct pathway, thereby facilitating faster learning. Therefore, what is being accomplished by two pathways could be accomplished by one pathway, given the initial bias of positive activity and an adequate increase in the synaptic learning rate to make up for the loss of a pathway. Granted, biological organisms employ a great deal of redundancy so that when one system fails, the other can maintain functionality. One could argue that the indirect pathway is redundant despite the clear asymmetry between the two pathways. Both the direct and indirect

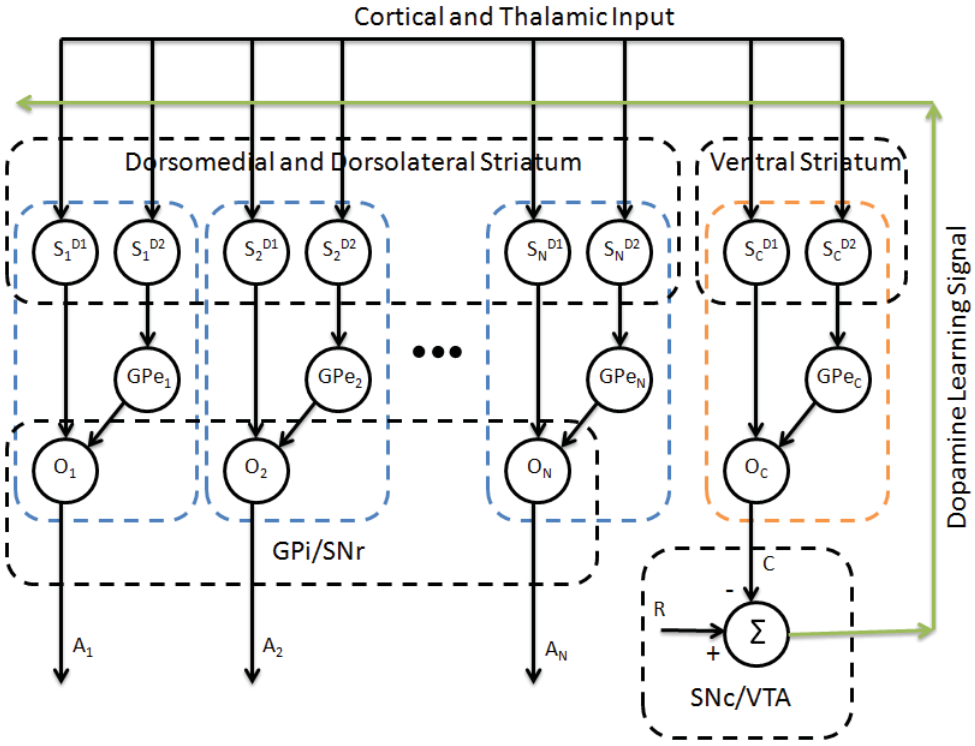

Figure 9.2: The actor-critic approach mapped to basal ganglia anatomy, based on Houk (2007). The critic, which is mapped to the ventral striatum, receives input that reflects the state of the system and learns to predict the value of future reward from the given state. The actor, mapped to the dorsolateral and dorsomedial striatum, receives input that represents the state and activates its output node according to its degree of preference for the associated action in that state. The critic provides the learning signal for both modules, but the individual actors only use it to update their associated action if they were recently employed.

pathways arrive at the same output nuclei and appear to provide cooperative signals. The direct pathway even provides a collateral projection to the GPe. Thus, both the direct and indirect pathway targets are both influenced by the D1 and D2 populations of MSpNs.

However, for such an elaborate system of distinct dopamine receptor types, contrast enhancement, learning rule, and projection path, a more parsimonious explanation seems to be that there is a distinct purpose for the indirect pathway. Houk et al.'s (1995) model suggested that the indirect pathway expresses the expectation of future reward for the present time step, whereas the direct pathway expressed the expected future reward for a previous time step. Since its proposal, we have learned that there is much asymmetry in MSpN learning between the direct and indirect pathways, making the idea that the pathways learn the same thing but for different time steps seem unlikely.

From this work, two novel purposes for having a second pathway are proposed that can cooperate, as will be shown in Section 9.5. Through the Dual Noisy OR model and DPR, it has been shown that a second pathway provides a way to improve generalization in the presence of lots of features and noise. A consequence of this is that the indirect pathway comes to specifically represent inhibitory features whereas the direct pathway represents excitatory features. Here, inhibitory features are not viewed as "negative" valued predictions per se, but as canceling positive predictions when present. This accommodates the notion that the direct pathway might learn the generally rewarded actions (e.g., grasping) and stimuli whereas the indirect pathway might learn the exceptions to the rule (e.g., grasping the air). The SLIM model provides an additional insight that encourages representing excitatory and inhibitory features in separate pathways. We have briefly shown that the dual pathway structure is capable of explaining negative patterning (i.e., the XOR problem) by representing individual excitatory stimuli in the direct pathway and the inhibitory combination of the stimuli in the indirect pathway. Thus, non-linear discriminations can be made using dual pathways. In Section 9.5, we show further evidence of this and provide promising preliminary data that suggest that this non-linear mechanism is very effective when integrated within DPR. I therefore submit that the purpose of the dual pathway structure may be to simultaneously help represent non-linear scenarios and

improve generalization.

9.3 Models with a Focus on Lateral Inhibition in the Striatum

In the case of the actor-critic, the actor must select from among its candidate actions in each state. Therefore, all but one of the actions controlling a particular motor output must be suppressed. Given that parallel channels in the basal ganglia represent different candidate actions, it has been suggested that the lateral inhibition between striatal neurons might implement this function (Barto, 1995; Suri & Schultz, 1999), and is frequently referred to as Winner-Take-All. However, physiological studies have not supported this notion (for a brief review, see Tepper, Koós, & Wilson, 2004). Instead of strong reciprocal connections between neurons, they found that individual lateral connections are usually weak and rarely reciprocated.

The interpretation of the lateral inhibition taken by SLIM is that it sculpts striatal activity resulting in an *ensemble* of active neurons that suppress others irrespective of specific actions. This interpretation has been referred to as "winnerless competition". In particular, Ponzi and Wickens (2010, 2012) have studied this approach, taking it one step further by suggesting that temporally evolving spatial patterns occur under specific circumstances that seem plausible for the striatum. A computational advantage to the winnerless competition paradigm is that certain neurons come to represent the "value" for certain stimuli and this generalizes between similar stimuli to a degree. This is similar to the notion of sparse coarse-coding, where specific learning elements are responsible for representing a "value" when "activated" by input (Sutton, 1996; Sutton & Barto, 1998). For example, let's say that a specific element is activated whenever an x-y coordinate input falls within its associated circular boundary. Other circular elements have boundaries of similar radius but have a different center position. Element boundaries overlap substantially such that a given x-y coordinate input falls within the boundaries of a fraction (e.g., 10%) of the elements' boundaries, activating them. The activated elements then share the responsibility of representing the learned value associated with the given coordinate. Because elements are activated by coordinate position, similar coordinate positions will activate a similar set of elements, thus largely generalizing what was learned for one coordinate to nearby coordinates. This is very similar to what happens in SLIM,

where neurons become the learning elements and the input stimuli replace the x-y coordinate input. The key difference is that SLIM has as many input dimensions as input elements rather than only the 2 dimensions of an x-y coordinate.

Bar-Gad et al. (2003) provide another interpretation. They show how the lateral inhibitory connections may implement a form of feature reduction.

9.3.1 Bar Gad et al.'s (2003) Dimensionality Reduction Model

This model expresses the possibility that the basal ganglia can accomplish dimensionality reduction, very similar to the principle components analysis approach briefly described in Chapter 1, and is referred to as the reinforcement-driven dimensionality reduction (RDDR) model. We follow the more recent and more mathematical description from Bar-Gad et al. (2003), although a slightly different version (Bar-Gad, Havazelet-Heimer, Goldberg, Ruppin, & Bergman, 2000) came first. From 8x8 pixel images of vertical and horizontal line pairs, the RDDR model's output nodes learn to activate exclusively for certain vertical or horizontal lines. The model essentially learns the most important components of the training images.

Figure 9.3 depicts the structure of the model. It is a fully connected, two-layer neural network (three layers in Bar-Gad et al., 2000) with more input nodes than hidden/output nodes. This forces it to learn to represent the input with fewer nodes and thereby perform reduction. Each output node becomes active according to

$$r_i = \sum_{j=1}^{n} w_{ij}^{I} x_j + \sum_{k=1}^{m} w_{ik}^{L} r_k \tag{9.1}$$

where r_i is the activity of an output node, x is a vector of inputs, n is the number of feed-forward inputs, m is the number of output nodes, and w^I and w^L are the input feed-forward and lateral weights, respectively. The lateral inhibitory connections are set to be asymmetric, where $w_{ik}^{L} = 0$ for $i < k$.

Although RDDR uses a reinforcement learning signal, it is unsupervised at the core. That is, it does not need labels or reinforcements to perform the dimensionality reduction. The model therefore also does not "predict" outcomes. Instead, it learns to transform the input data into a reduced representation. The reinforcement's role is to encourage the model to represent certain (rewarded) patterns over others. The

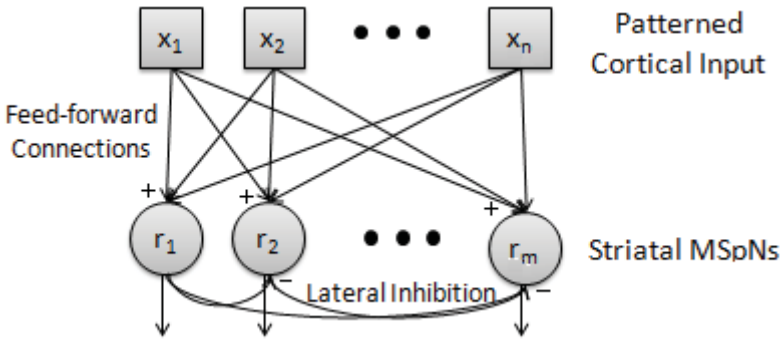

Figure 9.3: Bar-Gad et al.'s model of the basal ganglia. Patterned cortical input excites striatal neurons in a single pathway. Lateral inhibition reduces this activity. Lateral connectivity is asymmetric such that $\Delta w_{ik}^L = 0$ for $i < k$.

feed-forward input learning rule is

$$\Delta w_{ij}^I = \alpha\delta(r_i x_j - r_i^2 w_{ij}) \tag{9.2}$$

where δ is the reinforcement (i.e., the prediction error). In one of their experiments, Bar-Gad et al. (2003) presented 8x8 images each containing either horizontal or vertical lines. When reinforcement was associated with vertical lines, the output neurons specialized for the vertical lines. When reinforcement was subsequently associated with horizontal lines, the output neurons reorganized to specialize for horizontal lines. Lateral weights are updated with a very similar rule,

$$\Delta w_{ik}^L = -\alpha(r_i r_k + r_i^2 w_{ik}^L). \tag{9.3}$$

If two neurons fire together, they increase the inhibition between them. If one fires and the other does not, the firing neuron's incoming lateral inhibition is reduced. The inactive neuron sees no change. Unlike the feed-forward learning rule, this learning rule is *not* proportional to the reinforcement signal.

A key feature of this model is that it explains why the lateral weights appear to be generally weak. Although it is necessary to increase these weights initially to decorrelate the output neurons' activity, the feed-forward weights eventually become responsible for encoding the reduced input representations. Once this happens, the uncorrelated outputs will subsequently reduce the lateral inhibition between them according to Equation 9.3. So, in this model, lateral weights are strong only temporarily.

A Qualitative Comparison with SLIM

Generally speaking, the RDDR model is very different from SLIM in terms of its structure, but its form of lateral inhibition and lateral learning is very similar to that found in SLIM. Perhaps, the most profound difference is that the RDDR model is mainly unsupervised whereas SLIM is a supervised model.

The RDDR model's structure uses a fully connected two-layered network. In the hidden layer, however, the lateral inhibitory connectivity is only partial (systematically asymmetric, $\Delta w_{ik}^{L} = 0$ for $i < k$). SLIM, on the other hand, employs partial connectivity for both the forward and lateral inhibitory weights. The RDDR model only has one "pathway", whereas SLIM is broken down into two pathways, which we have seen may be useful for representing inhibitory and configural stimuli. In the case of RDDR, multiple pathways were not necessary to demonstrate the dimensionality reduction functionality. In SLIM, the synaptic weights are not permitted to change sign. However, in RDDR, the synaptic weights are able to change signs. The authors note that this is not very biologically plausible since occurrences of neurotransmitters having both positive and negative effects are rare and not likely to exist in the striatum.

The form of lateral inhibition here is very similar to SLIM, having specific neuron-to-neuron inhibitory connections. Another similarity between the two models' is that the feed-forward and lateral weights are generally updated in the same direction (sign). In SLIM, this is responsible for the retrospective revaluation results, since lateral learning in the opposite direction gives the opposite effect (i.e., mediated conditioning).

There are a few significant differences in the lateral learning rules. In SLIM, both the sending and receiving neurons must be active for any learning to take place, whereas learning may occur in RDDR when only the receiving neuron is active. This feature of RDDR appears to be responsible for the weakening of lateral inhibition following decorrelation of the output nodes. RDDR's lateral learning rule is not formally influenced by reinforcement, whereas it is in SLIM. This is important in SLIM to account for the retrospective revaluation, since this is what keeps the direction of the lateral learning the same as the direction of feed-forward learning.

Is the notion of dimensionality reduction and the decorrelation of output nodes

incompatible with SLIM? Not entirely. SLIM's lateral inhibition instead encourages multiple instances of stimulus-outcome learning to be spread out among the various neurons of the network. To be more concrete, as lateral weights between the neurons of an ensemble are increased during conditioning, these neurons will naturally become more difficult to activate by novel stimuli. So, the novel stimuli will gravitate toward being represented by a largely separate ensemble. Ultimately, this behaviour encourages a more thorough use of all of the neurons in the network. Although not evaluated here, this would presumably reduce destructive interference or degradation in its predictions as more and more stimulus-outcome relationships are learned.

9.4 Frank's (2005) Model

Frank (2005) proposes a model of the basal ganglia that shares the most in common with the present model. Here, we describe this model and provide additional details in Appendix B. Frank's rate-coding model, shown in Figure 9.4, is composed of cortical, striatal, pallidal, and thalamic regions representing the complete cortico-basal-ganglia-thalamo-cortical loop. It is a model of action selection, where each action has a separate, topologically organized, loop throughout the system. The cortical (patterned) input to the striatum and premotor cortex (PMC) are the only fully connected (not topologically organized) sets of connections in the system. Frank uses this system to simulate cognitive phenomena and, in particular, the difference between a healthy subject and one with Parkinson's disease (PD) by attenuating the dopaminergic effects on striatal neurons in the PD case.

