University of California
Santa Barbara

Algorithmic and Implementational Level Models of Liking, Flexibility, and Adaptive Learning

A dissertation submitted in partial satisfaction
of the requirements for the degree

Doctor of Philosophy
in
Dynamical Neuroscience

by

Jeffrey B. Inglis

Committee in charge:

Professor Greg Ashby, Chair
Professor Wendy Meiring
Professor Jeff Moehlis

March 2022

The Dissertation of Jeffrey B. Inglis is approved.

Professor Wendy Meiring

Professor Jeff Moehlis

Professor Greg Ashby, Committee Chair

December 2021

For my parents, thank you for always encouraging my curiosity.

Acknowledgements

First and Foremost I would like to thank my family: Brian, Darlene, and Bradley, for their invaluable love and support throughout graduate school and my journey leading up to it.

Thank you to my PhD advisor Greg Ashby who has taught me a tremendous amount through countless constructive discussions and collaborations. I am immensely grateful to have worked with such a prolific yet supportive scientist.

Thank you to my collaborators Vivian V. Valentin and James Bird. Thank you to my lab mates: Paul Kovacs, Luke Rosedahl, Stella von Meer, and Yi-Wen Wang, for all of the stimulating scientific conversation. I will miss it deeply. Thank you to my research assistants Paul Tang and Jasmine Feng for all your effort on the projects we worked on together. Thank you to my committee, Wendy Meiring and Jeff Moehlis. Thank you to Anna Spickard for your advising and help with administrative work throughout the PhD.

Thank you to James Bonaiuto, Sven Bestmann, and Tom de Graaf for mentoring me during my research masters and giving me my first real academic research experience.

Thank you to my friends and family back home for their love and support. Thank you for putting up with all these years of me being away, and for all the times I came home and bragged about living in sunny California. Thank you to my friends in Santa Barbara. You have made living here unforgettable. Thank you to my roommates Matthew Varble and Niklas Griessbaum for their friendship and for all of the fun times and interesting discussions we've had over the years.

Lastly, thank you to Ishany Balder for collaborating with me, for providing feedback on this dissertation, for listening to all my bad ideas about nonviable models, and for making life good even when research was bad.

Curriculum Vitæ

Jeffrey B. Inglis

Education

2021	Ph.D. in Dynamical Neuroscience (Expected), University of California, Santa Barbara.
2016	MSc(Res) Cognitive and Clinical Neuroscience, specialization: Neuroeconomics, Maastricht University
2014	BSc Economics, Dalhousie University, Halifax, Canada

Professional Employment

ENG3: Introduction to Programming Teaching Assistant, College of Engineering. Santa Barbara, CA. Fall and Summer 2021, and Fall 2020

PSY10B: Statistics, Psychological and Brain Science. Santa Barbara, CA. Spring 2021

Health Data Science Intern, Evidation Health, Inc., San Mateo, California. Summer 2020

PSY221B: Design and Measurement Teaching Assistant, Psychological and Brain Science. Santa Barbara, CA. Winter 2021,2020, and 2019

PSY10A: Research Methods Teaching Assistant, Psychological and Brain Science. Santa Barbara, CA. Spring and Fall 2018

Publications

Inglis, J.B., Valentin, V.V., & Ashby, F.G. (2021). Modulation of dopamine for adaptive learning: a neurocomputational model. *Computational Brain & Behavior* 4 (2021), 4(1), 34–52.

Crossley, M., Ashby, F. G., & Inglis, J. B. (2021). Mathematical models of human learning. In *The new handbook of mathematical psychology.* Cambridge University Press.

Inglis, J.B., Bird, J., & Ashby, F.G. (submitted). A General Recognition Theory Model for Identifying an Ideal Stimulus.

Inglis, J.B., Ashby, F.G. (in preparation). A Neurocomputational Model of Flexible Rule Learning.

Conference Talks

Inglis, J.B. & Ashby, F.G. *A Neurocomputational Generalization of the Actor-Critic Model of Reinforcement Learning* Presented at: Math Psych. July 21st-24th, 2018. Madison, WI

Scholarships/Fellowships

University of California Santa Barbara the Regents in Dynamical Neuroscience Fellowship, UC Santa Barbara 2016-2017

Awards

Institute for Perception Student Award 2020

Editorial Review (Ad Hoc)

Psychological Review 2021

Psychological Science 2018

Field of Study

Major Field: Computational Cognitive Neuroscience

Studies in adaptive learning, flexibility, category learning, and mathematical psychology with Prof. F. Greg Ashby

Abstract

Algorithmic and Implementational Level Models of Liking, Flexibility, and Adaptive Learning

by

Jeffrey B. Inglis

Computational modeling is indispensable in the pursuit of understanding how the brain generates some of our most intimate subjective experiences, and how it solves some of the most interesting problems posed by our environment. The first model presented in this dissertation attempts to improve our understanding of how humans generate subjective liking judgements of stimuli, while two additional models are presented that attempt to improve our understanding of how humans learn adaptively and flexibly in a changing environment. This dissertation begins by introducing the theoretical foundations underlying these models and then proceeds by introducing each of them independently.

The first model is a probabilistic multidimensional model that accounts for both sensory and hedonic ratings collected from the same experiment. The model combines a general recognition theory model of the sensory ratings with Coombs' unfolding model of the hedonic ratings. The model uses sensory ratings to build a probabilistic multidimensional representation of the sensory experiences elicited by exposure to each stimulus, and it also builds a similar representation of the hypothetical ideal stimulus in this same space. It accounts for hedonic ratings by measuring differences between the presented stimulus and the imagined ideal on each rated sensory dimension. Therefore, it provides precise estimates of the sensory qualities of the ideal on all rated sensory dimensions. The model is tested successfully against data from a novel experiment.

The second model is a neurocomputational model of the flexible learning of abstract

rules. The model is constructed from highly simplified building blocks that each represent a different brain region. It implements win-stay and lose-switch signals, and it computes and represents predicted rewards. Despite its simplicity, the model gives an impressively accurate qualitative and quantitative account of some challenging behavioral and neural data.

The third model uses a network of spiking neurons to represent activity within a neural circuit that implements adaptive learning rates by modulating the gain on the dopamine response to reward prediction errors. The model generates a dopamine signal that depends on the size of the tonically active dopamine neuron population and the phasic spike rate. The model was tested successfully against results from two single-neuron recording studies and a fast-scan cyclic voltammetry study. The general applicability of the model to dopamine-mediated tasks transcend the experimental phenomena it was initially designed to address.

This dissertation concludes with the most instrumental findings that surfaced through the process of creating the three models. It will give a behind the scenes view of the process of model invention and offer some practical advice for creating computational cognitive neuroscience models.

Contents

Chapter 1

Introduction

This dissertation has been heavily influenced by two key modeling approaches; David Marr's three levels of analysis framework for modeling information systems (Marr, 1982), and the modeling principles of Computational Cognitive Neuroscience (CCN) (Ashby, 2018). Accordingly, in this introduction I will first present these two approaches. I will then introduce three research questions that motivated the invention of three novel models. I will also discuss each model's relationship to Marr's hierarchy and the CCN approach where appropriate. Finally, I will discuss how the three models presented here likely share overlapping functional relationships and underlying neural architectures.

1.1 Theoretical Foundations of Modeling

1.1.1 Marr's three levels of analysis

In 1982, David Marr proposed a framework for modeling information processing systems that has been particularly applicable to the field of cognitive neuroscience. Marr proposed that in order to understand any information processing system it needs to be understood at three loosely coupled levels of analysis: the computational level, algo-

rithmic level, and implementational level. The computational level is concerned with *what* is *computed* and *why*. The algorithmic level - referred to as the process level in mathematical psychology - is concerned with *what* is *represented* and *how* the inputs are transformed into the outputs. The implementational level is concerned with *how* the representation and algorithm are physically instantiated by the underlying hardware.

We can illustrate Marr's hierarchy by applying it to an interesting hypothetical example originally presented by Hofstadter (2007). Consider a device that resembles a typical chain of dominoes, except in this case, each domino is spring loaded so that when it falls it shoots back up after a short refractory period. In this example, the domino device takes as input a positive number and uses the various chains, loops, bifurcations, and coincidences of falling dominos to implement computations. Ultimately, the computations inside the domino device terminate and the final output domino falls if the input number is not prime and remains standing if the number is prime. Hofstadter asks the reader to consider a naive observer who stumbles upon the domino device and notices that after an input to the device, the output domino does not fall. The observer then asks why the domino does not fall. Hofstadter asks us to consider two possible answers to this question. First, one could answer that the domino does not fall because none of its neighbors fell. Obviously, the observer will be quite unsatisfied with this answer as it necessitates a follow-up question concerning why each neighbor did not fall. A second possible answer is that the domino did not fall because the number 641 is prime. Hofstadter (2007) states "641's primality is the best explanation, perhaps even the only explanation, for why certain dominos did fall and certain other ones did not fall" (p. 39).

I would offer a slight revision to Hofstader's statement - 641's primality is a component of the best explanation for understanding why the domino did not fall. Indeed, it is the best explanation at the computational level. However, it seems unlikely to me that the first question from a naive observer who stumbles upon a domino computer will be "why

did that domino not fall?". I suspect the first question would be something more similar to "what is this?". In an attempt to answer this question, an observer who is aware of Marr's framework would begin asking several questions at many levels of analysis. At the computational level the observer would ask "what does it compute?" and "why does it compute that". Once the observer discovers that the domino computer determines whether an input is a prime number and that it does this because a previous observer similar to them has requested it, then the observer would ask "what is represented by the domino computer and how does it transform the inputs (numbers) into the outputs (decisions about primality)?". At this algorithmic level the observer may conjecture a number of different primality test algorithms. Once the observer has considered a few candidate algorithms, they will begin to ask questions at the implementational level such as "how do the dominoes instantiate the representations and the algorithm?". Therefore, with Marr's framework in mind I would suggest that the best explanation of why the output domino did not fall involves all three of Marr's levels of analysis.

In accordance with the kind of theoretical pluralism advocated for by Marr's framework, this dissertation is presented as an amalgamation of models from multiple levels of analysis. However, each of these models apply to different phenomena, and thus on their own only represent a component of the complete explanation of the phenomena. In some cases, models at other levels of analysis have already been invented and these models are briefly discussed in the chapter introductions and discussions. In other cases these models do not yet exist. In the general discussion section I provide some suggestions for integrating already existing models and creating novel models.

1.1.2 Computational Cognitive Neuroscience

Technological advances have resulted in the ability to noninvasively observe brain activity in the process of collecting behavioral data. This has led to the emergence of the field of cognitive neuroscience. By studying both mental processes (the mind) and neural substrates (the brain), cognitive neuroscientists found themselves navigating the space between abstract high-level models and biophysical low-level models. Historically, data from these domains have been explained by separate models from separate fields; with mathematical psychologists modeling behavior and computational neuroscientists modeling neurobiological data. Models in the field of mathematical psychology tend to exist at the algorithmic (process) level, while models in the field of computational neuroscience tend to exist at the implementational level. The growth of empirical data on the neural underpinnings of behavior has made it possible to create implementational level models that can account for both neural and behavioral data. The field of Computational Cognitive Neuroscience emerged to meet this demand (Ashby, 2018).

There are significant advantages to modeling at the implementational level as long as the data are available to do so. First, scientists have the ability to test their models against a variety of data types - from response times and accuracy curves to single neuron recordings and anything in between (e.g. measurements of neuromodulators, neurosurgical lesions, fMRI and EEG data, etc.). Requiring models to account for a vast array of data necessarily constrains the viable parameterizations of the model (Ashby, 2018). It also expedites the process of falsifying and eliminating models (Ashby, 2018).

Second, modeling at the implementational level leads to inflexible models (Ashby, 2018). When parameters are added to models at the computational level their sole purpose is almost always to fit more data, thus leading to greater flexibility of the model. As we descend Marr's hierarchy, the addition of parameters is not solely to account

for more data. At the algorithmic level it is typically added to model an additional process, while at the implementational level it is typically added to model the physical structure of the hardware. Therefore, on average, the addition of a parameter at the implementational level has less impact on overall model flexibility than the addition of a parameter at the computational level. Furthermore, the neuroanatomy of the brain puts constraints on possible models and parameter values. For example, if the neuroscience literature suggests that there is no anatomical connection between two brain regions, then the architecture of the model cannot assume that there is one. Alternatively, if there is a connection between two brain regions and the neuroscience literature suggests that it is excitatory, then the parameter representing that connection cannot be negative. Moreover, although CCN models tend to have lots of free parameters, many of these parameters are fixed after modeling a portion of the data and would require neurobiological justification to be modified for fitting additional data (Ashby, 2018).

Third, CCN models are more likely to converge over time due to the fact that all of the models must be constrained by the underlying neuroscience (Ashby, 2018). The models may initially disagree on what constitutes the important regions, connections, neuromodulators, etc. but as models are falsified some consensus should emerge.

Finally, the CCN approach can help unite models created in distinct fields of cognitive neuroscience because in principle the models should be able to be wired together to create composite models (Ashby, 2018). Furthermore, by inspecting CCN models in diverse fields, researchers may discover that processes that once seemed to be entirely distinct are, in fact, mediated by the same brain networks (Ashby, 2018). To ensure that researchers can reap the full range of advantages offered by the CCN approach there are a number of principles that inform the construction of these CCN models. Interested readers should consult (Ashby, 2018) for details. The models presented in chapters 3 and 4 are examples of CCN models and as such they are strongly influenced by CCN

principles.

1.2 Overview of Dissertation

The three models presented in this dissertation address the following three questions: 1) What algorithm does a human use to decide how much it likes a stimulus? 2) What algorithm does an agent use to flexibly learn abstract rules and how is this implemented by the neural architecture of the brain? 3) How does the neural architecture of the brain implement an algorithm that causes an agent to modulate its learning rate in a changing environment?

1.2.1 What algorithm does a human use to decide how much it likes a stimulus?

Chapter 2 presents a General Recognition Theory (GRT) (Ashby & Townsend, 1986) model for identifying an ideal stimulus. This is an algorithmic level model in that it specifies the process by which agents generate perceived liking for a stimulus. The entities in this model are multivariate normal distributions that represent an imagined ideal stimulus or a sensory percept generated by an agent when presented with a stimulus. The model proposes that agents generate perceived liking by computing the Mahalanobis distance between a sample from a multivariate normal distribution that represents their ideal stimulus and a sample from the multivariate normal distribution that represents the sensory perception of the to-be-rated stimulus.

1.2.2 What algorithm does an agent use to flexibly learn abstract rules and how is this implemented by the neural architecture of the brain?

Chapter 3 presents a neurocomputational model for flexible rule learning. This model straddles the boundary between the algorithmic and implementational levels in Marr's (1982) hierarchy. It is neurobiologically informed in the sense that all but one of its modules are mapped onto brain regions and are linked together in a circuit that specifies the inputs and outputs of each region. Furthermore, given that this model straddles the boundary between the algorithmic and implementational level, the modeling principles of CCN were influential throughout its creation. However, because the model does not neatly fall into either the algorithmic or implementational levels, I have developed additional modeling insights that informed its construction. I will discuss these in more detail in the general discussion section.

1.2.3 How does the neural architecture of the brain implement an algorithm that causes an agent to modulate its learning rate in a changing environment?

Chapter 4 presents a neurocomputational model for the modulation of dopamine for adaptive learning (MODAL). MODAL is an implementational level model in that all of the entities in the model are represented by phenomenological models of single neurons. Given input about the current state of the environment (represented by uncertainty, volatility, contingency, or any other modulating variable) interactions between model neurons in MODAL are able to modulate the release of dopamine in the brain in response to rewards or punishments. Furthermore, given that this model is located at

the implementational level, the modeling principles of CCN were particularly influential throughout its creation.

1.2.4 Relationship among models

In addition to being linked by the underlying theory of how to understand information processing systems, the models presented in this dissertation are also linked by their function and neural architecture. MODAL was created as an implementational level model of how dopamine release is modulated in a changing environment to adapt the rate of learning. Interestingly, the experimental paradigm modelled in chapter 3 relies on a changing environment. Therefore, the flexible rule learning model presented in chapter 3 must modulate dopamine release in response to changes in its environment. In this flexible rule learning model, modulation is implemented by a very simple mathematical model. In a more detailed and fully implementational level composite model, MODAL would be used to modulate the amount of dopamine released in the neural network.

Additionally, although the GRT model for identifying an ideal stimulus is at the algorithmic level and does not rely on any reference to the brain, it is highly likely that the underlying neural architecture of a implementational version of this model would have significant overlap with the other models presented in this dissertation. Learning one's preferences over a lifetime would likely depend on the brain's dopamine system as modelled by MODAL, and on areas that encode value such as the orbitofrontal cortex and core of the nucleus accumbens. These regions are key modules in the model of flexible rule learning.

1.3 Permissions and Attributions

1. The content of chapter 2 and appendix A is the result of a collaboration with F. Gregory Ashby and James Bird. It has been submitted to Attention, Perception, & Psychophysics.

2. The content of chapter 3 is the result of a collaboration with F. Gregory Ashby.

3. The content of chapter 4 is the result of a collaboration with Vivian V. Valentin and F. Gregory Ashby, and has previously appeared in Computational Brain & Behavior (Inglis, Valentin, & Ashby, 2021). Reprinted by permission from Springer Nature Customer Service Centre GmbH: Springer Nature, Computational Brain & Behavior. Modulation of Dopamine for Adaptive Learning: a Neurocomputational Model. Jeffrey B. Inglis, Vivian V. Valentin, F. Gregory Ashby. Society for Mathematical Psychology 2020.

Chapter 2

A General Recognition Theory Model for Identifying an Ideal Stimulus

This chapter and appendix A is the result of a collaboration with F. Gregory Ashby and James Bird, which has been submitted to Attention, Perception, & Psychophysics.

2.1 Introduction

Hedonic responses about a novel object are often based on the sensory characteristics of that object. Is the color pleasing? Does the curry have the right amount of heat? A popular model of such responses, called the unfolding model, was proposed more than 50 years ago by Coombs (1964). The unfolding model assumes that when judging one's hedonic responses to a set of objects – for example, foods, beverages, or paintings – the observer imagines their ideal object within that category and then compares each object in the set to this imagined ideal. The objects are then ordered by preference according to

their similarity to the ideal. So the most preferred object is the one that is most similar to the imagined ideal and the least preferred is the one that is least similar to the ideal. The unfolding model has been generalized in a variety of different ways (e.g., Borg, 2018; DeSarbo & Rao, 1984; De Soete, Carroll, & DeSarbo, 1986; Schönemann & Wang, 1972; Zinnes & Griggs, 1974; Mullen & Ennis, 1991; Ennis, 1993; Ennis & Johnson, 1994), and applied successfully in a wide variety of different domains (e.g., Andrich, 1989; Davison, 1979; DeSarbo, Young, & Rangaswamy, 1997; Ennis & Rousseau, 2020; J. S. Roberts, Donoghue, & Laughlin, 2000).

The unfolding model provides an accurate account of preference orderings, but it is less successful at identifying the sensory characteristics associated with the imagined ideal. Some multidimensional versions of the model produce a multidimensional scaling (MDS) solution that situates each of the to-be-judged objects and the hypothetical ideal as a single point or probability distribution in a multidimensional space (e.g., De Soete et al., 1986; Zinnes & Griggs, 1974). However, as in traditional MDS, no information is provided about the nature of these dimensions. Sometimes, by noting which stimuli are situated at one extreme on a dimension and which stimuli are situated at the other extreme, it is possible to speculate about the nature of one or more dimensions. For example, if an MDS representation of odors places lemon and cedar at opposite ends of some dimension then one might infer that that dimension measures arousal. But with many dimensions, no such obvious ordering will emerge, and whatever inferences are made are generally impossible to test.

One experimental method for estimating the sensory characteristics of a stimulus, which is popular within the field of perception, is called the concurrent-ratings task. In this paradigm participants rate the magnitude of each stimulus simultaneously on a number of sensory dimensions, and then the observed ratings are used to estimate the participant's sensory, perceptual, or cognitive impressions of the stimulus (Hirsch,

Hylton, & Graham, 1982; Olzak, 1986). For example, consider an experiment in which participants first taste cups of coffee that were prepared using different amounts of ground coffee and different amounts of sugar. Next, the participants are asked to rate each cup on its sweetness and on the richness of its aroma (e.g., on a 1 to 7 scale). In this case the ratings would be used to estimate the sweetness and aroma of each cup, and these representations could be used to judge whether sweetness interacts with aroma, and to understand the psychophysical transformations from amount of sugar to perceived sweetness and amount of ground coffee to aroma.

When stimuli are rated on a single sensory dimension – most commonly sensory magnitude – the resulting data often can be modeled accurately by a signal-detection theory analysis. In fact, this is a popular experimental method for estimating a receiver operating characteristic (ROC) curve (e.g., Ashby & Wenger, in press). When ratings are collected on multiple sensory dimensions, then the percepts are multivariate, rather than univariate, so the multidimensional generalization of signal-detection theory called general recognition theory (GRT; Ashby, 1988; Ashby & Townsend, 1986) is more appropriate. This analysis assumes that (1) the unobservable perceived values have a trial-by-trial (or participant-by-participant) multivariate normal distribution across the relevant sensory dimensions, (2) the participant establishes a set of criteria or cut-points on each rated dimension that partitions that dimension into intervals, and (3) a different numerical rating is assigned to each interval (Ashby, 1988; Wickens, 1992). This model assumes that on each trial, the participant determines in which interval the percept is in on each rated dimension and then selects the associated ratings.

Ashby and Ennis (2002) combined the unfolding model and the signal detection model of the ratings task to account for simultaneous sensory and liking ratings. This model used the participant's sensory ratings to estimate the sensory representation of the ideal. However, the model was only developed and applied to situations in which the various

stimuli all varied on a single sensory dimension. This article extends the model of Ashby and Ennis (2002) to more complex real-world stimuli that vary on many sensory dimensions. The resulting model estimates the distribution of imagined ideals (i.e., across trials and participants) by identifying the ideal mean on each rated sensory dimension and estimating the variance-covariance matrix of the ideal distribution across all rated dimensions.

The new model, which we call the GRT-unfolding model, is described in the next subsection. We then describe general methods for applying the model to data from an experiment that collects ratings on multiple sensory dimensions or attributes and on some hedonic dimension, such as liking. The methods and results sections describe an empirical test of the GRT-unfolding model against data from a new experiment. Finally, we discuss implications of our results and close with some brief conclusions.

2.2 The GRT-Unfolding Model

This section develops the GRT-unfolding model. An intuitive illustration of the assumptions underlying the model is provided in Figure 2.1 for one hypothetical trial of a coffee-tasting experiment similar to the one described earlier. The only difference is that in this experiment participants are asked to rate: 1) the sweetness of the coffee; 2) the richness of the aroma; and 3) how much they like the coffee – all on a 1 to 4 rating scale. The figure depicts hypothetical events on a trial in which the participant rates sweetness and liking, but not aroma. The circle in the top panel is a contour of equal likelihood from the bivariate normal distribution that represents all possible percepts that are elicited by the specific cup of coffee that the participant tastes on this trial. The star labeled $\underline{\boldsymbol{x}}_i$ represents the specific percept experienced by the participant when tasting the current cup of coffee – that is, the specific perceived sweetness and aroma of the current cup,

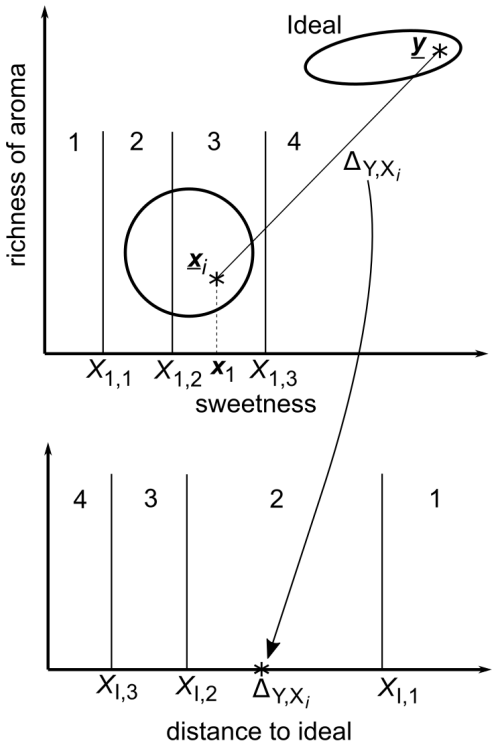

Figure 2.1: A schematic illustrating the GRT-unfolding model for a hypothetical trial in which a participant tastes a cup of coffee and then provides ratings (from 1 to 4) on the coffee's sweetness and on liking. The circle and ellipse in the top panel are contours of equal likelihood from the sensory and ideal distributions, respectively. The participant's responses on this trial are "3" on sweetness and "2" on liking.

which is the i^{th} cup of coffee in the experiment. Let $\boldsymbol{x}_1$ denote the perceived sweetness and $\boldsymbol{x}_2$ the richness of the aroma (i.e., so $\underline{\boldsymbol{x}}_i = [\boldsymbol{x}_1, \boldsymbol{x}_2]'$). The percept $\underline{\boldsymbol{x}}_i$ is assumed to be a random sample from the bivariate normal distribution that describes all possible percepts elicited by this cup. Note that the perceived sweetness of this particular cup (i.e., $\boldsymbol{x}_1$) falls in the interval assigned to a rating of 3, so in this hypothetical example, the participant responds with a rating of 3 when asked to judge sweetness.

The model assumes that if the participant had been asked to rate the richness of the coffee's aroma, rather than its sweetness, then the participant would have evaluated the position of the percept $\boldsymbol{x}_2$ relative to the positions of three criteria established on the aroma dimension (i.e., denoted $X_{2,1}, X_{2,2}$ and $X_{2,3}$, respectively). These are not shown

in Figure 2.1 to keep the figure as simple as possible.

The tilted ellipse in the top panel of Figure 2.1 is a contour of equal likelihood from the imagined ideal cup of coffee. Note that on average the imagined ideal coffee is sweeter and has a richer aroma than the current cup. The star labeled $\underline{\boldsymbol{y}}$ denotes the sensory values of the imagined ideal on this trial, which again is assumed to be a random sample from the bivariate normal distribution that describes all possible imagined ideals. So note that the model predicts that because of a variety of different sources of variability (e.g., in preference and memory), the imagined ideal changes from trial to trial. The model assumes that to respond with a liking rating, the participant imagines the ideal cup of coffee, computes the distance (or similarity) of the current cup to this imagined ideal, and then responds with a rating based on this distance, with greater distances (or lower similarities) eliciting lower levels of liking and therefore smaller ratings. In Figure 2.1, the distance falls in the interval assigned to a rating of 2 (see the bottom panel), so the participant responds with a liking rating of 2 on this trial.

More generally, consider an experiment in which participants are presented with N stimuli (one per trial) and each stimulus varies on D sensory dimensions. The goal is to collect ratings from 1 to r for each stimulus on the sensory strength for all D dimensions and on liking or some other hedonic response (with r representing maximum strength or maximum liking). In this general experiment, the GRT-unfolding model makes the following assumptions.

1) The sensory value on a trial when stimulus i is presented is represented by a $D \times 1$ random vector $\underline{\boldsymbol{x}}_i$ in which $\underline{\boldsymbol{x}}_i' = [\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_D]$, where $\boldsymbol{x}_d$ represents the sensory magnitude on stimulus dimension d. Because of stimulus and perceptual noise and individual difference, $\underline{\boldsymbol{x}}_i$ varies randomly over trials and participants. We assume $\underline{\boldsymbol{x}}_i$ has a multivariate normal distribution with mean vector $\underline{\boldsymbol{\mu}}_i$ and variance-covariance matrix Σ_i.

Note that the variance-covariance matrix Σ_i contains $D(D-1)/2$ covariances and D variances. For example, in the next section we consider an application of the GRT-unfolding model to an experiment in which participants rate the stimuli on 6 sensory dimensions. In this case, each Σ_i includes 15 covariances and 6 variances. If these are all free parameters then the model would include 27 parameters for each stimulus (15 covariances, 6 variances, and 6 means). These would require an enormous amount of data for accurate estimation. Furthermore, estimation of the covariances would require simultaneous ratings on all possible pairs of dimensions, plus the assumption that all of these ratings are based on the same sensory sample of the stimulus. Unfortunately, this assumption seems untenable. For example, if a participant is asked to rate a stimulus on 6 different dimensions then it seems likely that the participant would re-examine the stimulus one or more times before responding with all 6 ratings. According to the model, the sensory representation of the stimulus after each examination is represented by a new random sample $\underline{\boldsymbol{x}}_i$. If ratings on two dimensions are based on different $\underline{\boldsymbol{x}}_i$ samples then the correlation (e.g., across trials) between the ratings will not reflect the correlation between sensory dimensions.

For these reasons, we only consider applications of the model to experimental paradigms in which a single one of the $D+1$ ratings are requested on each trial, and each stimulus is presented to every participant on at least $D+1$ different trials to ensure that all the necessary ratings are collected. In this case, no information about covariances is available, and as a result, we assume that all covariances equal 0 and therefore that Σ_i is diagonal. Furthermore, we also assume, without loss of generality, that all variances equal 1. This just serves to set the arbitrary unit of measurement on each dimension. Collectively, these assumption mean that, for all stimuli, $\Sigma_i = \mathrm{I}$, where I is the identity matrix.

2) When asked to rate the sensory magnitude of the stimulus on dimension d, the

participant constructs $r-1$ response criteria, denoted $X_{d,1}, X_{d,2}, ...X_{d,r-1}$, and responds with rating j if and only if $X_{d,j-1} < \boldsymbol{x}_d \leq X_{d,j}$, where $X_{d,0} = -\infty$ and $X_{d,r} = \infty$. Note that in the Figure 2.1 example, the perceived value of stimulus i on dimension 1 of this hypothetical trial (i.e., x_1) lies between $X_{1,2}$ and $X_{1,3}$ and therefore the participant rates the sensory magnitude of this stimulus on dimension 1 as 3.

3) To generate a liking rating, the participant first imagines an ideal stimulus, which is represented by the $D \times 1$ random vector $\underline{\boldsymbol{y}}$. Because of variability in the imagining process (e.g., due to variability in memory and affective state) and individual difference, $\underline{\boldsymbol{y}}$ varies randomly over trials and participants. We assume $\underline{\boldsymbol{y}}$ has a multivariate normal distribution with mean vector $\underline{\boldsymbol{\mu}}_Y$ and variance-covariance matrix Σ_Y.

In the Figure 2.1 example, note that the imagined ideal distribution has greater variance on sweetness than on aroma, and that the values on these two dimensions have a slight positive correlation. The greater sweetness variance indicates that sweetness is less critical to liking than aroma because when participants imagine their ideal cup of coffee they are more consistent in their imagined aroma than in their imagined ideal level of sweetness.

4) The participant computes the Mahalanobis distance Δ_{Y,X_i} between the imagined ideal $\underline{\boldsymbol{y}}$ and the sensory value $\underline{\boldsymbol{x}}_i$ (from step 1). As we will see, this is just regular Euclidean distance except each dimension is weighted by its psychological importance to the ideal. So for example, if a participant cares more about sweetness than aroma, then the Mahalanobis distance between the current percept and the imagined ideal would weight differences between the perceived sweetness of the current cup of coffee and the imagined ideal cup more heavily than differences in aroma.