In the Weather Prediction task (Knowlton & Squire, 1994), 1, 2 or 3 of 4 possible symbols are presented to the subject. The subject predicts whether Sun or Rain will follow and the outcome is then revealed. In the model, a binary pattern representing the presence of the predictive symbols is given as input. The striatum and other regions of the model begin to activate and ultimately arrive at a decision, marked by the active PMC node. This initial activation of the system is referred to as the "minus" phase by Frank (2005). This is followed by the "plus" phase, in which correct actions earn a phasic increase in dopamine while incorrect actions earn a dip in dopamine. An increase or decrease of dopamine increases or decreases activity in direct pathway neurons respectively and has the opposite effect on the indirect

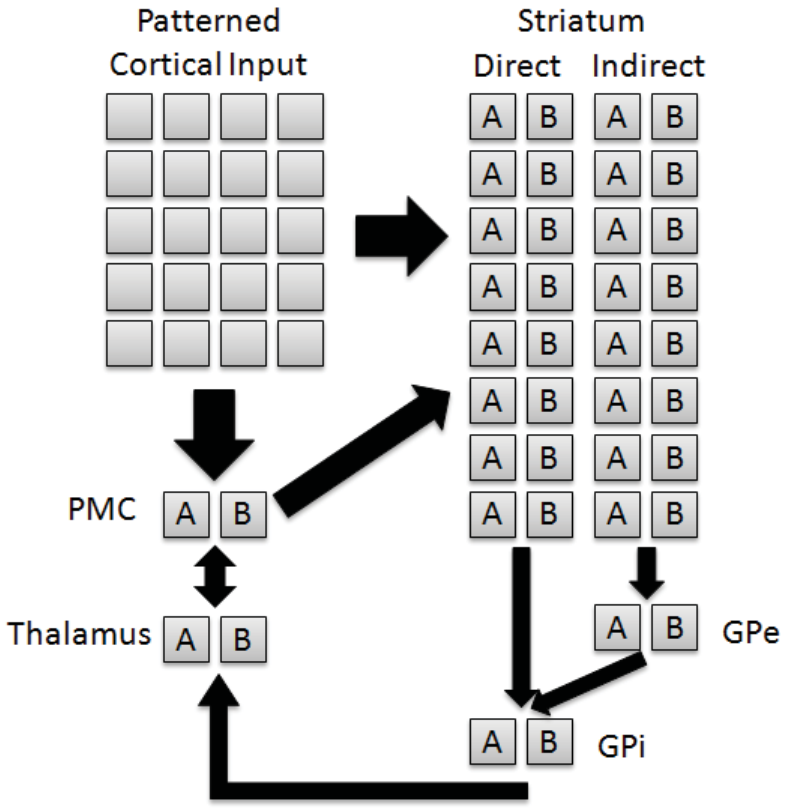

Figure 9.4: Frank's (2005) model of the basal ganglia with cortical and thalamic nuclei. Patterned cortical input excites striatal neurons in the direct and indirect pathways. Channels associated with actions A and B stay segregated throughout the downstream connections. Terms: PMC - premotor cortex, GPe - globus pallidus externa, GPi - globus pallidus interna.

pathway neurons.

There are a number of dynamics that contribute to the operation of the model and its learning. The striatal neurons are subdivided in two ways. First, each neuron is topographically related to one of the possible actions (Sun or Rain). Second, each neuron belongs either to the so called Go pathway (the direct pathway) or the NoGo pathway (indirect pathway). Following the anatomy, Go pathway neurons project directly to the GPi while NoGo pathway neurons project first to the GPe, which then projects to the GPi. Thus, when an input is provided, the difference in Go to NoGo pathway neuron activity is born out in the GPi, leading to the basal ganglia's action recommendation. This recommendation modulates activity in the thalamus, which is reciprocally connected to the PMC. The thalamus essentially becomes a shutoff valve on the activity of the non-preferred action, encouraging the preferred action. A key feature of the model is the PMC's topographic input connection to the striatum. As the basal ganglia begins to settle on a decision, the PMC encourages activity in stratial neurons that correspond to that decision. According to the contrastive Hebbian learning rule used to update striatal weights (see Appendix B.3), a neuron's weights are changed in proportion to its activity. So, the PMC's striatal projection appropriately encourages or focuses learning on the neurons responsible for making the decision. Without the PMC's striatal influence, it would seem that some sort of manual procedure would be needed to force learning in the neurons associated with the decision made (as proposed in an actor-critic model described by Barto (1995)).

Learning occurs in both the striatum and the PMC. In the striatum, the change in a neuron's weights in a certain trial is proportional to the difference in its minus and plus phase activity and the activity of the input unit from which it receives a connection. A striatal neuron's activity is also modulated by dopamine. Each neuron, regardless of pathway, is enhanced to the same degree by the baseline dopamine level in the minus phase. If the outcome matched the model's prediction, dopamine levels are increased. This enhances the Go neurons' activity and reduces the NoGo Pathway neurons' activity. When the outcome does not match the prediction, dopamine levels are decreased. This has the opposite effect on Go and NoGo neuron activity. The difference in activity between this latter (plus) phase and the earlier (minus) phase and the way that dopamine oppositely modulates the Go and NoGo pathways means that

these two pathways learn in opposite directions. In the PMC, learning is Hebbian in nature, increasing the weights connecting active inputs and active neurons. An important factor in this model is that the PMC learning *must* be made slow. Since the PMC significantly influences activity in the striatum, causing a positive feedback loop, a fast learning PMC may get an early reward or punishment that is statistically uncharacteristic and then get stuck in this belief, ignoring the basal ganglia's recommendations even as experience grows. With a proper, slow learning rate, the basal ganglia is eventually not needed to make a decision because the correct response has been encoded in the input-PMC weights, an encoding process that could not have been achieved with a Hebbian learning system alone.

9.4.1 A Qualitative Comparison with SLIM

The basal ganglia portion of Frank's model shares much in common with the present model, but there are a variety of differences as well. Both models involve multiple neurons in the striatum, although SLIM uses far more for statistical stability reasons[1]. Both models use a non-linear activation function, which translates internal values to neural activities. Both have dual pathways, where one pathway has a positive influence on the output whereas the other has a negative influence. SLIM takes advantage of this feature to be able to develop configurations so as to accomplish negative patterning (i.e., the XOR problem). Frank's model does not do this.

Frank's model is one of action selection (i.e., instrumental learning) rather than merely learning the association between inputs and an outcome (i.e., classical conditioning). As we mentioned earlier, the striatum appears to be involved in both forms of learning, where it may use a similar approach to encode the values of actions as well as states.

Both models have a form of lateral inhibition. SLIM has specific neuron-neuron lateral inhibitory connections whereas Frank's model uses a general global (striatum-wide) inhibition. The purpose of the lateral inhibition is different in each model. In Frank's model, lateral inhibition is simply used to sparsify the neural activity. In SLIM, the lateral inhibition sparsifies the neural activity but it also sets the stage

[1]With few neurons, SLIM can still behave correctly, but is more sensitive to the initial parameters used. However, having many neurons reduces this sensitivity and allows SLIM to consistently demonstrate conditioning findings as shown earlier.

for lateral learning which leads to the retrospective revaluation phenomena. In both Frank's model and SLIM, the striatum is the only basal ganglia region that employs lateral inhibition.

The learning rules for both models encourage neurons of the different pathways to learn in opposite directions – when direct pathway synapses are strengthened, indirect pathway synapses are weakened and vice versa. The exact way in which the models implement this is different. Frank's approach more accurately models the action of dopamine on the activities of the D1 and D2 neurons and uses a common learning rule based on the changes in neural activities between the minus and plus phases. SLIM abstracts these details by simply using opposite signs in the learning rule for the two pathways. In both models, having this opposite learning feature allows the two pathways to learn cooperatively, that is, the effects of weight updates in the two pathways add or work together rather than cancel out one another. In one way or another, both models' learning rules are proportional to the input, output, and level of dopamine.

9.5 Relating SLIM and DPR

In the preceding sections, we described the relationship between two related basal ganglia models and SLIM. We could have also drawn comparisons between these models, especially the Frank (2005) model, and DPR. However, SLIM and DPR are similar enough that to do so would be redundant. For example, SLIM and DPR use the same dual-pathway structure and have similar cortico-striatal plasticity rules in that the direction of learning is opposite for the direct pathway than for the indirect pathway and that learning is proportional to prediction error and the salience of the input. They differ in that DPR only uses 2 neurons (one per pathway) whereas SLIM uses many neurons per pathway. DPR also multiplies the contributions of the direct and indirect pathway whereas SLIM sums the contributions.

Because of the similarities, it seems possible to combine mechanisms from both models to increase the total capability of a single model. As a proof of concept, I extended DPR to use multiple neurons per pathway and a quadratic activation function. The learning rule also made each neuron's learning proportional to the square root of its activity, as in SLIM. The result not only enabled DPR to perform a non-linear

discrimination, but to do so in an extremely efficient manner. Figure 9.5 compares the results of an XOR task (negative patterning) between this non-linear DPR model and a Support Vector Machine (SVM). Here we are not performing regression but rather classification with 0 and 1 labels. In the task, there are two relevant features such that when either is present, but not both or neither, reinforcement is delivered. In the generated data, there is a 50% probability of activation for every feature, whether relevant or irrelevant. There are a variable number of irrelevant features (indicated in the figure legend). In this task, it is impossible to tell which of the features are relevant based only on the frequency of association with the reinforcement, making it a worst case scenario. We see that with few irrelevant features, the non-linear DPR and SVM can accurately classify test examples with few training examples. However, as the number of irrelevant features increases, DPR relatively quickly learns the relationship whereas SVM lags behind, requiring an order of magnitude more training examples to achieve comparable classification rates when there are only 50 irrelevant features.

Fully combining SLIM and DPR is a task saved for future work. Unfortunately, SLIM requires substantial computational time to run a single trial and has far fewer stimulus inputs than we have typically been using in tests of DPR.

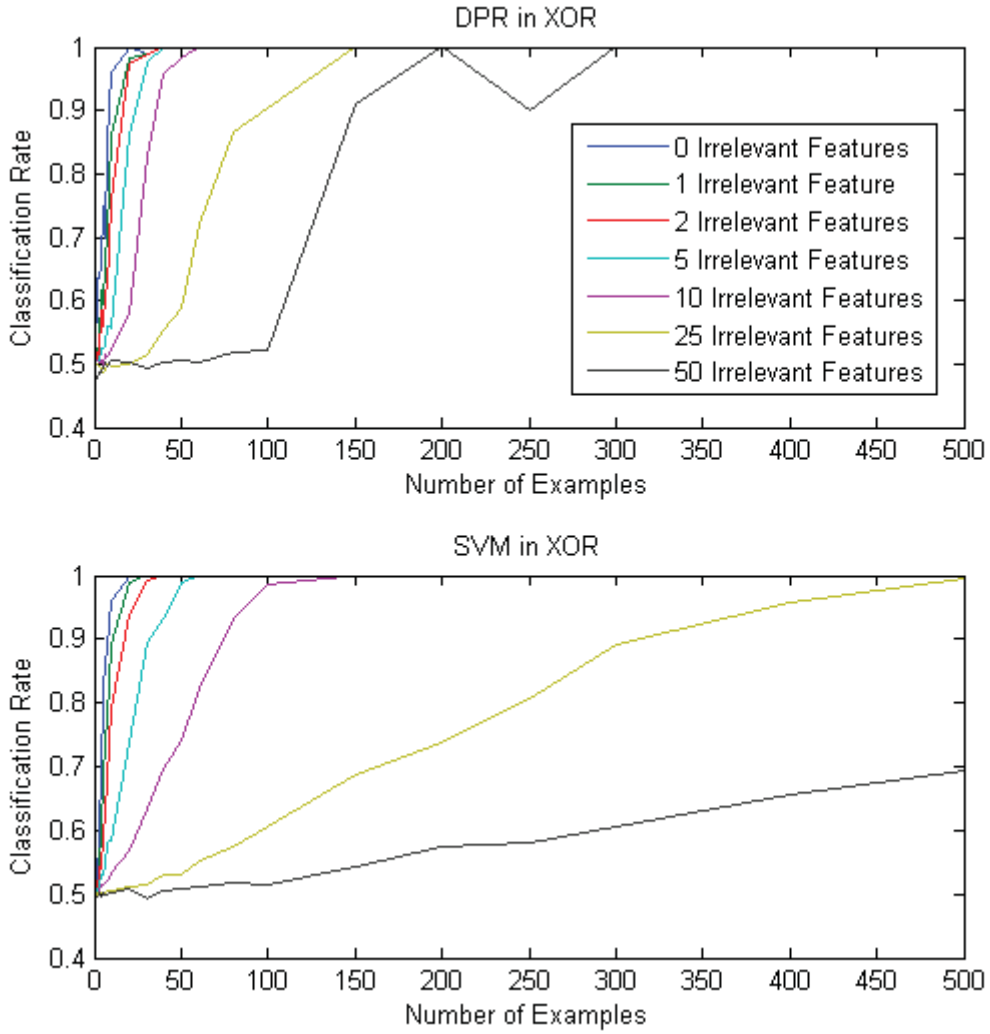

Figure 9.5: An XOR classification task comparing an SVM and DPR enriched with two mechanisms from SLIM. In the task, there are two relevant features such that when either is present, but not both or neither, reinforcement is delivered. Relevant and irrelevant features are present with 50% probability and, by nature of the task, reinforcement is given as frequently in the presence of an individual relevant feature as in the presence of an irrelevant feature. In this worst case type of scenario, we see that the SVM performs comparably to DPR so long as the number of irrelevant features is low. When there are a large number of irrelevant features, the SVM requires far more training examples to provide comparable classification accuracy.

Chapter 10

General Discussion

10.1 Thesis Summary and Integration

The aim of the present work was to understand and evaluate potential biologically plausible ways of solving the spatial credit assignment problem, especially those with a dual pathway structure. In so doing, it might reveal novel approaches that are beneficial for machine learning and may shed light on reasons for specific animal behaviours.

Toward this end, I characterized the spatial credit assignment problem and defined a probability density function from which the x and y training and test set data were generated for the main regression task. An important detail was that inhibitory stimuli were defined as being distinct from stimuli that predict a "negative" or oppositely valenced outcome. This is different than the Rescorla-Wagner and LMS approaches, which put reward and punishment on the same "number line", but with opposite signs. In that line of thinking, negative values can both cancel a positive prediction and suggest a negatively valenced outcome. This seems like a troublesome perspective. A direct prediction from this would be that an inhibitor which signals no food reward (due to conditioned inhibition) should lead to an expectation of punishment when presented alone because the prediction value should be very negative in this case. Despite such an unnatural prediction, using contemporary machine learning methods like MLP or SVM for predicting future reinforcement will do this. Additional evidence that the perspective taken in this thesis on inhibitory features is correct is that animals tend not to extinguish inhibitory features by non-reinforced presentation. The LMS and Rescorla-Wagner models, however, do extinguish inhibitory stimuli. The rLMS model, the optimal model for the task, and DPR do not extinguish inhibitory stimuli in this way, treating inhibitory stimuli as a means of canceling positive predictions only.

A next step was to identify potential strategies taken by biological systems. This

was done by evaluating a number of existing theoretical approaches. An obvious choice was simple linear regression. In so doing, I have discovered that a number of classical conditioning phenomena in addition to those explained by the Rescorla-Wagner model can be explained by merely repetitively reprocessing past experiences, which is equivalent to linear regression. An interesting feature of this discovery is that phenomena as different as retrospective revaluation and CS preexposure can be additionally explained with this simple approach. Rationale and logic are sometimes spoken of when verbal explanations of various conditioning phenomena are given. Essentially, rationale wants to best account for all of our experiences, or in other words, find the explanation(s) that maximizes the likelihood of our observed experiences. This is formalized in linear regression in computing the maximum likelihood estimate for our PDF model. Although the batch-learning associated with regression may initially appear biologically implausible, the classical conditioning findings herein suggest it is *behaviorally plausible*, and the ability of SLIM to explain retrospective revaluation phenomena supports the idea that it is also biologically possible.

Generally speaking, the field of machine learning has helped us formally recognize that learning runs into prediction problems when there is little data, many features, and noise. Formal simulations presented here have confirmed that this is the case even for the specific conditions represented in data sets fashioned after features of the real-world. The model that was used to generate the data became the optimal model with which to discover the underlying parameters, given enough data. Yet, imposing irrelevant features (i.e., underlying parameter values of zero) does not match the implicit assumption made by our "optimal model" that parameters are drawn from a uniform distribution. The so called "optimal" model is therefore not optimal. Nevertheless, this parallels the fact that the uniform distribution assumption is commonly made in machine learning problems. As a work around for making this incorrect assumption, feature selection is used to improve performance. This essentially corrects for the fact that many parameters have no influence on the output (i.e., have a parameter value of zero). The most extreme form of this is represented by the Bayesian optimal model results, where the rLMS model was told exactly which 2 features were relevant and regressed over these alone. Also, several different sub-optimal feature selection methods were evaluated. A common theme is that having many irrelevant

features, small data, and noise makes finding the relevant features more difficult for rLMS and related methods. An important contributor to high prediction errors are the residual values in parameters associated with irrelevant features. Regularization methods shrink these residual values, transferring the value to relevant feature parameters. The LASSO, a specific regularization technique was found to be the most effective of these overall in our simulations.

Classical conditioning models also exhibit problems eliminating residual values. In the relative validity procedure, an irrelevant stimulus, X, retained substantial associative strength after many AX+, BX- trials in simulations of the Rescorla-Wagner model. In contrast, no residual in X remained with an AX+, X- treatment (SFPD). In SFPD, relative correlation takes place, where the most highly correlated stimulus gains the associative strength, leaving little for the lesser correlated stimulus. Thus, the addition of an irrelevant stimulus, B, in relative validity leads to residual errors and a break-down in this relative correlation process. The problem seems to be in permitting B's strength to be negative because it then is able to support the residual association in X. The Dual Noisy OR model, an extension of the Noisy OR model, offered a solution. As reiterated in the formulation of DPR, instead of allowing stimulus B to take on a negative value and sum with X, B's inhibition is represented as a number between 0 and 1 (smaller for greater inhibition) and is multiplied by the positive value of X to make predictions. Importantly, the residual associative strength in relative validity could no longer be supported without summation and a negative associative strength for B and thus A and X reached full strength and zero strength, respectively. This represents a restoration of the relative correlation principle and a better fit for relative validity experimental data. The LASSO method was also able restore relative correlation as well. The LASSO diminishes all parameters, relevant or irrelevant. Truly relevant features are reinforced consistently enough to overwhelm the LASSO's effect on their associated parameters. The irrelevant features' parameter strengths, however, are reduced and ultimately transferred to the relevant features.

Whereas machine learning identified the key challenges to spatial credit assignment and possible high-level solutions, neuroscience helps to narrow the range of possible algorithms by demanding biological plausibility. Many of the algorithms investigated here have biologically plausible implementations. Specifically, we saw that

the regularization methods could be seen as neuron synaptic strengths being reduced over time and data augmentation as there being noise in the system during an update. Such general purpose qualities could be at work anywhere in the brain. In contrast, the proposed SLIM and DPR algorithms were centered around the basal ganglia and are admittedly less parsimonious but offer potential purposes for this complicated structure. Many machine learning algorithms do not have readily plausible neural implementations. For example, the process of feature selection by employing a wrapper method begins with a certain subset of features and then adds or removes a feature based on whether it improves the prediction error or not. Such a serial and highly combinatorial process would be too slow for a biological system to use effectively. Similarly, properly employing the Bayesian model selection would require evaluating the plethora of feature combinations to arrive at the best hypothesis, and is therefore considered implausible. As for general purpose learning machines, the MLP has an air of biological plausibility but has been criticized for having to transfer knowledge of synaptic weights in the output layer to the hidden layer (referred to as "weight transport"), and it is not clear how this may be done biologically. In contrast, SLIM and the enriched DPR model of Section 9.5 avoided the need for a second layer (and thus weight transport) but were still able to learn simple non-linear discriminations.

The basal ganglia is an elaborate system of several pathways, synaptic plasticity factors, and intranuclear connectivity, and not all of these features have strong, purposeful explanations. In this work, the dual pathways are stripped of the purpose that says the indirect pathway duplicates the functionality of the direct pathway. Instead, new purposes are proposed. The dual pathway structure represents excitatory and inhibitory features separately, which allows for: 1) improved generalization through a multiplicative integration of inhibition and 2) configuration through a non-linear activation function. DPR also suggests possible explanations for the direct pathway axon collaterals to the GPe and the reciprocating projection from the thalamus to the striatum. SLIM has shown how lateral inhibition in the striatum could be responsible for online expression of retrospective revaluation phenomena, allowing one aspect of regression to operate in an online setting.

In summary, mainly two novel biologically constrained algorithms have been proposed. DPR uses multiplicative inhibition to suppress irrelevant features, which

ultimately leads to fewer training data points being needed to learn the same discrimination (especially clear in Figure 9.5). SLIM offers a way of learning non-linear configurations and performing an aspect of regression online. Together, the models support an online means of efficiently learning which features are predictive of an outcome or, in other words, spatial credit assignment. As described in more detail in Section 10.3, machine learning stands to benefit in terms of improved linear regression, in online settings especially. This work also speaks to classical conditioning theory. From a large number of classical conditioning phenomena it appears that animals are performing a regression over past experience. Also, SLIM offers a novel elemental approach to performing retrospective revaluation phenomena, and DPR also explains conditioning phenomena beyond LMS. Finally, rLMS' and DPR's ability to properly avoid extinction of conditioned inhibition supports the stance that prediction-canceling inhibitory stimuli are qualitatively different from excitatory stimuli of opposite valence.

10.2 A 3-Stage Model of Reinforcement Learning

In the introduction, the spatial credit assignment problem was defined as being distinct from the unsupervised learning systems responsible for deriving high-level features from raw sensory data. So, we started by breaking the total reinforcement learning problem into two parts. As Figure 10.1 illustrates, the total reinforcement learning problem could also be divided into three parts: an unsupervised part, a supervised part, and a reinforcement learning part. As described in the introduction, the "unsupervised learning" module transforms raw sensory input into a vector of mid-to-high level features of the real world, where the salience of each feature is expressed as a scalar value between 0 (absent) and 1 (strongly present). This output is then provided to the "supervised learning" module, which performs spatial credit assignment. This module predicts the expected future salience of a range of stimuli (especially USs), although in the earlier simulations we only ever predict a single output for simplicity. This output is passed to the third, "reinforcement learning" module. This final module simply maps each of its inputs to an independent reinforcement value and sums the effects of the expected reinforcements. This is where

predictions of reward and punishment would mix to account for certain animal behavior. There are a few primary reasons why it may be helpful to make the 3-module distinction.

Firstly, *inhibitory stimuli cancel expectations but do not predict reinforcement of the opposite valence.* In reinforcement learning, the subfield of machine learning, reinforcements can be either positive or negative. Naturally, rewards have positive values and punishments have negative values. In reinforcement learning, inhibitors of reward would acquire a negative reinforcement value so that when coupled with the reward predictor they inhibit, there is less (or zero) expectation of reward. The problem comes when the inhibitor is presented alone. Having a negative reinforcement value should mean that the agent is being "punished" in some way, suggesting that the agent should avoid such a state. To draw an analogy in animal learning terms, this would mean that the presence of an inhibitor of food reward, which normally indicates disappointment, when presented alone should make the animal fearful of punishment. In both cases, the inhibitor is only meant to affect predictions when in the presence of the stimulus/stimuli it was conditioned to inhibit. By separating the supervised learning and reinforcement learning parts, the inhibitor is only used in the supervised learning part to help predict whether or not the US will appear. This *non-negative* prediction of the US is then funneled to the reinforcement learning module, which assesses and integrates the reinforcement values of all predicted USs or stimuli.

Secondly, *prediction of future reward is more flexible if a US' reinforcement changes.* After CS-US conditioning (e.g., tone $\rightarrow$ sucrose), it is possible that the US may be devalued (e.g., sucrose paired with sickness). However, if conditioning only encodes the CS' reinforcement value, then it will not notice the US devaluation and still predict substantial future reinforcement. In the 3-module distinction, as CS-US associations are made during conditioning, the supervised learning part would learn to predict that the US will follow the CS. During the US devaluation, the reinforcement learning module would update its direct mapping from US to reinforcement value. Subsequent presentations of the CS would lead to predictions of the US whose updated reinforcement value would be reported by the reinforcement learning module.

Thirdly, *this organization provides for a basal ganglia-based attentional mechanism that is useful in looking ahead* or, in other words, contributes to a model-based reinforcement learning architecture. The basal ganglia modulates topologically organized cortical activity through its outputs' influence on thalamo-cortical connections. If these basal ganglia outputs represent the prediction of important CSs, the cortical activity representing such CSs would be enhanced relative to other CSs, suppressing the distracting unimportant stimuli. Model-based reinforcement learning involves taking the present state/time and looking ahead through expected state transitions to potential outcomes and thereby computing the expected reward. If the basal ganglia took on the commonly believed reinforcement learning role by encoding "reward value" only, the basal ganglia would presumably enhance only stimuli that lead to reward and help direct a model-based lookahead process toward potential rewarding outcomes. However, if the basal ganglia instead takes a supervised learning role, it should influence model-based lookahead toward *punishing* as well as rewarding outcomes since it encodes importance or motivational *salience* rather than motivational value. It would seem that being able to additionally predict aversive outcomes would be necessary to an organism in the lookahead process. However, having a prediction error that treats rewarding and punishing predictors alike does not suit the development of action preferences believed to reside in the putamen and caudate areas of the striatum. Such a prediction error would encourage actions that lead not only to rewarding outcomes but also to *aversive* outcomes, which is not in the organism's best interests. There is some evidence suggesting that the prediction-error dopamine neurons can be divided into two groups: neurons that respond to unpredicted rewarding and punishing stimuli with dopamine burst and dips, respectively (i.e., motivational *value*) and neurons that respond to unpredicted rewarding and punishing stimuli with bursts alike (i.e., motivational *salience*). There is also evidence that these two groups of neurons preferentially target different brain areas (see Bromberg-Martin, Matsumoto, & Hikosaka, 2010 for a review). Motivational salience-reporting neurons would be relevant for assisting high-level cognitive areas that require knowledge of aversive as well as rewarding stimuli (e.g., lookahead), whereas motivational value-reporting neurons would be especially important for encouraging actions that lead to reward or away from punishment.

10.3 Tips for Building a Better Value Function Approximator

A simple way to do value-function approximation is to use a supervised learning algorithm with reinforcement learning data, which provides a state/time representation as input and a reward value as output. Not all supervised learning algorithms are conducive to representing value, however (Sutton & Barto, 1998). It is important that the methods used can learn effectively when values are being learned via bootstrapping. In bootstrapping, values are learned in relation to neighbouring values. In temporal difference learning, for example, the reward value is slowly passed backward through a sequence of states. So, the value for a given state is not usually provided as an output immediately, but is only discovered after a number of episodes. For this, an online-learning approach seems more appropriate or perhaps a batch-learning approach that puts more weight on recent data. Another caveat is that not all approaches will provide convergent predictions (with repeated episodes). So, one must take care to ensure that the method in use converges.

From the present work, several recommendations can be made to improve the spatial credit assignment nature of value-function approximation approaches.

- Consider dividing the reinforcement learning problem into 3 modules as noted above. The first stage of unsupervised learning discovers the regularities in the world and represents the salience of common features. This largely takes care of managing the non-linearities inherent in raw sensory data. Unfortunately, existing unsupervised learning methods are still not as good as the brain at extracting features that are invariant to common transformations. In the meantime, manually constructed high-level features may suffice. Distinguishing between the supervised and reinforcement learning stages is valuable. The key benefit here is that features which predict no reward (inhibitors) are distinguished from features that predict punishment. In this way, when the feature that predicts no reward is presented on its own, it is not equated with the expectation of punishment or negative reward. The interaction of punishment and reward predicting stimuli instead take place in the reinforcement learning module, where the reinforcement value of each expected US, whether positive or negative, is summed.

- Batch-learning gets you most of the way there but can be problematic. It has been shown how simply repeatedly retraining on previous data can increase the number of classical conditioning phenomena that LMS can explain. In this way, batch-learning both improves prediction errors while better representing the processes at work in the brain. However, when the predictive ability of features change or the reinforcement learning strategy is to use bootstrapping, an online-learning approach may fare better.

- Use a dual pathway strategy, which generalizes better than LMS when both have a simple uniform prior. Also, employ an appropriate prior when feasible. Even if all one can safely do is to assume that many features will be irrelevant, the zero-peak prior will at least replace the implicit uniform distribution with one that acknowledges that many features will have a parameter value of zero. Employing the right prior appears to reduce overfitting (data not shown) and may thus eliminate the need for a validation set.

- Use a dual pathway strategy with multiple elements per pathway and a non-linear activation function to model non-linear combinations of features (e.g., XOR, AND, etc.). This structure invokes non-linearity without the usual problem of avoiding local minima since the local minimum is also the global one (i.e., it is a convex optimization problem).

- Add lateral inhibition among the multiple neurons to provide an efficient multi-dimensional representation of the present stimuli. In coarse coding (Sutton, 1996; Sutton & Barto, 1998), linear and non-linear conjunctions of features are represented by an ensemble of elements. As the number of input dimensions increases, the number of elements necessary for a similar degree of coverage grows exponentially. Lateral inhibition encourages configurations of active ensembles to represent specific (multi-dimensional) combinations of inputs and this generalizes well to subsets of such combinations (i.e., generalizes well even as the number of dimensions/active inputs changes). Lateral inhibition also has the expected benefit of increasing the capacity of the system to store information since it forces fewer neurons to represent the same values as it would have without lateral inhibition. Adding even a small amount of lateral learning helps to

spread out these stimulus-outcome associations.

- If online-learning is important, add mechanisms like lateral learning that emulate aspects of what occurs in batch-learning. Classical conditioning models, besides SLIM, offer candidate mechanisms for this purpose.