5) The participant constructs $r-1$ response criteria, denoted $X_{\mathrm{I},1}, X_{\mathrm{I},2}, ...X_{\mathrm{I},r-1}$, and responds with rating j if and only if $X_{\mathrm{I},j} < \Delta_{Y,X_i} \leq X_{\mathrm{I},j-1}$, where $X_{I,0} = \infty$ and $X_{I,r} = 0$. Note that in the Figure 2.1 example, the distance between the imagined ideal and the

perceived stimulus (i.e., Δ_{Y,X_i}) lies between $X_{\mathrm{I},2}$ and $X_{\mathrm{I},1}$ and therefore the participant responds with a liking rating of 2.

2.2.1 Fitting the Model to Data

For each stimulus, the data can be collected as a $(D+1)\times r$ matrix in which the entry in row d and column j is the frequency that participants assigned rating j to the stimulus on dimension d, where row $D+1$ is liking. Note that each matrix has $(D+1)\times(r-1)$ degrees of freedom, since there is one constraint per row (i.e., each row sum equals the number of trials that participants rated the stimulus on the attribute associated with that row). There is one such matrix for each of the N stimuli, so overall, the data include $N\times(D+1)\times(r-1)$ degrees of freedom.

The model predicts that the probability that rating j is assigned to stimulus i on sensory dimension d equals the area under the dimension d marginal pdf of $\underline{\boldsymbol{x}}_i$ between $X_{d,j-1}$ and $X_{d,j}$. Because these marginal distributions are all normal, each of these probabilities can be computed via straightforward z transformations and appeal to the cumulative z distribution function.

Computing the predicted probabilities of various liking ratings is considerably more difficult. The predicted probability that participants assign stimulus i a liking rating of j equals

$$P_{\mathrm{L}}(j|S_i) = P(X_{\mathrm{I},j} < \Delta_{Y,X_i} \leq X_{\mathrm{I},j-1}), \tag{2.1}$$

where, as before, Δ_{Y,X_i} is the Mahalanobis distance between the imagined ideal $\underline{\boldsymbol{y}}$ and the sensory value $\underline{\boldsymbol{x}}_i$. Since Δ_{Y,X_i} is nonnegative, note that

$$\begin{aligned} P_{\mathrm{L}}(j|S_i) &= P(X_{\mathrm{I},j} < \Delta_{Y,X_i} \leq X_{\mathrm{I},j-1}) \\ &= P(X^2_{\mathrm{I},j} < \Delta^2_{Y,X_i} \leq X^2_{\mathrm{I},j-1}). \end{aligned} \tag{2.2}$$

Now

$$\begin{aligned}\Delta^2_{Y,X_i} &= (\underline{\boldsymbol{y}} - \underline{\boldsymbol{x}}_i)'\Sigma_Y^{-1}(\underline{\boldsymbol{y}} - \underline{\boldsymbol{x}}_i) \\ &= \underline{\boldsymbol{w}}'\Sigma_Y^{-1}\underline{\boldsymbol{w}}, \end{aligned} \tag{2.3}$$

where $\underline{\boldsymbol{w}} = \underline{\boldsymbol{y}} - \underline{\boldsymbol{x}}_i$ is a multivariate normally distributed random vector with mean vector $\underline{\boldsymbol{\mu}}_W = \underline{\boldsymbol{\mu}}_Y - \underline{\boldsymbol{\mu}}_i$ and variance-covariance matrix $\Sigma_W = \Sigma_Y + \mathrm{I}$.

The random variable Δ^2_{Y,X_i} defined by Eq. 2.3 has the distribution of a weighted sum of D non-central χ^2 random variables, each with one degree of freedom (Scheffé, 1999). In the application described in the next section, $D = 6$, which is large enough so that this weighted sum could be considered approximately normally distributed. Therefore, to implement the normal approximation to the Eq. 2.2 probability, we need only to compute the mean and variance of Δ^2_{Y,X_i}.

The Appendix shows that the Eq. 2.3 random variable has mean

$$\mu_{\Delta^2} = D + \mathrm{trace}(\Sigma_Y^{-1}) + (\underline{\boldsymbol{\mu}}_Y - \underline{\boldsymbol{\mu}}_i)'\Sigma_Y^{-1}(\underline{\boldsymbol{\mu}}_Y - \underline{\boldsymbol{\mu}}_i) \tag{2.4}$$

and variance

$$\sigma^2_{\Delta^2} = 2D + 4\mathrm{trace}(\Sigma_Y^{-1}) + 2\mathrm{trace}(\Sigma_Y^{-2}) + 4(\underline{\boldsymbol{\mu}}_Y - \underline{\boldsymbol{\mu}}_i)'\Sigma_Y^{-1}(\mathrm{I} + \Sigma_Y^{-1})(\underline{\boldsymbol{\mu}}_Y - \underline{\boldsymbol{\mu}}_i). \tag{2.5}$$

Therefore, we can approximate the predicted probability that rating j is assigned to stimulus i on the liking dimension by computing the area between $X^2_{\mathrm{I},j}$ and $X^2_{\mathrm{I},j-1}$ under the pdf of a normal distribution with mean and variance specified by Eqs. 2.4 and 2.5, respectively.

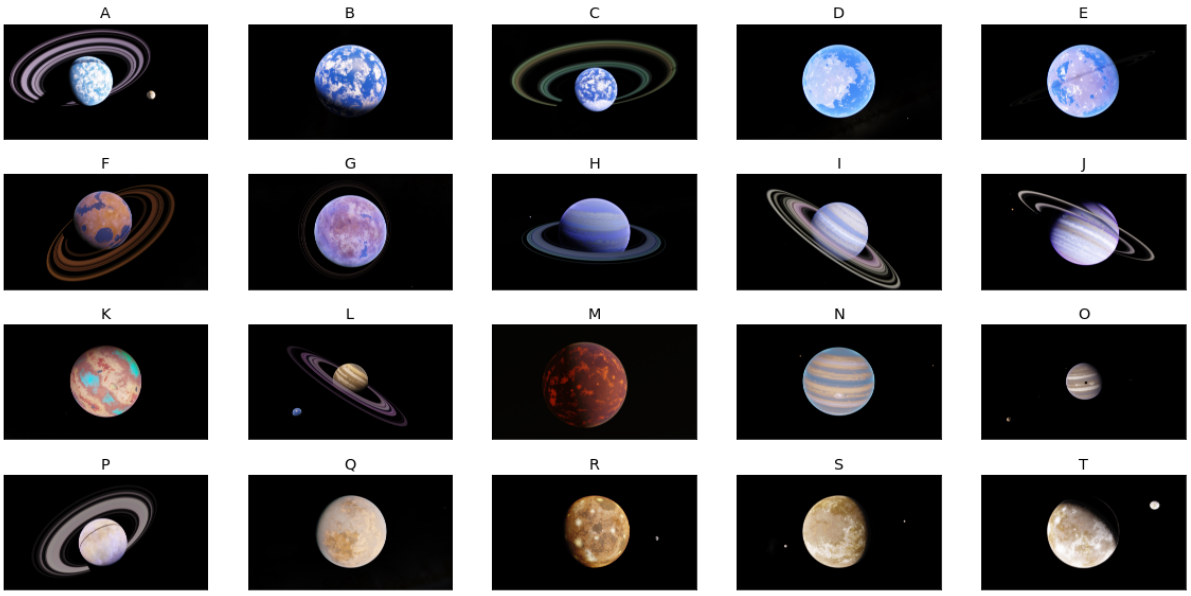

Figure 2.2: The 20 planets shown to each participant.

2.2.2 An Empirical Application

As an empirical test of the model, we ran an experiment in which 29 participants rated the 20 images of hypothetical planets shown in Figure 2.2 on six sensory dimensions and on an hedonic dimension. Specifically, participants were told to imagine that they were in a spaceship traveling through deep space, and that their mission was to rate planets they encountered (from 1 to 7) on the prominence of a number of sensory dimensions (water, clouds, rings, moons, blue-green, red-yellow) and on how important it was to retain a photograph of the planet and send it back to earth.

2.3 Methods

2.3.1 Stimuli

All images were gathered using SpaceEngine (SpaceEngine.org), a universe simulator that randomly generates a plethora of astronomical objects. The procedural generation

process creates 3-dimensional rendered planets, which are captured with extreme detail using a 3840×2160 $4K$ resolution and resulting in over 8 million pixels per image. Due to the stochastic nature of each image, the options for planetary features and combinations are nearly limitless. The stimuli used in this experiment are displayed in Figure 2.2.

2.3.2 Participants

Twenty-nine students at the University of California, Santa Barbara participated in an (approximately) one-hour experiment in exchange for course credit. All participants had normal color vision. All relevant ethical regulations were followed and the study protocol was approved by the Human Subjects Committee at UCSB. Informed consent was obtained from all participants, and every participant was allowed to quit the experiment at any time for any reason and still receive credit.

2.3.3 Procedure

Participants were told to imagine that they were in a spaceship traveling through deep space and that the ship automatically takes photos of planets that it encounters. They were also told that their mission was to rate each planet on a number of physical attributes and on how important they thought it was to send the image back to earth so that the rest of humanity would know of that planet's existence. Participants were presented the images in 5 phases. During each phase, the 20 images were displayed one-at-a-time in a random order. In phase 1, participants passively observed the images. In phases 2-5, each image was displayed with a ratings bar that ranged from 1 to 7 and participants were instructed to move the mouse and click the integer on the ratings bar that agreed with their rating. During phases 2 and 4, the image and ratings bar were accompanied by a word cue that specified the physical attribute to be rated, such as"water". Participants

rated 6 different attributes (or dimensions) and each attribute/image combination was presented once per phase, resulting in a total of 240 sensory judgments during phases 2 and 4 (2 sensory judgments per planet per dimension). During phases 3 and 5, the image and ratings bar were accompanied by the word cue "importance". Prior to each phase, participants were reminded that their job was to use the mouse to click the value on the scale that best reflected the prominence of the feature indicated by the word cue, with 1 being least prominent and 7 being most prominent. Each image was presented once during phases 3 and 5, resulting in 2 importance judgments per planet.

A few participants were not sufficiently engaged in some phases of the experiment. These participants tended to repeat the same rating, over and over. Therefore, any importance phase (3 and 5) in which the participant emitted 3 or fewer unique ratings was excluded from analysis. Six liking phases were excluded, leaving 52 liking phases for analysis. Additionally, any sensory phase (2 and 4) in which the participant gave the same rating on any dimension to all images was excluded from analysis. One sensory phase was excluded resulting in 57 sensory phases for analysis.

2.4 Results

The data from this experiment were aggregated across participants and then recorded in a 20 (planets) $\times$ 7 (dimensions) $\times$ 7 (ratings) frequency array, where importance was included as one of the 7 dimensions. The importance ratings for each planet are shown in Figure 2.3. For each planet and dimension, the frequency sum across the 7 ratings equals the number of trials participants were asked to rate that planet on that dimension. Therefore, the data include 6 degrees of freedom for each planet and dimension, and so the entire data set includes 840 degrees of freedom (i.e., $20 \times 7 \times 6$).

The GRT-unfolding model was fit to these data. The model included a total of 183

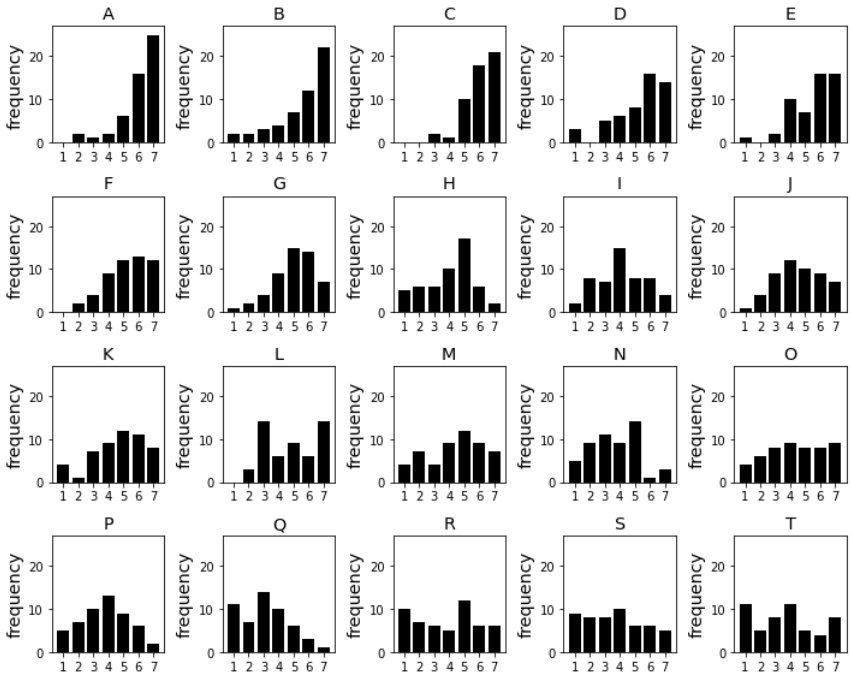

Figure 2.3: Importance ratings for each of the 20 planets.

free parameters. Without loss of generality, we fixed the mean vector for planet N, $\underline{\boldsymbol{\mu}}_N$, to the zero vector. In addition, as described earlier, to limit the number of free parameters we fixed the variance-covariance matrices of all sensory distributions to $\Sigma_i = \mathrm{I}$. The following parameters were all free to vary:

1) The remaining 19 mean vectors, $\underline{\boldsymbol{\mu}}_i$ for all $i \neq N$. Each $\underline{\boldsymbol{\mu}}_i$ is 6×1, so there were a total of 114 free mean parameters (i.e., 19×6).

2) Six criteria, $X_{d,j}$, on each of the 6 sensory dimensions, resulting in an additional 36 parameters.

3) Six means for the ideal distribution, $\underline{\boldsymbol{\mu}}_Y$,

4) The 6×6 ideal variance-covariance matrix, Σ_Y (21 free parameters).

5) Six criteria, $X_{\mathrm{I},j}$, on the squared-distance-to-ideal dimension.

All parameters were estimated via constrained optimization by linear approximation (COBYLA; Powell, 1994) using SciPy (Virtanen et al., 2020) in Python (Van Rossum

& Drake, 2011) by minimizing the sum of squared errors between the predicted and observed response frequencies.

Overall, the GRT-unfolding model accounted for 95.27% of the variance in the data (r^2). Although the model included 183 free parameters, because the data had 840 degrees of freedom, after parameter estimation, there were still 657 degrees of freedom left to test the model (i.e., $840-183$). So accounting for 95% of the variance in these 657 proportions seems impressive. Not surprisingly, however, the model was more successful at accounting for the sensory ratings than the liking ratings. Specifically, the GRT-unfolding model accounted for 96.09% of the variance in the sensory ratings data and 73.67% of the variance in the liking ratings.

Figure 2.4 shows estimated sensory distributions for each planet on each dimension as well as the estimated criteria. Note that, except for rings, the planets vary fairly continuously on all sensory dimensions. Not surprisingly, the perceived prominence of rings is approximately bimodal with some planets displaying prominent rings (e.g., planets A, C, and F) and other planets showing a prominent absence of rings (e.g., planets B, D, and S).

Table 2.1 shows the variance-covariance matrix of the estimated ideal distribution. The variances provide an inverse measure of how important each dimension is to the ideal. Note that the smallest variance is on the clouds dimension and the next smallest is on water. The small variances suggest that when ideal planets are imagined on different trials, participants always tend to imagine a planet with similar values on the water and cloud dimensions. In contrast, the variances on the red-yellow and moons dimensions are large, suggesting that the different imagined ideals vary widely on the red-yellow and moons dimensions. Therefore, for example, if the imagined ideal sometimes has a moon and sometimes does not, then the presence or absence of a moon is not an important attribute of the ideal planet.

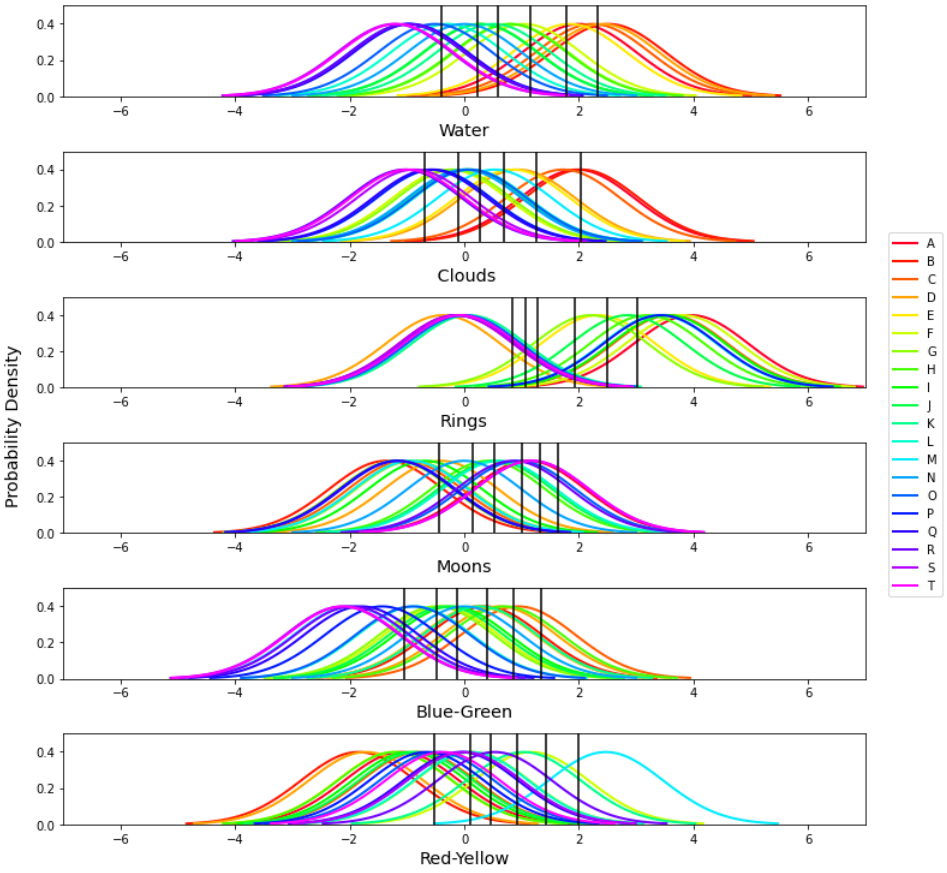

Figure 2.4: Estimated sensory distributions from the best-fitting version of the GRT-unfolding model, along with the estimated criteria on each dimension that participants used to assign ratings.

Figure 2.5 shows the ideal distribution and the mean of each planet distribution projected onto the plane defined by the two most important sensory dimensions – namely, water and clouds. The ellipses denote the contours of equal likelihood of the ideal distribution, so the ideal mean lies at the center of these ellipses. Note from Table 2.1 that water and clouds are negatively correlated in the ideal distribution, which is the reason that the ellipses in Figure 2.5 have a negative orientation. This makes sense because as cloud cover increases there is less available surface to display water. Note that planet B is closest to the ideal mean, closely followed by planets A and C, and that planets R, S, and T are furthest (i.e., see Figure 2.2 for ordering relative to the ideal when consid-

Table 2.1: Variance-covariance matrix of the ideal distribution from the best-fitting version of the GRT-unfolding model.

	Water	Clouds	Rings	Moons	Blue-Green	Red-Yellow
Water	33.65	-20.46	33.27	29.46	30.45	-10.09
Clouds		23.09	-23.32	-10.77	1.46	1.78
Rings			191.14	245.11	83.73	21.00
Moons				558.85	113.26	66.89
Blue-Green					145.61	20.25
Red-Yellow						214.29

ering all dimensions). Therefore, these data suggest that the ideal planet would have a greater prominence of cloud cover and water than any of the planets that were shown to participants.

2.5 Discussion

The GRT-unfolding model uses sensory ratings to build a probabilistic, multidimensional representation of the sensory experiences elicited by exposure to each stimulus. If participants rate the stimuli on D sensory dimensions then the sensory representations built by the model will be D dimensional. And the model will also build a representation of the ideal stimulus in this same space. It then attempts to account for hedonic ratings by measuring differences between the presented stimulus and the imagined ideal on each of these D sensory dimensions. This approach can only hope to account for the hedonic responses of participants if the rated sensory dimensions include all stimulus attributes that significantly affect the hedonic response. To take an extreme example, consider an experiment in which participants rate a set of stimuli on D sensory dimensions but that the participants' hedonic responses to those stimuli depend exclusively on some other, unrated sensory dimension. In this case, their hedonic responses will be independent of

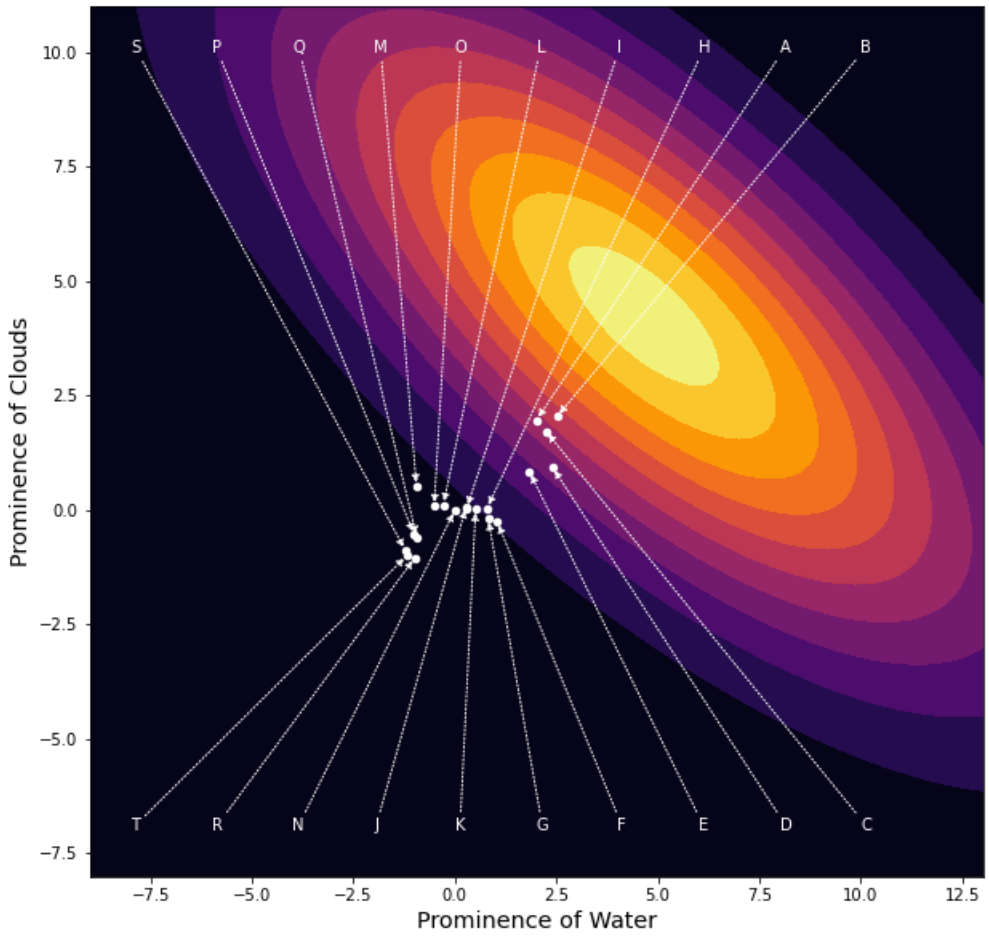

Figure 2.5: Contours of equal likelihood of the ideal distribution from the best-fitting version of the GRT-unfolding model (i.e., the ellipses) and the sensory means of each planet projected onto the plane defined by the water and cloud dimensions.

the stimulus value on any of the rated dimensions, and therefore a comparison of the stimulus to the ideal values on the D rated dimensions will not predict the participant's hedonic response. So the efficacy of the GRT-unfolding model depends strongly on the ability of the experimenter to identify all sensory dimensions that could significantly affect the hedonic responses of participants to the selected stimuli.

Given this, the default expectation should be that the model will account for sensory ratings better than it accounts for hedonic ratings. In the experiment described here, the GRT-unfolding model accounted for 96% of the variance in the sensory ratings and 74%

of the variance in the hedonic ratings. Therefore, we believe that one plausible account for this difference is that participants based their hedonic responses, at least in part, on some unrated dimension or attribute of the planets. Traditional multidimensional unfolding models that lack any sensory data could just add more unspecified dimensions to the model until goodness-of-fit is maximized (e.g., exactly as in MDS). Even so, note that a better fit by such a model would provide only vague information about the sensory qualities of the ideal. Given that the GRT-unfolding model provides precise estimates of the sensory qualities of the ideal on all rated sensory dimensions, we believe that accounting for 74% of the variance in the hedonic ratings is impressive, especially since the model was provided no information about how participants might make these judgments.

As described earlier, a number of multivariate generalizations of the unfolding model have been proposed (Zinnes & Griggs, 1974; De Soete et al., 1986; Mullen & Ennis, 1991; Ennis, 1993; Ennis & Johnson, 1994). Despite the theoretical value of these models, Ennis and Ennis (2013) suggested three reasons why these generalizations have not had a greater practical impact. First, the models require pairwise comparisons that can be expensive to obtain (e.g., "which do you prefer, A or B?"). The GRT-unfolding model avoids this criticism because it only requires hedonic and sensory ratings on single stimuli. For example, with the 20 planets used in our experiment, pairwise comparisons would require collecting ratings on 190 different pairs (i.e., $(20 \times 19)/2$), whereas the GRT-unfolding model only requires ratings on the 20 individual planets. Therefore, the GRT-unfolding model requires far fewer trials than previous models, and the data it does require can be readily collected remotely via any of several widely available current software packages.

Second, Ennis and Ennis (2013) noted that previous models are mathematically complex, which makes them difficult to apply. In contrast, the GRT-unfolding model is simple to apply due to the normal approximation to the squared Mahalanobis distance between

the sensory representation of the stimulus and the imagined ideal. This approximation reduces the complexity of the model significantly since the probability of the various sensory and liking ratings can be computed via straightforward z transformations and appeal to the cumulative z distribution function. Hopefully, these advantages will lead to more applications of the GRT-unfolding model to academic and industry data sets.

Third, Ennis and Ennis (2013) noted that another reason that previous multidimensional unfolding models are not more popular is because they do not generate individual-level ideal representations. This is largely due to the enormous amount of data they require (e.g., because they rely on paired-comparison experiments). As we just noted, the GRT-unfolding model requires much less data and therefore is much less susceptible to this problem. Nevertheless, because the GRT-unfolding model requires ratings on each identified sensory dimension, the amount of data it requires increases (linearly) with the number of rated sensory dimensions. Therefore, whereas individual ideal representations should be straightforward to estimate in applications where only a few sensory dimensions require ratings, estimating individual ideal representations is more problematic when many sensory dimensions are required. For example, in our empirical application to planets, we collected ratings on 6 sensory dimensions, which was too many to allow the model to be fit to individual participant data when each participant completed only a single 50-minute experimental session.

If individual ideal representations are desired, then there are several options. One, of course, is to collect sufficient data from each participant to allow the model to be fit to individual-participant data – either by increasing the length of the experimental session or increasing the number of sessions. A second option is to reduce the number of rated sensory dimensions, which would increase the number of ratings that could be collected on each dimension in a single session. The trick here is to eliminate dimensions that do not affect the participant's hedonic response. One approach might be to run an initial

group experiment with many dimensions, fit the model to the group data, and then use these results to identify the key sensory dimensions. For example, the variances listed in Table 2.1 indicate that in our experiment, Moons had little or no effect on hedonic ratings, and Rings and Red-Yellow had at most a minimal effect. Therefore, a follow-up experiment that asked for ratings only on the Water, Clouds, and Blue-Green sensory dimensions might be able to collect enough data to allow estimation of individual ideals at the cost of only a minimal decrease in goodness-of-fit. Another approach to reducing the number of sensory dimensions is to consult someone with expertise with the stimuli (e.g., a Master Sommelier in the case of wines).

Finally, a third option is to estimate an ideal representation, not for individual participants, but for groups of similar participants. This requires a separate experiment for each identified group, but each participant in these experiments only needs to complete a single experimental session. This approach seems especially relevant for product design, since industries do not create unique products for each individual, but they might create a product that is tailor-made for one particular segment of consumers.

As an empirical test of the GRT-unfolding model, we chose the planets shown in Figure 2.2 because they are interesting, real-world objects. However, the GRT-unfolding model could be applied to any stimuli. Because the model estimates the sensory values of the ideal stimulus, it has the potential to greatly benefit product development. In many cases, the sensation elicited by a stimulus on an identified sensory dimension is directly related to some underlying physical quantity. For example, the sweetness of a Merlot wine is related to its residual sugar content (among other factors). Therefore, identifying the ideal sweetness of a Merlot could facilitate the efforts of vintners to create more popular wines.

Chapter 3

A Neurocomputational Model of Flexible Rule Learning

The content of chapter 3 is the result of a collaboration with F. Gregory Ashby.

3.1 Introduction

Humans have a remarkable ability to adapt flexibly to changes in the environment. This flexibility is crucial for successful performance in tasks that are diverse as choosing the appropriate hunting grounds based on prey availability to abiding by appropriate social norms in different social settings. Flexible decision making has been studied extensively and significant progress has been made in understanding how flexible behavior is implemented in the brain.

The standard paradigm for investigating behavioral flexibility is reversal learning (A. Roberts, 2006; Izquierdo, Suda, & Murray, 2004; Murray & Wise, 2010; Dias, Robbins, & Roberts, 1996; Rygula, Walker, Clarke, Robbins, & Roberts, 2010; Monosov & Rushworth, 2021). In typical reversal-learning experiments, the animal must initially

learn stimulus-reward associations, which at some point then reverse. Often only one reversal occurs but some paradigms include repeated reversals.

Many human behaviors are rule guided – that is, they depend on a set of explicit instructions that can be generalized to a variety of different stimuli or scenarios (e.g., counting to add two numbers). Flexibility with respect to rule-guided behaviors is also critical. Perseverating with an old rule after an environmental change can have deadly consequences.

The goal of this article is to better understand how flexibility in rule-guided behaviors is implemented at the neural level. Toward this end, we propose a novel neurocomputational model of flexibility in rule learning and use. The model is constructed from highly simplified building blocks that each represent a different brain region and that implement win-stay and lose-switch signals, as well as compute and represent predicted rewards. However, as we will see, despite its simplicity, the model gives an impressively accurate qualitative and quantitative account of some challenging behavioral and neural data.

3.1.1 The Wisconsin Card Sorting Test (WCST) and its Analogs

The Wisconsin Card Sorting Test (WCST; Heaton, 1981) is among the most widely used experimental paradigms for studying the flexibility of rule-guided behaviors. The WCST is a well-known neuropsychological assessment that was designed originally to detect frontal dysfunction (e.g., Kimberg, D'Esposito, & Farah, 1997). Stimuli in this task are cards containing geometric patterns that vary in hue, shape, and the number of symbols that are depicted. The participant's task is to use trial-by-trial feedback to learn to assign each card to its correct category. In all cases, the correct strategy is a simple one-dimensional rule (e.g., choose the response that matches the hue of the symbols

on the stimulus card). The paradigm assesses flexibility because at certain times, and without any warning to the participant, the correct classification rule changes.

Simplified versions of the WCST have been developed for use with nonhumans – especially nonhuman primates (Mansouri & Tanaka, 2002; Mansouri, Matsumoto, & Tanaka, 2006; Buckley et al., 2009). In these WCST analogs, the animal must use feedback to switch among alternative abstract rules. Critically, no cues are presented that signal a change in the rewarded rule and these rules are abstract, that is, they are independent of the features of the stimuli. Therefore, this task is similar to traditional reversal-learning paradigms, in that the rewarded abstract rule reverses at some point during the experiment, but it differs in that it does not rely on associative learning.