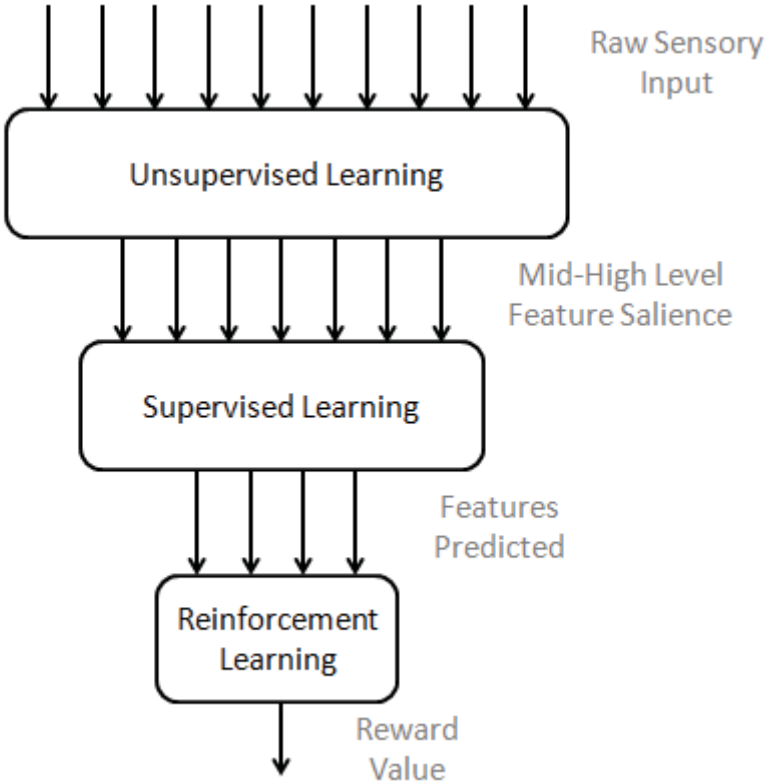

Figure 10.1: A 3-stage reinforcement learning model that computes a reward value prediction from raw sensory input. An unsupervised learning module transforms raw sensory data into the saliences of mid-high level features. A supervised learning module then uses this input to predict the future saliences of other stimuli, especially those with motivational value. Finally, a reinforcement learning module maps the prediction of future stimuli to reward values, where positive values represent rewards and negative values represent costs or punishments. These individual values are summed to give a final reward value or prediction.

Chapter 11

Future Work

The present work has highlighted the value of a dual pathway approach to spatial credit assignment and more specifically as the primary solution implemented by biological systems. There are a great number of possibilities, but with additional time, I feel it would be most worthwhile to pursue research in the following directions:

- Investigating further the configural properties of SLIM in light of the classical conditioning literature. Although preliminary work has suggested that the SLIM's configural mechanism is very robust, it has not yet been evaluated with respect to a number of related classical conditioning phenomena. Also, because of its robustness (and regardless of its relationship to biological systems), it seems worthwhile to further evaluate this mechanism as an extension of DPR, since we have seen that this can have a powerful effect in learning certain non-linear relationships and because it makes DPR a slightly more general purpose model.

- Formally evaluating the latent cause theory of conditioning from a rigourous spatial credit assignment problem perspective as done for the other approaches described here. It appears that such methods would inherently require a lot of data, but a formal analysis would confirm this.

- Evaluating the dual pathway approach on datasets generated with different assumptions (priors) to discern whether or not DPR is a good general purpose generalization approach. I would also like to combine DPR with the zero-peak prior to see if the combination of their residual-reducing strategies adds to further improve performance on the main regression task.

- Developing further or scaling up online approaches to regression. Although with SLIM it was suggested how retrospective revaluation phenomena might be explained in a biologically implementable online way, the same was not shown

for CS Preexposure or other revaluation phenomena. Earlier versions of SLIM included a weight-proportional learning mechanism that could accomplish this but significantly complicated the model and was set aside. This approach is very nearly the same as Pearce and Hall's model (1980) and is worth investigating further. There are also other models of classical conditioning phenomena that employ certain mechanisms to likewise generate retrospective revaluation and/or other revaluation phenomena without relying on batch processing (e.g., Kasprow et al., 1987; Jamieson et al., 2012; Kutlu & Schmajuk, 2012). The difficulty in making use of these approaches is in scaling them to be used in the same versatile way as LMS. Nevertheless, these approaches constitute a potentially fertile field of mostly untapped learning mechanisms for machine learning.

- Relating the dual pathway mechanisms of basal ganglia function described here to explain the cause or symptoms of various disorders localized in the basal ganglia, namely Parkinsonism, Huntington's disease, Schizophrenia, Attention Deficit Hyperactivity Disorder, etc. Frank's (2005) model already does this to a certain degree for Parkinsonism, but the multiplicative inhibitory aspect of DPR, the lateral inhibition of SLIM, etc. may also play a role in this and other ailments.

Chapter 12

Conclusions

In this thesis, the main regression task was designed to represent aspects of the real world. The optimal unbiased model (i.e., with uniform prior distribution) called rectified least mean squares (rLMS) ruminates over all of its previously seen data (i.e., batch-learning) until it converges to a low average (mean squared) error. This work has shown that revisiting previous trials in this way adds many classical conditioning phenomena to the list that can be explained by the Rescorla-Wagner model, which is otherwise equivalent. This supports the notion that biological systems are, in some way, regressing over past experiences to make the most sense of them (i.e., to maximize the likelihood of the observed events).

Simulating the main regression task has demonstrated that the spatial credit assignment problem is difficult with small data, many irrelevant features, and additive noise. Although optimal under certain assumptions, rLMS did not perform the best of all models for the data at hand. However, by introducing more accurate assumptions (e.g., through regularization/Bayesian priors, feature selection, etc.) performance was improved. It is expected that biological systems are employing similar means. Specifically, we saw that the conditioning phenomenon called relative validity could be explained when one of these additional assumptions were integrated with rLMS or through the use of the proposed Dual Noisy OR and Dual Pathway Regression (DPR) models.

Of all of the various machine learning approaches for reducing prediction error in the regression task, the LASSO regularization for rLMS appears to be the most generally useful method, assuming that rLMS can be employed effectively without having to accurately specify the variance of the additive noise. The regularization and dual pathway methods can all be implemented in biological terms. The LASSO and zero-peak priors could be realized as a rule for synaptic weight decay over time. DPR, a dual pathway model, can be realized in terms of basal ganglia structure and

the cortico-striatal synaptic plasticity details uncovered by neuroscientific studies.

A number of earlier basal ganglia models provide various interpretations for the function of the basal ganglia. This thesis offers a different view. In the likely event that most spatial credit assignment learning occurs in an online fashion, the Striatal Lateral Inhibition Model (SLIM) offers an online way of explaining and biologically implementing some of the aspects otherwise explained by repeatedly reviewing the data. The basal ganglia's dual pathway structure also provides a seat for integrating multiplicative inhibition and non-linear discriminations. Finally, the notion of the basal ganglia (at least the stream involving the ventral striatum) serving as a supervised learning module sandwiched between unsupervised and reinforcement learning modules appears to be beneficial.

So, how do biological systems solve the spatial credit assignment problem? This thesis supports the strategy that we implicitly maximize the likelihood of our experiences, as in regression, and that we employ certain assumptions about the real-world that enable us to learn with as little experience as possible. It appears that such a strategy and assumptions can be implemented in terms of the dual pathway nature of basal ganglia anatomy and function, even in the likely event that the learning process is largely online.

Appendix A

Least Mean Squares Regression Simulations of Classical Conditioning Phenomena

Here, it is demonstrated that LMS is capable of correctly simulating retrospective revaluation phenomena (Table 3.1) and other revaluation phenomena (Table 3.2). Each simulation is captured in its own figure with simulation specific details and a paper citation referencing the associated animal experiment. In each simulation, stimuli have a salience of 1, the context has a salience of 0.2, and all stimuli/contexts begin with zero associative strength. Each phase runs for 50 blocks of conditioning trials and the results show the associative strength at the end of the last conditioning phase.

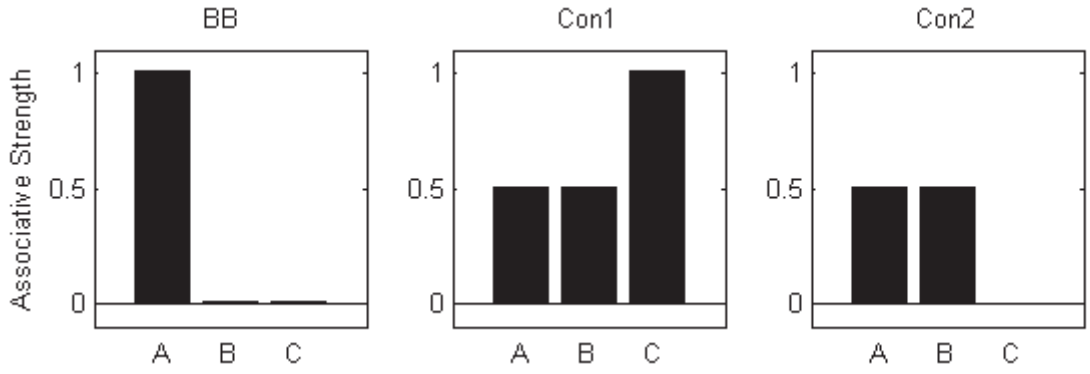

Figure A.1: Backward Blocking. Phase 1: AB+, Phase 2: A+ (BB); C+ (Con1); CXT- (Con2). Paradigm taken from Shanks (1985) and adds another control group. After the first phase, associative strength is split between the equally salient stimuli. After the second phase in the BB group, the B stimulus is extinguished because stimulus A can account for reinforcement in phase 1. The control groups show that only when the previously paired stimulus is conditioned in phase 2 will B be extinguished. Backward blocking is usually found to be a weak phenomenon in the animal learning literature. Our simulation, however, shows a very strong backward blocking effect.

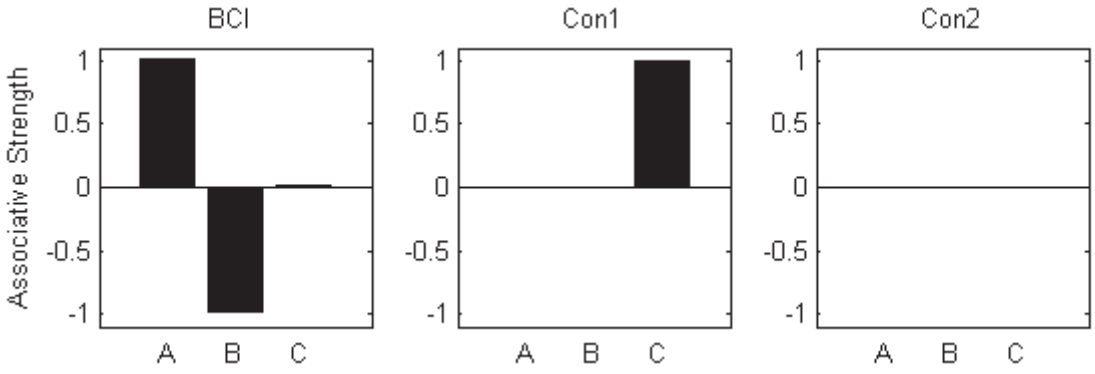

Figure A.2: Backward Conditioned Inhibition. Phase 1: AB-, Phase 2: A+ (BCI); C+ (Con1); CXT- (Con2). Taken from Chapman (1991), Experiment 5, but added an extra control group 1 (Con1). After the first phase, neither A nor B has any associative strength. After the second phase in the BCI group, the B stimulus gains substantial inhibitory strength because stimulus A was reinforced. B's inhibitory gain is used to account for the zero reinforcement given to compound AB in the first phase in light of A's excitatory gain. The control groups confirm that B becomes inhibitory only with the conditioning of the previously paired stimulus (A). The inhibitory gain in this simulation is substantially larger than in Chapman's (1991) human causal learning experiment.

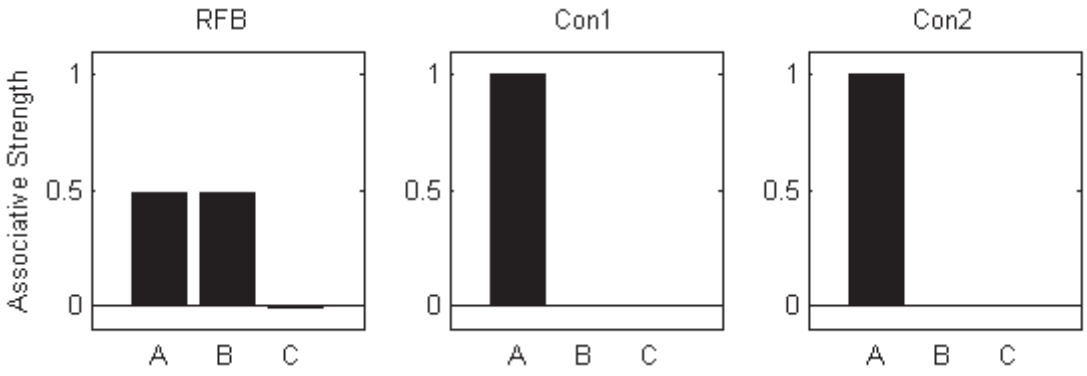

Figure A.3: Recovery from Forward Blocking. Phase 1: A+ Phase 2: AB+, Phase 3: A- (RFB); C- (Con1); CXT- (Con2). Simulation derived from Blaisdell, Gunther and Miller (1999), Experiment 3. Blaisdell et al. found that it takes a large number of extinction trials to detect recovery from forward blocking (compare Experiments 2 and 3), whereas in this simulation, far fewer are used to get a very substantial effect. With the 50 extinction trials (A-), the simulation does not get the thorough extinction that Blaisdell et al. gets with 800 trials. The controls show that the effect only occurs when the blocking stimulus is extinguished.

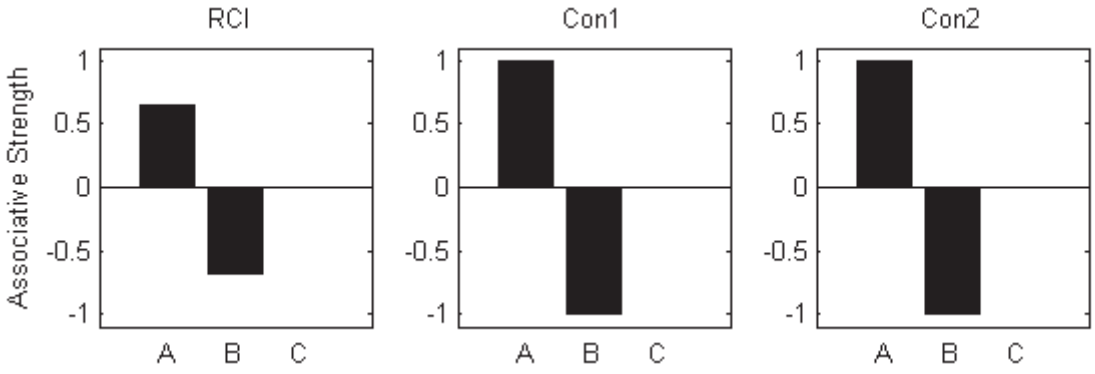

Figure A.4: Recovery from Conditioned Inhibition. Phase 1: A+, Phase 2: A+, AB-, Phase 3: A- (BCI); C- (Con1); CXT- (Con2). Taken from Lysle and Fowler (1985), Experiment 2. After the first two phases, A is seen as an excitor and B as an inhibitor (C is novel). After the third phase in the BCI group, the B stimulus loses inhibitory associative strength in proportion to the amount of excitatory strength lost by A's extinction. This contrasts with the two simulated control groups, where B's inhibitory strength is unaffected. In Lysle and Fowler, the extinction of A led to a nearly complete loss of inhibitory associative strength in B. Thus, this is a very potent effect. The matching effect in the simulation would be similarly potent given additional extinction trials in Phase 3 or a larger learning rate to complete the extinction of A as in Lysle and Fowler (1985).