An example of a WCST analog that was reported by Buckley et al. (2009) is illustrated in Figure 3.1. On each trial of this experiment, one of two possible abstract rules was active (rewarded). The animal's task was to match either the shape or the hue of the sample. There were 36 stimuli, each of which was constructed from one of 6 possible hues and one of 6 possible shapes. At the beginning of each trial, the animal was presented with a single randomly selected stimulus in the center of the screen, called the sample. After the animal touched the sample, 3 additional stimuli were then presented. One of these had the same hue as the sample, another had the same shape, and the third did not match the sample on hue or shape. If the animal selected the correct stimulus – that is, the stimulus that matched the sample on hue or shape depending on which rule was active, then a food reward was delivered and the correctly chosen stimulus remained on the screen for 1 second. If the animal chose incorrectly then no reward was delivered, the stimuli were removed from the screen, and a white circle remained on the screen for 1 second. The intertrial interval lasted 12 seconds following an error and 6 seconds following a correct response. The active rule changed when the animal reached an accuracy of 85% correct on the previous 20 trials.

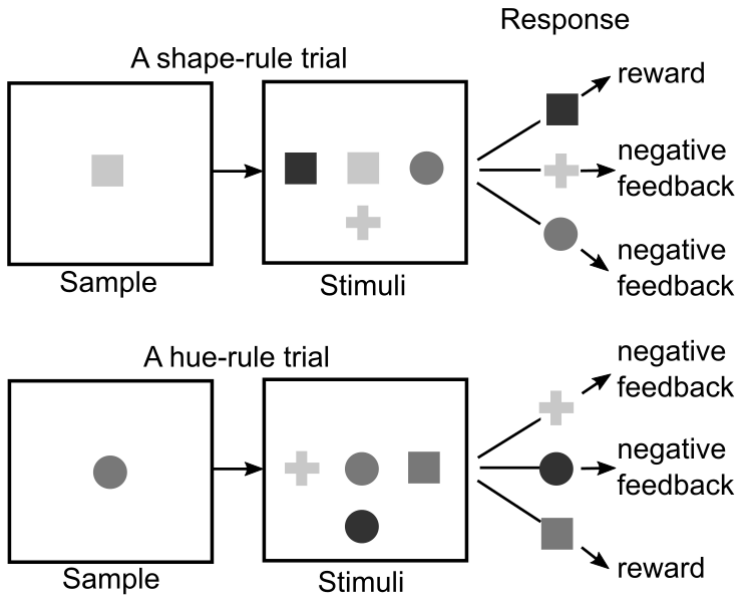

Figure 3.1: WCST analog. See text for details. Figure is adapted from Buckley et al. (2009).

Buckley et al. (2009) trained 14 macaque monkeys in this task. Each animal was initially trained for 15 daily sessions of 300 trials each. After these preoperative sessions, the animals were split into a variety of different groups. For our purposes, the most important groups were the dlPFC group ($n = 3$), which received aspiration lesions bilaterally to the principal sulcus within the inferior dorsolateral prefrontal cortex, the ACC group ($n = 4$), which received aspiration lesions bilaterally to the anterior cingulate cortex sulcus within the medial frontal cortex, the OFC group ($n = 3$), which received aspiration lesions bilaterally to the orbitofrontal cortex, and the control group ($n = 6$), which were not operated on.[1] For detailed information on all aspects of the experiment, see the supplementary information from Buckley et al. (2009).

[1]There were also two other groups. One received aspiration lesions bilaterally to the superior dorsolateral prefrontal cortex, and the other received bilateral lesions to the ventrolateral prefrontal cortex. We did not model results from either of these groups. We excluded the former group because there was no effect of these lesions on any performance measures, and the latter group was excluded because these animals only completed a task in which the rule changed at the beginning of each daily session and did not change within a session. Furthermore, although Buckley et al. (2009) stated that ventrolateral PFC lesions impaired the ability of the animals to implement rules they had learned during preoperative training, no data documenting this impairment were provided that we could model.

Buckley et al. (2009) reported detailed behavioral results from all these different groups, which included the total number of rule switches, the proportion of perseverative errors, the average proportion of perseverative errors in the three trials following each rule change, the proportion correct after an error, and the proportion correct after a string of consecutive correct responses (where the length of the string could take any value from 1 to 7). As a result, these data form an excellent testbed for any neurobiologically oriented model of flexible rule learning, and later in this article, we will test the validity of the new model we propose by examining its ability to account for this detailed data set.

3.2 A Neurocomputational Model of the WCST

This section describes a new neurocomputational model of the context-dependent rule learning that occurs in the WCST and especially in the simplified version illustrated in Figure 3.1.

The general problem for any agent trying to perform well in the Figure 3.1 task is as follows. On any trial, the agent can choose one of two possible categorization rules – either select the stimulus that matches the shape of the sample or matches the hue.[2] Call the rule in which the agent matches the shape M_S and the rule in which the agent matches the hue M_H. Also on each trial, the environment can be in one of two possible contexts – a reward is delivered if the agent selects the stimulus that matches the sample on either shape or hue. Call the former context C_S and the latter C_H. So the agent must learn which context is active on each trial and which rule is optimal in each context.

Figure 3.2 describes the architecture of the model, which includes five modules; the Spread-of-Effect Learning (SEL) module, the Context-Dependent Rule Learning (CDRL)

[2]Logically, other rules are possible. For example, the agent could guess randomly, or choose the stimulus that does not match the sample on hue or shape. However, in the Buckley et al. (2009) experiment, such choices were rare – even during pre-op training.

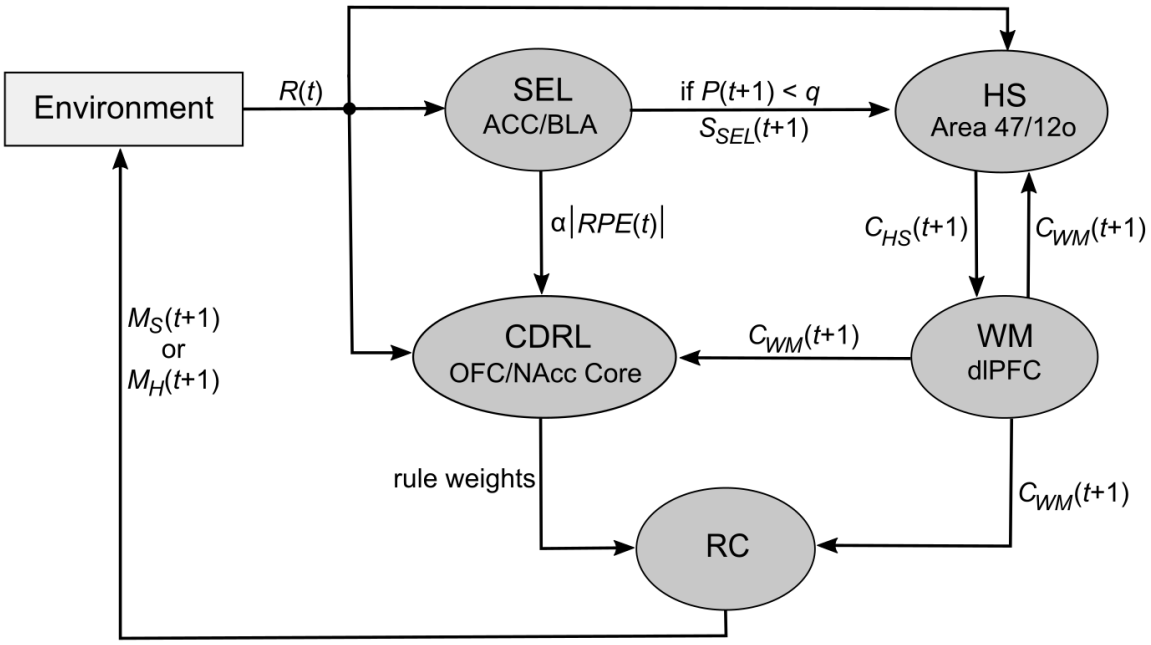

Figure 3.2: Model Architecture. SEL = spread of effect learning module, HS = heuristic strategy module, CDRL = context-dependent rule learning module, WM = working memory module, RC = rule choice module, ACC = anterior cingulate cortex, BLA = basolateral amygdala, Area 47/12o = an area in ventolateral prefrontal cortex, OFC = orbitofrontal cortex, NAcc = nucleus accumbens, dlPFC = dorsolateral prefrontal cortex, $R(t)$ = obtained reward on trial t, $M_S(t+1)$ = a choice to match the shape of the trial $t+1$ sample, $M_H(t+1)$ = a choice to match the hue of the trial $t+1$ sample. See the text for a description of the other inputs and outputs.

module, the Heuristic Strategy (HS) module, the Working Memory (WM) module, and the Rule Choice (RC) module.

The model uses the feedback from the current trial to: 1) select a strategy (either stay with the current context, or switch contexts), 2) estimate the context that will be active during the next trial, 3) update the weights associated with each rule in both contexts, and 4) select the rule to apply on the next trial. In brief, the HS module learns which context is currently active, and the CDRL module learns which rule to apply in each context. The SEL module computes the predicted reward on each trial and attempts to bias the HS module's decision about which context is currently active. The WM

module holds the current context in working memory during all delay periods, and the RC module uses the current context to select and implement a categorization rule.

The model straddles the boundary between the algorithmic and implementational levels in Marr's (1982) hierarchy. It is neurobiologically informed in the sense that all but one of its modules are mapped onto brain regions and are linked together in a circuit that specifies the inputs and outputs of each region.

The following subsections describe each module in detail. The main focus is to describe the computations performed by each module. We also give a brief justification of our neurobiological assumptions. However, these latter assumptions are considered in more detail in the general discussion.

3.2.1 Spread of Effect Learning (SEL) Module

The SEL module tracks reward according to the Rescorla-Wagner rule:

$$RPE(t) = R(t) - P(t), \tag{3.1}$$

$$P(t+1) = P(t) + \lambda\, RPE(t), \tag{3.2}$$

where $RPE(t)$ is the reward prediction error on trial t, $R(t)$ is the obtained reward on trial t, $P(t)$ is the predicted reward on trial t, and λ is the learning rate. We assume that $R(t) = 1$ on rewarded trials and $R(t) = 0$ on non-rewarded trials. As a result, note that $P(t)$ can also be interpreted as the predicted probability of reward.

The SEL module has two outputs. The first, which is input to the HS module, is a strategy signal, $S_{SEL}(t+1)$, that recommends either staying with the current context or switching contexts on the subsequent trial (i.e., on trial $t+1$). Following the feedback

received on trial t, the SEL module computes $P(t+1)$ according to Eq. 3.2. If $P(t+1) < q$ then the module sets $S_{SEL}(t+1) = stay$ with probability $P(t+1)$ and $S_{SEL}(t+1) = switch$ with probability $1-P(t+1)$, and projects this value to the HS module, where it attempts to bias ongoing decision making. If $P(t+1) \geq q$ then there is no output from the SEL module to HS. The strategy signal $S_{SEL}(t+1)$ models a spread of effect, such that decisions to stay or switch are made by integrating historical rewards, rather than solely on the basis of feedback from the most recent trial. The length of the memory for past rewards is determined by λ of Eq. 3.2.

The second output of the SEL module, which is sent to the CDRL module, is $\alpha\,|RPE(t)|$ – that is, the absolute value of the RPE modulated by a learning rate α. This value, which measures the surprise of the trial t outcome, is used by the CDRL module to update the rule weights. The SEL module has three free parameters (q, λ, and α; $P(0) = 0.5$ was fixed prior to fitting the model).

We assume the SEL module is mediated within the BLA and the ACC. The amygdala has been shown to play a role in a number of computational processes that are of relevance to flexible behavior in the WCST, including noncontingent learning (Jocham et al., 2016; Chau et al., 2015) and the encoding of lose-switch signals (Chau et al., 2015). Furthermore, the amygdala also plays a central role in learning from surprise (Pearce-Hall learning; Holland & Schiffino, 2016; Roesch, Esber, Li, Daw, & Schoenbaum, 2012). For example, learning rate changes in response to surprise were eliminated by BLA lesions (Stolyarova & Izquierdo, 2017). In addition, lesions to the BLA facilitate reversal learning in rats (Izquierdo et al., 2013).

The assumption that the SEL module learns according to the Rescorla-Wagner rule is consistent with the role of the amygdala in Pavlovian conditioning. One output of the SEL module is the α weighted absolute RPE which is later used by the CDRL module (OFC-NAc core synapses) to gate the amount of learning that occurs at OFC-NAc core

synapses. This is consistent with results showing that the amygdala may be tracking the value of associability, which is used to gate future learning (Li, Schiller, Schoenbaum, Phelps, & Daw, 2011; Holland & Schiffino, 2016; Roesch et al., 2012). Additionally, some evidence suggests that the ACC represents environmental volatility (Behrens, Woolrich, Walton, & Rushworth, 2007), which some models assume modulates the learning rate (Mathys, Daunizeau, Friston, & Stephan, 2011). The ACC also has a role in integrating feedback over time (Kennerley, Walton, Behrens, Buckley, & Rushworth, 2006). Furthermore, the amygdala has been shown to encode unsigned prediction errors that are then propagated to ACC where these errors are equipped with a sign and propagated back to the amygdala (Klavir, Genud-Gabai, & Paz, 2013). The other output of the SEL module is the strategy signal, $S_{sel}(t+1)$. This is consistent with findings that the BLA encodes win-stay, lose-shift signals (Chau et al., 2015). Importantly, however, Chau et al. (2015) reported that these BLA win-stay, lose-shift signals seemed to be driven primarily by the lose-shift signal and only on trials when accuracy was below 70 percent (averaged over the previous 5 trials). To model this finding we set the threshold, q, to 0.7 and the win-stay, lose-shift signal generated by the SEL module is only transmitted to the HS module on trials in which predicted accuracy does not exceed q.

3.2.2 Context Dependent Rule-Learning (CDRL) Module

The CDRL module learns which rule to apply in each context by tracking rewards using a Pearce-Hall/Rescorla-Wagner hybrid model that updates rule values within the context that the WM module assumes is currently active. The Pearce-Hall component of the model enables the CDRL module to use surprise to modulate the amplitude of prediction errors, where surprise is defined as the absolute value of the RPE from the previous trial as computed by the SEL module (modulated by α; so surprise equals

$\alpha\,|RPE(t-1)|$).

Let $W_{IJ}(t)$ denote the weight of rule M_I ($I = S$ or H) in context C_J ($J = S$ or H) on trial t. Without loss of generality, we assume that $W_{SJ}(t) + W_{HJ}(t) = 1$ for $J = S$ and for $J = H$, and for all t. We set the initial weights to $W_{SS}(0) = W_{HH}(0) = 0.7$ to reflect the fact that prior to surgery, the macaques underwent extensive training in which they were able to learn the value of the rules in each context (see supplementary materials in Buckley et al., 2009).

If the WM module predicts that the current context is $C_{WM}(t) = C_S$ then upon receiving feedback, the CDRL module updates the rule weights as follows:

$$W_{SS}(t+1) = W_{SS}(t) + R(t)\,\alpha\,|RPE(t-1)|\,[1 - W_{SS}(t)], \tag{3.3}$$

and

$$W_{HH}(t+1) = W_{HH}(t) - \beta\,W_{HH}(t). \tag{3.4}$$

Alternatively, if WM predicts that the current context is $C_{WM}(t) = C_H$ then upon receiving feedback the CDRL module updates the weights as follows:

$$W_{HH}(t+1) = W_{HH}(t) + R(t)\,\alpha\,|RPE(t-1)|\,[1 - W_{HH}(t)], \tag{3.5}$$

and

$$W_{SS}(t+1) = W_{SS}(t) - \beta\,W_{SS}(t). \tag{3.6}$$

In other words, the rules associated with the active context are updated via a Pearce-Hall/Rescorla-Wagner hybrid rule, but only when the reward is positive (because $R(t) = 0$ on error trials).[3] Therefore, negative reward is interpreted as evidence that the agent's

[3]These equations assume that $R(t)$ always equals 0 or 1 because reward magnitude was not varied in the Buckley et al. (2009) experiment. In applications to experiments with variable reward magnitudes,

assumption about the active context is wrong rather than its choice about the rule within that context. This is reasonable given the extensive prior training that allowed the animals to learn that both rules are valuable if used in the appropriate context. The weight of the correct rule associated with the inactive context decays at rate β. In the Buckley et al. (2009) experiment the active rule changes fast enough that $W_{SS}(t)$ and $W_{HH}(t)$ rarely decay to a value below 0.5 (the weights fell below 0.5 on 0.01%, 0.025%, 0.18%, and 0.006% percent of trials for the Control, OFC, ACC, and DLPFC lesion groups, respectively). In an experiment in which the environment is more stable, equations 3.4 and 3.6 should be modified so that when the context is inactive for a sufficiently long period the weights converge to 0.5. This would imply that the agent has forgotten the association between the rules and the inactive context.

The outputs of the CDRL are the four rule weights (two for each context). The CDRL module has two free parameters [β and the value of $W_{SS}(0) = W_{HH}(0)$].

The CDRL can be viewed as representing an actor-critic architecture. The actor is implemented by OFC-NAc core synapses, whereas the critic is implemented by a separate neural network that generates the dopamine required for strengthening the OFC-NAc core synapses. Support for the hypothesis that circuits within OFC and NAc core represent the actor comes from studies suggesting that these regions play a role in encoding prediction errors (Hart, Rutledge, Glimcher, & Phillips, 2014; Rutledge, Dean, Caplin, & Glimcher, 2010), value (Schoenbaum, Takahashi, Liu, & McDannald, 2011; Corbit, Muir, & Balleine, 2001), and rules (Sleezer, Castagno, & Hayden, 2016). Previous models also proposed that mappings between values and states are encoded in the PFC-NAc synapses (Takahashi, Schoenbaum, & Niv, 2008) and that OFC plays a role in separating states, such that only rules within the active context are updated (Wilson, Takahashi,

however, $R(t)$ can be any value on the real line. As a result, in such applications, Eqs. 3.3 and 3.5 would need to be modified so that the second term is positive on all reward trials and 0 on negative reward trials.

Schoenbaum, & Niv, 2014). We previously proposed a neurocomputational model of the critic, called the Modulation Of Dopamine For Adaptive Learning (MODAL) network (Inglis et al., 2021). Computations carried out by MODAL rely on the ventral subiculum, NAc shell, ventral pallidum, pendunculopontine nucleus, and lateral habenula.

3.2.3 Heuristic Strategy (HS) Module

The HS module learns which context is active by implementing a win-stay, lose-shift strategy, which is subject to interference from the strategy input that comes from the SEL module [i.e., from $S_{SEL}(t+1)$].

If the agent's trial t response is rewarded (i.e., if $R(t) = 1$) then the HS module initially sets the strategy for the next trial to $S_{HS}(t+1) = stay$. Alternatively, if $R(t) = 0$ then HS sets $S_{HS}(t+1) = switch$. In addition to implementing this win-stay, lose-shift strategy, the HS module must also suppress input from the SEL module that is possibly task irrelevant. In other words, input from the SEL module could change the value of $S_{HS}(t+1)$. The final value of $S_{HS}(t+1)$ remains at its initial value with probability h and changes to $S_{HS}(t+1) = S_{SEL}(t+1)$ with probability $1-h$. Note that h models the ability of the HS module to suppress input from the SEL module; if $h = 1$ then SEL input is completely suppressed and the HS module is in full control of the strategy choice. Alternatively, if $h = 0$ then SEL input overwhelms the HS module and takes full control of the strategy choice. Additionally, if there is no output from SEL (trials in which $P(t) \geq q$) then the HS module has full control over the strategy decision.

If the final value $S_{HS}(t+1) = stay$ and $C_{WM}(t) = C_S$ (C_H) then the module sets $C_{HS}(t+1) = C_S$ (C_H). Alternatively, if $S_{HS}(t+1) = switch$ and $C_{WM}(t) = C_S$ (C_H) then HS sets $C_{HS}(t+1) = C_H$ (C_S). Finally, the output of the HS module, $C_{HS}(t+1)$, projects to the WM module, where it is held throughout the delay period (intertrial

interval) until a decision is required. The HS module has one free parameter (h).

The HS module generates win-stay, lose-shift signals about the currently active context and uses them to generate a prediction for the context that will be active on the subsequent trial. Area 47/12o has been shown to encode win-stay, lose-shift signals in object discrimination paradigms (Chau et al., 2015), ventral area 12 has been shown to display switching signals (Fascianelli, Ferrucci, Tsujimoto, & Genovesio, 2020), and lesions to regions that include area 12 cause deficits in implementing strategies (Bussey, Wise, & Murray, 2001; Baxter, Gaffan, Kyriazis, & Mitchell, 2009). Finally, Area 47/12o has been shown to play a role in contingent learning (Rudebeck, Saunders, Lundgren, & Murray, 2017; Jocham et al., 2016), and in suppressing noncontingent learning (Jocham et al., 2016; Noonan, Chau, Rushworth, & Fellows, 2017; Chau et al., 2015).

3.2.4 Working Memory (WM) Module

The WM module receives the HS prediction for the active context on the next trial as input [i.e., $C_{HS}(t+1)$], attempts to maintain this value in working memory until a decision is required, and at the end of the delay period, it outputs its own prediction of which context is active on the next trial, which we denote as $C_{WM}(t+1)$. The model assumes that the memory of $C_{HS}(t+1)$ is maintained throughout the delay period with probability $p \times v$, and lost with probability $1 - pv$, where larger values of p model greater working memory capacity (dlPFC lesions result in diminished working memory capacity), and larger values of v model shorter delay periods and/or reduced working memory demands. In the model fits described below, $v = 0.55$ following an error trial (modeling the 12 second intertrial interval) and $v = 1$ following a correct trial (modeling the shorter 6-second delay). If the memory of $C_{HS}(t+1)$ is retained then $C_{WM}(t+1) = C_{HS}(t+1)$. However, if the memory of $C_{HS}(t+1)$ is lost then $C_{WM}(t+1)$ is set to the contrasting

context (the context opposite to $C_{HS}(t+1)$; recall that in this application there are only two contexts).[4] Lastly, the output of the WM module, $C_{WM}(t+1)$, projects to the RC, HS, and CDRL modules. As input to the RC module, it is used to specify the subset of rule weights (context C_S or context C_H) to use to make the rule choice on trial $t+1$ (details in the following section). As input to the HS module, it is used to implement the win-stay, lose-shift strategy following feedback on trial $t+1$. Finally, as input to the CDRL module, it is used to specify the context under which the weights are updated following feedback on the next trial. The WM module has 3 free parameters $(p, v = 0.55, v = 1)$.

There is substantial evidence that WM is mediated by a broad neural network centered in the PFC (for a review, see e.g., Lara & Wallis, 2015), but extending to many other regions including the caudate nucleus (e.g., Hikosaka, Sakamoto, & Usui, 1989), globus pallidus (e.g., Mushiake & Strick, 1995), medial dorsal nucleus of the thalamus (e.g., Fuster & Alexander, 1971), and regions of posterior cortex (e.g., Constantinidis & Steinmetz, 1996).

3.2.5 Rule Choice (RC) Module

The RC module receives $C_{WM}(t+1)$ from the WM module and the associated weights of each rule in each context (from the CDRL module) as input. If the WM input is $C_{WM}(t+1) = C_S$ then the agent chooses rule M_S with probability $W_{SS}(t+1)$ and rule M_H with probability $W_{HS}(t+1) = 1 - W_{SS}(t+1)$. Alternatively, if the WM input is $C_{WM}(t+1) = C_H$ then the RC module chooses rule M_S with probability $W_{SH}(t+1) = 1 - W_{HH}(t+1)$ and rule M_H with probability $W_{HH}(t+1)$. The RC module has zero free

[4]Logically, it makes more sense that the WM module would guess the context on trials when it failed to maintain $C_{HS}(t+1)$. However, note that this guessing version of the model is mathematically equivalent to the version we implemented (in the sense that it makes identical predictions). The only difference is that the two versions have different values of the parameter p.

parameters.

3.3 Methods

The proposed model was evaluated using numerical simulations. The parameters for each model fit are shown in Table 3.1. The goal of the simulations was to test the architecture of the model, rather than our ability to optimize parameter estimation. Hence, it is important to note that although this model includes 9 free parameters, the majority of these were fixed after modeling the control data. It is noteworthy that only one parameter each was changed from the control condition to fit the dlPFC and ACC lesion data and only two parameters were changed to fit the OFC lesion data (with one exception being the data summarized in Figure 3.8). Furthermore, these modifications provide insight into the consequences of the various prefrontal lesions for task performance. In order to fit the ACC lesion data, α was decreased relative to the control condition, modeling a reduced learning rate. In order to fit the dlPFC lesion data, p was decreased relative to the control condition, modeling a decrease in working memory capacity. In order to fit the OFC lesion data, α was decreased relative to the control condition, once again modeling a reduced learning rate, and h was decreased relative to the control condition, modeling a decreased ability of the HS module to suppress the noncontingent strategy signal from the SEL module. Finally, to fit the model to performance following a short interrupt (Figure 3.8), a single additional parameter was modified for all conditions ($v = 0.64$) to model the 11-second intertrial interval.

Each simulation included 300 trials of post-operative performance. We did not model preoperative learning, since all monkeys performed similarly and it is likely that different mechanisms are at play during initial learning. On postoperative trial 1, the model chose a context at random, and then, given this context, chose a rule according to the initial rule

Table 3.1: Model Parameters. Bold text represents parameter values that changed relative to the control condition

Parameter	CON	DLPFC	ACC	OFC
λ	0.35	0.35	0.35	0.35
q	0.7	0.7	0.7	0.7
α	0.27	0.27	**0.12**	**0.16**
β	0.008	0.008	0.008	0.008
$W_{SS}(0), W_{CC}(0)$	0.7	0.7	0.7	0.7
h	1	1	1	**0**
p	1	**0.85**	1	1
v (error trials)	0.55	0.55	0.55	0.55
v (reward trials)	1	1	1	1
v (unfilled delay)	0.64	0.64	0.64	0.64

weights. Also on postoperative trial 1, the objectively correct rule was chosen at random and once the model achieved 85% correct on the previous 20 trials, the correct rule changed. The model was run for 2000 simulations of 300 trials per condition (comparable to approximately 133 monkeys undergoing 15 daily sessions of 300 trials).

3.4 Results

The results for the CON, dlPFC, ACC, and OFC groups are shown in Figures 3.3-3.9, along with predictions of the model.[5] These results are particularly difficult to model for at least two salient reasons. First, the sheer volume of empirical data from a single experiment implies that all results from a single condition must be fit by a single parameter set. Second, in accordance with presenting a neurocomputational model, the model architecture and the changes in parameter values between conditions must be neurobiologically justified. This constrains the space of possible parameterizations for fitting such a wide range of data.

[5]The results from Buckley et al. (2009) presented in Figures 3.3-3.9 were derived directly from the published manuscript using WebPlotDigitizer (Rohatgi, 2021).

The results that consist of one data point per lesion group (Figures 3.3,3.5,3.6,3.7,3.9) are summarized in the following text. All lesion groups exhibited less rule switches per daily session (300 trials) relative to the CON group but there were no differences in number of rule switches between lesion groups (Figure 3.3). There were no differences between groups regarding perseveration following a rule switch (Figure 3.5). The OFC lesion group shows a performance deficit relative to all other groups on trials following a single correct response (Figure 3.6). According to Buckley et al. (2009) there were no differences between preoperative and post-operative performance for any group on trials following a single error and all groups performed near 50% accuracy (Figure 3.7)[6]. On select trials, once the monkey reached 85% accuracy a short unfilled delay was implemented which caused a performance deficit for the DLPFC lesion group relative to the CON group but not relative to ACC or OFC lesion groups. Furthermore, there were no differences in performance between the CON, ACC and OFC lesion groups for the delay manipulation. Note that any parameter change that is used to fit the results from one of the scenarios above cannot interfere with the model's ability to fit to the results from one of the other scenarios. For example, if we change a parameter of the model to fit the performance deficit for the OFC lesion group in Figure 3.6 then this parameter change must not significantly change OFC performance in Figure 3.5. It is impressive that the model was not only able to fit the qualitative features of this data but it was also able to generate good quantitative fits to the data in almost all scenarios - all while manipulating only one or two parameters per condition[7].

The same parameter choices that were used to fit the results consisting of a single data point per lesion group must also fit the empirical results from the fine-grained accuracy analysis in Figure 3.8. Buckley et al. (2009) reported accuracy for each group

[6]Note that we do not model preoperative performance but the model simulations do result in post operative performance that is close to 50% accuracy for all groups.

[7]With one exception being figure 3.9 which required a single parameter change across all conditions.

on trials that followed an error (E) and then anywhere between 1 and 7 succeeding correct responses. They referred to these trials as ECn, where n denotes the number of correct responses following the error. The DLPFC and ACC lesion groups both appear to have a generalized deficit relative to the control group across all n in the fine-grained accuracy analysis. Alternatively, the OFC lesion group appears to have a deficit relative to the control group for low n but seems to mostly recover as n increases suggesting that the OFC is necessary for integrating positive feedback early in learning but becomes less important once a new context is sufficiently well established. In order to fit the OFC lesion data in Figure 3.8 the model must treat positive feedback differently depending on whether the preceding trial resulted in positive or negative feedback. One could simply prescribe that in order for learning to occur in the CDRL module the agent must receive 2 consecutive trials of positive feedback but this seems arbitrary, post-hoc and neuroscientifically unjustified. The OFC deficits on display in Figures 3.6 and 3.8 emerge naturally from our neurocomputational model. For trials EC1 and EC2 the average predicted reward generated by the SEL module is below 0.7 and therefore a strategy signal, S_{SEL}, is sent to the HS module. For trials EC1 and EC2 the percentage of trials in which $S_{SEL} = switch$ is approximately 40% and 13%, respectively. Alternatively, the percentage of trials in which $S_{HS} = switch$ is 0% for each of these trials. The S_{SEL} signal has no effect on CON, DLPFC, and ACC lesion groups because they have an intact HS module that is able to suppress the non-contingent strategy signal from the SEL module. On the other hand, the OFC lesion group has a compromised HS module that is not able to suppress the non-contingent signal and the SEL strategy signal overrules the HS strategy signal until the predicted reward generated by the SEL module exceeds the threshold. Recall that trial EC1 represents accuracy on trials following the sequence "Error, Correct". Also note that on trial EC1 the correct strategy is stay[8]. Therefore,

[8]Assuming that the agent choose the correct context on the rewarded trial.

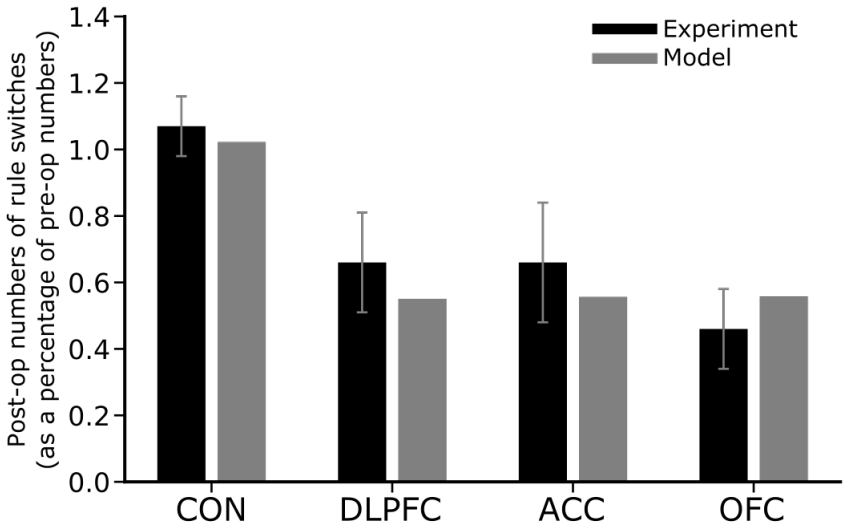

Figure 3.3: Rule switching data (black bars) from Buckley et al. (2009) and model simulations of the same experiment (gray bars). CON = control, dlPFC = dorsal lateral prefrontal cortex, ACC = anterior cingulate cortex, OFC = orbitofrontal cortex. Error bars represent standard error of the mean.

the application of a switch signal on approximately 40% of trials implies that in order to choose correctly the agents that choose to switch must also choose the wrong rule within the wrong context[9]. Importantly, this neurobiologically informed solution does not interfere with the models ability to fit all of the other empirical data. In the following sections we provide more details on the model fits to the empirical data.

3.4.1 Number of Rule Switches

Figure 3.3 shows the number of post-op rule switches per session as a percentage of pre-op switches for each animal in the four conditions of the Buckley et al. (2009) experiment. The model assumed identical pre-op performance for all groups, so the pre-

[9]Also note that a large proportion of these errors will occur early in learning and therefore some portion the of the correct stay trials will also result in incorrect rule choices and lead to the low accuracy for EC1 shown in Figures 3.6 and 3.8.

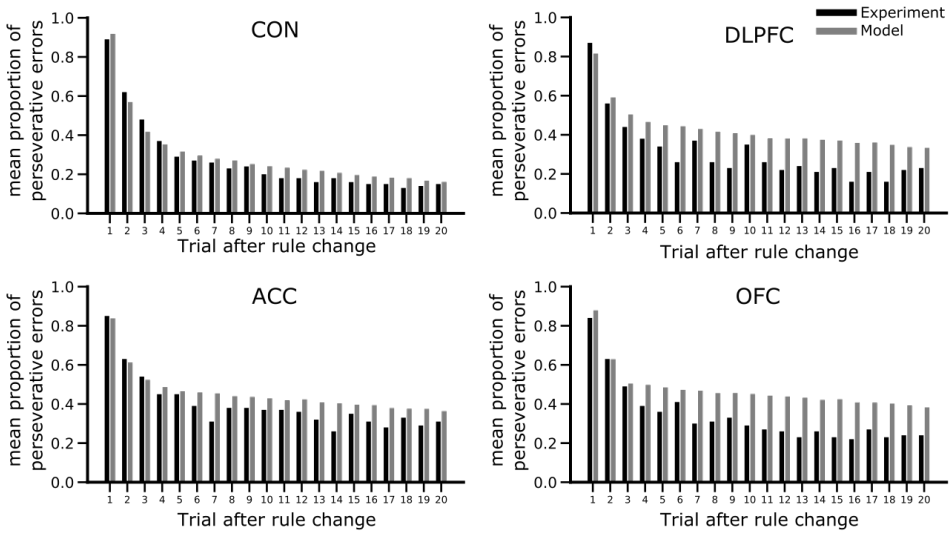

Figure 3.4: Perseverative error data (black bars) from Buckley et al. (2009) and model simulations of the same experiment (gray bars) for each of the 20 trials following a rule change. CON = control, dlPFC = dorsal lateral prefrontal cortex, ACC = anterior cingulate cortex, OFC = orbitofrontal cortex.

op baseline was the same for all groups. Note that dlPFC, ACC, and OFC lesions all decreased the number of rule switches, and that the model accurately accounts for this effect. In paricular, the model fits all fall within the standard error of the mean in every condition.

3.4.2 Distribution of Perseverative Errors Following a Rule Change

We fit the model to the post-op distribution of perseverative errors for each of the 20 trials following a rule change. These data and model fits are shown in Figure 3.4.[10] The model fits the data from all of these conditions reasonably well, although note that it

[10]In the case of the Control group, these data are pre-operative since Buckley et al. (2009) reported there were no significant performance differences between the pre- and post-operative (unoperated) control monkeys.

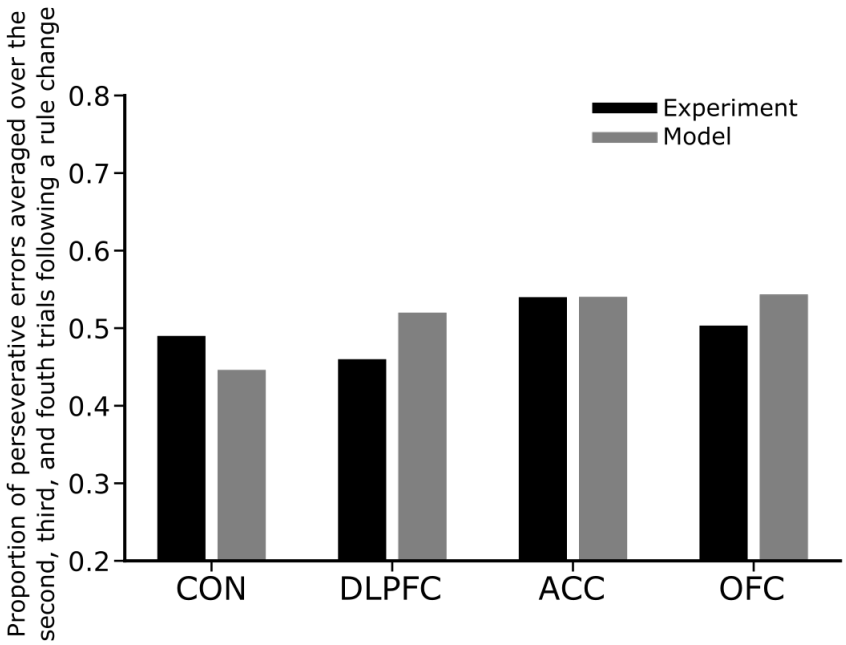

Figure 3.5: Perseverative error data (black bars) from Buckley et al. (2009) and model simulations of the same experiment (gray bars) averaged over the first 3 trials following a rule change. CON = control, dlPFC = dorsal lateral prefrontal cortex, ACC = anterior cingulate cortex, OFC = orbitofrontal cortex.

tends to over predict the number of perseverative errors on trials that occur long after the switch – especially in the dlPFC and OFC conditions. These consistent over predictions suggest that the animals may have used some compensatory strategy that the model fails to include. For our purposes, however, the most important feature of these data occur on trials 2, 3, and 4 following the rule change. The results from these three trials are summarized in Figure 3.5. A key finding from Buckley et al. (2009) was that none of the groups made more perseverative errors on these trials than the control group. The model provides an impressive fit to the data from these trials, suggesting that it correctly accounts for the failure of dlPFC, ACC, and OFC lesions to cause more perseverative errors.

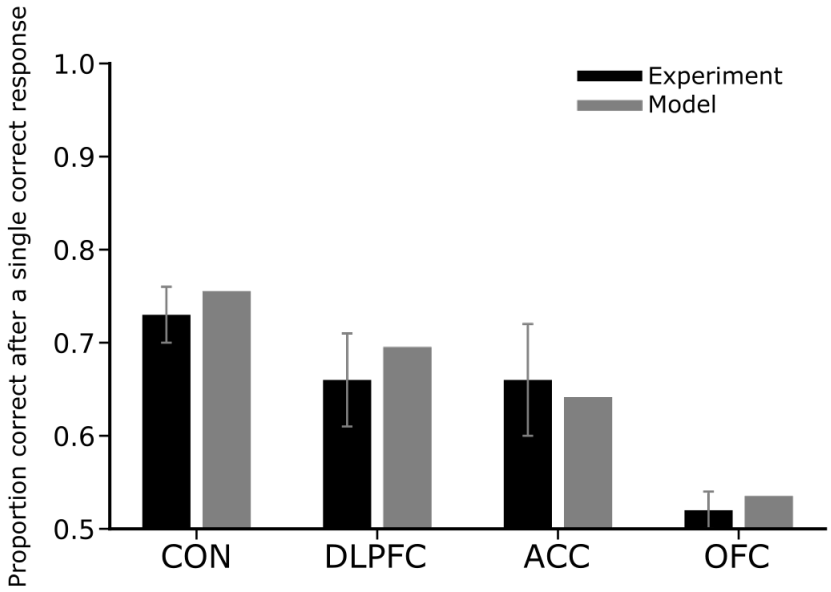

Figure 3.6: Proportion correct after a single correct response from Buckley et al. (2009) (black bars) and model simulations of the same experiment (gray bars). CON = control, dlPFC = dorsal lateral prefrontal cortex, ACC = anterior cingulate cortex, OFC = orbitofrontal cortex. Error bars represent standard error of the mean.

3.4.3 Performance Following a Single Correct Response

Third, we fit the model to accuracy data for trials that immediately followed a single correct response. The first 3 trials following a rule switch were excluded from this analysis. These data and model fits are shown in Figure 3.6. Note that OFC lesions severely impaired performance on trials that followed a single correct response. Also note that the model accurately accounts for this effect, and it also provides good fits to the data from the other conditions. Specifically, for all four groups, the model fits fall within the standard error of the mean of the experimental data.

3.4.4 Performance Following a Single Error

Fourth, we fit the model to accuracy data from trials that immediately followed a single error. These results and model fits are shown in Figure 3.7. Note that all groups

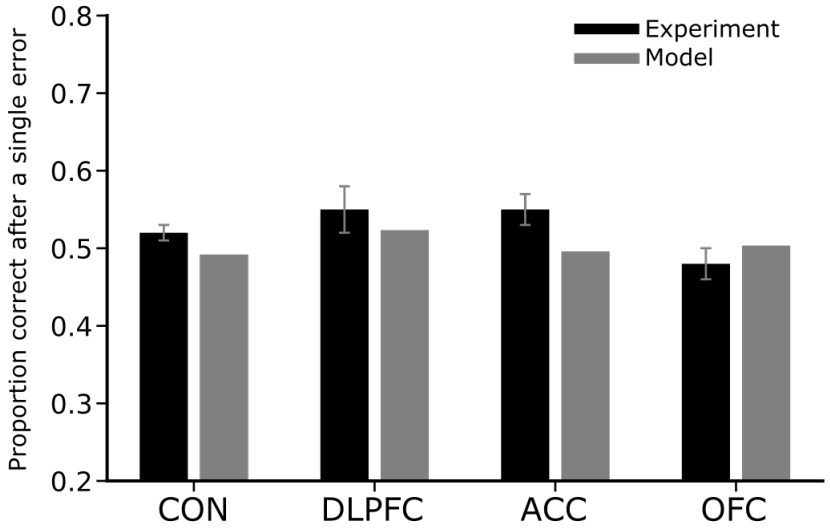

Figure 3.7: Proportion correct after a single error from Buckley et al. (2009) (black bars) and model simulations of the same experiment (gray bars). CON = control, dlPFC = dorsal lateral prefrontal cortex, ACC = anterior cingulate cortex, OFC = orbitofrontal cortex. Error bars represent standard error of the mean.

performed near chance on these trials (i.e., 50% correct), and therefore there were no significant differences among groups. This may have been a floor effect, but regardless, note also that the model accurately accounts for the results from all four groups.

3.4.5 EC Analysis

Our fifth test of the model was against a more fine-grained accuracy analysis. Specifically, Buckley et al. (2009) reported accuracy for each group on trials that followed an error (E) and then anywhere between 1 and 7 succeeding correct responses. They referred to these trials as ECn, where n denotes the number of correct responses following the error. In other words, EC3 denotes the sequence ECCC – that is, an error followed by 3 successive correct responses. So the values depicted at $n = 3$ are average accuracies on the trial that succeeded the sequence ECCC. The first 3 trials following a rule switch

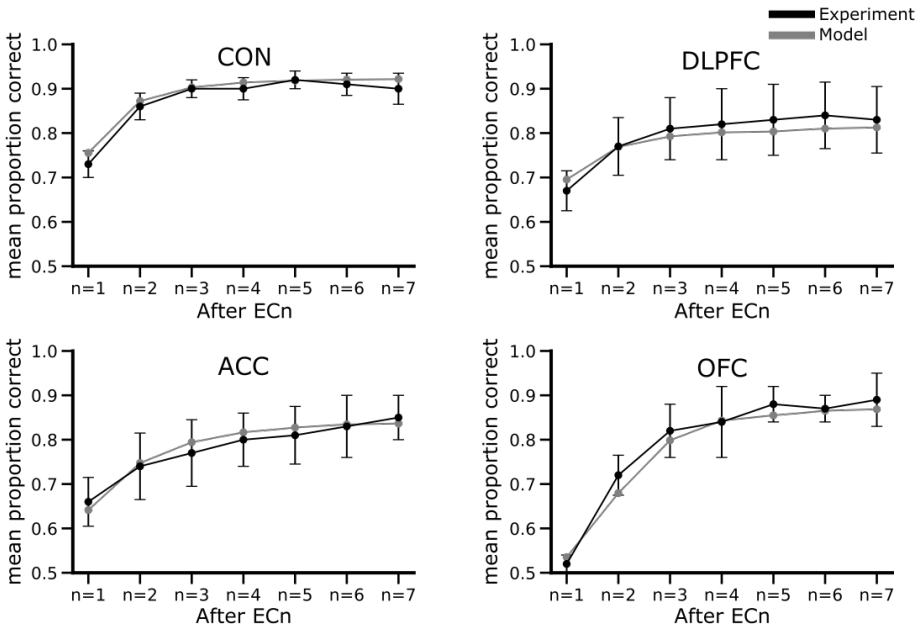

Figure 3.8: Fine-grained accuracy analysis from Buckley et al. (2009) (black bars) and model simulations of the same experiment (gray bars). Values plotted at $n = i$ (for $i = 1, ..., 7$) are average accuracies on trials that succeeded a sequence of $i+1$ trials in which the first response was an error and the next i responses were correct. CON = control, dlPFC = dorsal lateral prefrontal cortex, ACC = anterior cingulate cortex, OFC = orbitofrontal cortex. Error bars represent standard error of the mean.

were excluded from this analysis. Results for each group, and fits of the model are shown in Figure 3.8. Note that the model provides extremely accurate accounts of these data for all groups.

3.4.6 Performance Following a Short Unfilled Interrupt

As a final test of the model, we examined its ability to account for accuracy after a short, unfilled interrupt that Buckley et al. (2009) introduced after each animal reached 85% accuracy. In these conditions, rather than changing the correct rule after criterion accuracy was reached, Buckley et al. (2009) increased the delay before the next trial began by 5 seconds, which previous research had shown was long enough to impair

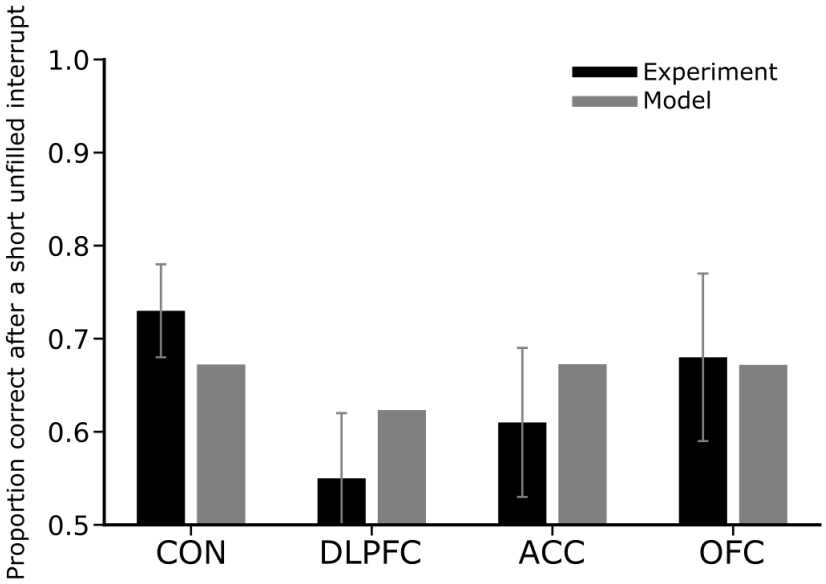

Figure 3.9: Proportion correct following a short unfilled interrupt (black bars) from Buckley et al. (2009) and model simulations of the same experiment (gray bars). CON = control, dlPFC = dorsal lateral prefrontal cortex, ACC = anterior cingulate cortex, OFC = orbitofrontal cortex. Error bars represent standard error of the mean.

working memory. So the goal of this manipulation was to assess the contribution of each lesioned area to the ability of the animals to maintain the current rule in working memory.

For each 300 trial session by collecting the trials in which a switch occurred and reinitialized the model with the conditions at the time of the switch. We then simulated the subsequent trial without a switch and with a 5 second additional delay (11 second intertrial interval, $v = 0.64$). Mean performance was calculated for each session and overall average performance was calculated from these values. The results and model predictions for each group are shown in Figure 3.9. As expected, the dlPFC group performed worst on these extended delay trials, which is consistent with the prominent role that the PFC is known to play in working memory (e.g., Lara & Wallis, 2015). Note that the model slightly underestimates the performance of the control group and slightly

overestimates the performance of the dlPFC lesion group. Even so, it correctly predicts that the poorest performance is by the dlPFC group, and its predicted accuracy falls nearly within the standard error of the mean for each condition.

3.5 Discussion

This article proposed a new neurocomputational model of flexibility in rule-guided behavior. The model includes five interacting modules that determine the current environmental context and then select an appropriate explicit rule to guide behavior. We tested the model against a wide variety of detailed behavioral and neuroscientific data from a simplified version of the WCST that was reported by Buckley et al. (2009). The model gave an impressive account of the varied behavioral results from this study and it successfully predicted the effects of selective lesions to dlPFC, ACC, and OFC on each behavioral measure.

Buckley et al. (2009) proposed that their findings provide evidence supporting the hypotheses that the dlPFC helps mediate the maintenance of abstract rules in working memory, that the OFC helps mediate the representation and rapid updating of the value of abstract rules based on feedback, and that the ACC helps mediate response selection and the degree to which recent feedback should be integrated to influence future decisions. Our results find significant support for these hypotheses. Furthermore, as the next subsection describes, the neuroanatomical assumptions of our model are also bolstered by a variety of results that were reported after the publication of Buckley et al. (2009).

3.5.1 The Role of the OFC and Area 47/12o in Flexible Behavior

The role of OFC in flexible behavior has become particularly controversial. OFC lesions impair performance in reversal paradigms, a standard test of behavioral flexibility (Clark, Cools, & Robbins, 2004; A. Roberts, 2006; Izquierdo et al., 2004; Murray & Wise, 2010; Dias et al., 1996; Rygula et al., 2010), but only when the lesions are made by aspiration (Kazama & Bachevalier, 2009; Rudebeck & Murray, 2011; Rudebeck, Saunders, Prescott, Chau, & Murray, 2013). Neurotoxic lesions have not been found to cause impairments. These findings suggest that impairments from aspiration lesions are not caused by a loss of function in traditional OFC regions, but are instead a result of damage to passing fibers from nearby regions (Kazama & Bachevalier, 2009; Monosov & Rushworth, 2021; Rudebeck & Murray, 2011; Rudebeck et al., 2013). The most likely such region is lateral to the lateral orbitofrontal sulcus (Monosov & Rushworth, 2021). This region is often included in the ventrolateral prefrontal cortex and sometimes referred to as lateral OFC or area 47/12o (Carmichael & Price, 1994; Mackey & Petrides, 2010; Monosov & Rushworth, 2021; Saleem, Miller, & Price, 2014). Area 47/12o has strong connectivity to other PFC regions and aspiration lesions to OFC compromise this connectivity (by destroying white matter tracts), even when such lesions spare area 47/12o itself (Sallet et al., 2020). In summary, evidence suggests that aspiration lesions of central OFC may compromise the ability of area 47/12o to mediate the computations necessary for successful performance in discrimination reversal-learning tasks.[11]

Recent research has implicated area 47/12o in a number of findings that are partic-

[11]Even so, it is important to note that Rudebeck et al. (2017) found that object discrimination reversal learning was not impaired following fiber-sparing lesions of OFC and vlPFC (including area 47/12o), and as a result, they suggested that reversal learning may depend on disconnection of a number of regions from the PFC, including medial striatum and medial dorsal thalamus (H. Clarke, Dalley, Crofts, Robbins, & Roberts, 2004; H. F. Clarke, Robbins, & Roberts, 2008; Groman et al., 2013; Iversen & Mishkin, 1970).

ularly relevant to our model. First, area 47/12o has been shown to represent win-stay and lose-switch signals in object discrimination reversal paradigms (Chau et al., 2015). Second, area 11, ventral area 12, and area 13 have also been shown to display switching signals when the switch was cued by the experimenter (Fascianelli et al., 2020). Third, Bussey et al. (2001) lesioned a region that included areas 11, 12, 13, and 45 and showed a deficit in implementing abstract strategies. Fourth, vlPFC lesions (including area 12) but not OFC lesions (areas 11, 13 and 14) have been shown to be crucial for successful performance in a strategy implementation task (Baxter et al., 2009). Taken together, this research seems to suggest that although the encoding of strategy signals may be spread widely throughout OFC and vlPFC, it appears that area 47/12o may be a crucial locus for strategy representation and implementation. Finally, area 47/12o has been shown to have a role in credit assignment by encoding the contingency between feedback and choice (Rudebeck et al., 2017; Jocham et al., 2016) and suppressing noncontingent learning or Thorndike's "spread of effect", that is, statistical learning in which current choices may be influenced by feedback and choices from trials before and after the current trial (Jocham et al., 2016; Noonan et al., 2017; Chau et al., 2015).

When the Buckley et al. (2009) article was published, the distinction between the effects of aspiration and neurotoxic lesions to OFC had not been discovered. We propose that the OFC lesions performed by Buckley et al. (2009) likely caused two different types of impairment. The intentional lesions of OFC (areas 11, 13 and 14) caused a decreased rate of learning from positive feedback (Murray & Wise, 2010; Rudebeck & Murray, 2008; Stolyarova & Izquierdo, 2017), whereas the unintended disruption of processing in area 47/12o compromised the animals' ability for contingent learning and reduced their ability to suppress statistical learning information coming from the BLA (Jocham et al., 2016; Noonan et al., 2017; Chau et al., 2015).

3.5.2 Predictions and Future Experiments

The model proposed here makes a number of novel predictions. First, it proposes that impairments in the rapid updating of rule values could occur for two separate reasons; intentional lesions to OFC cause an overall reduction in the learning rate, whereas any unintentional disruption of processing in area 47/12o interferes with its ability to suppress spread of effect input from BLA. Therefore, the model predicts that lesions of area 47/12o should result in similar impairments to those reported by Buckley et al. (2009) (though possibly more dramatic). Although it is likely that the aspiration lesions of vlPFC performed by Buckley et al. (2009) may have included some of area 47/12o, a more targeted lesion of area 47/12o is necessary to test this prediction.

Second, the model predicts that neurotoxic lesions of OFC should impair performance by reducing the learning rate. Since neurotoxic lesions should not disrupt area 47/12o, such lesions should not impair the animal's ability to suppress BLA activity. This should result in a performance profile that resembles the Buckley et al. (2009) ACC lesioned animals, rather than the OFC lesioned animals.

Third, our model proposes a role for amygdala in propagating spread of effect information, which can interfere with performance if it is not suppressed. Therefore, the model predicts that amygdala lesions may improve performance on the WCST analog. Another possibility is that the effect of OFC aspiration lesions (or the predicted effect of area 47/12o lesions) on performance could be reversed by amygdala lesions. In fact, Stalnaker, Franz, Singh, and Schoenbaum (2007) have already demonstrated this in rats performing a reversal-learning task. These researchers showed that BLA lesions eliminated the impairment caused by OFC lesions. Although these experiments used a discrimination reversal-learning task, our model predicts that BLA lesions should cause a similar reduction in WCST impairments that occur after lesions of OFC (or area 47/12o).

3.5.3 Relation to Other Models

A number of models have been developed to account for results of experiments with the WCST, and some of these include considerable biological detail (e.g., Amos, 2000; Bishara et al., 2010; Dehaene & Changeux, 1991; Monchi, Petrides, Petre, Worsley, & Dagher, 2001; Moustafa & Gluck, 2011; Steinke, Lange, & Kopp, 2020). Even so, to our knowledge, none have enough detail to fit all of the behavioral results described in Figures 3.3 – 3.9, and most of them are more than 20 years old and therefore, do not incorporate the interesting findings regarding area 47/12o and its interactions with the amygdala. Furthermore, the current model is simpler than the biologically detailed WCST models, and despite this simplicity, provides a comprehensive account of performance on the WCST analog in healthy and lesioned animals.

Another related model proposed that the OFC serves as a cognitive map of task space (Wilson et al., 2014). This is not a model of the WCST, but it has been tested against reversal-learning data. In these applications, the model uses the Rescorla-Wagner rule for learning and the Luce choice rule for response selection. In reversal-learning tasks there are two contexts; one in which the single rewarded option is A and the other in which option B is rewarded. The Wilson et al. (2014) model assumes that OFC lesions impair the ability to learn which context is currently active. Therefore, the model predicts that prior to a reversal, control and OFC lesioned animals should perform similarly. However, following a reversal, the control animals will recognize that the context has changed, whereas OFC lesioned animals will not.

Our model could be viewed as an extension of the Wilson et al. (2014) model to the WCST. In fact, the CDRL module is almost identical to the Rescorla-Wagner learning algorithm assumed by Wilson et al. (2014), except that the CDRL module modulates the effect of RPEs by the absolute value of the prediction error on the previous trial

(i.e., according to the Rescorla-Wagner/Pearce-Hall hybrid learning rule). Both models also only update the rule weights for the active context. Our model uses a separate HS module that implements the win-stay, lose-shift strategy to change the agent's belief about the current context, which is subject to interference from the SEL module and must be maintained by the WM module throughout any delays. In contrast, the Wilson et al. (2014) model uses the temperature parameter in the Luce choice model to determine how long it takes to reverse its behavior.

Although there are a number of differences between our model and the Wilson et al. (2014) model, perhaps the most crucial is the proposed effect of OFC lesions. Wilson et al. (2014) proposed that OFC lesions cause the animal to become incapable of separating contexts. Alternatively, we propose that OFC lesions reduce the learning rate from positive feedback and make it impossible to suppress noncontingent learning (due to the unintended disruption of area 47/12o caused by aspiration lesions to OFC). In the Buckley et al. (2009) experiment, the animals received extensive training prior to surgery, during which time they became adept at identifying the active context. It is unclear whether or not the Wilson et al. (2014) model would predict that this ability would be lost following OFC lesions. If the ability to identify context was lost, it seems unlikely that the model would be able to account for the Buckley et al. (2009) results. For example, under these conditions, it seems that the model would predict that the animal would have to unlearn and relearn action values repeatedly following reversals. If so, then one would expect that Figure 3.5 should show a significant difference between the OFC group and other groups (i.e., in the frequency of perseverative errors), which it does not. On the other hand, if the ability to identify context is not lost, then the model would predict no effect of OFC lesions on any dependent measure, which conflicts with the results shown, for example, in Figure 3.6.

3.5.4 Future Modeling Considerations

Our model gave an impressive account of the varied behavioral results from this study and it successfully predicted the effects of selective lesions to dlPFC, ACC, and OFC on each behavioral measure. As previously mentioned, our model straddles the boundary between the algorithmic and implementational levels in Marr's (1982) hierarchy. It is neurobiologically informed in the sense that all but one of its modules are mapped onto brain regions and are linked together in a circuit that specifies the inputs and outputs of each region. However, the model is not composed of model neurons or synapses and as a result there are no within trial dynamics, hence the only data from Buckley et al. (2009) that our model was not applied to was reaction time data. A more detailed implementational level version of our model that generates intratrial dynamics should be able to account for reaction time data. Furthermore, this lower level model should be able to account for single-cell and population level neural data. Nonetheless, our model should be regarded as laying the foundation for an even better explanation that could emerge from the more fine-grained lower implementational level model.

In modeling the data presented in this paper we viewed our task as constructing the simplest possible model architecture that maintains a reasonable mapping from parameters to neural circuit entities. We do not necessarily believe that this is the *appropriate* level of analysis for the flexible learning of abstract rules but we do believe that it provides a clear road map for the invention of lower level implementational models. This can be made clear by superficially considering how one might approach this task. Note that this model is primarily composed of modules that predict reward probabilities and implement win-stay/low-switch strategies that must be maintained in working memory. It is not difficult to construct neural networks that compute these values and strategies via activation patterns and synaptic weights. Additionally, recurrent neural networks

can be constructed that hold strategies in working memory throughout delay periods. These networks can be substituted for the modules that we've presented in our model and could be tested against additional neuroscientific and behavioral data in an effort to make progress toward a satisfying explanation of flexible abstract rule learning.

Chapter 4

Modulation of Dopamine for Adaptive Learning: a Neurocomputational Model

The content of chapter 4 is the result of a collaboration with Vivian V. Valentin and F. Gregory Ashby, and has previously appeared in Computational Brain & Behavior (Inglis et al., 2021). Reprinted by permission from Springer Nature Customer Service Centre GmbH: Springer Nature, Computational Brain & Behavior. Modulation of Dopamine for Adaptive Learning: a Neurocomputational Model. Jeffrey B. Inglis, Vivian V. Valentin, F. Gregory Ashby. Society for Mathematical Psychology 2020.

4.1 Introduction

Normative and machine-learning models of learning have been integral to development and progress in a wide range of fields, including computer science (Sutton & Barto, 1998), neuroscience (Maia, 2009; Dayan & Abbott, 2001), and psychology (Rescorla &

Wagner, 1972; Bush & Mosteller, 1951; Berridge, 2000). For example, reinforcement learning algorithms have provided successful models of how predicted reward estimates are updated when new rewards are encountered in the environment. In these models, the amount of learning on each trial is proportional to the reward prediction error (RPE), which is defined as the obtained reward (R) minus the predicted reward (P).

The standard assumption is that dopamine (DA) neurons in the ventral tegmental area (VTA) and the substantia nigra pars compacta (SNpc) encode the RPE via their response to rewarding events and to cues that predict rewards (Montague, Dayan, & Sejnowski, 1996; Schultz, Dayan, & Montague, 1997). Even so, it is also well known that RPE is an imperfect predictor of the DA response. For example, DA neurons also respond to novel events and to salient stimuli with no reward-related associations (Horvitz, 2002). In addition, there are large individual differences in the DA response to any given RPE, which depend, at least in part, on personality type (Pickering & Pesola, 2014).

To account for such variability in the DA response to RPE, reinforcement learning models typically include an additional learning-rate parameter – denoted by λ_n in Equation 4.1 below – that controls the amount of learning that occurs for any given value of RPE. When fitting reinforcement learning models to data, λ_n is typically treated as a free parameter, which allows the models to account for unexplained variability in the learning effects of any given RPE, albeit only via *post hoc* curve fitting. A complete theory of learning must describe a neural account of these changes in λ_n. This article takes a significant step towards this goal by describing a neural network that modulates the DA response to RPE under a wide variety of environmental conditions.