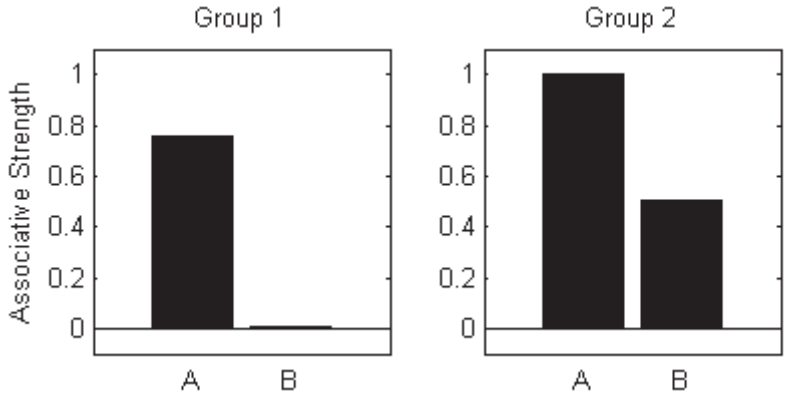

Figure A.5: Hall-Pearce Negative Transfer. Phase 1: A+ (G1), B+ (G2), Phase 2: A++. Adapted from Hall and Pearce (1979), Experiment 1. The first phase establishes the associative strength of the A and B stimuli at 0.5 (the strength of the reinforcement represented by a single "+" sign). In the second phase, a full strength reinforcement follows presentation of A. The associative strength of A in Group 1 lags that of Group 2. This appears to be a fairly strong effect in Hall and Pearce (1979) and is relatively strong in the simulations as well.

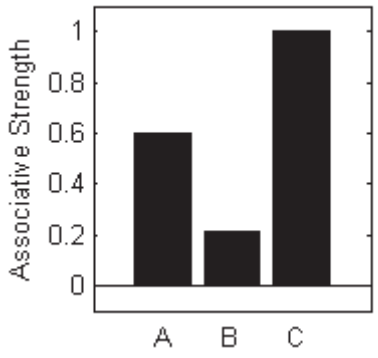

Figure A.6: Retardation Test for Conditioned Inhibition. Phase 1: A+, Phase 2: A+, AB-, Phase 3: B+, C+. See Rescorla (1969) for a review of many instances of the usage of this paradigm. After the first two phases, A is seen as an excitor and B as an inhibitor. In the third phase, the inhibitor is instead conditioned alongside a control stimulus (C). The general finding is that the novel stimulus C will condition more readily than the conditioned inhibitor. This is also seen in the simulation such that after 50 phase 3 trials, the associative strength of stimulus C is much greater than for stimulus B. Note that in this simulation, the context stimulus was given 0.0 salience because in the third phase, it was gaining a lot of associative strength and obscuring the final results, though it did not change the ordinal relationship of the findings. This phenomenon can also be simulated with the Rescorla-Wagner model because an inhibitory feature will have a larger difference between its associative strength after phase 2 and the US than does the difference between a novel stimulus and the US.

Appendix B

Details of Frank's (2005) Model of the Basal Ganglia

The following subsections represent a formal description of Frank's 2005 model based on information gleaned from the original paper, related references, a book about Leabra modeling (O'Reilly & Munakata, 2000), the Emergent software implementation of the model and its documentation, as well as personal communication with Michael Frank. The goal was to determine the details of the model that are primarily responsible for its outward behaviour.

B.1 Connectivity

Each layer is composed of a set number of neurons, which receive a certain pattern of inputs from incoming projections. Most of the layers are topographic, having a channel to represent one action/choice versus another. The system input pattern is the exception, sending 20 projections (5 per cue) to fully connect with the striatal neurons and PMC neurons.

For some connections, strengths are uniformly distributed random numbers with a certain range. These connections are Input-PMC (0.345, 0.355), Input-STR (0.25, 0.75), and PMC-STR (0.44, 0.56). All other connections between regions have a certain mean strength (usually 1.0) but no randomness.

B.2 Activation

For each neuron in a layer or region, the membrane potential is computed based on three facets,

$$\frac{dV_m(t)}{dt} = g_e(t)\overline{g_e}(E_e - V_m(t)) + g_l(t)\overline{g_l}(E_l - V_m(t)) + g_i(t)\overline{g_i}(E_i - V_m(t)) \qquad \text{(B.1)}$$

where $g_e(t)$ represents the excitatory input, $\overline{g_e}$ represents the maximum expected value of $g_e(t)$, and E_e is the reversal potential of the excitatory input. Corresponding

terms labeled with subscripts l and i represent the contribution of the leak current and the inhibitory input respectively. Frank includes an equation for computing the equilibrium potential,

$$V_m^{\infty}(t) = \frac{g_e(t)e}{g_e(t)\overline{g_e} + g_l(t)\overline{g_l} + g_i(t)\overline{g_i}} \tag{B.2}$$

However, since there is a feedback loop in the system, the input to the neurons will change over time and thus this equation will not hold.

A neuron fires with a rate according to

$$y(t) = (1 + \frac{1}{\gamma[V_m(t) - \Theta]_+})^{-1} \tag{B.3}$$

where Θ is the firing threshold and γ is an activation gain control. Under most circumstances, $\Theta = 0.25$. The exception to this is that it increases to $\Theta = 0.254$ in neurons being additionally enhanced by a burst (Go neurons) or dip (NoGo neurons) in dopamine[1]. The gain control is defined by the level of dopamine and the pathway to which a neuron belongs. By default, $\gamma = 600$. When dopamine levels are high (as in a burst), Go neurons are excited ($\gamma = 10000 * k$, where k represents the percentage of living dopamine cells, a value between 0 and 1) whereas when dopamine levels are low, they are enhanced less by dopamine and thus γ is set to less than the default $\gamma = 600 - 300 * k$). Combined with the learning rules described later, this leads to an increase of synaptic strengths when Go neurons are rewarded with high dopamine and a decrease when Go neurons are punished with a dopamine dip. The scenario is opposite for NoGo neurons, and thus the γ values above are set in the opposite arrangement. Note that the values for γ reported here are different from the current Emergent (Leabra) version of this model, but follow the original paper more closely.

The time-varying conductances used to compute the membrane potential can be computed as follows. For the leak current, the conductance is actually not time varying, but constant, $g_l(t) = 1.0$. Note that the GPe and GPi only receive inhibitory inputs, and thus require some excitatory input or conductance to become active. This is accomplished by adding a large positive number (20) to E_l, bringing it above the

[1]Note that Θ is not manipulated in Frank's more recent models of this kind, because it apparently introduces complicated dynamics into the model. In the present model, the threshold increase printed in the original paper associated with a correct choice in the WP Task cause most or all of the direct pathway neurons to be silenced during the plus phase, which means that their weights would never be increased, leading to negative effects.

threshold, which is interpreted as the leak current letting in positive ions instead of negative ions and thereby encouraging activity in the neuron rather than attenuating it. For the excitatory input conductances are

$$g_e(t) = (1 - dt_{net})g_e(t-1) + dt_{net}(\frac{\beta}{N} + \frac{1}{n_p}\sum_k \frac{1}{\alpha_k}\sum_i x_{ik}w_{ijk}) \tag{B.4}$$

where $dt_{net} = 0.7$ is the proportion of the previously computed value mixed with that computed at the present time. The constants n_p and α_k are the number of projections to a layer and each (k) one's total expected value, respectively. These help to normalize and balance input projections with differing numbers of neurons and levels of activity. Finally, the constant β represents a bias that has the weight of an additional projection.

The inhibition used in this model is a k-winners-take-all form. To get k winners, the inhibition used is computed to suppress all but k neurons. This is done by computing conductances for the k^{th} and $k+1$ neurons and taking their average

$$g_i(t) = g^{\Theta}_{k+1} + 0.25(g^{\Theta}_k - g^{\Theta}_{k+1}) \tag{B.5}$$

where the conductance for a particular neuron (i.e. the k^{th} or $k+1$) is

$$g^{\Theta} = \frac{g_e(t)\overline{g_e}(E_e - \Theta) + g_l(t)\overline{g_l}(E_l - \Theta)}{\Theta - E_i} \tag{B.6}$$

Not every layer uses k-winners-take-all inhibition. According to Frank (personal communication), the GPe and GPi layers do not make use of it because they have inhibitory input from other layers. However, in so doing, the GPe basically outputs either all or nothing. This means that the indirect pathway tends to make a tentative decision based on the striatal activity rather than expressing the degree to which indirect pathway neurons are active. Using the k-winners-takes-all inhibition in the GPe avoids this.

B.3 Learning

Learning is done in two ways: Hebbian learning with Oja normalization and contrastive Hebbian learning. Hebbian learning with Oja normalization is computed as

$$\Delta w_{ij} = y_j(x_i - w_{ij}) \tag{B.7}$$

whereas contrastive Hebbian learning is

$$\Delta w_{ij} = x_i^+ y_j^+ - x_i^- y_j^- \tag{B.8}$$

where the superscript + indicates a quantity associated with the plus phase and the − subscript for minus phase quantities. Note that in cases where this learning is used, there is no change for the connections of neurons that are inactive in the plus phase. Learning occurs in the striatum and PMC. In the striatum, a mix of Hebbian and contrastive Hebbian is used with a "mixing factor" that describes the relative proportions of each. This is used so that when contrastive Hebbian learning is no longer effective (i.e., the difference between Go neural activations in the plus and minus phases is very small), the simple Hebbian learning can still reinforce these weights.

Appendix C

Notices of Permission to Use Excerpts from Author's Publications

In this thesis, large and small excerpts were taken verbatim from two of the author's own papers (Connor et al., 2013; Connor & Trappenberg, 2013). Both Springer (publisher of the journal Learning and Behavior) and IEEE (publisher of the proceedings of the International Joint Conference on Neural Networks) who accepted these articles state in the documents reproduced on the following pages that use of the author's work in their own dissertation is allowed.

Copyright Transfer Statement

The copyright to this article, including any graphic elements therein (e.g. illustrations, charts, moving images), is hereby assigned for good and valuable consideration to Springer effective if and when the article is accepted for publication and to the extent assignable if assignability is restricted for by applicable law or regulations (e.g. for U.S. government or crown employees). Author warrants (i) that he/she is the sole owner or has been authorized by any additional copyright owner to assign the right, (ii) that the article does not infringe any third party rights and no license from or payments to a third party is required to publish the article and (iii) that the article has not been previously published or licensed.

The copyright assignment includes without limitation the exclusive, assignable and sublicensable right, unlimited in time and territory, to reproduce, publish, distribute, transmit, make available and store the article, including abstracts thereof, in all forms of media of expression now known or developed in the future, including pre- and reprints, translations, photographic reproductions and microform. Springer may use the article in whole or in part in electronic form, such as use in databases or data networks for display, print or download to stationary or portable devices. This includes interactive and multimedia use and the right to alter the article to the extent necessary for such use.

Authors may self-archive the Author's accepted manuscript of their articles on their own websites. Authors may also deposit this version of the article in any repository, provided it is only made publicly available 12 months after official publication or later. He/she may not use the publisher's version (the final article), which is posted on **SpringerLink** and other Springer websites, for the purpose of self-archiving or deposit. Furthermore, the Author may only post his/her version provided acknowledgement is given to the original source of publication and a link is inserted to the published article on Springer's website. The link must be accompanied by the following text: „The final publication is available at **link.springer.com**".

Prior versions of the article published on non-commercial **pre-print servers** like arXiv.org can remain on these servers and/or can be updated with Author's accepted version. The final published version (in pdf or html/xml format) cannot be used for this purpose. Acknowledgement needs to be given to the final publication and a link must be inserted to the published article on Springer's website, accompanied by the text "The final publication is available at link.springer.com". Author retains the right to use his/her article for his/her further scientific career by including the final published journal article in other publications such as dissertations and postdoctoral qualifications provided acknowledgement is given to the original source of publication.

Author is requested to use the appropriate DOI for the article. Articles disseminated via **link.springer.com** are indexed, abstracted and referenced by many abstracting and information services, bibliographic networks, subscription agencies, library networks, and consortia.

After submission of the agreement signed by the corresponding author, changes of authorship or in the order of the authors listed will not be accepted by Springer.

Journal:

Title of article:

Author(s):

Author's signature:

Date:

62482 wmxdesign 03/13

Springer

Frequently Asked Questions Regarding IEEE Permissions

- **When is permission to reuse IEEE required?**

As a general rule, IEEE requires permission be sought to reproduce any substantial part of its intellectual property, including any text, illustrations, charts, tables, photographs, or other material from previously published sources used. IEEE also requires that all references or sources used be credited, whether or not permission is required. For further guidance, please contact pubs-permissions@ieee.org.

- **From whom do I need permission?**

Permission must be sought from IEEE to reuse its intellectual property. In most cases this will mean locating the material you wish to reuse in IEEE Xplore, where you will find a "request permission" link either on the Table of Contents or on the Article Abstract Page.

- **What if I do not see the "Request Permission" link on either the Table of Contents or the Abstract Page in Xplore?**

If you do not see a permission link on the Abstract Page, we recommend you review the front cover and/or the copyright page in the document itself (often, these pages are freely available for viewing in Xplore) in order to determine copyright owner. If you are unsure, please contact pubs-permissions@ieee.org.

- **Does IEEE require individuals working on a thesis or dissertation to obtain formal permission for reuse?**

The IEEE does not require individuals working on a thesis to obtain a formal reuse license, however, you must follow the requirements listed below:

Textual Material

Using short quotes or referring to the work within these papers) users must give full credit to the original source (author, paper, publication) followed by the IEEE copyright line © 2011 IEEE.

In the case of illustrations or tabular material, we require that the copyright line © [Year of original publication] IEEE appear prominently with each reprinted figure and/or table.

If a substantial portion of the original paper is to be used, and if you are not the senior author, also obtain the senior author's approval.

Full-Text Article

If you are using the entire IEEE copyright owned article, the following IEEE copyright/ credit notice should be placed prominently in the references: © [year of original publication] IEEE. Reprinted, with permission, from [author names, paper title, IEEE publication title, and month/year of publication]

Only the accepted version of an IEEE copyrighted paper can be used when posting the paper or your thesis on-line.

In placing the thesis on the author's university website, please display the following message in a prominent place on the website: In reference to IEEE copyrighted material which is used with permission in this thesis, the IEEE does not endorse any of [university/educational entity's name goes here]'s products or services. Internal or personal use of this material is permitted. If interested in reprinting/republishing IEEE copyrighted material for advertising or promotional purposes or for creating new collective works for resale or redistribution, please go to http://www.ieee.org/publications_standards/publications/rights/rights_link.html to learn how to obtain a License from RightsLink.

If applicable, University Microfilms and/or ProQuest Library, or the Archives of Canada may supply single copies of the dissertation.

- **If I want to republish an article in another language do I still need to obtain a license from IEEE?**

If you are republishing IEEE intellectual property, we do require you obtain a license that includes any translations. The required translation disclaimer and other translation guidelines are available in the IEEE Terms and Conditions contained in the license provided by the Copyright Clearance Center (RightsLink service).