If the learning-rate λ_n is too small, learning is slower than necessary and the learner is insensitive to changes in the reward structure of the environment. If λ_n is too large, learning is unstable. The optimal value of λ_n changes adaptively in response to environmental changes in the statistical structure of rewards (Daw & O'Doherty, 2014;

Dayan, Kakade, & Montague, 2000; Dayan & Long, 1998). Additionally, a number of investigators have proposed a variety of factors that may affect λ_n such as expected and unexpected uncertainty (Dayan & Yu, 2003; Yu & Dayan, 2005), volatility (Behrens et al., 2007), outcome, informational, and environmental uncertainty (Mathys et al., 2011), covariance between predictions and past RPEs, estimation, and unexpected uncertainty (Payzan-LeNestour & Bossaerts, 2011; Preuschoff & Bossaerts, 2007), and state-feedback contingency (Crossley, Ashby, & Maddox, 2013). For detailed reviews of some of these taxonomies of uncertainty and the relationships between them, see Bland and Schaefer (2012) and Soltani and Izquierdo (2019).

In the language of Marr (1982), almost all of these models are computational. Thus, they make little or no attempt to describe the neural circuitry that implements the proposed computations. In particular, there are few current hypotheses about the neural mechanisms that modulate the amount of learning that occurs for any given RPE (for exceptions see Bernacchia, Seo, Lee, & Wang, 2011; Franklin & Frank, 2015; Iigaya, 2016; Farashahi et al., 2017).

This article proposes such a mechanism. Specifically, we describe a biologically detailed computational model of how the adaptive learning rate proposed in the models described above could be implemented at the neural level. We describe the neural circuit that mediates this modulation and model activity at the level of spiking neurons. The input to the network is a computed value of some relevant theoretical variable such as unexpected uncertainty, volatility, or feedback contingency and the output is spiking activity in a population of DA neurons. The resulting DA release is presumed to then affect tonic and phasic DA levels in target brain regions. The model is agnostic about which factors modulate learning rates and how they are computed. The neuroanatomy of the network we propose is consistent with many of the alternative proposals about how learning rates are modulated. Thus, the proposed model should be of widespread

interest.

Furthermore, the model can be applied to a variety of DA mediated tasks, many of which transcend the experimental phenomena that it was created to address. Potential applications of the model extend to working memory, creative problem solving, cognitive flexibility, and category learning. The general applicability of the model to paradigms that extend beyond the implementation of learning rates in simple reinforcement learning tasks follows from the fact that the network predicts changes in tonic and phasic DA in all brain regions that are targets of VTA DA neurons, and thus is applicable to any model of behavior that depends on these regions and assigns a specific functional role to DA.

The article begins with a brief review of a simple and common reinforcement learning algorithm. We then discuss the benefits of an adaptive learning rate and briefly review many of the factors that have been proposed that influence this rate. We refer to these factors as modulating variables. Next, we describe our neurocomputational model of how the modulating variable controls DA neuron firing and therefore also DA release and learning. The computational principle implemented by the network is to control the gain on the DA response to any given RPE in addition regulating tonic levels via the modulating variable. This new theory is formulated as a biologically detailed computational model that we refer to as the Modulation of Dopamine for Adaptive Learning (MODAL) model. Finally, we close with a discussion of the relationship between our implementational-level model and other levels of analysis and possible directions for future research.

4.1.1 Reinforcement Learning Algorithms

This article proposes a neural interpretation of learning rates. Virtually all learning algorithms include a learning rate parameter and the network described below could pro-

vide a neural interpretation of that parameter in many of these algorithms. To keep the presentation concrete however, we focus on one simple reinforcement learning algorithm that is ubiquitous in the literature and that formalizes the notion of a learning rate – namely, the single-operator model of Bush and Mosteller (1951) (also see Rescorla & Wagner, 1972).

The single-operator model assumes that the predicted reward value on trial n, denoted by P_n, equals:

$$\begin{aligned} P_n &= P_{n-1} + \lambda_n(R_n - P_{n-1}) \\ &= P_{n-1} + \lambda_n RPE_n, \end{aligned} \tag{4.1}$$

where R_n is value of the obtained reward on trial n and λ_n is the learning rate on trial n. It is well known that in a stable environment, P_n converges asymptotically to the mean reward value and the rate of convergence increases with λ_n (i.e., for all λ_n in the range $0 < \lambda_n < 1$). So any variable that increases λ_n, increases the learning rate.

Note that even simple algorithms that set all λ_n to the same constant value predict a form of cooling because the magnitude of the RPEs will decrease as learning progresses. Even so, many algorithms change λ_n with n (e.g., Sutton, 1992). For example, it is common to decrease λ_n as n increases – a process that accelerates cooling. In addition, there have been many proposals that other factors also dynamically adjust learning rates in the brain. The remainder of this section briefly reviews various modulating variables that have been proposed to affect λ_n.

Although λ_n is often treated as a free parameter in many applications, its optimal value can be determined trial-by-trial by considering the iterative updates in reinforcement learning as a statistical problem of how best to integrate previous estimates with

new evidence. Taking a Bayesian approach, it has been shown that under certain assumptions, the optimal way to integrate past predictions with new data is to set the learning rate to (Daw & O'Doherty, 2014; Dayan & Long, 1998; Dayan et al., 2000):

$$\lambda_n = \frac{\sigma_n}{\sigma_n + Var(R)}, \tag{4.2}$$

where σ_n represents the variance or uncertainty in our current estimate of the predicted reward and $Var(R)$ represents the variance in the reward values. Therefore, if the obtained reward values are not changing very much (i.e., $Var(R)$ is small), then λ_n should be large, which will cause the predicted reward estimate to converge quickly to the (mean) obtained reward value. On the other hand, if the obtained reward values are noisy (i.e., $Var(R)$ is large) then we should set λ_n to be small to avoid over-reacting to an unexpectedly large or small reward value.

Several researchers have argued that learning rates in the brain are also affected by volatility – that is, by how quickly the reward contingencies change in the environment (Behrens et al., 2007; Mathys et al., 2011). The idea is that increases in volatility should increase λ_n because agents should learn faster in a rapidly changing environment in order to track the fluctuations. Alternatively, when the environment is stable, the agent should learn more slowly to ensure it uses as much data as possible in order to converge upon the true stable reward probabilities.

Dayan and Yu proposed that learning rates depend on what they called expected and unexpected uncertainty (Yu & Dayan, 2005; Dayan & Yu, 2003). Expected uncertainty arises as a result of the unreliability of the cue that signals reward and the agent should suppress the use of the cue when expected uncertainty is high. Unexpected uncertainty is similar to the Behrens et al. (2007) notion of volatility and the Mathys et al. (2011) notion of environmental uncertainty, that is, unexpected uncertainty is high when the agent

is confident in their top-down model but these expectations are nonetheless violated by the bottom-up sensory data. This may be an indication that although the model was accurate, the environment has changed and therefore learning from bottom-up data should be more heavily weighted than the top-down model. However, according to Bland and Schaefer (2012), unexpected uncertainty differs from volatility in that volatility is related to the frequency with which stimulus-response-outcome (SRO) contingencies change. For example, in a probabilistic reversal task where SRO contingencies reverse every 30 trials, unexpected uncertainty will increase following the reversal. Furthermore, this environment would be characterized as having higher volatility relative to an environment in which the SRO contingencies only reversed every 100 trials.

Payzan-LeNestour and Bossaerts (2011) proposed that λ_n depends on unexpected uncertainty and estimation uncertainty and that prediction risk scales the RPE (Preuschoff & Bossaerts, 2007). Prediction risk is the irreducible uncertainty due to outcome uncertainty. Estimation uncertainty is measured as the entropy of the posterior distribution (similar to the uncertainty of the prior in the above equation), whereas unexpected uncertainty is high when SRO contingencies change abruptly, as described above. Preuschoff and Bossaerts (2007) also proposed that the covariance between past predictions and reward prediction errors may contribute to λ_n, as derived from least-squares learning theory.

Finally, empirical evidence suggests that state-feedback contingency, defined as the covariance between rewards and predictions, has a significant effect on the learning rate (Ashby & Vucovich, 2016; Crossley et al., 2013). The intuition here is that a measure of the covariance between rewards and predictions enables a parsimonious method for the agent to infer the degree to which its actions play a role in determining its rewards. If state-feedback contingency is high, the agent recognizes that its behavior plays a significant role in determining its rewards and takes advantage of this by increasing the

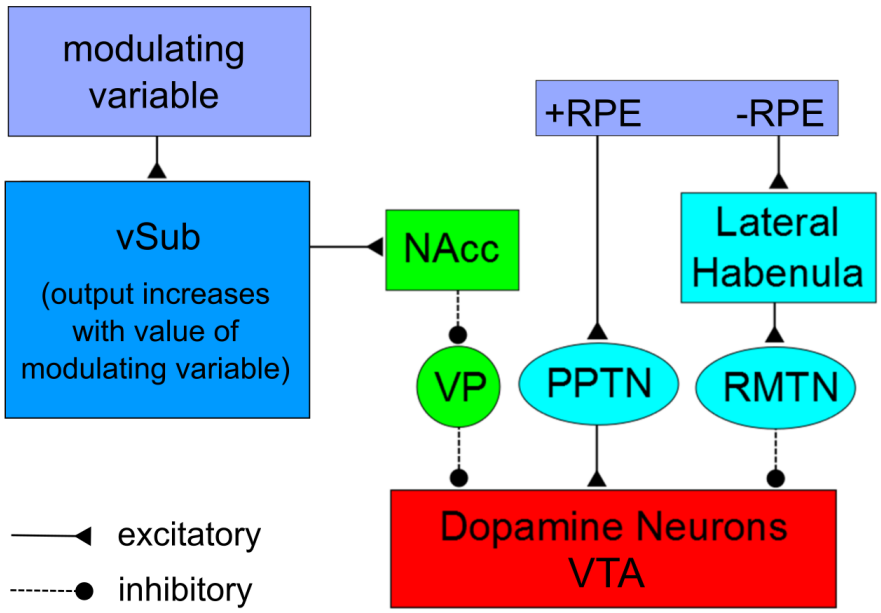

Figure 4.1: Neural architecture of the proposed MODAL model of how DA activity is regulated by one of the modulating variables described in the previous section. RPE = reward prediction error; vSub = ventral subiculum; NAcc = nucleus accumbens; VP = ventral pallidum; PPTN = pedunculopontine tegmental nucleus; RMTN = rostromedial tegmental nucleus; VTA = ventral tegmental nucleus; SNpc = substantia nigra pars compacta

learning rate. Alternatively, if state-feedback contingency is low, the agent recognizes that its behavior does not play a significant role in determining its rewards and therefore it can conserve resources and preserve previous learning by decreasing the learning rate.

The next section proposes a neural network that could implement any of these modulating effects on learning rate.

4.1.2 Neuroanatomy of MODAL

Reward and feedback processing recruit diverse brain networks that include the limbic system and prefrontal and sensory cortices (Liu, Hairston, Schrier, & Fan, 2011; Watabe-Uchida, Zhu, Ogawa, Vamanrao, & Uchida, 2012; Tian & Uchida, 2015; Haber,

2016; Faget et al., 2016; Takahashi, Langdon, Niv, & Schoenbaum, 2016). Multiple brain regions respond to reward and compute predicted rewards (Sesack & Grace, 2010; Bromberg-Martin, Matsumoto, & Hikosaka, 2010; Humphries & Prescott, 2010), and this redundancy inspired many alternative theories of how DA neuron firing is modulated by RPE (Houk, Adams, & Barto, 1995; Schultz et al., 1997; Sutton & Barto, 1998; Schultz, 1998; Tan & Bullock, 2008; Kawato & Samejima, 2007; Morita, Morishima, Sakai, & Kawaguchi, 2013; Brown, Bullock, & Grossberg, 1999; Hazy, Frank, & O'Reilly, 2010; Joel, Niv, & Ruppin, 2002; Stuber et al., 2008; Contreras-Vidal & Schultz, 1999; O'Reilly, Frank, Hazy, & Watz, 2007). In contrast to all this work, we do not know of any neurocomputational models that attempt to account for any modulating effects of the DA response to RPE. We propose that dynamic changes in learning rate are mediated by changes in the size of the population of tonically firing DA neurons. As the size of this population grows, more DA neurons become available to respond to any given RPE, which has the computational effect of increasing the learning rate.

The neural architecture of the model is described in Figure 4.1. The inputs to the network are from regions that compute RPE and the value of the relevant modulating variable. Whereas the alternative modulating variables that have been proposed might recruit somewhat different neural networks, they all depend on temporal integration or continuous updating of feedback and reward information. Therefore, they are likely to depend on similar networks that include regions in orbitofrontal, medial prefrontal, anterior cingulate, parahippocampal, and entorhinal cortices. We make no attempt to describe this network in detail, but we assume that it sends a prominent projection to the ventral subiculum (vSub), which is the main output structure of the hippocampus. vSub receives input from a variety of regions, including CA1 of the hippocampus (Fanselow & Dong, 2010), parahippocampal cortex, and entorhinal cortex (Kerr, Agster, Furtak, & Burwell, 2007). The entorhinal cortex encodes general properties of the current context

(Jacobs, Kahana, Ekstrom, Mollison, & Fried, 2010), and the parahippocampal cortex has a general role in contextual binding (Aminoff, Kveraga, & Bar, 2013). Additionally, the entorhinal cortex receives almost all of its cortical inputs from polymodal association areas, including cingulate, orbitofrontal, and parahippocampal cortices, making it well situated for integrating diverse inputs (Insausti, Amaral, & Cowan, 1987). Given the positioning of the vSub as an interface between the contextual information processing in the hippocampus and cortical and subcortical regions implicated in reward processing, learning, and motivation (Quintero, Díaz, Vargas, de la Casa, & López, 2011), we propose that the vSub is a likely target of the complex neural networks that mediate processing of the alternative modulating variables that have been proposed.

The right half of the Figure 4.1 network instantiates the standard RPE model. The idea is that reward sensitive units in regions such as prefrontal and orbitofrontal cortex contribute to the RPE DA signal by providing excitatory inputs to the pedunculopontine tegmental nucleus (PPTN) (Hong & Hikosaka, 2014; Kobayashi & Okada, 2007; Okada & Kobayashi, 2013) and lateral habenula (LH) (Tian & Uchida, 2015; Hong, Jhou, Smith, Saleem, & Hikosaka, 2011; M. Matsumoto & Hikosaka, 2007, 2009). Through these circuits, positive RPEs excite VTA DA neurons via the PPTN, whereas negative RPEs inhibit VTA DA activity via the LH [and the rostromedial tegmental nucleus (RMTN)].

The more novel features of the Figure 4.1 model are presented in the left half of the figure. First, factors thought to influence the modulating variable are integrated in the vSub, which results in an output signal to the NAcc that is proportional to the value of the modulating variable.

The next component of the model builds on the work of Grace, Floresco, Goto, and Lodge (2007), who proposed that the pathway vSub $\rightarrow$ NAcc $\rightarrow$ VP $\rightarrow$ VTA controls the number of VTA DA neurons that fire tonically. The NAcc $\rightarrow$ VP and VP $\rightarrow$ VTA projections are both GABAergic, but the tonic firing rate of VP neurons is much higher

than the tonic firing rate of NAcc neurons. As a result, many DA neurons in VTA are silent due to tonic inhibition by VP. Estimates suggest that because of this inhibition, only about half of VTA DA neurons are spontaneously active under control conditions, and these tonically firing neurons are the only ones available for phasic bursts when excited by PPTN (Lodge & Grace, 2006). When the value of the modulating variable is high, vSub excites NAcc neurons, which inhibit VP neurons. This releases VTA DA neurons from tonic inhibition, which increases the number of tonically firing VTA DA neurons, thereby enlarging the pool of DA neurons that can respond to excitatory input from PPTN. In this way, increasing the value of the modulating variable amplifies the RPE-induced VTA DA response. Thus, the Figure 4.1 network proposes a neural mechanism via which the modulating variable can control the gain of the DA response to any given RPE.

To test this theory more rigorously, we built a biologically detailed computational model of the Figure 4.1 network, and examined its ability to account for RPE and learning rate effects on DA release. Our model is consistent with known neuroanatomy and neurophysiology and accurately accounts for neuroscientific data.

4.2 Neurocomputational Details

We built a computational cognitive neuroscience model of the Figure 4.1 network that includes spiking neurons as the basic units and that obeys the relevant neuroscience constraints (e.g., Ashby, 2018; Ashby & Helie, 2011). The model was programmed using the Python programming language (Van Rossum & Drake, 2011).

4.2.1 Model Architecture and Activation Equations

As described earlier, a rough schematic of MODAL is shown in Figure 4.1. Our main goal was to understand how changes in the value of the modulating variable affect the

rate of learning via modulations of VTA DA neuron activity. As a result, we made no attempt to model neural firing in hippocampus or upstream cortical regions. Modeling these complex structures is beyond the scope of the current project. Furthermore, we modeled activation in vSub, PPTN, and LH as either on or off (via square waves). Because our hypothesis is that the value of the modulating variable affects VTA DA neuron activity via the NAcc $\rightarrow$ VP $\rightarrow$ VTA pathway, we modeled activity in all these structures using spiking-neuron models; specifically Izhikevich (2007) medium spiny neuron (MSN) models for NAcc, quadratic integrate-and-fire models for VP (Ermentrout, 1996), and Izhikevich (2003) regular spiking neuron models for VTA. Parameter values for the Izhikevich units were set equal to the values used by Izhikevich (2007) and parameter values for the quadratic integrate-and-fire units were identical to those used in Ashby (2018), except when otherwise noted.[1]

Postsynaptic effects of a spike were modeled via the α-function (e.g., Rall, 1967). Specifically, when the presynaptic unit spikes, the input projected to the postsynaptic unit is (with spiking time $t = 0$):

$$\alpha(t) = \frac{t}{\delta}\exp\left(1 - \frac{t}{\delta}\right). \tag{4.3}$$

The parameter δ, which models temporal delays in synaptic transmission, was set to 123 for NAcc and VP units, and 225 for VTA units.

The following subsections describe additional details about how we modeled activity in NAcc, VP, and VTA. Table 4.1 lists values of all connectivity parameters. These parameter values were based on biological constraints (e.g., excitatory versus inhibitory). In its current form, MODAL does not exhibit any synaptic plasticity; therefore, all con-

[1]However, note that in Izhikevich (2007) and Ashby (2018), the β parameter controls the rate of tonic spiking. Each region in our model has a different tonic firing rate; therefore, $\beta = 0$ in NAcc, $\beta = 20$ in VP, and $\beta = 62$ in VTA.

Table 4.1: Connectivity parameter values between layers of MODAL

Parameter	Value
$w_{PPTN \to VTA_i}$	125
$w_{LH \to VTA_i}$	-125
$w_{NAcc_i \to VP_i}$	-10
$w_{VP_i \to VTA_i}$	-1000

nection weight parameters in this network were fixed throughout the simulations.

NAcc

The NAcc layer was modeled with 100 Izhikevich (2007) MSNs with input to NAcc_i (for $i = 1, 2, ..., 100$):

$$I_{NAcc_i}(t) = vSub(t), \tag{4.4}$$

where $vSub(t)$ represents activation in vSub as a square wave with amplitude equal to the value of the modulating variable. For simplicity, the tonic firing rate of NAcc in the absence of input was chosen to be 0 Hz, which is reasonable considering that Fabbricatore, Ghitza, Prokopenko, and West (2009) reported a tonic rate of 0.53 Hz.

Braganza and Beck (2018) hypothesized that the disinhibition motif that characterizes the basal ganglia plays the computational role of gating. However, in addition to gating DA via disinhibition, MODAL does this in a continuous fashion such that as the value of the modulating variable increases, the size of the population of VTA DA neurons also increases. In other words, whereas the disinhibition motif implements gating at the single synapse level, at the population level it can implement a gain or amplification of the signal. The striatal MSNs provide an excellent candidate for implementing the amplification. The MSNs exhibit bistable dynamics consisting of up and down states. In

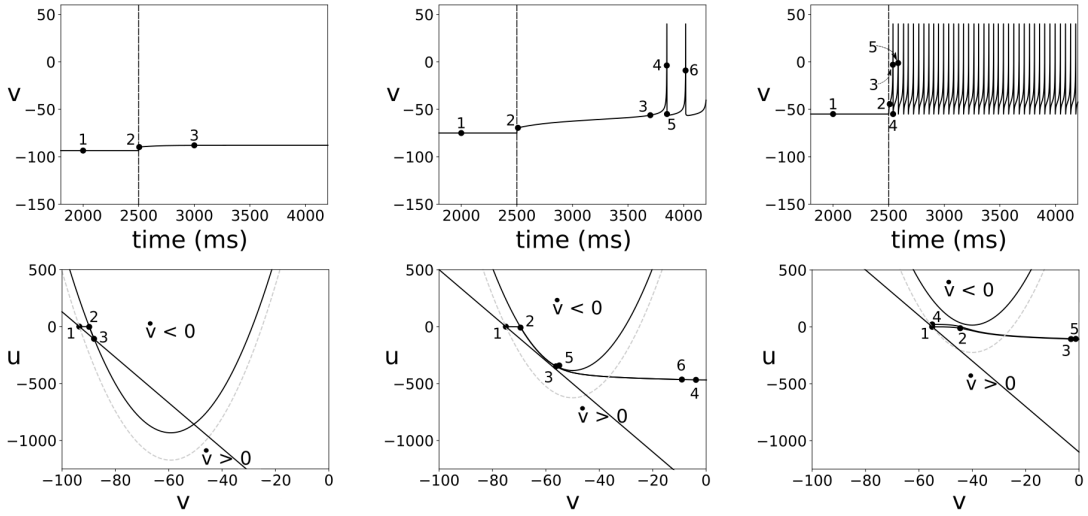

Figure 4.2: Top: spiking behavior of three NAcc medium spiny neurons with low (-93.5 mV, left), intermediate (-75 mV, center) and high (-55 mV, right) resting potentials. The vertical dashed line represents the time when the current is injected into the neuron. Bottom: Phase portraits with v and u nullclines and state trajectories for each of the neurons in the top panel. The v-nullcline is represented by the gray dashed line (no current) and the black u-shaped curve (current on). The u-nullcline is represented by the straight black line. When the current is turned on the v-nullcline shifts upwards. The numbers in the bottom plots correspond to the time points indicated by the numbers in the plots directly above.

the up state these neurons are responsive to inputs and will fire spikes, while in down states they tend not to fire in response to inputs.

In MODAL, the NAcc neurons play a critical role in controlling the size of the population of tonically active VTA DA neurons (i.e., partly because of their one-to-one connectivity through the VP). When a NAcc neuron transitions from its down state to its up state, the VP neuron it projects to is silenced due to NAcc inhibition. This releases the corresponding VTA neuron from tonic inhibition, causing it to fire tonically and become responsive to inputs from PPTN and LH. Therefore, although the NAcc neurons are also responsive to inputs while in their up state, their key role in MODAL is to determine the appropriate size of the active VTA DA neuron population.

A key property of MODAL is that the size of the population of tonically firing DA

neurons grows with increases in the value of the modulating variable. This requires NAcc neurons to transition from their down states to their up states at different levels of input to the vSub. To implement this property, each NAcc neuron in the model has a different resting state drawn randomly from a uniform distribution between -93.5 and -55 mV. Figure 4.2 shows the effect of the different resting states on the nullclines and state trajectories of three NAcc neurons (for more detail on the application of dynamical systems theory and phase portraits to neural modeling, interested readers may consult Izhikevich, 2007 or Ashby, 2018). The top row of Figure 4.2 shows predicted intracellular voltage levels for three NAcc neurons and the bottom row shows the corresponding phase portraits. The neurons are all identical, except for their resting membrane potential, which is low in column 1 (-93.5 mV), medium in column 2 (-75 mV), and high in column 3 (-55 mV). Notice that increasing the level of vSub activation causes all the v-nullclines to shift upwards. For the neuron with the lowest resting state (-93.5 mV, left), this upward shift is not sufficient to cause the neuron to undergo a saddle-node bifurcation (collision and annihilation of its fixed points) and therefore the state moves from the fixed point (1) to point 2 on the v-nullcline and it slides down the v-nullcline until it reaches the new fixed point (3) and the down state persists due to insufficient input from vSub (the neuron does not spike). Alternatively, for the neuron with the intermediate resting state (-75 mV, center), the upward shift in the v-nullcline is sufficient to cause the neuron to undergo a saddle-node bifurcation, moving the state from the fixed point (1) to point 2 on the v-nullcline. Due to the ghost of the saddle-node, the state slides slowly along the v-nullcline until point 3 when it leaves the v-nullcline and the derivative of v goes positive causing the voltage to increase rapidly, leading to a transition to the up state and spiking behavior. Once a spike is registered (4), the voltage is reset (5) below the ghost of the saddle-node leading to shorter latency spikes (6). The neuron with the highest resting state undergoes similar behavior to the intermediate neuron, except the latency to the

spike is substantially shorter. This is because the upward shift of the v-nullcline results in point 2 being below the v-nullcline, immediately leading to a positive derivative of v and rapid transition to the up state and spiking (3). Furthermore, following a spike the voltage is reset below the v-nullcline (and the ghost) (4) and therefore subsequent spikes are rapid (5).

The key result of this network architecture is that there is a continuum where neurons with lower resting states require substantially more current to undergo the saddle-node bifurcation, relative to neurons with higher resting states. This has the desired effect of increasing the size of the VTA DA population as the value of the modulating variable increases by transitioning more NAcc neurons into their up states.

VP

The VP layer was modeled with 100 quadratic integrate-and-fire units. The input to VP_i $(i = 1, 2, ..., 100)$ was equal to:

$$I_{VP_i}(t) = w_{NAcc_i \to VP_i} \times \alpha_{NAcc_i}(t) \tag{4.5}$$

where $w_{NAcc_i \to VP_i}$ is the connection weight between NAcc_i and VP_i and $\alpha_{NAcc_i}(t)$ is the integrated α-function generated by spikes in NAcc_i.

The tonic firing rate for VP units was set to approximately 7 Hz, which is consistent with measurements reported by Root et al. (2012). Despite NAcc inhibition, the higher tonic firing rate of the VP units relative to the NAcc units has the effect of ensuring that the VP units still fire spikes at low values of the modulating variable. As the modulating variable increases, more NAcc neurons transition to their up states, silencing more VP units. Each VP unit is connected to one VTA unit.

VTA

The DA neurons in the VTA were modeled with 100 Izhikevich regular-spiking neurons. Call these units VTA_i ($i = 1, 2, ..., 100$). All 100 units received identical input from PPTN and LH. In addition, the VTA_i unit received input from the corresponding VP_i unit.

The input to unit VTA_i was:

$$I_{VTA_i}(t) = \begin{cases} w_{VP_i \to VTA_i} \alpha_{VP_i}(t) \\ \qquad + w_{PPTN \to VTA} PPTN(t) + w_{LH \to VTA} LH(t) \end{cases} \tag{4.6}$$

where $w_{VP_i \to VTA_i}$ denotes the synaptic strength between VP_i and VTA_i, $\alpha_{VP_i}(t)$ denotes the output of unit VP_i at time t (i.e., the α-function), $w_{PPTN \to VTA}$ denotes the synaptic strength between PPTN and all VTA neurons and $w_{LH \to VTA}$ denotes the synaptic strength between LH and all VTA neurons.

Activation in PPTN was modeled as follows:

$$PPTN(t) = \begin{cases} RPE & \text{if } RPE > 0 \text{ and } 7000 \leq t < 7100 \\ 0 & \text{otherwise} \end{cases} \tag{4.7}$$

This results in a square wave with amplitude equal to RPE (for positive RPE) lasting 100ms (Bayer, Lau, & Glimcher, 2007). Activation in LH was:

$$LH(t) = \begin{cases} 1 & \text{if } RPE < 0 \text{ and } 7000 \leq t < (7000 - 400RPE) \\ 0 & \text{otherwise} \end{cases} \tag{4.8}$$

This results in a square wave of amplitude equal to 1 (for negative RPE) of varying duration (0 – 400 ms) that elicits pauses in VTA units with the length of the pause proportional to the magnitude of the negative RPE[2]. This formulation of activity in PPTN and LH produces DA neuron firing that is proportional to RPE (Bayer & Glimcher, 2005) and results in symmetric encoding of positive and negative RPE by extracellular DA concentrations (Hart et al., 2014).

4.2.2 Single-Neuron Dynamics of MODAL

The dynamics of three neurons in each layer of MODAL are illustrated in Figure 4.3. The level of input to vSub increases linearly from 0.0001 to 1 in increments of 0.0001. At the beginning of the 10-second interval, vSub activation is low and all of the NAcc neurons are in their down state. However, as vSub input increases slightly, the first NAcc unit from the left transitions to its up state and increases its firing rate, which silences the first VP unit. This disinhibits the first VTA unit and it begins to fire tonically, making it possible for this unit to burst or pause in response to input from PPTN or LH. Similarly, as vSub input increases further, the second (center) and then the third (right) NAcc units also transition to their up states, which first silences the second and then the third VP units, respectively. This causes the second and then the third VTA units to become disinhibited and fire tonically, making them available for bursting or pausing as well. This network structure creates the desired effect of having a larger pool of VTA DA units available for bursting and pausing as the level of vSub input rises. The level of vSub activation that determines the size of the VTA DA neuron population depends on the preferred resting states of each of the NAcc neurons. The PPTN activation in this

[2]However, for Figures 4.3, 4.4, 4.5 (left and center), and 4.6, the PPTN square wave lasted 1000ms and the LH square wave lasted a maximum of 1000ms. This was done to ensure a sufficiently long interval to extract accurate measurements of firing rate and active population size. Figures showing dopamine output used the parameters described in the text.

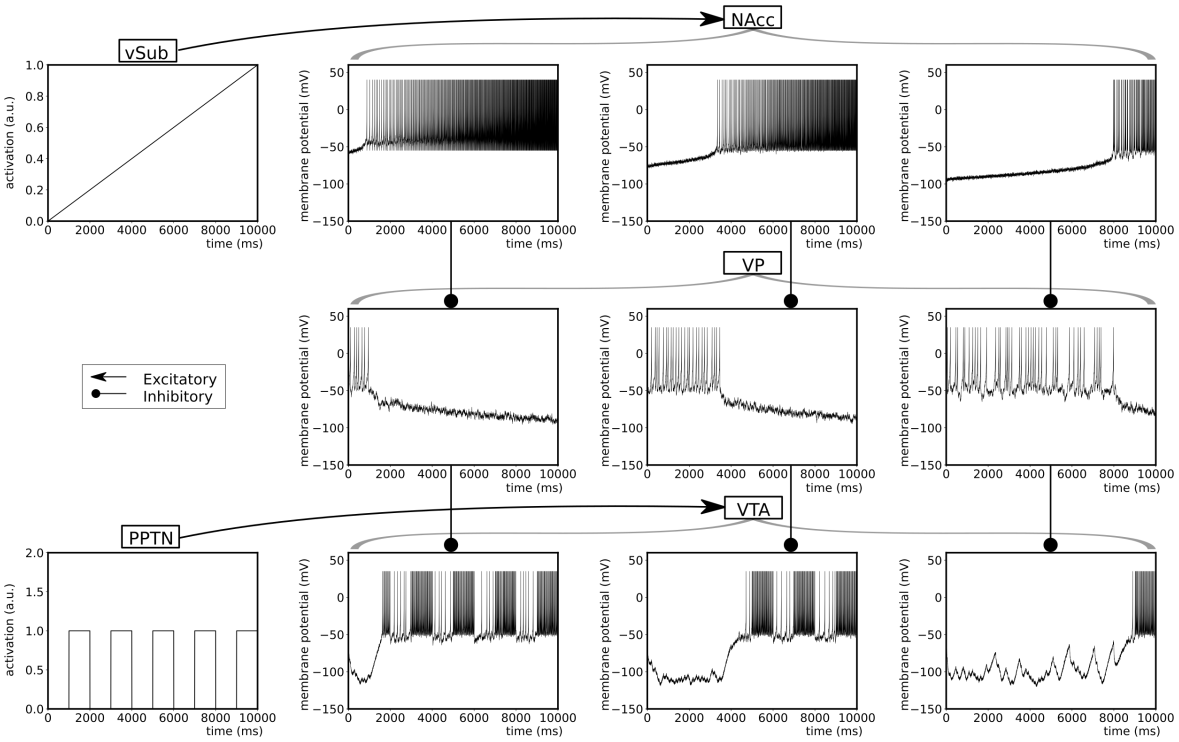

Figure 4.3: Spike dynamics in the simplified network for a 10 second interval. As activation in vSub increases this causes more NAcc neurons to transition into their up state, thereby inhibiting more VP neurons and disinhibiting more VTA neurons. The result of this architecture is a recruitable pool of VTA DA neurons that can respond to inputs from PPTN and LH (not shown). See text for more details on the dynamics. vSub = ventral subiculum, NAcc = nucleus accumbens, VP = ventral pallidum, VTA = ventral tegmental area, PPTN = pendunculopontine nucleus.

simulation alternates between 0 and 1 for 1000 msec intervals. Notice that as long as the corresponding NAcc neuron is silent, the VTA neuron is unresponsive to inputs from PPTN. However, once the NAcc firing rate is high enough to disinhibit VTA neurons, they now alternate between periods of tonic firing and phasic bursting (or tonic firing and phasic pausing, not shown in Figure 4.3).