- **How do I obtain permission to use photographs or illustrations?**

IEEE does not always own reproduction rights to photographs or illustrations; rather, such rights may have been retained by the photographer or illustrator. If the source from which the material is borrowed does not indicate who owns reproduction rights, users of these photographs or illustrations are required to locate the rightsholder, directly.

- **Do I need to obtain permission to use IEEE material posted on its website?**

Yes. As a general rule, most material found on the internet is protected by copyright law even if a notice is not displayed. IEEE does require that you inquire about such permission before using any material found on all IEEE copyright owned websites.

- **Does IEEE require certain rights when requesting permission to use material in an IEEE work?**

IEEE does allow permission to reuse small portions of text in another IEEE copyright owned document only (e.g., the equivalent of several paragraphs only) and figures. Our only requirement is that you 1) provide full credit information pertaining to the original IEEE publications (e.g., author name, paper title, publication title, month and year of original publication). Requests for permission to reuse larger portions should be sent to pubs-permissions@ieee.org .

- **What is Rightslink®?**

Rightslink® is the Copyright Clearance Center's automated permissions granting service, which is used by IEEE along with many other STM publishers such as Springer, Elsevier, and Taylor & Francis. Through this permission service, customers can request permission for IEEE Periodical and Conference content from the point of access; (normally found on the abstract page of the individual article, in IEEE Xplore).

- **Is IEEE an STM signatory publisher?**

No, IEEE is not a signatory to the STM (International Association of Scientific, Technical & Medical Publishers) Permissions Guidelines, last updated February 2012.

April 2013, nbd

Bibliography

Aitken, M., & Dickinson, A. (2005). Simulations of a modified SOP model applied to retrospective revaluation of human causal learning. *Learning & Behavior*, *33*(2), 147–159.

Albin, R. L., Young, A. B., & Penney, J. B. (1989). The functional anatomy of basal ganglia disorders. *Trends in Neuroscience*, *12*(10), 366-375.

Alexander, G. E., DeLong, M. R., & Strick, P. L. (1986). Parallel organization of functionally segregated circuits linking basal ganglia and cortex. *Annual Review of Neuroscience*, *9*(1), 357-381.

Ambroggi, F., Ishikawa, A., Fields, H. L., & Nicola, S. M. (2008). Basolateral amygdala neurons facilitate reward-seeking behavior by exciting nucleus accumbens neurons. *Neuron*, *59*(4), 648–661.

Amundson, J., Escobar, M., & Miller, R. (2003). Proactive interference between cues trained with a common outcome in first-order Pavlovian conditioning. *Journal of Experimental Psychology: Animal Behavior Processes*, *29*(4), 311.

Back, A. D., & Trappenberg, T. P. (2001). Selecting inputs for modeling using normalized higher order statistics and independent component analysis. *Neural Networks, IEEE Transactions on*, *12*(3), 612–617.

Baetu, I., & Baker, A. G. (2010). Extinction and blocking of conditioned inhibition in human causal learning. *Learning & Behavior*, *38*, 394–407.

Baldassarre, G. (2002). A modular neural-network model of the basal ganglia's role in learning and selecting motor behaviours. *Cognitive Systems Research*, *3*(1), 5–13.

Balleine, B. W., & O'Doherty, J. P. (2009). Human and rodent homologies in action control: Corticostriatal determinants of goal-directed and habitual action. *Neuropsychopharmacology*, *35*(1), 48–69.

Bar-Gad, I., Havazelet-Heimer, G., Goldberg, J. A., Ruppin, E., & Bergman, H. (2000). Reinforcement-driven dimensionality reduction-a model for information processing in the basal ganglia. *Journal of basic and clinical physiology and pharmacology*, *11*(4), 305–320.

Bar-Gad, I., Morris, G., & Bergman, H. (2003). Information processing, dimensionality reduction and reinforcement learning in the basal ganglia. *Progress in neurobiology*, *71*(6), 439–473.

Barto, A. G. (1995). Adaptive critics and the basal ganglia. In J. Houk, J. Davis, & D. Beiser (Eds.), *Models of information processing in the basal ganglia* (pp. 215–232). MIT Press.

Bengio, S., & Bengio, Y. (2000). Taking on the curse of dimensionality in joint distributions using neural networks. *Neural Networks, IEEE Transactions on*, *11*(3), 550–557.

Berridge, K. C. (2007). The debate over dopamine's role in reward: the case for incentive salience. *Psychopharmacology*, *191*(3), 391–431.

Bi, G., & Poo, M. (1998). Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type. *The Journal of Neuroscience*, *18*(24), 10464–10472.

Bi, G., & Poo, M. (2001). Synaptic modification by correlated activity: Hebb's postulate revisited. *Annual review of neuroscience*, *24*(1), 139–166.

Bishop, C. M. (1995). Training with noise is equivalent to Tikhonov regularization. *Neural computation*, *7*(1), 108–116.

Blaisdell, A. P., Bristol, A. S., Gunther, L. M., & Miller, R. R. (1998). Overshadowing and latent inhibition counteract each other: Support for the comparator hypothesis. *Journal of Experimental Psychology: Animal Behavior Processes*, *24*(3), 335.

Blaisdell, A. P., Gunther, L. M., & Miller, R. R. (1999). Recovery from blocking achieved by extinguishing the blocking CS. *Animal Learning Behavior*, *27*(1), 63–76.

Blaisdell, A. P., Savastano, H. I., & Miller, R. R. (1999). Overshadowing of explicitly unpaired conditioned inhibition is disrupted by preexposure to the overshadowed inhibitor. *Animal Learning & Behavior*, *27*(3), 346–357.

Bliss, T. V., & Lømo, T. (1973). Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetized rabbit following stimulation of the perforant path. *The Journal of physiology*, *232*(2), 331–356.

Bradfield, L. A., & McNally, G. P. (2010). The role of nucleus accumbens shell in learning about neutral versus excitatory stimuli during Pavlovian fear conditioning. *Learning & Memory*, *17*(7), 337–343.

Brandon, S. E., & Wagner, A. R. (1998). Occasion setting: Influences of conditioned emotional responses and configural cues. In N. A. Schmajuk & P. C. Holland (Eds.), *Occasion setting: Associative learning and cognition in animals* (pp. 343–382). American Psychological Association, Washington, DC.

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and regression trees*. Belmont, California, U.S.A.: Wadsworth Publishing Company.

Bromberg-Martin, E. S., Matsumoto, M., & Hikosaka, O. (2010). Dopamine in motivational control: rewarding, aversive, and alerting. *Neuron*, *68*(5), 815–834.

Bush, R. R., & Mosteller, F. (1955). *Stochastic models for learning*. New York: John Wiley and Sons.

Chapman, G. (1991). Trial order affects cue interaction in contingency judgment. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*(5), 837.

Chinta, L. V., & Tweed, D. B. (2012). Adaptive optimal control without weight transport. *Neural Computation*, *24*(6), 1487–1518.

Clugnet, M.-C., & LeDoux, J. E. (1990). Synaptic plasticity in fear conditioning circuits: induction of LTP in the lateral nucleus of the amygdala by stimulation of the medial geniculate body. *The Journal of neuroscience*, *10*(8), 2818–2824.

Cole, R., Barnet, R., & Miller, R. (1995). Effect of relative stimulus validity: Learning or performance deficit? *Journal of Experimental Psychology: Animal Behavior Processes*, *21*(4), 293.

Cole, R., Denniston, J., & Miller, R. (1996). Reminder-induced attenuation of the effect of relative stimulus validity. *Learning & Behavior*, *24*(3), 256–265.

Connor, P., LoLordo, V., & Trappenberg, T. (2013). An elemental model of retrospective revaluation without within-compound associations. *Learning and Behavior*.

Connor, P., & Trappenberg, T. (2011). Characterizing a brain-based value-function approximator. In C. Butz & P. Lingras (Eds.), *Advances in artificial intelligence* (Vol. 6657, pp. 92–103). Springer Berlin Heidelberg.

Connor, P., & Trappenberg, T. (2013). Biologically plausible feature selection through relative correlation. In *Proceedings of the 2013 international joint conference on neural networks (IJCNN)* (pp. 759–766).

Corlett, P. R., Aitken, M. R., Dickinson, A., Shanks, D. R., Honey, G. D., Honey, R. A., ... Fletcher, P. C. (2004). Prediction error during retrospective revaluation of causal associations in humans: fMRI evidence in favor of an associative model of learning. *Neuron*, *44*(5), 877–888.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning*, *20*(3), 273–297.

Courville, A. C., Daw, N. D., & Touretzky, D. S. (2006). Bayesian theories of conditioning in a changing world. *Trends in cognitive sciences*, *10*(7), 294–300.

Cowan, R. L., & Wilson, C. J. (1994). Spontaneous firing patterns and axonal projections of single corticostriatal neurons in the rat medial agranular cortex. *J Neurophysiol*, *71*(1), 17-32.

Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. *Mathematics of control, signals and systems*, *2*(4), 303–314.

Dawson, M. (2008). Connectionism and classical conditioning. *Comparative Cognition and Behavior Reviews*, *3*, 1–115.

De Houwer, J., & Beckers, T. (2002). Higher-order retrospective revaluation in human causal learning. *The Quarterly Journal of Experimental Psychology. B, Comparative and Physiological Psychology*, *55*(2), 137–151.

Delamater, A. R., Sosa, W., & Katz, M. (1999). Elemental and configural processes in patterning discrimination learning. *The Quarterly Journal of Experimental Psychology: Section B*, *52*(2), 97–124.

Delgado, M. R., Li, J., Schiller, D., & Phelps, E. A. (2008). The role of the striatum in aversive learning and aversive prediction errors. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1511), 3787–3800.

DeLong, M. R. (1990). Primate models of movement disorders of basal ganglia origin. *Trends in Neuroscience*, *13*(7), 281-285.

Denniston, J., Miller, R., & Matute, H. (1996). Biological significance as a determinant of cue competition. *Psychological Science*, *7*(6), 325–331.

Denniston, J., Savastano, H., & Miller, R. (2001). The extended comparator hypothesis: Learning by contiguity, responding by relative strength. In R. Mowrer & S. Klein (Eds.), *Handbook of contemporary learning theories* (pp. 65–117). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

DeVito, P. L., & Fowler, H. (1986). Effects of contingency violations on the extinction of a conditioned fear inhibitor and a conditioned fear excitor. *Journal of Experimental Psychology: Animal Behavior Processes*, *12*(2), 99.

DeVito, P. L., & Fowler, H. (1987). Enhancement of conditioned inhibition via an extinction treatment. *Animal Learning & Behavior*, *15*, 448–454.

DiCarlo, J. J., Zoccolan, D., & Rust, N. C. (2012). How does the brain solve visual object recognition? *Neuron*, *73*(3), 415–434.

Dickinson, A., & Burke, J. (1996). Within-compound associations mediate the retrospective revaluation of causality judgements. *The Quarterly Journal of Experimental Psychology. B, Comparative and Physiological Psychology*, *49*(1), 60–80.

Doig, N. M., Moss, J., & Bolam, J. P. (2010). Cortical and thalamic innervation of direct and indirect pathway medium-sized spiny neurons in mouse striatum. *The Journal of Neuroscience*, *30*(44), 14610–14618.

Dopson, J., Pearce, J., & Haselgrove, M. (2009). Failure of retrospective revaluation to influence blocking. *Journal of Experimental Psychology: Animal Behavior Processes*, *35*(4), 473.

Escobar, M., Pineño, O., & Matute, H. (2002). A comparison between elemental and compound training of cues in retrospective revaluation. *Learning & Behavior*, *30*(3), 228–238.

Espinet, A., Iraola, J., Bennett, C., & Mackintosh, N. (1995). Inhibitory associations between neutral stimuli in flavor-aversion conditioning. *Animal Learning & Behavior*, *23*(4), 361–368.

Evangelista, P. F., Embrechts, M. J., & Szymanski, B. K. (2006). Taming the curse of dimensionality in kernels and novelty detection. In *Applied soft computing technologies: The challenge of complexity* (pp. 425–438). Springer.

Fadok, J., Darvas, M., Dickerson, T., & Palmiter, R. (2010). Long-term memory for Pavlovian fear conditioning requires dopamine in the nucleus accumbens and basolateral amygdala. *PloS one*, *5*(9), e12751.

Fanselow, M. S., & Poulos, A. M. (2005). The neuroscience of mammalian associative learning. *Annual Review of Psychology*, *56*(1), 207–234.

Fiorillo, C. D. (2003). Discrete coding of reward probability and uncertainty by dopamine neurons. *Science*, *299*(5614), 1898–1902.

FitzHugh, R. (1961). Impulses and physiological states in theoretical models of nerve membrane. *Biophysical journal*, *1*(6), 445–466.

Flagel, S., Clark, J., Robinson, T., Mayo, L., Czuj, A., Willuhn, I., ... Akil, H. (2010). A selective role for dopamine in stimulus-reward learning. *Nature*, *469*(7328), 53–57.

Flaherty, A. W., & Graybiel, A. M. (1991). Corticostriatal transformations in the primate somatosensory system. Projections from physiologically mapped body-part representations. *J Neurophysiol*, *66*(4), 1249-1263.

Fodor, I. K. (2002). *A survey of dimension reduction techniques.* Technical Report UCRL-ID-148494, Lawrence Livermore National Laboratory.

Frank, M. J. (2005). Dynamic dopamine modulation in the basal ganglia: a neurocomputational account of cognitive deficits in medicated and nonmedicated Parkinsonism. *Journal of Cognitive Neuroscience*, *17*(1), 51–72.

Frank, M. J., & Fossella, J. A. (2011). Neurogenetics and pharmacology of learning, motivation, and cognition. *Neuropsychopharmacology: official publication of the American College of Neuropsychopharmacology*, *36*(1), 133–152.

Gallistel, C. R., Fairhurst, S., & Balsam, P. (2004). The learning curve: Implications of a quantitative analysis. *Proceedings of the national academy of sciences of the United States of America*, *101*(36), 13124–13131.

Gershman, S., Markman, A., & Otto, A. (2012). Retrospective revaluation in sequential decision making: A tale of two systems. *Journal of experimental psychology. General*.

Gershman, S. J., & Niv, Y. (2012). Exploring a latent cause theory of classical conditioning. *Learning & behavior*, *40*(3), 255–268.

Ghirlanda, S. (2005). Retrospective revaluation as simple associative learning. *Journal of experimental psychology. Animal behavior processes*, *31*(1), 107–111.

Gruber, A. J., Powell, E. M., & O'Donnell, P. (2009). Cortically activated interneurons shape spatial aspects of Cortico-Accumbens processing. *Journal of Neurophysiology*, *101*(4), 1876–1882.

Gurney, K., Prescott, T. J., & Redgrave, P. (2001). A computational model of action selection in the basal ganglia. I. a new functional anatomy. *Biological Cybernetics*, *84*(6), 401-410.

Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *The Journal of Machine Learning Research*, *3*, 1157–1182.

Guyon, I., Gunn, S., Nikravesh, M., & Zadeh, L. (2006). *Feature extraction* (Vol. 207). Springer.

Hall, G., & Pearce, J. M. (1979). Latent inhibition of a CS during CS–US pairings. *Journal of Experimental Psychology: Animal Behavior Processes*, *5*(1), 31.