4.3 Methods

The proposed model was evaluated using numerical simulations on two different types of data from nonhuman animals: single-unit recordings and fast-scan cyclic voltammetry. The goal of the simulations was to test the neural architecture of the model, not parameter optimization. Hence, it is important to note that although this network includes a number of parameters, the majority of these were fixed after modeling each level of experimental data. In particular, the parameters that were estimated when fitting the single-unit data of Lodge and Grace (2006) then remained fixed at those values in all future simulations. This process ensured that the network is able to account simultaneously for experimental data at many levels of analysis and implicitly implements a significant degree of inflexibility into the model's structure by constraining it by the lower levels of analysis. The key parameters that were modified will be described in each section. For all simulations, the voltage for each unit was estimated for each msec of a 10,000-msec trial.

In the neural simulations, there was no learning and noise was minimal; therefore, results are from a single simulation. We ran the neural network through simulations with 100 levels of the value of the modulating variable (from 0.01 to 1 by increments of 0.01) and 201 values of RPE (from -1 to 1 by increments of 0.01). The amount of DA released by the network was computed for each combination of learning rate and RPE, resulting in a total of 20,100 DA measurements.

4.4 Results

MODAL was subjected to three neural benchmark tests. First, we explored whether it could account for the Lodge and Grace (2006) results showing that vSub activation

increases the number of tonically active VTA DA neurons, whereas activation of the PPTN induces burst firing of VTA DA neurons. Second, we examined whether the model was consistent with the data of Bayer and Glimcher (2005), which showed that DA neuron firing increases linearly with RPE between minimal and maximal values. Third, we tested whether the model could account for the data of Hart et al. (2014), which showed that DA release (i.e., extracellular DA concentration) is a linear function of RPE and that positive and negative RPEs are encoded symmetrically.

4.4.1 Neural Tests of MODAL

Benchmark test 1: Distinct pathways in PPTN and NAcc

Lodge and Grace (2006) provided evidence that distinct interacting pathways exhibit differential influences on VTA DA neurons. In this experiment, they activated vSub, PPTN, or both structures via NMDA infusion, and then they counted the number of VTA DA neurons that were firing tonically (i.e., per electrode track), and they also estimated the average firing rate of all VTA DA neurons. Their results are shown in the left column of Figure 4.4. Note that vSub activation increased the number of tonically active VTA DA neurons but did not affect the population firing rate. On the other hand, activation of the PPTN caused an increase in the population firing rate as a result of burst firing, but did not affect the size of the tonically active population. Finally, simultaneous activation of vSub and PPTN caused a significant increase in both burst firing and the size of the tonically active population.

MODAL fits are shown in the right column of Figure 4.4. We simulated the control condition of Lodge and Grace (2006) by setting the square-wave activations of vSub, LH, and PPTN to 0.27, -0.31, and 0, respectively. Activation of vSub by NMDA infusion was simulated by changing the amplitude of the square wave activation of vSub to 1. Acti-

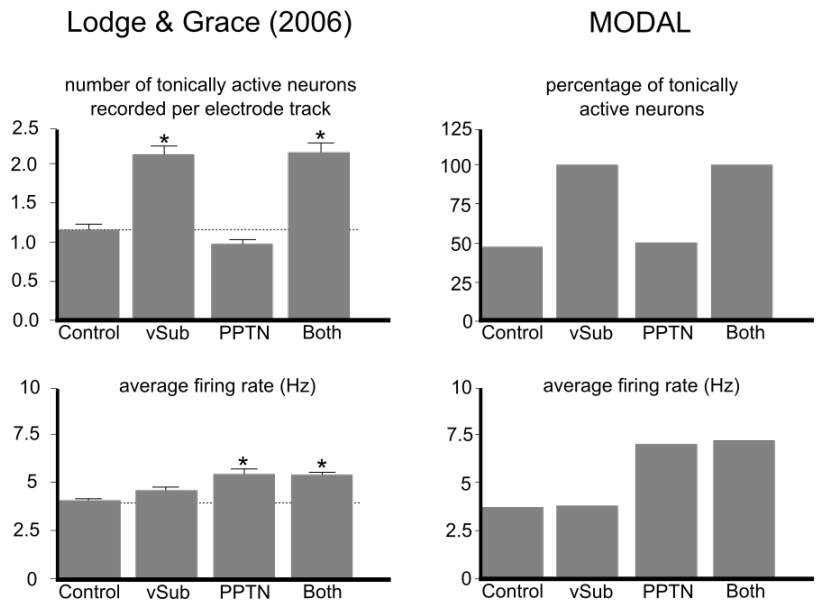

Figure 4.4: Benchmark Test 1. Left panel: Experimental data (left) from Lodge and Grace (2006). Right panel: MODAL simulations of the same experiment. Plots of the experimental data are reprinted and modified from Lodge and Grace (2006). vSub = ventral subiculum, PPTN = pedunculopontine nucleus.

vation of PPTN was simulated by changing the amplitude of the square-wave activation of PPTN to 0.05. Note that the model accurately captures all qualitative features of the data. It should also be noted that the connection weights between PPTN and the VTA units were set so that excitatory input from PPTN to a tonically firing VTA neuron was sufficient to result in burst firing according to the criteria used by Lodge and Grace (2006) [i.e., an interspike interval (ISI) of $\leq$ 80ms and bursting that persists until the ISI exceeds 160 ms; Grace & Bunney, 1983].

Figure 4.5 shows heat-maps that depict the number of active neurons and population firing rate as a function of vSub activation and RPE (Figure 4.5 left and center, respectively). These plots show that the overall qualitative behavior reported by Lodge and Grace (2006) is implemented in MODAL across a wide range of input values for vSub activation and RPE. The population size of tonically firing DA neurons increases with

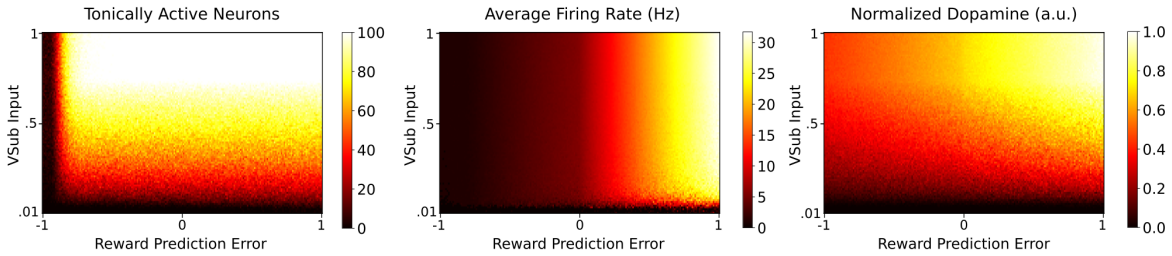

Figure 4.5: Heatmaps showing the number of tonically active DA neurons (left), average population firing rate (center) and normalized DA release as a function of vSub input and RPE (right).

vSub activation, but is relatively independent of RPE, and therefore of PPTN/LH activation while population firing rate increases with PPTN/LH activation but is relatively independent of vSub activation. Furthermore, Figure 4.5 (right) shows the predicted extracellular DA concentration for the Lodge and Grace (2006) experiment, with minimal DA release when both vSub and PPTN activation are low, limited DA release when only one of vSub or PPTN has high activation, and maximal DA release for concurrent high activity in vSub and PPTN. We assumed extracellular DA concentrations would be proportional to the total postsynaptic effects of all DA units in the model and so we estimated extracellular DA concentrations as the integral of each VTA neuron's α-function for a 1-second period following reward and summed over all neurons in the population.

Benchmark test 2: Single-unit recordings from DA neurons

Bayer and Glimcher (2005) recorded from single midbrain DA neurons (VTA and SNpc) while monkeys performed a task that required them to learn to make appropriately timed eye movements. Correct responses were rewarded with a small amount of juice. Bayer and Glimcher (2005) found that the response of the midbrain DA neurons was proportional to an estimate of the RPE (the difference between obtained reward value and a weighted average of previous reward values). Their results from a population of

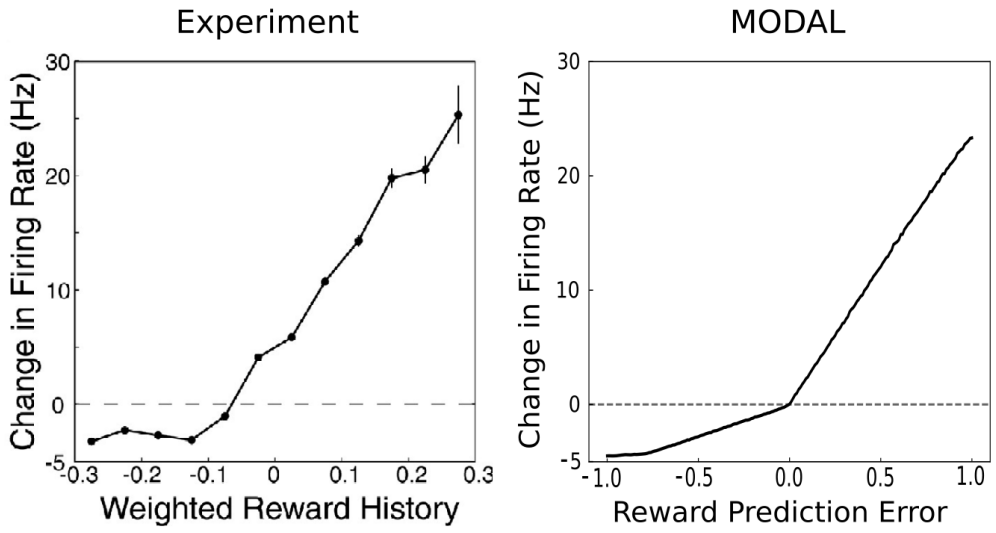

Figure 4.6: Benchmark Test 2. Left panel: Firing rate of a population of midbrain DA neurons as a function of RPE (from Bayer & Glimcher, 2005). Right panel: MODAL simulations of the same experiment. Plots of experimental data are reprinted and modified from Bayer and Glimcher (2005).

midbrain DA neurons are shown in the left panel of Figure 4.6. Note that the increase in firing rate is linear after RPE exceeds a minimum value (i.e., of around -0.1).

Simulations of the model under similar conditions are shown in the right panel of Figure 4.6. Note that the model accurately captures the qualitative properties of the data. In these simulations we set the tonic firing rate of VTA neurons to approximately 5 Hz (to match data reported by Bayer et al., 2007). For simplicity, the LH $\rightarrow$ VTA and PPTN $\rightarrow$ VTA connection weights were set to be equal, which was sufficient to cause VTA neurons to pause in response to negative RPEs. The right panel of Figure 4.6 shows the average firing rate of the VTA population.

Benchmark test 3: Extracellular DA levels in NAcc

The DA neuron firing-rate data shown in Figure 4.6 suggest a more limited dynamic range for encoding negative as opposed to positive RPEs. In particular, the amount

of increase in firing rate observed for positive RPEs was considerably greater than the amount of decrease seen for negative RPEs. Bayer and Glimcher (2005) speculated that negative RPEs might also be encoded by pause duration, and Bayer et al. (2007) later reported evidence supporting this hypothesis. Of course, synaptic effects of DA are more closely related to extracellular DA levels than to DA neuron spiking. For this reason, Hart et al. (2014) used fast-scan cyclic voltammetry to examine how extracellular DA levels in the rat NAcc varied as a function of RPE. Their results are summarized in the left panel of Figure 4.7. Note that the phasic bursting and pausing of midbrain DA neurons results in symmetric encoding of positive and negative RPEs in extracellular DA concentrations.

Our third benchmark test was to ask whether a model constrained by benchmark tests 1 and 2 could also account for the symmetric encoding of positive and negative RPEs shown in Figure 4.7. Therefore, we simulated performance of the model in the Hart et al. (2014) experiment by choosing the maximum duration of LH activation to be 400ms (see Equation 4.8). All other parameter estimates from benchmark tests 1 and 2 were fixed. As in benchmark 1, we assumed that extracellular DA concentrations would be proportional to the total postsynaptic effects of all VTA units in the model and so we estimated extracellular DA concentrations as the integral of each VTA neuron's α-function for a 1-second period following reward and summed over all neurons in the population.

The results are shown in the right panel of Figure 4.7 for a variety of different levels of vSub activation. Note that MODAL accounts for the symmetric encoding of positive and negative RPEs seen in the Hart et al. (2014) data, and it does this for all levels of vSub activation. But note that the model also makes an important novel, and to our knowledge, untested prediction – decreasing the level of vSub activation (via decreases in the value of the modulating variable) should decrease the slope of the regression line

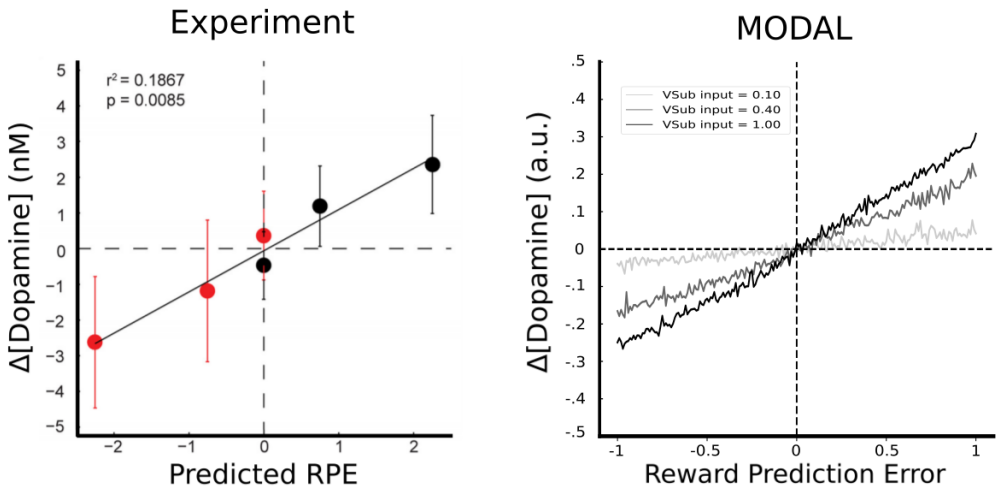

Figure 4.7: Benchmark Test 3. Left panel: Experimental measurements of extracellular DA concentrations in NAcc as a function of RPE (from Hart et al., 2014). Right panel: MODAL simulations of the Hart et al. (2014) experiment for a variety of different levels of feedback contingency.

that best fits the observed extracellular DA concentrations.

The Hart et al. (2014) results shown in the left panel of Figure 4.7 were averaged across results from three conditioning tasks – two that used probabilistic feedback and one that used deterministic feedback. Note that for many of the proposed modulating variables, probabilistic and deterministic feedback would likely lead to predictable differences. Therefore, our model predicts that the linear relationship evident in Figure 4.7 is likely a result of averaging across three distinct linear curves.

Although not evident in Figure 4.7, note that MODAL also predicts that decreases in vSub activation should result in a downward shift in baseline or tonic concentrations of extracellular DA. This is because a reduction in vSub activation reduces the number of VTA units that are tonically firing, which reduces the number of VTA units contributing to the baseline concentration of extracellular DA.

4.5 Discussion

This article proposed a neurobiologically detailed spiking neural network model that varies the size of the population of tonically firing DA neurons in response to environmental changes. The model makes specific quantitative predictions about how changes in the size of this population alter baseline DA levels and the gain on the DA response to any given RPE. This new model successfully accounts for two single-cell recording data sets and results from a fast-scan cyclic voltammetry study.

A strong theory should make novel predictions. We highlighted two novel predictions of the model proposed here. First, any experimental change that reduces the value of the modulating variable should reduce the magnitude of change in NAcc extracellular DA concentrations for any given change in RPE. This prediction is illustrated in Figure 4.7. Second, the model predicts that decreasing the value of the modulating variable should decrease tonic extracellular DA levels due to the decreased size of the active VTA DA neuron population. To our knowledge neither of these predictions of MODAL have been tested. It should be noted however, that recent evidence suggests that testing this latter prediction may be complicated by effects of local mechanisms on extracellular DA concentrations (Berke, 2018).

4.5.1 Behavioral Applications

The MODAL network illustrated in Figure 4.1 includes no motor units, nor any units associated with motor planning or decision making. As a result, in its current form, MODAL produces no behavior and therefore, without some significant augmentation, it cannot be tested against behavioral data. Even so, MODAL makes strong predictions about how DA levels will vary trial-by-trial in any brain region that is a target of VTA DA neurons. This includes regions such as prefrontal cortex, hippocampus, amyg-

dala, ventral striatum and the most anterior portions of the dorsal striatum (e.g., head of the caudate nucleus). Therefore, MODAL could be combined with any model that accounts for behavior with a neural network that includes these regions and assigns a functional role to DA. The result should be a more powerful model of the behavior that can dynamically adjust tonic and phasic DA release in response to environmental changes in some modulating variable such as volatility, environmental uncertainty, or feedback contingency. Many such models have been proposed – far too many to review here. Furthermore, because DA projections are diffuse, rather than synapse specific, MODAL should be able to interface with a wide variety of computational models – not just those that include a high level of biological detail.

This section briefly discusses three qualitatively different types of behavioral applications of MODAL: 1) to models of value learning that could benefit from a more accurate model of reward-driven phasic DA firing; 2) to models of executive function that posit a modulatory role for cortical DA; and 3) to models of procedural learning in which synaptic plasticity depends on DA neuron activity in the substantia nigra pars compacta (SNpc).

The primary motivation for the creation of MODAL is to provide a neurocomputational mechanism for how changes in the environment can modulate the learning rate (i.e., λ_n in Eq. 4.1). Takahashi et al. (2008) proposed a model in which the ventral striatum encodes state values similar to those generated by Eq. 4.1. MODAL could be conjoined with this model since the ventral striatum is a primary target of VTA DA neurons. Furthermore, it has also been reported that the ventral striatum plays a key role in probabilistic reversal learning (Cools, Clark, Owen, & Robbins, 2002). Behrens et al. (2007) proposed a Bayesian model of reversal learning in which the learning rate changes with the volatility of the environment. Their model is purely computational and makes no attempt to describe any of the underlying neural circuitry. Therefore, MODAL

could be integrated with the Behrens et al. (2007) model to produce a more biologically detailed model of reversal learning.

Within the striatum, DA is quickly cleared from synapses by DA active transporter (DAT) and, as a result, the temporal resolution of DA in the striatum is high enough so that DA levels roughly track phasic DA neuron firing. Unlike the striatum however, DAT concentrations in frontal cortex are low (e.g., Seamans & Robbins, 2010). As a result, cortical DA levels change slowly – too slowly to track phasic DA activity. Even so, MODAL could be used in conjunction with almost any model of executive function that assigns a functional role to cortical DA levels. For example, Ashby, Valentin, and Turken (2002) proposed a connectionist network model of creative problem solving that mapped loosely onto the anterior cingulate, prefrontal cortex, and head of the caudate nucleus. Although the model included little neuroanatomical detail, it made specific quantitative predictions about the effects of changing DA levels on cognitive flexibility and creative problem solving. No model of DA release was included, so MODAL could be used to fill this role.

Although MODAL could be used to predict changes in DA levels in any VTA DA target region and in virtually any task, it is important to note that how these changes affect behavior might be task and brain-region dependent. For example, in some tasks that depend on executive function, performance is an inverted U-shaped function of DA level. This includes creative problem solving and cognitive flexibility (Ashby, Isen, & Turken, 1999; Cools & D'Esposito, 2011; Cools & Robbins, 2004; Cools, 2006). Cools (2006) suggested that optimal levels of prefrontal DA facilitate the maintenance of stable representations, whereas optimal levels of striatal DA underlay cognitive flexibility. Therefore, changing global levels of DA can have different implications for task performance depending on local dynamics. Accordingly, although MODAL modulates DA input to these regions, the implications of changing global DA levels for behavior and

performance will require region-specific models that consider local baseline DA levels, re-uptake mechanisms, receptor dynamics, and interactions between regions.

A more challenging goal is to extend MODAL to the DA neurons in the substantia nigra pars compacta (SNpc). The vSub → NAcc → VP pathway shown in Figure 4.1 projects to VTA but not to SNpc, so accounting for changes in SNpc DA firing when environmental uncertainty or feedback contingency changes requires a different neuroanatomical model. This problem is complicated by recent evidence suggesting that despite many similarities, VTA and SNpc DA have dissociable roles (Keiflin, Pribut, Shah, & Janak, 2019). Furthermore, VTA and SNpc DA neurons project to different (but overlapping) targets. In particular, the dorsal striatum receives its DA projection almost exclusively from the SNpc (e.g., Smith & Kieval, 2000). This is important because there is overwhelming evidence that procedural learning is mediated within the basal ganglia, and especially at cortical-striatal synapses in the dorsal striatum (e.g., Ashby & Ennis, 2006; Houk et al., 1995; Mishkin, Malamut, & Bachevalier, 1984; Willingham, 1998). Therefore, to interface MODAL with models of the dorsal striatum and/or models of procedural learning, the model must be generalized to include the SNpc. One possibility is to model the spiraling architecture of the basal ganglia that enables activity in the ventral striatum (i.e., the NAcc) to influence the central striatum, which then influences the dorsolateral striatum (Takahashi et al., 2008; Haber, Fudge, & McFarland, 2000; Belin & Everitt, 2008). In fact, an existing actor-critic model of the basal ganglia already relies on this spiraling architecture (Takahashi et al., 2008).

4.5.2 Relation to RPE Models

The model proposed here describes how changes in some modulating variable affect the DA response to RPE. But note that the model makes no assumptions about the

neural networks that compute RPE. Many models of these circuits have been proposed (Brown et al., 1999; Cohen, Haesler, Vong, Lowell, & Uchida, 2012; Contreras-Vidal & Schultz, 1999; Eshel et al., 2015; Hazy et al., 2010; Houk et al., 1995; Joel et al., 2002; Humphries & Prescott, 2010; Kawato & Samejima, 2007; Morita, Morishima, Sakai, & Kawaguchi, 2012; Morita et al., 2013; O'Reilly et al., 2007; Salum, da Silva, & Pickering, 1999; Schultz et al., 1997; Schultz, 1998; Stuber et al., 2008; Sutton & Barto, 1998; Tan & Bullock, 2008; Vitay & Hamker, 2014). MODAL does not generate behavior; therefore, rather than compute RPE using one of these proposed circuits, we chose to project hypothetical values of RPE (ranging from -1 to +1) to the VTA units via the PPTN or LH. This network architecture is consistent with accounts that midbrain DA neurons receive the signals necessary for computing RPE from upstream regions via the PPTN (Hong & Hikosaka, 2014; Kobayashi & Okada, 2007; Okada & Kobayashi, 2013), LH (Tian & Uchida, 2015; Hong et al., 2011; M. Matsumoto & Hikosaka, 2007, 2009), and RMTN (Jhou, Fields, Baxter, Saper, & Holland, 2009).

We chose not to model the neural networks that compute RPE in an effort to provide a stronger test of the hypothesis that effects of the modulating variable on the DA response to RPE are mediated by a circuit that includes vSub, NAcc, and VP. Adding neural structures to compute RPE would increase the complexity of the model, thereby making it more difficult to attribute a success or failure of the overall network to one specific subnetwork. Even so, one advantage of the modeling approach followed here is the potential to develop 'plug-and-play' models of different neural networks (Ashby, 2018; Cantwell, Riesenhuber, Roeder, & Ashby, 2017). Because MODAL is consistent with known neuroanatomy and neurophysiology, it should be possible to wire it into an existing similarly constrained model of the networks that compute RPE or networks that compute the modulating variable. This exercise is beyond the scope of the current application.

4.5.3 Relation to Existing Neural Models of Learning Rates

Several alternative neural accounts of how learning rates are modulated have been proposed. None of these include DA neurons however, and thus, to our knowledge, none can account for any of the neural data we considered in our benchmark tests. This section briefly discusses the more prominent of these alternative accounts, with a special emphasis on their relation to MODAL.

Bernacchia et al. (2011) reported single-unit recording results from monkeys that showed evidence that different neurons in ACC, PFC, and lateral intraparietal cortex are differentially sensitive to the time since the last reward. Based on these results, they proposed a neural network model in which a reservoir of such neurons could be used to dynamically alter learning rates, depending on how quickly environmental reward probabilities are changing.

Similarly, Farashahi et al. (2017) proposed that the ACC adjusts learning rates in response to environmental changes in reward probabilities via synaptic metaplasticity, which is a synaptic change that alters the plasticity of the synapse to future events, without altering the efficacy of current synaptic transmission. Specifically, they proposed that the ACC may be endowed with metaplastic synapses that can switch between strong and weak meta states, effectively changing the learning rate.

It is important to note, however, that neither of these proposals made any attempt to describe how the learning rate selected from the reservoir or computed in the ACC via metaplasticity, modulates neural plasticity in other brain networks. Thus, rather than competing with MODAL, these models could be viewed as candidate models for the network (or part of the network) that computes the modulating variable that serves as input to vSub in MODAL.

In contrast, a model that more directly competes with MODAL was proposed by

Franklin and Frank (2015). According to this account, the pause duration of tonically active cholinergic neurons (TANs) in the striatum signals uncertainty and modulates learning by controlling the activity of the MSN population through a feedback loop. In this model, the TAN pauses are driven by input from the striatal MSNs. High entropy in the MSN population leads to long TAN pauses, which result in fast initial learning, whereas low entropy in the MSN population leads to short pauses, which result in slow initial learning. The result is a neural network that implements a dynamic learning rate that enables rapid learning after a reversal.

This is an interesting hypothesis that deserves further testing. Even so, it faces several significant challenges. First, Franklin and Frank (2015) acknowledged that they are unaware of any empirical support for the claim that TAN pause durations are modulated by MSN activity. Second, their model omits the strongest excitatory glutamatergic inputs to the TANs, which come from the caudal intralaminar nuclei of the thalamus (Cornwall & Phillipson, 1988; Sadikot, Parent, & Francois, 1992). Furthermore, simultaneous single-unit recordings from these thalamic neurons and from TANs show that thalamic activity is required for the TANs to pause (N. Matsumoto, Minamimoto, Graybiel, & Kimura, 2001). Third, there is evidence (acknowledged by Franklin & Frank, 2015) that DA also modulates the duration of TAN pauses (Deng, Zhang, & Xu, 2007; Doig, Magill, Apicella, Bolam, & Sharott, 2014; Ding, Guzman, Peterson, Goldberg, & Surmeier, 2010).

An alternative account of TAN activity was proposed by Ashby and Crossley (2011), who hypothesized that the main functional role of the TANs is to serve as a gate between cortex and the striatum. The TANs tonically inhibit cortical inputs to the striatum, so the default state of the gate is closed. However, environmental cues that signal reward cause the TANs to pause (via excitatory input from thalamus), which opens the gate and allows cortical-striatal plasticity. Furthermore, Crossley et al. (2013) proposed a model that included this role for the TANs, which protects cortical-striatal synapses when state-

feedback contingency is low (e.g., as when the feedback is random), by eliminating the TAN pause to cues that formerly predicted reward (and thereby closing the gate). In this model, decreases in DA are necessary for the TANs to unlearn the pause response. Crossley et al. (2013) made no attempt to describe a neural circuit via which state-feedback contingency could modulate the amount of DA released, so MODAL could be combined with the Crossley et al. (2013) model to provide a more complete description of these contingency-related phenomena.

In summary, there are few true competitors to MODAL, but many models that could be combined with MODAL to produce a more powerful model than any that currently exists. Models in this latter class are of two types. One type, which includes the models of Bernacchia et al. (2011) and Farashahi et al. (2017), could be used to compute the value of the modulating variable that is the input to vSub in MODAL (see Figure 4.1). Another type, which includes the Crossley et al. (2013) model, could use MODAL to compute the amount of DA released to feedback during each trial of some learning task. When combined in this way, MODAL would act as the critic, and the other model as the actor in an actor-critic architecture.

4.5.4 Neural Basis of Modulating Variables

MODAL proposes a neural account of how some modulating variable could affect the DA response to RPE and therefore learning rates in the brain. Many such variables have been proposed. MODAL does not require that the computation of all these putative variables are mediated by the same neural network, but it does require that the variable, whatever it is, is mediated by a network that sends a prominent projection to the vSub. Fortunately, almost all hypothesized modulating variables seem to meet this requirement. For a review of the numerous brain regions involved in coding uncertainty, see Soltani

and Izquierdo (2019).

The vSub receives dense projections from the hippocampal CA1 subfield and from entorhinal cortex (Kerr et al., 2007), and these regions receive input from many areas of frontal cortex, including large portions of PFC, orbitofrontal cortex, and ACC (e.g., Gloor, 1997). For example, entorhinal cortex receives almost all of its cortical inputs from polymodal association areas, including cingulate, orbitofrontal and parahippocampal cortices (Insausti et al., 1987; Jones & Witter, 2007).

Almost all modulating variables are thought to depend on one or more of these regions. For example, the ACC seems to play a significant role in encoding volatility (Behrens et al., 2007), uncertainty (Rushworth & Behrens, 2008), and valence-specific uncertainty (Monosov, 2017). Activity in orbitofrontal cortex has been shown to correlate with uncertainty (Jo & Jung, 2016; O'Neill & Schultz, 2010) and additional evidence suggests that it may play a role in unexpected uncertainty and volatility (Riceberg & Shapiro, 2012).

The encoding of unexpected uncertainty has been found in the posterior cingulate cortex, a portion of the postcentral gyrus and posterior insular cortex, the left middle temporal gyrus, the left hippocampus and the locus coeruleus (Payzan-LeNestour, Dunne, Bossaerts, & O'Doherty, 2013). The encoding of estimation uncertainty has been found in the ACC extending to the posterior dorsomedial PFC, bilateral dorsolateral PFC and a portion of the inferior parietal lobule (Payzan-LeNestour et al., 2013). The encoding of risk was found in the inferior frontal gyrus (Payzan-LeNestour et al., 2013; Huettel, Song, & McCarthy, 2005) and a portion of the lingual gyrus (Payzan-LeNestour et al., 2013), the adjacent anterior insula (Huettel et al., 2005; Preuschoff, Quartz, & Bossaerts, 2008) and the ACC (Christopoulos, Tobler, Bossaerts, Dolan, & Schultz, 2009). Preuschoff et al. (2008) found that activation in the insula encodes risk and risk prediction errors and Jo and Jung (2016) found that the anterior insula encodes signals related to reward

uncertainty. Activity in the hippocampus has been found to correlate with uncertainty (Harrison, Duggins, & Friston, 2006; Vanni-Mercier, Mauguiere, Isnard, & Dreher, 2009; Strange, Duggins, Penny, Dolan, & Friston, 2005). Furthermore, Payzan-LeNestour et al. (2013) noted the similarity between unexpected uncertainty and novelty detection and therefore the role played by the hippocampus in novelty detection (Rutishauser, Mamelak, & Schuman, 2006) may be relevant. This proposal is particularly interesting when considering how the role of the hippocampus in mismatch detection may relate to the detection of changes in the environment (Kumaran & Maguire, 2006). Dayan and Yu (2003) proposed that the effects of expected and unexpected uncertainty are mediated in cortex by acetylcholine and norepinephrine, respectively (Yu & Dayan, 2005). This is relevant because Lipski and Grace (2013) showed that norepinephrine and locus coeruleus activation can modulate the activity of neurons in vSub and Bortz and Grace (2018) showed that the modulation of VTA DA population size depends on cholinergic mechanisms in vSub. Additionally, lesions to the ventral striatum in monkeys have been shown to reduce learning rates in stochastic tasks, which is consistent with the role of NAcc in our model (Taswell, Costa, Murray, & Averbeck, 2018). Finally, the medial septum has been shown to play a role in reversal learning in rats by controlling the size of active midbrain DA neurons and this effect was mediated via projections from medial septum to vSub (Bortz, Gazo, & Grace, 2019).

The architecture of MODAL implies that tonic DA encodes the learning rate. Therefore, our model is consistent with the proposal by Friston et al. (2012) suggesting that tonic DA encodes precision, that is, the learning rate in Bayesian models of learning under uncertainty (Mathys et al., 2011). Furthermore, using precision as a modulating variable in MODAL would enable our network to implement precision-weighted prediction errors.

Niv, Daw, Joel, and Dayan (2007) proposed that tonic DA levels encode the average rate of reward in free-operant tasks. In the data used to test this model, pigeons and rats

were trained in steady-state environments in which reward contingencies did not vary. Thus, the main modulating variables considered in this article, including uncertainty, volatility, and feedback contingency, are likely to have remained constant as well. As a result, more research is needed to distinguish between the average-reward-rate hypothesis and MODAL. Another possibility, however, is that average reward rate could be treated as a modulating variable that serves as input to MODAL.

4.5.5 The Benefit of Multiple levels of Analysis and Future Research

We proposed an implementational-level model of how any of a variety of different modulating variables could control the gain on the DA response to RPE, and therefore implement dynamic learning rates. Although computational and algorithmic levels of analysis have been successful in accounting for behavioral phenomena, moving to the implementational level allows us to further constrain the models by the underlying neuroanatomy and neurophysiology, and brings to light questions that may not have been proposed at higher levels of analysis. Some questions that arise due to the implementational-level modeling are: (1) what are the various computations encoded in cortical circuits that may act as inputs to MODAL? (2) In neural models of RPE, tonic DA levels often represent zero RPE; however, what happens when the tonic DA levels change? Note that MODAL predicts that increases in the value of the modulating variable should increase tonic concentrations of extracellular DA, even though it will not increase tonic firing rates in active DA neurons (however, note that local control mechanisms may also need to be considered; Berke, 2018). To our knowledge, this prediction is untested and therefore should be investigated in detail. (3) What is the cellular or molecular mechanism that causes silent DA neurons to begin firing tonically? We modeled this transition

by assuming there is variability in the resting state potential across the population of NAcc neurons. Topologically, this assumption caused some neurons to be further from a saddle-node bifurcation, and therefore to require more input current for the fixed points to collide and annihilate. However, because the Izhikevich MSNs are phenomenological models of neural spiking, there are other possible mechanisms that could lead to similar dynamical behavior. Future research should test our proposed mechanism and investigate other possibilities.

Modeling at the implementational level has significant implications for disease states. Knowledge of the neurophysiology of disease can lead to hypotheses for models at the computational and algorithmic levels. For example, empirical evidence indicates that in schizophrenia, the DA system is in overdrive due to aberrant regulation of midbrain DA neurons by the vSub (Grace, 2010). If the predictions of MODAL are considered, this kind of knowledge has implications for performance in a variety of behavioral tasks in people with schizophrenia.

Future research should extend our model upwards by investigating how MODAL could be integrated with the various circuits that have been proposed to monitor contextual and statistical aspects of the environment (e.g., as in Bernacchia et al., 2011; Farashahi et al., 2017). Greater specification of these circuits will enable us to take full advantage of the computational cognitive neuroscience approach by combining the circuits in a plug-and-play fashion. Finally, the model presented here was derived from neurobiological principles and meant to account for neurophysiological data and to serve as a foundation for the successful application of the model to behavioral data. Accordingly, future work should explore the application of the model to a variety of behavioral paradigms in which performance relies on DA levels, such as working memory, creative problem solving, reversal learning, task-set switching, category learning, and instrumental conditioning.

Chapter 5

General Discussion

Three computational models were presented that advance our understanding of how the brain generates some of our most intimate subjective experiences, and how it solves some of the most interesting problems posed by our environment. The first model accounted for sensory ratings by building a probabilistic, multidimensional representation of the sensory experiences elicited by exposure to each stimulus. It also builds a similar representation of the hypothetical ideal stimulus in this same space and generates hedonic ratings by measuring the Mahalanobis distance between the presented stimulus and the imagined ideal. The second model accounted for results from a flexible abstract rule learning experiment using a model that combines highly simplified building blocks that implement win-stay and lose-switch signals, as well as compute and represent predicted rewards. The third model uses a network of spiking neurons to represent activity within a neural circuit that implements adaptive learning rates by modulating the gain on the dopamine response to reward prediction errors. This model was able to account for a wide array of neuroscientific data.

I have gained some insights in the process of building these models, and in the following section I will provide some commentary on the process of model invention. I will

follow this with some practical advice that may help with future model development.

5.1 Behind the Scenes of Model Invention

Each of the models presented in this dissertation resulted from a long and arduous process that is rarely discussed in any scientific publications (for an excellent exception consult Shiffrin and Nobel (1997)). Throughout this process, countless nonviable models are disposed of in the process of inventing the models that are finally published. In this section I discuss the process of inventing models and give some practical tips to make this process less arduous. This section is particularly relevant to implementational level models, such as MODAL, and models that straddle the boundary between implementational and algorithmic levels, such as the model of flexible rule learning. Nonetheless, much of this advice is still applicable to models at the algorithmic level, such as the GRT model for identifying an ideal stimulus.

Deutsch (2011) compares the evolution of scientific theories with the successive improvements that occur in biological evolution. He discusses two significant differences that I consider crucial for understanding how models are invented. First, in biological evolution, random mutations from one generation to the next only result in slight deviations from the dominant strain. The evolution of scientific theories is similarly gradual but the vast majority of the invention and falsification of bad explanations takes place within the brain of the scientist. The gradual improvements in science rarely meet the high bar of publication, and thus, when one of these improvements is published the process by which it was invented seems obscure. Second, while random mutations in evolution must improve the fitness of the organism in order to increase the likelihood that they will propagate to the next generation, this is not a requirement for the mutations of scientific theories. The countless models invented in the intermediate stages between

good explanations do not have any viability requirements at all.

Shiffrin and Nobel (1997) note that the vast majority of effort put forth by the theorist requires skills that are "highly abstract and subjective, and these are seldom discussed in the literature" (p. 6). They estimate that in the timeline of model development perhaps 1% of the theorists time is spent on parameter estimation and analysis of the final model, and simulation of programs for other viable model variations. They suggest that the remaining 99% of the time is occupied by other "stuff". This stuff is the creation of intermediate nonviable models.

The creation of these nonviable models results from a continuous interaction between the theorist's intuitions, the simulation results, and the empirical and theoretical literature. For example, assume the theorist already has a vague idea of a candidate model and decides to program and simulate it for the phenomena of interest. At this point the model should be consistent with other empirical data and theory. The theorist will run the simulation with a very limited search of the parameter space. The purpose of this simulation is not to fit the empirical data for the phenomena in question; it is to understand how the components of the model interact and what kinds of behavior it may produce. Upon inspecting the results, the theorist may decide to change something about the model (either parameter values or structure) because they have some intuition and a prediction about how this change will alter the results of the simulation. However, prior to implementing this change the theorist will reference the literature to determine whether the parameter or structural change is empirically supported. For example, they may want to change a synaptic connection parameter from positive to negative, thus changing the nature of the connection from excitatory to inhibitory. In order to justify this change the theorist would likely have to change the structure (architecture) of the model so that it includes different brain regions that are connected by an inhibitory synapse. If justification is found the theorist runs the modified simulation. In fact, if

justification is not found, the theorist probably still runs the simulation because they will undoubtedly learn something about the behavior of the model.

This process - generating an idea for modification, justifying (or not) the modification, running the modified simulation, inspecting and contemplating the resulting behavior, and generating a new idea for modification - repeats itself until the model architecture and parameter values converge to a refined model. The refined model is understood logically rather than intuitively by the theorist, it is justified by other empirical and theoretical research, and, based on the theorists intuition, could potentially fit the empirical data. Finally, after all this "stuff" is done, the model parameters may be optimized to fit the data. It's quite likely that the model fit will be unsatisfactory and the theorist will return to searching through nonviable models with better tuned intuitions.

5.2 Practical Advice for Model Invention

A number of tutorials have been written on computational cognitive neuroscience (Ashby, 2018; Ashby & Helie, 2011). These outline the principles and tools of the field and are indispensable in the creation of CCN models. In the following section I will present some practical tips and heuristics that can accompany the CCN principles and tools. This advice is in the form of rules of thumb and as such there will be many exceptions. However, I do believe keeping them in mind at the onset of model invention can save time and frustration even if one decides to eventually ignore them.

5.2.1 Stop simulating models in one's head

The model architecture first emerges in the theorists head and at some point they write it down with rough box and arrow diagrams. As soon as these diagrams exist the theorist should start planning and writing a simulation program. I would estimate that

I have created thousands of box and arrow diagrams on paper throughout my graduate studies, and that does not include the thousands of diagrams I have created in my head.

The reason one should start running simulations on a computer as soon as possible is that when one thinks about a model or writes it on paper it is a simulation in the theorists head. Brain simulations, that is, simulations that run on human brains, are prone to error due to constraints on working memory, long term memory, and computational resources.

5.2.2 Balance the pursuit of justified and unjustified parameter or structure changes

Occasionally, a theorist will have some intuition about a change in parameters or structure that may cause the model to produce interesting behavior. It can be quite time intensive to find justification for a change, especially if that change causes the theorist to have to comb through unfamiliar literature. If justification is not found it could be because it does not exist or it could be because the theorist has not found it. Therefore, sometimes it is best to explore changes in parameters or structure even if they are not justified because most changes create nonviable models anyway, and the model can be discarded without consulting the literature. This saves time and enables the theorist to generate more intuition about the model. Of course, this is a balancing act because the theorist could spend an extensive amount of time exploring a model that is not justified by empirical data - ultimately, this model will fail to meet the principles of CCN.

5.2.3 Avoid adding noise unless absolutely necessary

Sometimes the presence of random variation in model entities (e.g. spiking neurons) is necessary for the explanation or even to fit data. However, especially early on in model development, it would be wise to keep the noise out of the model for as long as possible.

Recall that the search through nonviable models is strongly dependent on the theorist's intuitions. If random variations causes changes in model behavior that obscure the effects of changes to parameters or structure chosen by the theorist, then it will be incredibly difficult for the theorist to generate any intuition about the model. If noise is absolutely necessary, consider adding it after some intuition has been developed and add as little as possible.

5.2.4 Use particle swarm optimization while simulating intermediate models

Consider using particle swarm optimization (PSO) (Clerc, 2012) while simulating intermediate level models because it can enable the theorist to quickly get some intuition for the model's behavior. Importantly, these parameter searches do not need to be exhaustive as they will undoubtedly be more efficient than tuning the parameters by hand. PSO feels conducive to 'human-in-the-loop' (Rothrock & Narayanan, 2011) computing which is absolutely crucial when the invention of these models depends so heavily on the interaction between human intuition and computer simulation. Furthermore, when running PSO the theorist should consider making use of virtual machines on cloud servers that scale computing resources to meet the demands of the simulations.

5.2.5 Straddle the boundary between algorithmic and implementational levels throughout the model invention process

I recommend an integrated approach to algorithm and implementational level model design. That is, one should not search for candidate algorithms and spend years attempt-

ing to justify these models by fitting them to the data that they are well-suited to fit (while ignoring the data that they cannot explain). Start by constructing the simplest possible model architecture that maintains a reasonable mapping from parameters to neural circuit entities. If one already has an algorithmic model in mind then the goal should already be to find an implementational level model architecture that will implement the algorithm. If one does not explicitly have an algorithmic model in mind (it is there implicitly), then one will emerge from this simplest possible model architecture. This will happen because the simplest possible model is likely going to model *functions* of brain regions rather than the activity of neural populations. Accordingly, the resulting model will likely straddle the boundary between algorithmic and implementational levels, as is the case for the model of flexible rule learning.

This simple model architecture will provide a clear road map for the invention of lower level implementational models because entities in the model map to brain regions. Each of these brain regions will be performing some function in the model and lower level neural networks can be created that implement these functions via activation patterns, recurrent activity, and synaptic weights. Eventually, these lower level neural networks can be substituted into the simple model. Let's refer to the simplest possible model architecture that straddles the boundary between algorithmic and implementational levels as the superordinate model and to the detailed neural network models as subordinate models.

Once the superordinate model exists, the parameter estimation and model fitting process can be made substantially simpler for the subordinate models. This is because they can be fit to the outputs of each module in the superordinate model prior to fitting the entire subordinate model to the empirical data. If the subordinate networks can mimic the outputs generated by the superordinate modules given similar inputs, then they will likely require minimal parameter tuning in order to fit the empirical behavioral data. Once these subordinate networks are created it should be possible to fit them to a

wide variety of data; from the behavioral and lesion data presented in this dissertation to fMRI and EEG data, and neural recordings. A failure of the subordinate model to predict neural data could be viewed as a failure of the chosen neural implementation (that is, there are many subordinate network architectures that could generate the superordinate level circuit outputs), or it could suggest that the superordinate level model should be modified. Therefore, the creation and testing of these subordinate and superordinate models would ideally occur in parallel in an effort to make progress toward a satisfying explanation of flexible abstract rule learning.

5.3 Closing Remarks

This dissertation presented three models of phenomena in psychology and cognitive neuroscience that exist at and between the algorithmic and implementational levels of analysis. Each one represents an incomplete explanation of the phenomena in question due to the fact that the models are only a component of an explanation that should span all levels of analysis. A model that has reached the stage at which it is ready for publication is simply a statement about it being sufficiently viable. By sufficiently viable I simply mean that its mechanisms are clear enough and its fits to data are good enough for it to escape the theorists brain and make its way to the rest of the scientific community. These models will be falsified and discarded entirely or modified accordingly. My only hope for them is that they remain sufficiently viable long enough to point in the direction of better explanations.

Appendix A

Derivation of mean and variance for the normal approximation

The mean of the Eq. 2.3 random variable is (e.g., (Khatri, 1980; Paolella, 2018))

$$\begin{aligned}\mu_{\Delta^2} &= \text{trace}(\Sigma_Y^{-1}\Sigma_{\underline{w}}) + \underline{\boldsymbol{\mu}}_{\underline{w}}'\Sigma_Y^{-1}\underline{\boldsymbol{\mu}}_{\underline{w}} \\ &= \text{trace}[\Sigma_Y^{-1}(\Sigma_Y + \text{I})] + (\underline{\boldsymbol{\mu}}_Y - \underline{\boldsymbol{\mu}}_i)'\Sigma_Y^{-1}(\underline{\boldsymbol{\mu}}_Y - \underline{\boldsymbol{\mu}}_i) \\ &= D + \text{trace}(\Sigma_Y^{-1}) + (\underline{\boldsymbol{\mu}}_Y - \underline{\boldsymbol{\mu}}_i)'\Sigma_Y^{-1}(\underline{\boldsymbol{\mu}}_Y - \underline{\boldsymbol{\mu}}_i). \end{aligned} \tag{A.1}$$

Furthermore, if we let λ_i denote the i^{th} eigenvalue of Σ_Y, then Eq. 2.4 reduces to

$$\mu_{\Delta^2} = D + \sum_{i=1}^{D}\lambda_i^{-1} + (\underline{\boldsymbol{\mu}}_Y - \underline{\boldsymbol{\mu}}_i)'\Sigma_Y^{-1}(\underline{\boldsymbol{\mu}}_Y - \underline{\boldsymbol{\mu}}_i). \tag{A.2}$$

Note that the last term is just the squared Mahalanobis distance between the means of the two distributions.

The variance of the Eq. 2.3 random variable is (e.g., (Khatri, 1980; Paolella, 2018))

$$\begin{aligned}
\sigma^2_{\Delta^2} &= 2\text{trace}[(\Sigma_Y^{-1}\Sigma_W)^2] + 4\underline{\boldsymbol{\mu}}'_W\Sigma_Y^{-1}\Sigma_W\Sigma_Y^{-1}\underline{\boldsymbol{\mu}}_W \\
&= 2\text{trace}[\Sigma_Y^{-1}(\Sigma_Y + \mathrm{I})\Sigma_Y^{-1}(\Sigma_Y + \mathrm{I})] + 4(\underline{\boldsymbol{\mu}}_Y - \underline{\boldsymbol{\mu}}_i)'\Sigma_Y^{-1}(\Sigma_Y + \mathrm{I})\Sigma_Y^{-1}(\underline{\boldsymbol{\mu}}_Y - \underline{\boldsymbol{\mu}}_i) \\
&= 2\text{trace}(\mathrm{I} + 2\Sigma_Y^{-1} + \Sigma_Y^{-2}) + 4(\underline{\boldsymbol{\mu}}_Y - \underline{\boldsymbol{\mu}}_i)'\Sigma_Y^{-1}(\mathrm{I} + \Sigma_Y^{-1})(\underline{\boldsymbol{\mu}}_Y - \underline{\boldsymbol{\mu}}_i) \\
&= 2D + 4\text{trace}(\Sigma_Y^{-1}) + 2\text{trace}(\Sigma_Y^{-2}) + 4(\underline{\boldsymbol{\mu}}_Y - \underline{\boldsymbol{\mu}}_i)'\Sigma_Y^{-1}(\mathrm{I} + \Sigma_Y^{-1})(\underline{\boldsymbol{\mu}}_Y - \underline{\boldsymbol{\mu}}_i). \quad \text{(A.3)}
\end{aligned}$$

If expressed in terms of the eigenvalues of Σ_Y, then Eq. 2.5 becomes

$$\begin{aligned}
\sigma^2_{\Delta^2} &= 2D + 4\sum_{i=1}^{D}\lambda_i^{-1} + 2\sum_{i=1}^{D}\lambda_i^{-2} + 4(\underline{\boldsymbol{\mu}}_Y - \underline{\boldsymbol{\mu}}_i)'\Sigma_Y^{-1}(\mathrm{I} + \Sigma_Y^{-1})(\underline{\boldsymbol{\mu}}_Y - \underline{\boldsymbol{\mu}}_i) \\
&= 2D + 2\sum_{i=1}^{D}(2\lambda_i^{-1} + \lambda_i^{-2}) + 4(\underline{\boldsymbol{\mu}}_Y - \underline{\boldsymbol{\mu}}_i)'\Sigma_Y^{-1}(\mathrm{I} + \Sigma_Y^{-1})(\underline{\boldsymbol{\mu}}_Y - \underline{\boldsymbol{\mu}}_i). \quad \text{(A.4)}
\end{aligned}$$

Therefore, we can approximate the predicted probability that rating j is assigned to stimulus i on the liking dimension by computing the area between $X^2_{\mathrm{I},j}$ and $X^2_{\mathrm{I},j-1}$ under the pdf of a normal distribution with mean and variance specified by Eqs. A.1 and A.3, respectively (or alternatively, by Eqs. A.2 and A.4, respectively).

References

Aminoff, E. M., Kveraga, K., & Bar, M. (2013). The role of the parahippocampal cortex in cognition. *Trends in Cognitive Sciences*, *17*(8), 379–390.

Amos, A. (2000). A computational model of information processing in the frontal cortex and basal ganglia. *Journal of Cognitive Neuroscience*, *12*(3), 505–519.

Andrich, D. (1989). A probabilistic IRT model for unfolding preference data. *Applied Psychological Measurement*, *13*(2), 193–216.

Ashby, F. G. (1988). Estimating the parameters of multidimensional signal detection theory from simultaneous ratings on separate stimulus components. *Perception & Psychophysics*, *44*(3), 195–204.

Ashby, F. G. (2018). Computational cognitive neuroscience. In W. Batchelder, H. Colonius, E. Dzhafarov, & J. Myung (Eds.), *New handbook of mathematical psychology, volume 2* (pp. 223–270). New York: New York: Cambridge University Press.

Ashby, F. G., & Crossley, M. J. (2011). A computational model of how cholinergic interneurons protect striatal-dependent learning. *Journal of Cognitive Neuroscience*, *23*(6), 1549–1566.

Ashby, F. G., & Ennis, D. M. (2002). A Thurstone-Coombs model of concurrent ratings with sensory and liking dimensions. *Journal of Sensory Studies*, *17*(1), 43–59.

Ashby, F. G., & Ennis, J. M. (2006). The role of the basal ganglia in category learning. *Psychology of Learning and Motivation*, *46*, 1–36.

Ashby, F. G., & Helie, S. (2011). A tutorial on computational cognitive neuroscience: Modeling the neurodynamics of cognition. *Journal of Mathematical Psychology*, *55*(4), 273-289.

Ashby, F. G., Isen, A. M., & Turken, A. (1999). A neuropsychological theory of positive affect and its influence on cognition. *Psychological Review*, *106*(3), 529-550.

Ashby, F. G., & Townsend, J. T. (1986). Varieties of perceptual independence. *Psychological Review*, *93*, 154–179.

Ashby, F. G., Valentin, V. V., & Turken, A. U. (2002). The effects of positive affect and arousal and working memory and executive attention: Neurobiology and computational models. In S. Moore & M. Oaksford (Eds.), *Emotional cognition: From brain to behaviour* (p. 245-287). Amsterdam: John Benjamins Publishing Company.

Ashby, F. G., & Vucovich, L. E. (2016). The role of feedback contingency in perceptual category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(11), 1731.

Ashby, F. G., & Wenger, M. J. (in press). Statistical decision theory. In F. G. Ashby, H. Colonius, & E. N. Dzhafarov (Eds.), *New handbook of mathematical psychology, Volume III.* Cambridge University Press.

Baxter, M. G., Gaffan, D., Kyriazis, D. A., & Mitchell, A. S. (2009). Ventrolateral prefrontal cortex is required for performance of a strategy implementation task but not reinforcer devaluation effects in rhesus monkeys. *European Journal of Neuroscience*, *29*(10), 2049–2059.

Bayer, H. M., & Glimcher, P. W. (2005). Midbrain dopamine neurons encode a quantitative reward prediction error signal. *Neuron*, *47*(1), 129-141.

Bayer, H. M., Lau, B., & Glimcher, P. W. (2007). Statistics of midbrain dopamine neuron spike trains in the awake primate. *Journal of Neurophysiology*, *98*(3), 1428–1439.

Behrens, T. E., Woolrich, M. W., Walton, M. E., & Rushworth, M. F. (2007). Learning the value of information in an uncertain world. *Nature Neuroscience*, *10*(9), 1214–1221.

Belin, D., & Everitt, B. J. (2008). Cocaine seeking habits depend upon dopamine-dependent serial connectivity linking the ventral with the dorsal striatum. *Neuron*, *57*(3), 432–441.

Berke, J. D. (2018). What does dopamine mean? *Nature Neuroscience*, *21*(6), 787-793.

Bernacchia, A., Seo, H., Lee, D., & Wang, X.-J. (2011). A reservoir of time constants for memory traces in cortical neurons. *Nature Neuroscience*, *14*(3), 366–372.

Berridge, K. C. (2000). Reward learning: Reinforcement, incentives, and expectations. In D. Medin (Ed.), *Psychology of learning and motivation* (Vol. 40, pp. 223–278). Elsevier.

Bishara, A. J., Kruschke, J. K., Stout, J. C., Bechara, A., McCabe, D. P., & Busemeyer, J. R. (2010). Sequential learning models for the Wisconsin card sort task: Assessing processes in substance dependent individuals. *Journal of Mathematical Psychology*, *54*(1), 5–13.

Bland, A. R., & Schaefer, A. (2012). Different varieties of uncertainty in human decision-making. *Frontiers in Neuroscience*, *6*, 85.

Borg, I. (2018). *Applied multidimensional scaling and unfolding*. Amsterdam: Springer.

Bortz, D. M., Gazo, K. L., & Grace, A. A. (2019). The medial septum enhances reversal learning via opposing actions on ventral tegmental area and substantia nigra dopamine neurons. *Neuropsychopharmacology*, 1–9.

Bortz, D. M., & Grace, A. A. (2018). Medial septum differentially regulates dopamine neuron activity in the rat ventral tegmental area and substantia nigra via distinct pathways. *Neuropsychopharmacology*, 1.

Braganza, O., & Beck, H. (2018). The circuit motif as a conceptual tool for multilevel neuroscience. *Trends in Neurosciences*, *41*(3), 128–136.

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

Brown, J., Bullock, D., & Grossberg, S. (1999). How the basal ganglia use parallel excitatory and inhibitory learning pathways to selectively respond to unexpected rewarding cues. *Journal of Neuroscience*, *19*(23), 10502–10511.

Buckley, M. J., Mansouri, F. A., Hoda, H., Mahboubi, M., Browning, P. G., Kwok, S. C., ... Tanaka, K. (2009). Dissociable components of rule-guided behavior depend on distinct medial and prefrontal regions. *Science*, *325*(5936), 52–58.

Bush, R. R., & Mosteller, F. (1951). A model for stimulus generalization and discrimination. *Psychological Review*, *58*(6), 413-423.

Bussey, T. J., Wise, S. P., & Murray, E. A. (2001). The role of ventral and orbital

prefrontal cortex in conditional visuomotor learning and strategy use in rhesus monkeys (Macaca mulatta). *Behavioral Neuroscience*, *115*(5), 971–982.

Cantwell, G., Riesenhuber, M., Roeder, J. L., & Ashby, F. G. (2017). Perceptual category learning and visual processing: An exercise in computational cognitive neuroscience. *Neural Networks*, *89*, 31–38.

Carmichael, S., & Price, J. (1994). Architectonic subdivision of the orbital and medial prefrontal cortex in the macaque monkey. *Journal of Comparative Neurology*, *346*(3), 366–402.

Chau, B. K., Sallet, J., Papageorgiou, G. K., Noonan, M. P., Bell, A. H., Walton, M. E., & Rushworth, M. F. (2015). Contrasting roles for orbitofrontal cortex and amygdala in credit assignment and learning in macaques. *Neuron*, *87*(5), 1106–1118.

Christopoulos, G. I., Tobler, P. N., Bossaerts, P., Dolan, R. J., & Schultz, W. (2009). Neural correlates of value, risk, and risk aversion contributing to decision making under risk. *Journal of Neuroscience*, *29*(40), 12574–12583.

Clark, L., Cools, R., & Robbins, T. (2004). The neuropsychology of ventral prefrontal cortex: Decision-making and reversal learning. *Brain and Cognition*, *55*(1), 41–53.

Clarke, H., Dalley, J., Crofts, H., Robbins, T., & Roberts, A. (2004). Cognitive inflexibility after prefrontal serotonin depletion. *Science*, *304*(5672), 878–880.

Clarke, H. F., Robbins, T. W., & Roberts, A. C. (2008). Lesions of the medial striatum in monkeys produce perseverative impairments during reversal learning similar to those produced by lesions of the orbitofrontal cortex. *Journal of Neuroscience*, *28*(43), 10972–10982.

Clerc, M. (2012). Standard particle swarm optimisation. *Open access archive HAL*.

Cohen, J. Y., Haesler, S., Vong, L., Lowell, B. B., & Uchida, N. (2012). Neuron-type-specific signals for reward and punishment in the ventral tegmental area. *Nature*, *482*(7383), 85–88.

Constantinidis, C., & Steinmetz, M. A. (1996). Neuronal activity in posterior parietal area 7a during the delay periods of a spatial memory task. *Journal of Neurophysiology*, *76*(2), 1352–1355.

Contreras-Vidal, J. L., & Schultz, W. (1999). A predictive reinforcement model of dopamine neurons for learning approach behavior. *Journal of Computational Neuroscience*, *6*(3), 191–214.

Cools, R. (2006). Dopaminergic modulation of cognitive function-implications for l-dopa treatment in parkinson's disease. *Neuroscience & Biobehavioral Reviews*, *30*(1), 1–23.

Cools, R., Clark, L., Owen, A. M., & Robbins, T. W. (2002). Defining the neural mechanisms of probabilistic reversal learning using event-related functional magnetic resonance imaging. *Journal of Neuroscience*, *22*(11), 4563–4567.

Cools, R., & D'Esposito, M. (2011). Inverted U–shaped dopamine actions on human working memory and cognitive control. *Biological Psychiatry*, *69*(12), e113–e125.

Cools, R., & Robbins, T. W. (2004). Chemistry of the adaptive mind. *Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and*

Engineering Sciences, *362*(1825), 2871–2888.

Coombs, C. H. (1964). *A theory of data*. New York: Wiley.

Corbit, L. H., Muir, J. L., & Balleine, B. W. (2001). The role of the nucleus accumbens in instrumental conditioning: Evidence of a functional dissociation between accumbens core and shell. *Journal of Neuroscience*, *21*(9), 3251–3260.

Cornwall, J., & Phillipson, O. (1988). Afferent projections to the parafascicular thalamic nucleus of the rat, as shown by the retrograde transport of wheat germ agglutinin. *Brain Research Bulletin*, *20*(2), 139–150.

Crossley, M. J., Ashby, F. G., & Maddox, W. T. (2013). Erasing the engram: The unlearning of procedural skills. *Journal of Experimental Psychology: General*, *142*(3), 710–741.

Davison, M. L. (1979). Testing a unidimensional, qualitative unfolding model for attitudinal or developmental data. *Psychometrika*, *44*(2), 179–194.

Daw, N. D., & O'Doherty, J. P. (2014). Multiple systems for value learning. In P. W. Glimcher & E. Fehr (Eds.), *Neuroeconomics: Decision making and the brain, Second edition* (pp. 393–410). Amsterdam, The Netherlands: Elsevier.

Dayan, P., & Abbott, L. F. (2001). *Theoretical neuroscience: Computational and mathematical modeling of neural systems*. Cambridge, MA: MIT Press.

Dayan, P., Kakade, S., & Montague, P. R. (2000). Learning and selective attention. *Nature Neuroscience*, *3*(11s), 1218–1223.

Dayan, P., & Long, T. (1998). Statistical models of conditioning. In M. I. Jordan, M. J. Kearns, & S. A. Solla (Eds.), *Advances in neural information processing systems: Proceedings of the 1997 conference* (pp. 117–123). Cambridge, MA: MIT Press.