Hallam, S. C., Grahame, N. J., Harris, K., & Miller, R. R. (1992). Associative structures underlying enhanced negative summation following operational extinction of a Pavlovian inhibitor. *Learning and Motivation*, *23*(1), 43–62.

Haralambous, T., & Westbrook, R. F. (1999). An infusion of bupivacaine into the nucleus accumbens disrupts the acquisition but not the expression of contextual fear conditioning. *Behavioral Neuroscience*, *113*(5), 925–940.

Harris, J. A. (2006). Elemental representations of stimuli in associative learning. *Psychological Review*, *113*(3), 584–605.

Harris, J. A., Gharaei, S., & Moore, C. A. (2009). Representations of single and compound stimuli in negative and positive patterning. *Learning & Behavior*, *37*(3), 230–245.

Hernandez-Lopez, S., Bargas, J., Surmeier, D. J., Reyes, A., & Galarraga, E. (1997). D1 receptor activation enhances evoked discharge in neostriatal medium spiny neurons by modulating an L-type Ca2+ conductance. *Journal of Neuroscience*, *17*(9), 3334–3342.

Hernandez-Lopez, S., Tkatch, T., Perez-Garci, E., Galarraga, E., Bargas, J., Hamm, H., & Surmeier, D. J. (2000). D2 dopamine receptors in striatal medium spiny neurons reduce L-type Ca2+ currents and excitability via a novel PLC[beta]1-IP3-calcineurin-signaling cascade. *Journal of Neuroscience*, *20*(24), 8987–8995.

Hodgkin, A. L., & Huxley, A. F. (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. *The Journal of physiology*, *117*(4), 500.

Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, *12*(1), 55–67.

Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural networks*, *2*(5), 359–366.

Horvitz, J. C., Choi, W. Y., Morvan, C., Eyny, Y., & Balsam, P. D. (2007). A "Good parent" function of dopamine: Transient modulation of learning and performance during early stages of training. *Annals of the New York Academy of Sciences*, *1104*(1), 270–288.

Houk, J. (2007). Models of basal ganglia. *Scholarpedia*, *2*(10), 1633.

Houk, J., Adams, J. L., & Barto, A. G. (1995). A model of how the basal ganglia generate and use neural signals that predict reinforcement. In *Models of information processing in the basal ganglia* (pp. 249–270). MIT Press.

Huerta-Ocampo, I., Mena-Segovia, J., & Bolam, J. P. (2013). Convergence of cortical and thalamic input to direct and indirect pathway medium spiny neurons in the striatum. *Brain Structure and Function*, 1–14.

Humphries, M. D., Stewart, R. D., & Gurney, K. N. (2006). A physiologically plausible model of action selection and oscillatory activity in the basal ganglia. *The Journal of neuroscience*, *26*(50), 12921–12942.

Iordanova, M., McNally, G., & Westbrook, R. (2006). Opioid receptors in the nucleus accumbens regulate attentional learning in the blocking paradigm. *The Journal of neuroscience*, *26*(15), 4036–4045.

Iordanova, M., Westbrook, R., & Killcross, A. (2006). Dopamine activity in the nucleus accumbens modulates blocking in fear conditioning. *European Journal of Neuroscience*, *24*(11), 3265–3270.

Ito, R., Robbins, T. W., Pennartz, C. M., & Everitt, B. J. (2008). Functional interaction between the hippocampus and nucleus accumbens shell is necessary for the acquisition of appetitive spatial context conditioning. *Journal of Neuroscience*, *28*(27), 6950–6959.

Jamieson, R., Crump, M., & Hannah, S. (2012). An instance theory of associative learning. *Learning & Behavior*, *40*(1), 61–82.

Jenkins, W. O., & Stanley Jr, J. C. (1950). Partial reinforcement: a review and critique. *Psychological Bulletin*, *47*(3), 193.

Ji, D., & Wilson, M. A. (2006). Coordinated memory replay in the visual cortex and hippocampus during sleep. *Nature neuroscience*, *10*(1), 100–107.

Joel, D., Niv, Y., & Ruppin, E. (2002). Actor–critic models of the basal ganglia: New anatomical and computational perspectives. *Neural networks*, *15*(4), 535–547.

Joel, D., & Weiner, I. (2000). The connections of the dopaminergic system with the striatum in rats and primates: an analysis with respect to the functional and compartmental organization of the striatum. *Neuroscience*, *96*(3), 451–474.

Jolliffe, I. T. (1986). *Principal component analysis* (Vol. 487). Springer-Verlag New York.

Kamin, L. J. (1968). "Attention-like" processes in classical conditioning. In M. R. Jones (Ed.), *Miami symposium on the prediction of behavior, 1967: Aversive stimulation* (pp. 9–31). Coral Gables, Florida: University of Miami Press.

Kamin, L. J. (1969). Predictability, surprise, attention and conditioning. In B. A. Campbell & R. M. Church (Eds.), *Punishment and aversive behavior* (pp. 279–296). New York: Appleton-Century-Crofts.

Kasprow, W. J., Cacheiro, H., Balaz, M., & Miller, R. (1982). Reminder-induced recovery of associations to an overshadowed stimulus. *Learning and Motivation*, *13*(2), 155–166.

Kasprow, W. J., Catterson, D., Schachtman, T., & Miller, R. (1984). Attenuation of latent inhibition by post-acquisition reminder. *The Quarterly Journal of Experimental Psychology*, *36*(1), 53–63.

Kasprow, W. J., Schachtman, T. R., & Miller, R. R. (1987). The comparator hypothesis of conditioned response generation: Manifest conditioned excitation and inhibition as a function of relative excitatory strengths of CS and conditioning context at the time of testing. *Journal of Experimental Psychology. Animal Behavior Processes*, *13*(4), 395–406.

Kaufman, M. A., & Bolles, R. C. (1981). A nonassociative aspect of overshadowing. *Bulletin of the Psychonomic Society*, *18*(6), 318-320.

Kawaguchi, Y., Wilson, C. J., & Emson, P. C. (1990). Projection subtypes of rat neostriatal matrix cells revealed by intracellular injection of biocytin. *The Journal of Neuroscience*, *10*(10), 3421–3438.

Knowlton, B. J., & Squire, . G. M. A., L. R. (1994). Probabilistic category learning in amnesia. *Learning and Memory*(1), 106-120.

Kraemer, P., Lariviere, N., & Spear, N. (1988). Expression of a taste aversion conditioned with an odor-taste compound: Overshadowing is relatively weak in weanlings and decreases over a retention interval in adults. *Learning & Behavior*, *16*(2), 164–168.

Krogh, A., & Hertz, J. A. (1992). A simple weight decay can improve generalization. In *Advances in neural information processing systems 4* (pp. 950–957). Morgan Kaufmann.

Kutlu, M., & Schmajuk, N. (2012). Solving Pavlov's puzzle: Attentional, associative, and flexible configural mechanisms in classical conditioning. *Learning & Behavior*, *40*(3), 269–291.

Lansink, C. S., Goltstein, P. M., Lankelma, J. V., McNaughton, B. L., & Pennartz, C. M. (2009). Hippocampus leads ventral striatum in replay of place-reward information. *PLoS biology*, *7*(8), e1000173.

LeDoux, J. (2007). The amygdala. *Current Biology*, *17*(20), R868–R874.

Lei, W., Jiao, Y., Del Mar, N., & Reiner, A. (2004). Evidence for differential cortical input to direct pathway versus indirect pathway striatal projection neurons in rats. *J. Neurosci.*, *24*(38), 8289-8299.

Le Pelley, M. E. (2004). The role of associative history in models of associative learning: a selective review and a hybrid model. *The Quarterly Journal of Experimental Psychology. B, Comparative and Physiological Psychology*, *57*(3), 193–243.

Le Pelley, M. E., & McLaren, I. P. L. (2001). Retrospective revaluation in humans: Learning or memory? *The Quarterly Journal of Experimental Psychology Section B*, *54*(4), 311–352.

Lex, B., & Hauber, W. (2010). The role of nucleus accumbens dopamine in outcome encoding in instrumental and Pavlovian conditioning. *Neurobiology of learning and memory*, *93*(2), 283–290.

Liljeholm, M., & Balleine, B. W. (2006). Stimulus salience and retrospective revaluation. *Journal of Experimental Psychology: Animal Behavior Processes*, *32*(4), 481–487.

Liljeholm, M., & Balleine, B. W. (2009). Mediated conditioning versus retrospective revaluation in humans: The influence of physical and functional similarity of cues. *The Quarterly Journal of Experimental Psychology*, *62*(3), 470–482.

Lotz, A., & Lachnit, H. (2009). Extinction of conditioned inhibition: effects of different outcome continua. *Learning & Behavior*, *37*(1), 85–94.

Lubow, R. E., & Moore, A. U. (1959). Latent inhibition: the effect of nonreinforced pre-exposure to the conditional stimulus. *Journal of Comparative and Physiological Psychology*, *52*, 415–419.

Luque, D., Flores, A., & Vadillo, M. A. (2013). Revisiting the role of within-compound associations in cue-interaction phenomena. *Learning & Behavior*, *41*(1), 61–76.

Lysle, D. T., & Fowler, H. (1985). Inhibition as a "slave" process: deactivation of conditioned inhibition through extinction of conditioned excitation. *Journal of Experimental Psychology. Animal Behavior Processes*, *11*(1), 71–94.

Maia, T. V. (2009). Reinforcement learning, conditioning, and the brain: Successes and challenges. *Cognitive, Affective, & Behavioral Neuroscience*, *9*(4), 343–364.

Maren, S. (2001). Neurobiology of Pavlovian fear conditioning. *Annual Review of Neuroscience*, *24*(1), 897–931.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. WH San Francisco: Freeman and Company.

Matamales, M., Bertran-Gonzalez, J., Salomon, L., Degos, B., Deniau, J.-M., Valjent, E., ... Girault, J.-A. (2009). Striatal medium-sized spiny neurons: identification by nuclear staining and study of neuronal subpopulations in bac transgenic mice. *PloS one*, *4*(3), e4770.

Matsumoto, M., & Hikosaka, O. (2009). Two types of dopamine neuron distinctly convey positive and negative motivational signals. *Nature*, *459*(7248), 837–841.

Matzel, L. D., Gladstein, L., & Miller, R. R. (1988). Conditioned excitation and conditioned inhibition are not mutually exclusive. *Learning and Motivation*, *19*(2), 99–121.

Matzel, L. D., Schachtman, T., & Miller, R. R. (1985). Recovery of an overshadowed association achieved by extinction of the overshadowing stimulus. *Learning and Motivation*, *16*, 398–412.

McDonald, A. (1991). Topographical organization of amygdaloid projections to the caudatoputamen, nucleus accumbens, and related striatal-like areas of the rat brain. *Neuroscience*, *44*(1), 15–33.

McGaugh, J. L. (2000). Memory–a century of consolidation. *Science*, *287*(5451), 248–251.

McKernan, M., & Shinnick-Gallagher, P. (1997). Fear conditioning induces a lasting potentiation of synaptic currents in vitro. *Nature*, *390*(6660), 607–611.

McLaren, I. P. L. (1993). APECS: A solution to the sequential learning problem. In *Proceedings of the xvth annual convention of the cognitive science society* (pp. 717–722).

McLaren, I. P. L. (2011). APECS: An adaptively parameterised model of associative learning and memory. In E. Alonso & E. Mondragon (Eds.), *Computational neuroscience for advancing artificial intelligence: Models, methods and applications* (pp. 145–164). Hershey: IGI Global.

McLaren, I. P. L., Forrest, C., & McLaren, R. (2012). Elemental representation and configural mappings: Combining elemental and configural theories of associative learning. *Learning & Behavior*, *40*(3), 320–333.

McNally, G., & Westbrook, R. (2006). Predicting danger: the nature, consequences, and neural mechanisms of predictive fear learning. *Learning & Memory*, *13*(3), 245–253.

Melchers, K., Lachnit, H., & Shanks, D. (2004). Within-compound associations in retrospective revaluation and in direct learning: A challenge for comparator theory. *The Quarterly Journal of Experimental Psychology Section B*, *57*(1), 25–54.

Miller, R. R., Barnet, R., & Grahame, N. (1992). Responding to a conditioned stimulus depends on the current associative status of other cues present during training of that specific stimulus. *Journal of Experimental Psychology: Animal Behavior Processes; Journal of Experimental Psychology: Animal Behavior Processes*, *18*(3), 251.

Miller, R. R., Barnet, R. C., & Grahame, N. J. (1995). Assessment of the Rescorla-Wagner model. *Psychological Bulletin*, *117*(3), 363–386.

Miller, R. R., & Matzel, L. D. (1988). The comparator hypothesis: A response rule for the expression of associations. *Psychology of Learning and Motivation*, *22*, 51–92.

Mink, J. W. (1996). The basal ganglia: focused selection and inhibition of competing motor programs. *Prog Neurobiol.*, *50*(4), 381-425.

Montague, P. R., Dayan, P., & Sejnowski, T. J. (1996). A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *Journal of Neuroscience*, *16*(5), 1936–1947.

Morris, C., & Lecar, H. (1981). Voltage oscillations in the barnacle giant muscle fiber. *Biophysical journal*, *35*(1), 193–213.

Morris, G., Arkadir, D., Nevet, A., Vaadia, E., & Bergman, H. (2004). Coincident but distinct messages of midbrain dopamine and striatal tonically active neurons. *Neuron*, *43*(1), 133–143.

Nagaishi, T., & Nakajima, S. (2008). Further evidence for the summation of latent inhibition and overshadowing in rats' conditioned taste aversion. *Learning and Motivation*, *39*(3), 221–242.

Nagumo, J., Arimoto, S., & Yoshizawa, S. (1962). An active pulse transmission line simulating nerve axon. *Proceedings of the IRE*, *50*(10), 2061–2070.

Nakajima, S., Ka, H., & Imada, J. (1999). Summation of overshadowing and latent inhibition in rats» conditioned taste aversion: Scapegoat technique works for familiar meals. *Appetite*, *33*(3), 299–307.

Nakajima, S., & Nagaishi, T. (2005). Summation of latent inhibition and overshadowing in a generalized bait shyness paradigm of rats. *Behavioural processes*, *69*(3), 369–377.

Nambu, A. (2007). Globus pallidus internal segment. *Progress in Brain Research*, *160*, 135–150.

Nambu, A., Tokuno, H., & Takada, M. (2002). Functional significance of the cortico-subthalamo-pallidal" hyperdirect" pathway. *Neuroscience research*.

Niv, Y. (2009). Reinforcement learning in the brain. *Journal of Mathematical Psychology*, *53*(3), 139–154.

O'Reilly, R. C., & Munakata, Y. (2000). *Computational explorations in cognitive neuroscience: Understanding the mind by simulating the brain*. MIT press.

Pakhotin, P., & Bracci, E. (2007). Cholinergic interneurons control the excitatory input to the striatum. *The Journal of neuroscience*, *27*(2), 391–400.

Pan, W., Schmidt, R., Wickens, J., & Hyland, B. (2008). Tripartite mechanism of extinction suggested by dopamine neuron activity and temporal difference model. *The Journal of Neuroscience*, *28*(39), 9619–9631.