Dayan, P., & Yu, A. J. (2003). Expected and unexpected uncertainty: ACh and NE in the neocortex. In S. Becker, S. Thrun, & K. Obermayer (Eds.), *Advances in neural information processing systems: Proceedings of the 2002 conference* (pp. 173–180). Cambridge, MA: MIT Press.

Dehaene, S., & Changeux, J.-P. (1991). The Wisconsin Card Sorting Test: Theoretical analysis and modeling in a neuronal network. *Cerebral Cortex*, *1*(1), 62–79.

Deng, P., Zhang, Y., & Xu, Z. C. (2007). Involvement of I_h in dopamine modulation of tonic firing in striatal cholinergic interneurons. *Journal of Neuroscience*, *27*(12), 3148–3156.

DeSarbo, W. S., & Rao, V. R. (1984). GENFOLD2: A set of models and algorithms for the GENeral unFOLDing analysis of preference/dominance data. *Journal of Classification*, *1*(1), 147–186.

DeSarbo, W. S., Young, M. R., & Rangaswamy, A. (1997). A parametric multidimensional unfolding procedure for incomplete nonmetric preference/choice set data in marketing research. *Journal of Marketing Research*, *34*(4), 499–516.

De Soete, G., Carroll, J. D., & DeSarbo, W. S. (1986). The wandering ideal point model: A probabilistic multidimensional unfolding model for paired comparisons data. *Journal of Mathematical Psychology*, *30*(1), 28–41.

Deutsch, D. (2011). *The beginning of infinity: Explanations that transform the world*. Penguin UK.

Dias, R., Robbins, T. W., & Roberts, A. C. (1996). Dissociation in prefrontal cortex of affective and attentional shifts. *Nature*, *380*(6569), 69–72.

Ding, J. B., Guzman, J. N., Peterson, J. D., Goldberg, J. A., & Surmeier, D. J. (2010). Thalamic gating of corticostriatal signaling by cholinergic interneurons. *Neuron*, *67*(2), 294–307.

Doig, N. M., Magill, P. J., Apicella, P., Bolam, J. P., & Sharott, A. (2014). Cortical and thalamic excitation mediate the multiphasic responses of striatal cholinergic interneurons to motivationally salient stimuli. *Journal of Neuroscience*, *34*(8), 3101–3117.

Ennis, D. M. (1993). A single multidimensional model for discrimination, identification and preferential choice. *Acta Psychologica*, *84*(1), 17–27.

Ennis, D. M., & Ennis, J. M. (2013). Mapping hedonic data: A process perspective. *Journal of Sensory Studies*, *28*(4), 324–334.

Ennis, D. M., & Johnson, N. L. (1994). A general model for preferential and triadic choice in terms of central F distribution functions. *Psychometrika*, *59*(1), 91–96.

Ennis, D. M., & Rousseau, B. (2020). *Tools and applications of sensory and consumer science*. Institute for Perception.

Ermentrout, G. B. (1996). Type I membranes, phase resetting curves, and synchrony. *Neural Computation*, *8*(5), 979–1001.

Eshel, N., Bukwich, M., Rao, V., Hemmelder, V., Tian, J., & Uchida, N. (2015). Arithmetic and local circuitry underlying dopamine prediction errors. *Nature*, *525*(7568), 243–246.

Fabbricatore, A. T., Ghitza, U. E., Prokopenko, V. F., & West, M. O. (2009). Electrophysiological evidence of mediolateral functional dichotomy in the rat accumbens during cocaine self-administration: tonic firing patterns. *European Journal of Neuroscience*, *30*(12), 2387–2400.

Faget, L., Osakada, F., Duan, J., Ressler, R., Johnson, A. B., Proudfoot, J. A., ... Hnasko, T. S. (2016). Afferent inputs to neurotransmitter-defined cell types in the ventral tegmental area. *Cell reports*, *15*(12), 2796–2808.

Fanselow, M. S., & Dong, H. W. (2010). Are the dorsal and ventral hippocampus functionally distinct structures? *Neuron*, *65*(1), 7–19.

Farashahi, S., Donahue, C. H., Khorsand, P., Seo, H., Lee, D., & Soltani, A. (2017). Metaplasticity as a neural substrate for adaptive learning and choice under uncertainty. *Neuron*, *94*(2), 401–414.

Fascianelli, V., Ferrucci, L., Tsujimoto, S., & Genovesio, A. (2020). Neural correlates of strategy switching in the macaque orbital prefrontal cortex. *Journal of Neuroscience*, *40*(15), 3025–3034.

Franklin, N. T., & Frank, M. J. (2015). A cholinergic feedback circuit to regulate striatal population uncertainty and optimize reinforcement learning. *Elife*, *4*, e12029.

Friston, K. J., Shiner, T., FitzGerald, T., Galea, J. M., Adams, R., Brown, H., ... Best-

mann, S. (2012). Dopamine, affordance and active inference. *PLoS Computational Biology*, *8*(1).

Fuster, J. M., & Alexander, G. E. (1971). Neuron activity related to short-term memory. *Science*, *173*(3997), 652–654.

Gloor, P. (1997). *The temporal lobe & limbic system*. New York: Oxford University Press.

Grace, A. A. (2010). Dopamine system dysregulation by the ventral subiculum as the common pathophysiological basis for schizophrenia psychosis, psychostimulant abuse, and stress. *Neurotoxicity Research*, *18*(3-4), 367–376.

Grace, A. A., & Bunney, B. S. (1983). Intracellular and extracellular electrophysiology of nigral dopaminergic neurons-1. Identification and characterization. *Neuroscience*, *10*(2), 301–315.

Grace, A. A., Floresco, S. B., Goto, Y., & Lodge, D. J. (2007). Regulation of firing of dopaminergic neurons and control of goal-directed behaviors. *Trends in Neurosciences*, *30*(5), 220–227.

Groman, S. M., James, A. S., Seu, E., Crawford, M. A., Harpster, S. N., & Jentsch, J. D. (2013). Monoamine levels within the orbitofrontal cortex and putamen interact to predict reversal learning performance. *Biological Psychiatry*, *73*(8), 756–762.

Haber, S. N. (2016). Corticostriatal circuitry. *Dialogues in Clinical Neuroscience*, *18*(1), 7.

Haber, S. N., Fudge, J. L., & McFarland, N. R. (2000). Striatonigrostriatal pathways in primates form an ascending spiral from the shell to the dorsolateral striatum. *Journal of Neuroscience*, *20*(6), 2369–2382.

Harrison, L. M., Duggins, A., & Friston, K. J. (2006). Encoding uncertainty in the hippocampus. *Neural Networks*, *19*(5), 535–546.

Hart, A. S., Rutledge, R. B., Glimcher, P. W., & Phillips, P. E. (2014). Phasic dopamine release in the rat nucleus accumbens symmetrically encodes a reward prediction error term. *Journal of Neuroscience*, *34*(3), 698–704.

Hazy, T. E., Frank, M. J., & O'Reilly, R. C. (2010). Neural mechanisms of acquired phasic dopamine responses in learning. *Neuroscience & Biobehavioral Reviews*, *34*(5), 701–720.

Heaton, R. K. (1981). *A manual for the Wisconsin Card Sorting Test*. Odessa, FL: Psychological Assessment Resources.

Hikosaka, O., Sakamoto, M., & Usui, S. (1989). Functional properties of monkey caudate neurons. III. Activities related to expectation of target and reward. *Journal of Neurophysiology*, *61*(4), 814–832.

Hirsch, J., Hylton, R., & Graham, N. (1982). Simultaneous recognition of two spatial-frequency components. *Vision Research*, *22*(3), 365–375.

Hofstadter, D. R. (2007). *I am a strange loop*. Basic books.

Holland, P. C., & Schiffino, F. L. (2016). Mini-review: Prediction errors, attention and associative learning. *Neurobiology of learning and memory*, *131*, 207–215.

Hong, S., & Hikosaka, O. (2014). Pedunculopontine tegmental nucleus neurons provide

reward, sensorimotor, and alerting signals to midbrain dopamine neurons. *Neuroscience*, *282*, 139–155.

Hong, S., Jhou, T. C., Smith, M., Saleem, K. S., & Hikosaka, O. (2011). Negative reward signals from the lateral habenula to dopamine neurons are mediated by rostromedial tegmental nucleus in primates. *Journal of Neuroscience*, *31*(32), 11457–11471.

Horvitz, J. C. (2002). Dopamine gating of glutamatergic sensorimotor and incentive motivational input signals to the striatum. *Behavioural Brain Research*, *137*(1-2), 65–74.

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

Huettel, S. A., Song, A. W., & McCarthy, G. (2005). Decisions under uncertainty: probabilistic context influences activation of prefrontal and parietal cortices. *Journal of Neuroscience*, *25*(13), 3304–3311.

Humphries, M. D., & Prescott, T. J. (2010). The ventral basal ganglia, a selection mechanism at the crossroads of space, strategy, and reward. *Progress in Neurobiology*, *90*(4), 385–417.

Iigaya, K. (2016). Adaptive learning and decision-making under uncertainty by metaplastic synapses guided by a surprise detection system. *Elife*, *5*, e18073.

Inglis, J. B., Valentin, V. V., & Ashby, F. G. (2021). Modulation of dopamine for adaptive learning: A neurocomputational model. *Computational Brain & Behavior*, *4*(1), 34–52.

Insausti, R., Amaral, D., & Cowan, W. (1987). The entorhinal cortex of the monkey: II. cortical afferents. *Journal of Comparative Neurology*, *264*(3), 356–395.

Iversen, S. D., & Mishkin, M. (1970). Perseverative interference in monkeys following selective lesions of the inferior prefrontal convexity. *Experimental Brain Research*, *11*(4), 376–386.

Izhikevich, E. M. (2003). Simple model of spiking neurons. *IEEE Transactions on Neural Networks*, *14*(6), 1569-1572.

Izhikevich, E. M. (2007). *Dynamical systems in neuroscience*. Cambridge, MA: MIT Press.

Izquierdo, A., Darling, C., Manos, N., Pozos, H., Kim, C., Ostrander, S., ... Rudebeck, P. H. (2013). Basolateral amygdala lesions facilitate reward choices after negative feedback in rats. *Journal of Neuroscience*, *33*(9), 4105–4109.

Izquierdo, A., Suda, R. K., & Murray, E. A. (2004). Bilateral orbital prefrontal cortex lesions in rhesus monkeys disrupt choices guided by both reward value and reward contingency. *Journal of Neuroscience*, *24*(34), 7540–7548.

Jacobs, J., Kahana, M. J., Ekstrom, A. D., Mollison, M. V., & Fried, I. (2010). A sense of direction in human entorhinal cortex. *Proceedings of the National Academy of Sciences*, *107*(14), 6487–6492.

Jhou, T. C., Fields, H. L., Baxter, M. G., Saper, C. B., & Holland, P. C. (2009).

The rostromedial tegmental nucleus (RMTg), a GABAergic afferent to midbrain dopamine neurons, encodes aversive stimuli and inhibits motor responses. *Neuron*, *61*(5), 786–800.

Jo, S., & Jung, M. W. (2016). Differential coding of uncertain reward in rat insular and orbitofrontal cortex. *Scientific Reports*, *6*, 24085.

Jocham, G., Brodersen, K. H., Constantinescu, A. O., Kahn, M. C., Ianni, A. M., Walton, M. E., ... Behrens, T. E. (2016). Reward-guided learning with and without causal attribution. *Neuron*, *90*(1), 177–190.

Joel, D., Niv, Y., & Ruppin, E. (2002). Actor–critic models of the basal ganglia: New anatomical and computational perspectives. *Neural Networks*, *15*(4-6), 535–547.

Jones, B. F., & Witter, M. P. (2007). Cingulate cortex projections to the parahippocampal region and hippocampal formation in the rat. *Hippocampus*, *17*(10), 957–976.

Kawato, M., & Samejima, K. (2007). Efficient reinforcement learning: computational theories, neuroscience and robotics. *Current Opinion in Neurobiology*, *17*(2), 205–212.

Kazama, A., & Bachevalier, J. (2009). Selective aspiration or neurotoxic lesions of orbital frontal areas 11 and 13 spared monkeys' performance on the object discrimination reversal task. *Journal of Neuroscience*, *29*(9), 2794–2804.

Keiflin, R., Pribut, H. J., Shah, N. B., & Janak, P. H. (2019). Ventral tegmental dopamine neurons participate in reward identity predictions. *Current Biology*, *29*(1), 93–103.

Kennerley, S. W., Walton, M. E., Behrens, T. E., Buckley, M. J., & Rushworth, M. F. (2006). Optimal decision making and the anterior cingulate cortex. *Nature Neuroscience*, *9*(7), 940–947.

Kerr, K. M., Agster, K. L., Furtak, S. C., & Burwell, R. D. (2007). Functional neuroanatomy of the parahippocampal region: the lateral and medial entorhinal areas. *Hippocampus*, *17*(9), 697–708.

Khatri, C. G. (1980). Quadratic forms in normal variables. In P. R. Krishnaiah (Ed.), *Handbook of statistics, volume 1* (pp. 443–469). Amsterdam, North-Holland.

Kimberg, D. Y., D'Esposito, M., & Farah, M. J. (1997). Effects of bromocriptine on human subjects depend on working memory capacity. *Neuroreport*, *8*(16), 3581–3585.

Klavir, O., Genud-Gabai, R., & Paz, R. (2013). Functional connectivity between amygdala and cingulate cortex for adaptive aversive learning. *Neuron*, *80*(5), 1290–1300.

Kobayashi, Y., & Okada, K. (2007). Reward prediction error computation in the pedunculopontine tegmental nucleus neurons. *Annals of the New York Academy of Sciences*, *1104*(1), 310–323.

Kumaran, D., & Maguire, E. A. (2006). An unexpected sequence of events: mismatch detection in the human hippocampus. *PLoS Biology*, *4*(12), e424.

Lara, A. H., & Wallis, J. D. (2015). The role of prefrontal cortex in working memory: A mini review. *Frontiers in Systems Neuroscience*, *9*, 173.

Li, J., Schiller, D., Schoenbaum, G., Phelps, E. A., & Daw, N. D. (2011). Differential roles of human striatum and amygdala in associative learning. *Nature Neuroscience*,

14(10), 1250–1252.

Lipski, W. J., & Grace, A. A. (2013). Activation and inhibition of neurons in the hippocampal ventral subiculum by norepinephrine and locus coeruleus stimulation. *Neuropsychopharmacology*, *38*(2), 285.

Liu, X., Hairston, J., Schrier, M., & Fan, J. (2011). Common and distinct networks underlying reward valence and processing stages: a meta-analysis of functional neuroimaging studies. *Neuroscience & Biobehavioral Reviews*, *35*(5), 1219–1236.

Lodge, D. J., & Grace, A. A. (2006). The hippocampus modulates dopamine neuron responsivity by regulating the intensity of phasic neuron activation. *Neuropsychopharmacology*, *31*(7), 1356–1361.

Mackey, S., & Petrides, M. (2010). Quantitative demonstration of comparable architectonic areas within the ventromedial and lateral orbital frontal cortex in the human and the macaque monkey brains. *European Journal of Neuroscience*, *32*(11), 1940–1950.

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

Mansouri, F. A., Matsumoto, K., & Tanaka, K. (2006). Prefrontal cell activities related to monkeys' success and failure in adapting to rule changes in a Wisconsin Card Sorting Test analog. *Journal of Neuroscience*, *26*(10), 2745–2756.

Mansouri, F. A., & Tanaka, K. (2002). Behavioral evidence for working memory of sensory dimension in macaque monkeys. *Behavioural Brain Research*, *136*(2), 415–426.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information.* New York: Freeman.

Mathys, C., Daunizeau, J., Friston, K. J., & Stephan, K. E. (2011). A Bayesian foundation for individual learning under uncertainty. *Frontiers in Human Neuroscience*, *5*, 39.

Matsumoto, M., & Hikosaka, O. (2007). Lateral habenula as a source of negative reward signals in dopamine neurons. *Nature*, *447*(7148), 1111–1115.

Matsumoto, M., & Hikosaka, O. (2009). Representation of negative motivational value in the primate lateral habenula. *Nature Neuroscience*, *12*(1), 77–84.

Matsumoto, N., Minamimoto, T., Graybiel, A. M., & Kimura, M. (2001). Neurons in the thalamic CM-Pf complex supply striatal neurons with information about behaviorally significant sensory events. *Journal of Neurophysiology*, *85*(2), 960–976.

Mishkin, M., Malamut, B., & Bachevalier, J. (1984). Memories and habits: Two neural systems. In G. Lynch, J. L. McGaugh, & N. M. Weinberger (Eds.), *Neurobiology of human learning and memory* (pp. 65–77). New York: Guilford Press.

Monchi, O., Petrides, M., Petre, V., Worsley, K., & Dagher, A. (2001). Wisconsin card sorting revisited: distinct neural circuits participating in different stages of the task identified by event-related functional magnetic resonance imaging. *Journal of Neuroscience*, *21*(19), 7733–7741.

Monosov, I. E. (2017). Anterior cingulate is a source of valence-specific information about value and uncertainty. *Nature Communications*, *8*(1), 134.

Monosov, I. E., & Rushworth, M. F. (2021). Interactions between ventrolateral prefrontal and anterior cingulate cortex during learning and behavioural change. *Neuropsychopharmacology*, 1–15.

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.

Morita, K., Morishima, M., Sakai, K., & Kawaguchi, Y. (2012). Reinforcement learning: Computing the temporal difference of values via distinct corticostriatal pathways. *Trends in Neurosciences*, *35*(8), 457–467.

Morita, K., Morishima, M., Sakai, K., & Kawaguchi, Y. (2013). Dopaminergic control of motivation and reinforcement learning: a closed-circuit account for reward-oriented behavior. *Journal of Neuroscience*, *33*(20), 8866–8890.

Moustafa, A. A., & Gluck, M. A. (2011). A neurocomputational model of dopamine and prefrontal-striatal interactions during multicue category learning by Parkinson patients. *Journal of Cognitive Neuroscience*, *23*(1), 151–167.

Mullen, K., & Ennis, D. M. (1991). A simple multivariate probabilistic model for preferential and triadic choices. *Psychometrika*, *56*(1), 69–75.

Murray, E. A., & Wise, S. P. (2010). Interactions between orbital prefrontal cortex and amygdala: Advanced cognition, learned responses and instinctive behaviors. *Current Opinion in Neurobiology*, *20*(2), 212–220.

Mushiake, H., & Strick, P. L. (1995). Pallidal neuron activity during sequential arm movements. *Journal of Neurophysiology*, *74*(6), 2754–2758.

Niv, Y., Daw, N. D., Joel, D., & Dayan, P. (2007). Tonic dopamine: opportunity costs and the control of response vigor. *Psychopharmacology*, *191*(3), 507–520.

Noonan, M. P., Chau, B. K., Rushworth, M. F., & Fellows, L. K. (2017). Contrasting effects of medial and lateral orbitofrontal cortex lesions on credit assignment and decision-making in humans. *Journal of Neuroscience*, *37*(29), 7023–7035.

Okada, K., & Kobayashi, Y. (2013). Reward prediction-related increases and decreases in tonic neuronal activity of the pedunculopontine tegmental nucleus. *Frontiers in Integrative Neuroscience*, *7*, 36.

Olzak, L. A. (1986). Widely separated spatial frequencies: Mechanism interactions. *Vision Research*, *26*(7), 1143–1153.

O'Neill, M., & Schultz, W. (2010). Coding of reward risk by orbitofrontal neurons is mostly distinct from coding of reward value. *Neuron*, *68*(4), 789–800.

O'Reilly, R. C., Frank, M. J., Hazy, T. E., & Watz, B. (2007). PVLV: The primary value and learned value Pavlovian learning algorithm. *Behavioral Neuroscience*, *121*(1), 31–49.

Paolella, M. S. (2018). *Linear models and time-series analysis: Regression, anova, arma and garch*. Hoboken, NJ: John Wiley & Sons.

Payzan-LeNestour, E., & Bossaerts, P. (2011). Risk, unexpected uncertainty, and esti-

mation uncertainty: Bayesian learning in unstable settings. *PLoS Computational Biology*, *7*(1), e1001048.

Payzan-LeNestour, E., Dunne, S., Bossaerts, P., & O'Doherty, J. P. (2013). The neural representation of unexpected uncertainty during value-based decision making. *Neuron*, *79*(1), 191–201.

Pickering, A. D., & Pesola, F. (2014). Modeling dopaminergic and other processes involved in learning from reward prediction error: Contributions from an individual differences perspective. *Frontiers in Human Neuroscience*, *8*, 740.

Powell, M. J. (1994). A direct search optimization method that models the objective and constraint functions by linear interpolation. In *Advances in optimization and numerical analysis* (pp. 51–67). Springer.

Preuschoff, K., & Bossaerts, P. (2007). Adding prediction risk to the theory of reward learning. *Annals of the New York Academy of Sciences*, *1104*(1), 135–146.

Preuschoff, K., Quartz, S. R., & Bossaerts, P. (2008). Human insula activation reflects risk prediction errors as well as risk. *Journal of Neuroscience*, *28*(11), 2745–2752.

Quintero, E., Díaz, E., Vargas, J. P., de la Casa, G., & López, J. C. (2011). Ventral subiculum involvement in latent inhibition context specificity. *Physiology & Behavior*, *102*(3-4), 414–420.

Rall, W. (1967). Distinguishing theoretical synaptic potentials computed for different soma-dendritic distributions of synaptic input. *Journal of Neurophysiology*, *30*(5), 1138-1168.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning ii: Current research and theory* (pp. 64–99). New York: Appleton-Century-Crofts.

Riceberg, J. S., & Shapiro, M. L. (2012). Reward stability determines the contribution of orbitofrontal cortex to adaptive behavior. *Journal of Neuroscience*, *32*(46), 16402–16409.

Roberts, A. (2006). Primate orbitofrontal cortex and adaptive behaviour. *Trends in cognitive sciences*, *10*(2), 83–90.

Roberts, J. S., Donoghue, J. R., & Laughlin, J. E. (2000). A general item response theory model for unfolding unidimensional polytomous responses. *Applied Psychological Measurement*, *24*(1), 3–32.

Roesch, M. R., Esber, G. R., Li, J., Daw, N. D., & Schoenbaum, G. (2012). Surprise! Neural correlates of Pearce-Hall and Rescorla-Wagner coexist within the brain. *European Journal of Neuroscience*, *35*(7), 1190–1200.

Rohatgi, A. (2021). *Webplotdigitizer: Version 4.5.* Retrieved from `https://automeris.io/WebPlotDigitizer`

Root, D. H., Fabbricatore, A. T., Pawlak, A. P., Barker, D. J., Ma, S., & West, M. O. (2012). Slow phasic and tonic activity of ventral pallidal neurons during cocaine self-administration. *Synapse*, *66*(2), 106–127.

Rothrock, L., & Narayanan, S. (2011). *Human-in-the-loop simulations.* Springer.

Rudebeck, P. H., & Murray, E. A. (2008). Amygdala and orbitofrontal cortex lesions differentially influence choices during object reversal learning. *Journal of Neuroscience*, *28*(33), 8338–8343.

Rudebeck, P. H., & Murray, E. A. (2011). Dissociable effects of subtotal lesions within the macaque orbital prefrontal cortex on reward-guided behavior. *Journal of Neuroscience*, *31*(29), 10569–10578.

Rudebeck, P. H., Saunders, R. C., Lundgren, D. A., & Murray, E. A. (2017). Specialized representations of value in the orbital and ventrolateral prefrontal cortex: Desirability versus availability of outcomes. *Neuron*, *95*(5), 1208–1220.

Rudebeck, P. H., Saunders, R. C., Prescott, A. T., Chau, L. S., & Murray, E. A. (2013). Prefrontal mechanisms of behavioral flexibility, emotion regulation and value updating. *Nature Neuroscience*, *16*(8), 1140–1145.

Rushworth, M. F., & Behrens, T. E. (2008). Choice, uncertainty and value in prefrontal and cingulate cortex. *Nature Neuroscience*, *11*(4), 389–397.

Rutishauser, U., Mamelak, A. N., & Schuman, E. M. (2006). Single-trial learning of novel stimuli by individual neurons of the human hippocampus-amygdala complex. *Neuron*, *49*(6), 805–813.

Rutledge, R. B., Dean, M., Caplin, A., & Glimcher, P. W. (2010). Testing the reward prediction error hypothesis with an axiomatic model. *Journal of Neuroscience*, *30*(40), 13525–13536.

Rygula, R., Walker, S. C., Clarke, H. F., Robbins, T. W., & Roberts, A. C. (2010). Differential contributions of the primate ventrolateral prefrontal and orbitofrontal cortex to serial reversal learning. *Journal of Neuroscience*, *30*(43), 14552–14559.

Sadikot, A., Parent, A., & Francois, C. (1992). Efferent connections of the centromedian and parafascicular thalamic nuclei in the squirrel monkey: a PHA-L study of subcortical projections. *Journal of Comparative Neurology*, *315*(2), 137–159.

Saleem, K. S., Miller, B., & Price, J. L. (2014). Subdivisions and connectional networks of the lateral prefrontal cortex in the macaque monkey. *Journal of Comparative Neurology*, *522*(7), 1641–1690.

Sallet, J., Noonan, M. P., Thomas, A., O'Reilly, J. X., Anderson, J., Papageorgiou, G. K., ... Rushworth, M. F. (2020). Behavioral flexibility is associated with changes in structure and function distributed across a frontal cortical network in macaques. *PLoS biology*, *18*(5), e3000605.

Salum, C., da Silva, A. R., & Pickering, A. (1999). Striatal dopamine in attentional learning: A computational model. *Neurocomputing*, *26*, 845–854.

Scheffé, H. (1999). *The analysis of variance*. New York: John Wiley & Sons.

Schoenbaum, G., Takahashi, Y., Liu, T.-L., & McDannald, M. A. (2011). Does the orbitofrontal cortex signal value? *Annals of the New York Academy of Sciences*, *1239*, 87.

Schönemann, P. H., & Wang, M. M. (1972). An individual difference model for the multidimensional analysis of preference data. *Psychometrika*, *37*(3), 275–309.

Schultz, W. (1998). Predictive reward signal of dopamine neurons. *Journal of Neuro-*

physiology, *80*(1), 1–27.

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

Seamans, J. K., & Robbins, T. W. (2010). Dopamine modulation of the prefrontal cortex and cognitive function. In K. A. Neve (Ed.), *The dopamine receptors, 2nd edition* (pp. 373–398). New York: Springer.

Sesack, S. R., & Grace, A. A. (2010). Cortico-basal ganglia reward network: Microcircuitry. *Neuropsychopharmacology*, *35*(1), 27–47.

Shiffrin, R. M., & Nobel, P. A. (1997). The art of model development and testing. *Behavior Research Methods, Instruments, & Computers*, *29*(1), 6–14.

Sleezer, B. J., Castagno, M. D., & Hayden, B. Y. (2016). Rule encoding in orbitofrontal cortex and striatum guides selection. *Journal of Neuroscience*, *36*(44), 11223–11237.

Smith, Y., & Kieval, J. Z. (2000). Anatomy of the dopamine system in the basal ganglia. *Trends in Neurosciences*, *23*, S28–S33.

Soltani, A., & Izquierdo, A. (2019). Adaptive learning under expected and unexpected uncertainty. *Nature Reviews Neuroscience*, *20*(10), 635–644.

Stalnaker, T. A., Franz, T. M., Singh, T., & Schoenbaum, G. (2007). Basolateral amygdala lesions abolish orbitofrontal-dependent reversal impairments. *Neuron*, *54*(1), 51–58.

Steinke, A., Lange, F., & Kopp, B. (2020). Parallel model-based and model-free reinforcement learning for card sorting performance. *Scientific Reports*, *10*(1), 1–18.

Stolyarova, A., & Izquierdo, A. (2017). Complementary contributions of basolateral amygdala and orbitofrontal cortex to value learning under uncertainty. *Elife*, *6*, e27483.

Strange, B. A., Duggins, A., Penny, W., Dolan, R. J., & Friston, K. J. (2005). Information theory, novelty and hippocampal responses: unpredicted or unpredictable? *Neural Networks*, *18*(3), 225–230.

Stuber, G. D., Klanker, M., De Ridder, B., Bowers, M. S., Joosten, R. N., Feenstra, M. G., & Bonci, A. (2008). Reward-predictive cues enhance excitatory synaptic strength onto midbrain dopamine neurons. *Science*, *321*(5896), 1690–1692.

Sutton, R. S. (1992). Adapting bias by gradient descent: An incremental version of delta-bar-delta. In *Proceedings of the tenth national conference on artificial intelligence* (pp. 171–176). Cambridge, MA: MIT Press.

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction*. Cambridge, MA: MIT Press.

Takahashi, Y. K., Langdon, A. J., Niv, Y., & Schoenbaum, G. (2016). Temporal specificity of reward prediction errors signaled by putative dopamine neurons in rat VTA depends on ventral striatum. *Neuron*, *91*(1), 182–193.

Takahashi, Y. K., Schoenbaum, G., & Niv, Y. (2008). Silencing the critics: Understanding the effects of cocaine sensitization on dorsolateral and ventral striatum in the context of an actor/critic model. *Frontiers in Neuroscience*, *2*, 14.

Tan, C. O., & Bullock, D. (2008). A local circuit model of learned striatal and dopamine cell responses under probabilistic schedules of reward. *Journal of Neuroscience*, *28*(40), 10062–10074.

Taswell, C. A., Costa, V. D., Murray, E. A., & Averbeck, B. B. (2018). Ventral striatum's role in learning from gains and losses. *Proceedings of the National Academy of Sciences*, *115*(52), E12398–E12406.

Tian, J., & Uchida, N. (2015). Habenula lesions reveal that multiple mechanisms underlie dopamine prediction errors. *Neuron*, *87*(6), 1304–1316.

Vanni-Mercier, G., Mauguiere, F., Isnard, J., & Dreher, J.-C. (2009). The hippocampus codes the uncertainty of cue–outcome associations: an intracranial electrophysiological study in humans. *Journal of Neuroscience*, *29*(16), 5287–5294.

Van Rossum, G., & Drake, F. L. (2011). *The Python language reference manual*. Network Theory Ltd.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., ... SciPy 1.0 Contributors (2020). SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, *17*, 261–272. doi: 10.1038/s41592-019-0686-2

Vitay, J., & Hamker, F. H. (2014). Timing and expectation of reward: A neurocomputational model of the afferents to the ventral tegmental area. *Frontiers in Neurorobotics*, *8*, 4.

Watabe-Uchida, M., Zhu, L., Ogawa, S. K., Vamanrao, A., & Uchida, N. (2012). Whole-brain mapping of direct inputs to midbrain dopamine neurons. *Neuron*, *74*(5), 858–873.

Wickens, T. D. (1992). Maximum-likelihood estimation of a multivariate Gaussian rating model with excluded data. *Journal of Mathematical Psychology*, *36*(2), 213–234.

Willingham, D. B. (1998). A neuropsychological theory of motor skill learning. *Psychological Review*, *105*, 558–584.

Wilson, R. C., Takahashi, Y. K., Schoenbaum, G., & Niv, Y. (2014). Orbitofrontal cortex as a cognitive map of task space. *Neuron*, *81*(2), 267–279.

Yu, J., Angela, & Dayan, P. (2005). Uncertainty, neuromodulation, and attention. *Neuron*, *46*(4), 681–692.

Zinnes, J. L., & Griggs, R. A. (1974). Probabilistic, multidimensional unfolding analysis. *Psychometrika*, *39*(3), 327–350.