Parent, A., & Hazrati, L. N. (1995a). Functional anatomy of the basal ganglia. II. the place of subthalamic nucleus and external pallidum in basal ganglia circuitry. *Brain Research. Brain Research Reviews*, *20*(1), 128–154.

Parent, A., & Hazrati, L. N. (1995b). Functional anatomy of the basal ganglia. I. the cortico-basal ganglia-thalamo-cortical loop. *Brain Research. Brain Research Reviews*, *20*(1), 91–127.

Pavlov, I. P. (1927). *Conditioned reflexes: an investigation of the physiological activity of the cerebral cortex*. Oxford, England: Oxford University Press.

Pearce, J. M., Dopson, J. C., Haselgrove, M., & Esber, G. R. (2012). The fate of redundant cues during blocking and a simple discrimination. *Journal of Experimental Psychology: Animal Behavior Processes*, *38*(2), 167.

Pearce, J. M., & Hall, G. (1980). A model for Pavlovian learning: Variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychological Review*, *87*(6), 532–552.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan Kaufmann.

Pennartz, C., Ito, R., Verschure, P., Battaglia, F., & Robbins, T. (2011). The hippocampalstriatal axis in learning, prediction and goal-directed behavior. *Trends in Neurosciences*, *34*(10), 548–559.

Phillips, G. D., Setzu, E., & Hitchcott, P. K. (2003). Facilitation of appetitive Pavlovian conditioning by d-amphetamine in the shell, but not the core, of the nucleus accumbens. *Behavioral Neuroscience*, *117*(4), 675–684.

Pineo, O., Urushihara, K., & Miller, R. R. (2005). Spontaneous recovery from forward and backward blocking. *Journal of experimental psychology Animal behavior processes*, *31*(2), 172–183.

Plenz, D. (2003). When inhibition goes incognito: feedback interaction between spiny projection neurons in striatal function. *Trends in Neurosciences*, *26*(8), 436–443.

Ponzi, A., & Wickens, J. (2010). Sequentially switching cell assemblies in random inhibitory networks of spiking neurons in the striatum. *The Journal of Neuroscience*, *30*(17), 5894–5911.

Ponzi, A., & Wickens, J. (2012). Input dependent cell assembly dynamics in a model of the striatal medium spiny neuron network. *Frontiers in systems neuroscience*, *6*.

Popescu, A., Popa, D., & Paré, D. (2009). Coherent gamma oscillations couple the amygdala and striatum during learning. *Nature Neuroscience*, *12*(6), 801–807.

Ratcliff, R. (1990). Connectionist models of recognition memory: constraints imposed by learning and forgetting functions. *Psychological review*, *97*(2), 285.

Redgrave, P. (1999). Is the short-latency dopamine response too short to signal reward error? *Trends in Neurosciences*, *22*(4), 146–151.

Redgrave, P., & Gurney, K. (2006). The short-latency dopamine signal: a role in discovering novel actions? *Nature Reviews Neuroscience*, *7*(12), 967–975.

Redhead, E. S., & Pearce, J. M. (1995). Stimulus salience and negative patterning. *The Quarterly Journal of Experimental Psychology*, *48*(1), 67–83.

Rescorla, R. A. (1969). Pavlovian conditioned inhibition. *Psychological Bulletin*, *72*, 77–94.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In B. AH & P. WF (Eds.), *Classical conditioning II: Current research and theory* (pp. 64–99). New York: Appleton Century Crofts.

Reynolds, J., & Wickens, J. (2002). Dopamine-dependent plasticity of corticostriatal synapses. *Neural Networks*, *15*(4-6), 507–521.

Ross, R. T., & Holland, P. C. (1981). Conditioning of simultaneous and serial feature-positive discriminations. *Learning & Behavior*, *9*(3), 293–303.

Rueda-Orozco, P. E., Mendoza, E., Hernandez, R., Aceves, J. J., Ibanez-Sandoval, O., Galarraga, E., & Bargas, J. (2009). Diversity in long-term synaptic plasticity at inhibitory synapses of striatal spiny neurons. *Learning & Memory*, *16*(8), 474–478.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning internal representations by error propagation. In D. E. Rumelhart, J. L. McClelland, & the PDP research group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition, volume 1: Foundations.* MIT Press.

Russchen, F., Bakst, I., Amaral, D., & Price, J. (1985). The amygdalostriatal projections in the monkey. An anterograde tracing study. *Brain Research*, *329*(1-2), 241–257.

Saeys, Y., Inza, I., & Larrañaga, P. (2007). A review of feature selection techniques in bioinformatics. *Bioinformatics*, *23*(19), 2507–2517.

San-Galli, A., Marchand, A. R., Decorte, L., & Di Scala, G. (2011). Retrospective revaluation and its neural circuit in rats. *Behavioural Brain Research*, *223*(2), 262–270.

Schachtman, T., Gee, J., Kasprow, W. J., & Miller, R. (1983). Reminder-induced recovery from blocking as a function of the number of compound trials. *Learning and Motivation*, *14*(2), 154–164.

Schultz, W. (1998). Predictive Reward Signal of Dopamine Neurons. *Journal of Neurophysiology*, *80*(1), 1–27.

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science*, *275*(5306), 1593–1599.

Setlow, B., Holland, P. C., & Gallagher, M. (2002). Disconnection of the basolateral amygdala complex and nucleus accumbens impairs appetitive Pavlovian second-order conditioned responses. *Behavioral Neuroscience*, *116*(2), 267–275.

Shanks, D. (1985). Forward and backward blocking in human contingency judgement. *The Quarterly Journal of Experimental Psychology Section B*, *37*(1), 1–21.

Shen, W., Flajolet, M., Greengard, P., & Surmeier, D. J. (2008). Dichotomous dopaminergic control of striatal synaptic plasticity. *Science*, *321*(5890), 848–851.

Shevill, I., & Hall, G. (2004). Retrospective revaluation effects in the conditioned suppression procedure. *Quarterly Journal of Experimental Psychology Section B*, *57*(4), 331–347.

Shiflett, M., & Balleine, B. (2010). At the limbic–motor interface: disconnection of basolateral amygdala from nucleus accumbens core and shell reveals dissociable components of incentive motivation. *European Journal of Neuroscience*, *32*(10), 1735–1743.

Smith, Y., & Bolam, J. (1990). The output neurones and the dopaminergic neurones of the substantia nigra receive a GABA-containing input from the globus pallidus in the rat. *Journal of Comparative Neurology*, *296*(1), 47–64.

Smith, Y., & Bolam, J. P. (1989). Neurons of the substantia nigra reticulata receive a dense GABA-containing input from the globus pallidus in the rat. *Brain research*, *493*(1), 160–167.

Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and computing*, *14*(3), 199–222.

Stout, S. C., & Miller, R. R. (2007). Sometimes-competing retrieval (SOCR): A formalization of the comparator hypothesis. *Psychological Review*, *114*(3), 759.

Stuber, G. D., Sparta, D. R., Stamatakis, A. M., van Leeuwen, W. A., Hardjoprajitno, J. E., Cho, S., ... Bonci, A. (2011). Excitatory transmission from the amygdala to nucleus accumbens facilitates reward seeking. *Nature*, *475*(7356), 377–380.

Suri, R. E., & Schultz, W. (1999). A neural network model with dopamine-like reinforcement signal that learns a spatial delayed response task. *Neuroscience*, *91*(3), 871–890.

Surmeier, D. J., Song, W.-J., & Yan, Z. (1996). Coordinated expression of dopamine receptors in neostriatal medium spiny neurons. *J. Neurosci.*, *16*(20), 6579-6591.

Sutton, R. S. (1996). Generalization in reinforcement learning: Successful examples using sparse coarse coding. *Advances in neural information processing systems*, 1038–1044.

Sutton, R. S., & Barto, A. G. (1990). Time-derivative models of Pavlovian reinforcement. In M. Gabriel & J. Moore (Eds.), *Learning and computational neuroscience: Foundations of adaptive networks* (pp. 497–537). MIT Press.

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction* (Vol. 1) (No. 1). Cambridge Univ Press.

Taverna, S., Ilijic, E., & Surmeier, D. J. (2008). Recurrent collateral connections of striatal medium spiny neurons are disrupted in models of Parkinson's disease. *Journal of Neuroscience*, *28*(21), 5504–5512.

Tepper, J. M., Bolam, J. P., et al. (2004). Functional diversity and specificity of neostriatal interneurons. *Current opinion in neurobiology*, *14*(6), 685–692.

Tepper, J. M., Koós, T., & Wilson, C. J. (2004). GABAergic microcircuits in the neostriatum. *Trends in neurosciences*, *27*(11), 662–669.

Tepper, J. M., Wilson, C. J., & Koos, T. (2008). Feedforward and feedback inhibition in neostriatal GABAergic spiny neurons. *Brain Research Reviews*, *58*(2), 272–281.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 267–288.

Tibshirani, R. (2011). Regression shrinkage and selection via the lasso: a retrospective. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *73*(3), 273–282.

Tikhonov, A. (1963). Solution of incorrectly formulated problems and the regularization method. In *Soviet math. dokl.* (Vol. 5, p. 1035).

Tobler, P. N., Dickinson, A., & Schultz, W. (2003). Coding of predicted reward omission by dopamine neurons in a conditioned inhibition paradigm. *Journal of Neuroscience*, *23*(32), 10402–10410.

Tobler, P. N., Fiorillo, C. D., & Schultz, W. (2005). Adaptive coding of reward value by dopamine neurons. *Science*, *307*(5715), 1642–1645.

Tsai, H.-C., Zhang, F., Adamantidis, A., Stuber, G. D., Bonci, A., de Lecea, L., & Deisseroth, K. (2009). Phasic firing in dopaminergic neurons is sufficient for behavioral conditioning. *Science*, *324*(5930), 1080–1084.

Tunstall, M. J., Oorschot, D. E., Kean, A., & Wickens, J. R. (2002). Inhibitory interactions between spiny projection neurons in the rat striatum. *Journal of Neurophysiology*, *88*(3), 1263–1269.

Tweed, D. (2011). *Cosmo 2011, Tweed's notes on learning* (Tech. Rep.). Queen's University.

Ungless, M. (2004). Dopamine: the salient issue. *Trends in Neurosciences*, *27*(12), 702–706.

Urcelay, G., Perelmuter, O., & Miller, R. (2008). Pavlovian backward conditioned inhibition in humans: Summation and retardation tests. *Behavioural Processes*, *77*(3), 299–305.

Urcelay, G. P., & Miller, R. R. (2006). Counteraction between overshadowing and degraded contingency treatments: Support for the extended comparator hypothesis. *Journal of Experimental Psychology: Animal Behavior Processes*, *32*(1), 21.

Usher, M., & McClelland, J. (2001). The time course of perceptual choice: the leaky, competing accumulator model. *Psychological Review*, *108*(3), 550.

Van Hamme, L., & Wasserman, E. A. (1994). Cue competition in causality judgments: The role of nonpresentation of compound stimulus elements. *Learning and Motivation*, *25*(2), 127–151.

Vapnik, V. N. (1995). *The nature of statistical learning theory*. New York, NY, USA: Springer-Verlag New York, Inc.

Waelti, P., Dickinson, A., & Schultz, W. (2001). Dopamine responses comply with basic assumptions of formal learning theory. *Nature*, *412*(6842), 43–48.

Wagner, A. R. (1981). SOP: A model of automatic memory processing in animal behavior. In N. E. Spears & R. R. Miller) (Eds.), *Information processing in animals: Memory mechanisms* (pp. 5–47). Earlbaum, Hillsdale, NJ.

Wagner, A. R. (2003). Context-sensitive elemental theory. *The Quarterly Journal of Experimental Psychology: Section B*, *56*(1), 7–29.

Wagner, A. R., & Brandon, S. E. (2001). A componential theory of Pavlovian conditioning. In R. R. Mowrer & S. B. Klein (Eds.), *Handbook of contemporary learning theories* (pp. 23–64). Earlbaum, Mahwah, NJ.

Wagner, A. R., Logan, F. A., Haberlandt, K., & Price, T. (1968). Stimulus selection in animal discrimination learning. *Journal of Experimental Psychology*, *76*(2), 171–180.

Wagner, A. R., Rudy, J. W., & Whitlow, J. W. (1973). Rehearsal in animal conditioning. *Journal of Experimental Psychology*, *97*(3), 407.

Wasserman, E., & Berglan, L. (1998). Backward blocking and recovery from overshadowing in human causal judgement: The role of within-compound associations. *The Quarterly Journal of Experimental Psychology Section B*, *51*(2), 121–138.

Wasserman, E. A. (1974). Stimulus-reinforcer predictiveness and selective discrimination learning in pigeons. *Journal of Experimental Psychology*, *103*(2), 284.

Williams, D. A. (1986). On extinction of inhibition: do explicitly unpaired conditioned inhibitors extinguish? *The American journal of psychology*, *99*(4), 515–525.

Williams, D. A., Travis, G. M., & Overmier, J. B. (1986). Within-compound associations modulate the relative effectiveness of differential and Pavlovian conditioned inhibition procedures. *Journal of Experimental Psychology: Animal Behavior Processes*, *12*(4), 351.

Wilson, C. J. (2004). Basal ganglia. In G. M. Shepherd (Ed.), *In the synaptic organization of the brain* (Fifth ed., pp. 361–413). Oxford University Press.

Witcher, E. S., & Ayres, J. J. (1984). A test of two methods for extinguishing Pavlovian conditioned inhibition. *Animal Learning & Behavior*, *12*(2), 149–156.

Witnauer, J. E., & Miller, R. R. (2011). The role of within-compound associations in learning about absent cues. *Learning & Behavior*, *39*(2), 146–162.

Woodbury, C. B. (1943). The learning of stimulus patterns by dogs. *Journal of Comparative Psychology*, *35*, 29–40.

Wu, Y., Richard, S., & Parent, A. (2000). The organization of the striatal output system: a single-cell juxtacellular labeling study in the rat. *Neuroscience Research*, *38*(1), 49–62.

Young, A. M. (2004). Increased extracellular dopamine in nucleus accumbens in response to unconditioned and conditioned aversive stimuli: studies using 1 min microdialysis in rats. *Journal of Neuroscience Methods*, *138*(1-2), 57–63.

Young, A. M., Ahier, R., Upton, R., Joseph, M., & Gray, J. (1998). Increased extracellular dopamine in the nucleus accumbens of the rat during associative learning of neutral stimuli. *Neuroscience*, *83*(4), 1175–1183.

Young, A. M., Moran, P., & Joseph, M. (2005). The role of dopamine in conditioning and latent inhibition: What, when, where and how? *Neuroscience & Biobehavioral Reviews*, *29*(6), 963–976.

Yung, K. K., Smith, A. D., Levey, A. I., & Bolam, J. P. (1996). Synaptic connections between spiny neurons of the direct and indirect pathways in the neostriatum of the rat: evidence from dopamine receptor and neuropeptide immunostaining. *The European Journal of Neuroscience*, *8*(5), 861–869.

Zimmer-Hart, C. L., & Rescorla, R. A. (1974). Extinction of Pavlovian conditioned inhibition. *Journal of Comparative and Physiological Psychology*, *86*, 837–845.

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *67*(2), 301–320.